Der Grenzwert – Zentralbegriff der Analysis (Langversion)

REINHARD WINKLER (TU WIEN)

Der Begriff des Grenzwertes ist zentral in der Analysis und einer der wichtigsten der Mathematik überhaupt. Nicht nur an sich ist eine sorgfältige Behandlung dieses Begriffs im Unterricht wünschenswert, sie macht auch beträchtliche Teile der obligatorischen Analysis viel verständlicher. Analogien zwischen verschiedenen Ausprägungen des Grenzwertes (Folgen, Reihen, Funktionen und Stetigkeit, Ableitung, Integral) erleichtern es, über einen nur vagen „intuitiven Grenzwertbegriff" hinauszukommen und zu einem präzisen Begriff vorzudringen. Dabei wird klar, wie die fünf genannten Stoffkapitel auf einer einzigen gemeinsamen Grundidee fußen. Wie gleichfalls im vorliegenden Artikel gezeigt werden soll, lässt sich damit der Lehrstoff zur Analysis viel besser überschauen.

1. Einleitung

Der rote Faden im vorliegenden Artikel ist der für die Analysis zentrale und darüber hinaus in die gesamte Mathematik ausstrahlende Begriff des Grenzwertes in fünf für den Schulstoff wichtigen Ausprägungen: Folgen, Reihen, Funktionen samt Stetigkeit, Differential- und Integralrechnung. An jedes dieser Themen schließen sich bestimmte Inhalte als Minimalprogramm an, ohne die ein adäquates Verständnis der Grundideen der Analysis unmöglich ist. Hier sollen die Grundbegriffe erarbeitet und ein solches Minimalprogramm präsentiert werden.

Da diese Inhalte den vorgegebenen Rahmen mehr als ausfüllen, muss die Darstellung sehr knapp gehalten werden. So gibt es keine Abbildungen, auch wenn mit ihnen die meisten der im Text verbal formulierten Ideen noch überzeugender dargestellt werden könnten. Ähnliches gilt für die äußerst interessanten historischen Bezüge. Auch Beweise fehlen im Haupttext. Sie sind in einen Anhang ausgelagert, der allerdings nur in der Langversion des Artikels enthalten ist, nicht jedoch in der kürzeren gedruckten Fassung. Das durch den Text vermittelte Minimalprogramm zur Analysis (ergänzt durch gelegentliche Verweise auf andere Artikel) ist inhaltlich zu verstehen und nicht als Anweisung für die konkrete Unterrichtsgestaltung. Folglich richtet sich der vorliegende Text an Lehrende und nicht direkt an Schüler. Ziel ist es, Hintergründe und Zusammenhänge herauszuarbeiten, die den Lehrenden jederzeit bewusst und im Unterricht (vor allem, aber nicht nur) für die zehnte bis zwölfte Schulstufe als Leitgedanken wirksam sein sollten.

Den mathematischen Inhalten sei lediglich die folgende pauschale Empfehlung für den Unterricht vorangestellt: Mathematisches Tun strebt nach klaren Begriffen und strengen Beweisen von interessanten Aussagen über diese Begriffe. Heuristische, anschauliche und intuitive Motivationen sollen also möglichst in klare Definitionen und Beweise münden, Ausnahmen brauchen gute Begründungen. Dabei kommt es nicht auf Formalisierung an, sondern auf Schlüssigkeit der Gedanken. Der Aufbau des vorliegenden Artikels versucht diese Schlüssigkeit zu gewährleisten, auch wenn das oft erst in den Beweisen (siehe Anhang der Langversion) deutlich wird. Fast allen Beweisen liegen sehr griffige, oft bildhafte Ideen zugrunde, die es lohnen, in adäquater Form in den Unterricht übersetzt zu werden. Wie genau in einer konkreten Klasse der Unterricht im Einzelnen zu gestalten ist, kann und soll durch einen Artikel wie den vorliegenden der Verantwortung der Lehrerin oder des Lehrers aber nicht entzogen werden.

Es folgen nun fünf thematischen Hauptteile (Kapitel 2 bis 6) und ein Resümee als Kapitel 7. Die Langversion Winkler (2018/19) enthält überdies noch einen Anhang mit Beweisen als letztes Kapitel.

2. Folgen

Das Kapitel gliedert sich in drei Abschnitte. Im ersten (2.1) werden die Definitionen von Monotonie, Beschränktheit, Konvergenz und Häufungspunkten von Folgen erarbeitet. Abschnitt 2.2 bringt typische Beispiele und die wichtigsten allgemeinen Beziehungen zwischen diesen Begriffen. Den Abschluss des

Schriftenreihe zur Didaktik der Mathematik der Österreichischen Mathematischen Gesellschaft (ÖMG), Heft 51, 2018/19, ??-?? S. 1–26
Erweiterte online-Version vom 19.10.2018

Kapitels bilden in 2.3 die Grenzwertsätze für Summe, Produkt etc. konvergenter Folgen mit Anwendung auf gebrochen rationale Folgen sowie das besonders wichtige Beispiel der geometrischen Folge.

2.1. Monotonie, Beschränktheit, Grenzwert, Häufungspunkt

Wir wollen diese Begriffe präzise fassen. Das ist nicht völlig trivial. Auch historisch war es ein langer Prozess, bis sich gewisse Definitionen als adäquat herauskristallisierten. Das lag weniger an der Kompliziertheit der Begriffe, als an der sich erst nach und nach etablierenden Einsicht, dass es gewisse logische Strukturen sind, die taugliche und unverzichtbare Grundlagen für das mächtige Lehrgebäude der modernen Mathematik auszeichnen. Heutzutage stehen wir Zwerge auf den Schultern eines Riesen, nämlich des Kollektivs der größten Mathematiker vergangener Epochen, und genießen einen bequemen Blick auf eine übersichtliche Landschaft, auf den zu verzichten es keinen Grund gibt und den wir uns nun gönnen wollen.

Als Einstieg soll anhand von drei sehr einfachen, aber typischen Beispielen eine Intuition für die vier Begriffe geschaffen werden. Exemplarisch betrachten wir die drei Folgen[1] $\mathbf{a} = (a_n)_n$, $\mathbf{b} = (b_n)_n$ und $\mathbf{c} = (c_n)_n$ mit den Gliedern $a_n = \frac{1}{n}$, $b_n = (-1)^n(1 + \frac{1}{n})$ und $c_n = n$. Von diesen Folgen sind sind **a** und **c** *monoton* (sogar streng monoton, **a** fallend und **c** wachsend), weil die Glieder immer kleiner bzw. immer größer werden, was für **b** nicht der Fall ist. Die Folgen **a** und **b** sind *beschränkt*, weil ihre Glieder (im Gegensatz zu den c_n) betragsmäßig nicht beliebig groß werden. *Konvergent* ist nur **a**, weil sich die a_n einem sogenannten Grenzwert α (im Beispiel $\alpha = 0$, man spricht von einer *Nullfolge*) beliebig annähern, während es für **b** und **c** kein solches α gibt. Die Folge **b** hat aber immerhin *Häufungspunkte* (und zwar zwei, nämlich 1 und -1), denen sich die Folge gleichfalls beliebig annähert – wenn auch nicht ausnahmslos für alle n, so doch jeweils unendlich oft. Nun zur Präzisierung dieser noch etwas ungenauen Beschreibungen.

Jeder der vier Begriffe ist mit einer Ungleichung verknüpft: (wachsende) Monotonie mit $a_n \leq a_{n+1}$, Beschränktheit mit $|a_n| \leq S$ (S steht für „Schranke“), Konvergenz gegen α und Häufungspunkt α mit $|a_n - \alpha| < \varepsilon$ ($\varepsilon > 0$ ist dabei als klein zu denken). Zwei Fragen tauchen auf: Welche Rollen spielen erstens die in diesen Formeln auftretenden Parameter n, S und ε, und worin genau besteht zweitens der Unterschied zwischen Grenzwert und Häufungspunkt, wenn dieselbe Ungleichung für beide zuständig ist? Die Antworten auf beide Fragen sind ähnlich und typisch für die gesamte Mathematik. Und zwar müssen die Parameter auf eine ganz bestimmte Weise spezifiziert werden, die mit Hilfe der beiden logischen *Quantoren*, dem *Allquantor* $\forall$ und dem *Existenzquantor* $\exists$, formalisiert werden kann. Das soll nun geschehen.

Beginnen wir mit dem einfachsten, offensichtlichen Fall, der (wachsenden) Monotonie: Auf die Frage, für welche n die Ungleichung $a_n \leq a_{n+1}$ gelten soll, lautet die Antwort schlicht *für alle* Indizes, als Formel: $\forall n: \ a_n \leq a_{n+1}$.

Auch für die Beschränktheit soll die entsprechende Ungleichung $|a_n| \leq S$ für alle n gelten. Was aber ist S? Offenbar können wir nicht ein beliebiges $S \in \mathbb{R}$ nehmen; das wäre zu viel verlangt. Uns genügt es, wenn irgendein S das leistet, wenn es also (irgend)ein geeignetes $S \in \mathbb{R}$ gibt, als Formel: $\exists S\, \forall n: \ |a_n| \leq S$. Um den Begriff zu exaktifizieren, sind also bereits zwei (unterschiedliche) Quantoren notwendig.[2]

[1] Zur Wiederholung: Eine Folge ist eine Abbildung (Synonym: Funktion) $n \mapsto a_n$ mit Definitionsbereich $\mathbb{N} = \{0, 1, 2, \ldots\}$ (oder einer Teilmenge $D \subseteq \mathbb{N}$, oft ist $D = \mathbb{N}^+ = \{1, 2, \ldots\}$). Liegen die sogenannten *Glieder* a_n der Folge in der Menge $\mathbb{R}$ der reellen Zahlen, so sprechen wir von einer reellen Folge. Wir wollen uns hier auf reelle Folgen beschränken. Sind die Glieder a_n der Folge $\mathbf{a}: D \to \mathbb{R}$ durch einen Funktionsterm gegeben und ist sonst nichts vermerkt, so nehmen wir als Definitionsbereich D die Menge aller $n \in \mathbb{N}$, für die dieser Term definiert ist, z.B. $D = \mathbb{N}^+ = \{1, 2, \ldots\}$ für $a_n = \frac{1}{n}$. Wir schreiben auch $\mathbf{a} = (a_n)_{n \in D}$ oder kürzer $\mathbf{a} = (a_n)_n$.

[2] Man kann sich fragen, ob eine geschicktere Definition von Beschränktheit möglich wäre, die mit nur einem Quantor auskommt. Das ist aber nicht der Fall. Denn Beschränktheit kann weder bestätigt noch widerlegt werden, solange man nur endlich viele Glieder der Folge kennt. Jede reine Allaussage (wie etwa Monotonie) kann aber durch ein einziges Gegenbeispiel widerlegt, jede reine Existenzaussage durch ein einziges positives Beispiel bewiesen werden. Mit mehr Aufwand ließe sich auch für die anschließende Definition des Grenzwertes zeigen, dass die Anzahl der logischen Quantoren (nämlich drei) nicht reduziert werden kann.

Für den Grenzwert wird die Aufgabe etwas komplizierter. Wenn wir sie lösen (es wird nur wenige Zeilen brauchen) werden wir aber den entscheidenden Schritt ins Zentrum der Analysis geschafft haben: In der Ungleichung $|a_n - \alpha| < \varepsilon$ ist α fest vorgegeben und verlangt deshalb keinen Quantor. Der Folgenindex n ist wieder mit einem „für alle“ zu versehen, allerdings mit Einschränkungen. Denn für $a_n = \frac{1}{n}$ und $\varepsilon = \frac{1}{10}$ beispielsweise stört es uns nicht, dass die Ungleichung erst ab dem Folgenindex $n = 11$ gilt und für noch kleinere $\varepsilon > 0$ erst entsprechend später. Entscheidend ist, dass die Ungleichung *für alle n ab einem geeigneten* n_0 gilt. (Man sagt auch *für fast alle n*, wenn man *für alle bis auf höchstens endlich viele Ausnahmen* meint.) Als Formel: $\exists n_0\ \forall n \geq n_0 :\ |a_n - \alpha| < \varepsilon$, was ausführlicher $\exists n_0\ \forall n :\ n \geq n_0 \Rightarrow |a_n - \alpha| < \varepsilon$ bedeutet. Unbestimmt (man sagt in diesem Zusammenhang *nicht durch Quantoren gebunden*) ist in dieser Formel aber noch das ε. Entscheidend für den Grenzwertbegriff ist, dass die Formel *für alle* $\varepsilon > 0$ gelten soll, also $\forall \varepsilon > 0\ \exists n_0\ \forall n \geq n_0 :\ |a_n - \alpha| < \varepsilon$. Verbal: Zu jedem positiven reellen ε gibt es einen Index n_0, ab dem alle Folgenglieder sich vom Grenzwert um weniger als ε unterscheiden. Nennt man die Abweichung der Folgenglieder vom Grenzwert „Fehler“, so kann man auch formulieren: *Jede beliebige vorgegebene Fehlerschranke wird garantiert eingehalten, sofern nur der Folgenindex hinreichend groß ist*. Auf das hier auftretende Wortpaar *beliebig – hinreichend* werden wir uns noch des öfteren beziehen. Damit ist der entscheidende Schritt bereits geschafft, und es hat gar nicht wehgetan!

Darüber hinaus steht noch die Unterscheidung zwischen Grenzwert und Häufungspunkt auf unserem Programm. Zur Illustration ziehen wir nochmals die Folgen $\mathbf{a}$ und $\mathbf{b}$ mit den Gliedern $a_n = \frac{1}{n}$ bzw. $b_n = (-1)^n(1 + \frac{1}{n})$ heran. Der entscheidende Unterschied zwischen den Ungleichungen $|a_n - 0| < \varepsilon$ bzw. $|b_n - 1| < \varepsilon$ und $|b_n - (-1)| < \varepsilon$ besteht darin, dass (sofern $\varepsilon < 2$) nur für erstere ein n_0 existiert, so dass die Ungleichung wirklich *für alle* $n \geq n_0$ gilt. Für die zweite und dritte treten nur immer wieder solche n (d.h. unendlich viele) auf, also zu jedem n_0 ein $n \geq n_0$, das die Ungleichung erfüllt. Formal wird das durch unterschiedliche Setzung der Quantoren sichtbar: $\exists n_0\ \forall n \geq n_0$ für den Grenzwert und $\forall n_0\ \exists n \geq n_0$ für den Häufungspunkt. Ein Häufungspunkt α einer Folge mit den Gliedern a_n ist also durch die Formel $\forall \varepsilon > 0\ \forall n_0\ \exists n \geq n_0 :\ |a_n - \alpha| < \varepsilon$ charakterisiert. Formal lauten die Definitionen demnach:

Definition 2.1 *Sei* $\mathbf{a} = (a_n)_{n \in \mathbb{N}}$ *eine reelle Folge.*

1. *Die Folge* $\mathbf{a}$ *heißt* monoton wachsend, *wenn*

$$\forall n :\ a_n \leq a_{n+1}$$

gilt. Ersetzt man in dieser Formel das Symbol $\leq$ *durch* $<$, *so heißt die Folge* streng monoton wachsend, *analog* monoton fallend *bei* $\geq$ *und* streng monoton fallend *bei* $>$. *Eine Folge heißt* monoton, *wenn sie eine dieser vier Eigenschaften hat.*

2. *Die Folge* $\mathbf{a} = (a_n)_n$ *heißt* beschränkt,[3] *wenn gilt:*

$$\exists S\ \forall n :\ |a_n| \leq S$$

3. *Eine Folge* $\mathbf{a} = (a_n)_n$ *heißt* konvergent, *wenn sie einen* Grenzwert *oder auch* Limes $\alpha \in \mathbb{R}$ *hat (man sagt auch: sie* konvergiert gegen α, *symbolisch* $\lim_{n\to\infty} a_n = \alpha$ *oder* $a_n \to \alpha$ *für* $n \to \infty$*), was definitionsgemäß*

$$\forall \varepsilon > 0\ \exists n_0 \forall n \geq n_0 :\ |a_n - \alpha| < \varepsilon$$

bedeutet. Konvergiert die Folge nicht, so heißt sie divergent.

4. *Die Zahl* $\alpha \in \mathbb{R}$ *heißt* Häufungspunkt *der Folge* $\mathbf{a} = (a_n)_{n \in \mathbb{N}}$, *wenn gilt:*

$$\forall \varepsilon > 0\ \forall n_0\ \exists n \geq n_0 :\ |a_n - \alpha| < \varepsilon$$

[3] Gelegentlich verwenden wir auch die Begriffe *nach oben* bzw. *nach unten beschränkt*, die durch die Formeln $\exists S\ \forall n :\ a_n \leq S$ bzw. $\exists S\ \forall n :\ a_n \geq S$ beschrieben werden.

2.2. Beziehungen zwischen diesen Begriffen und weitere Beispiele

Die wichtigsten Beziehungen zwischen den Begriffen aus Abschnitt 2.1 sind im folgenden Satz zusammengefasst:

Satz 1 *(***Eigenschaften von Folgen***)*

1. *Jeder Grenzwert einer Folge ist Häufungspunkt dieser Folge.*
2. *Eine Folge mit mehr als einem Häufungspunkt ist divergent.*
3. *Der Grenzwert einer konvergenten Folge ist eindeutig bestimmt.*
4. *Jede konvergente Folge ist beschränkt.*
5. *Jede Folge, die monoton und beschränkt ist, konvergiert.*
6. *Jede beschränkte Folge hat mindestens einen Häufungspunkt. (Satz von Bolzano Weierstraß 1)*
7. *Zu jedem Häufungspunkt einer Folge gibt es eine Teilfolge, die gegen diesen Häufungspunkt konvergiert. (Satz von Bolzano Weierstraß 2)*

Man mache sich klar, dass von den a priori 16 kombinatorischen Möglichkeiten für eine Folge, jede der vier Eigenschaften (monoton, beschränkt, konvergent und Existenz mindestens eines Häufungspunktes) zu haben oder auch nicht, zehn durch Satz 1 ausgeschlossen werden. Die verbleibenden sechs Möglichkeiten können tatsächlich realisiert werden, wie die Beispiele aus Satz 2 belegen:

Satz 2 (Typische Beispiele von Folgen) *Für eine reelle Folge $(a_n)_n$ sind, wie die jeweils angegebenen Beispiele zeigen, die folgende Kombinationen von Eigenschaften möglich:*

1. monoton, beschränkt, konvergent, Häufungspunkt: $a_n = \frac{1}{n}$
2. nicht monoton, beschränkt, konvergent, Häufungspunkt: $a_n = (-1)^n \frac{1}{n}$
3. nicht monoton, beschränkt, divergent, Häufungspunkt(e): $a_n = (-1)^n$.
4. monoton, unbeschränkt, divergent, kein Häufungspunkt: $a_n = n$
5. nicht monoton, unbeschränkt, divergent, ein Häufungspunkt: $a_n = n + (-1)^n n$.
6. nicht monoton, unbeschränkt, divergent, kein Häufungspunkt: $a_n = (-1)^n n$

2.3. Arbeit mit dem Grenzwertbegriff für Folgen

Zur expliziten Berechnung von Folgengrenzwerten ist es von besonderem Interesse, dass gliedweise Summe, Differenz, Produkt und Quotient (bei Nenner $\neq 0$) konvergenter Folgen wieder konvergieren, und zwar gegen Summe, Differenz, Produkt bzw. Quotient der einzelnen Grenzwerte. (In anderer Terminologie könnte man auch sagen: Die vier Grundrechnungsarten sind stetig.) Der intuitive Hintergrund ist klar: Wenn sich für $n \to \infty$ die Zahlen a_n einem Wert α und die b_n einem Wert β beliebig annähern, dann auch die Summen $a_n + b_n$ dem Wert $\alpha + \beta$, analog für Differenz, Produkt und Quotienten. Also:

Satz 3 (Grenzwertsätze) *Seien $(a_n)_n$ und $(b_n)_n$ konvergente Folgen mit $\lim_{n\to\infty} a_n = \alpha$ und $\lim_{n\to\infty} b_n = \beta$. Dann sind auch die Folgen $(c_n)_n$ mit den jeweils angegebenen Gliedern c_n konvergent gegen den jeweils angegebenen Grenzwert γ, im letzten Fall unter der angegebenen Zusatzbedingung.*

1. $c_n = a_n \pm b_n$, $\gamma = \alpha \pm \beta$ *(Summen- bzw. Differenzsatz)*
2. $c_n = a_n b_n$, $\gamma = \alpha\beta$ *(Produktsatz). Ist $\beta = 0$, so genügt schon Beschränktheit (statt Konvergenz) der a_n, um die Konvergenz $\lim_{n\to\infty} a_n b_n = 0$ zu garantieren.*
3. $c_n = \frac{a_n}{b_n}$, $\gamma = \frac{\alpha}{\beta}$, *sofern $\beta \neq 0$ (Quotientensatz)*

Die gewonnenen Sätze lassen sich in altbewährter Weise verwenden, um den Grenzwert gebrochen rationaler Folgen zu bestimmen:

Satz 4 (gebrochen rationale Folgen) *Folgen mit Gliedern der Form* $a_n = \frac{p(n)}{q(n)}$ *mit Polynomen* $p(n) = \sum_{i=0}^{k} \alpha_i n^i$ *und* $q(n) = \sum_{j=0}^{l} \beta_j n^j$ *vom Grad k bzw. l (also mit* $\alpha_i, \beta_j \in \mathbb{R}$, $\alpha_k, \beta_l \neq 0$*) konvergieren für* $k \leq l$ *und divergieren für* $k > l$. *Im Fall* $k = l$ *ist* $\lim_{n\to\infty} a_n = \frac{\alpha_k}{\beta_k}$, *im Fall* $k < l$ *ist* $\lim_{n\to\infty} a_n = 0$.

Besonders wichtig sind auch die sogenannten *geometrischen* Folgen:

Satz 5 (geometrische Folge) *Sei* $q \in \mathbb{R}$, *dann ist die Folge mit den Gliedern* $a_n = q^n$ *konvergent gegen* 0 *für* $|q| < 1$, *konvergent gegen* 1 *für* $q = 1$ *und sonst divergent.*

In Abschnitt 4.4 werden wir auch noch rekursiv definierte Folgen behandeln, die durch ein Anfangsglied a_0 und die Forderung $a_{n+1} = T(a_n)$ für alle n mit einer gegebenen Transformation $T : \mathbb{R} \to \mathbb{R}$ (oder $T : D \to D$ mit $a_0 \in D \subseteq \mathbb{R}$) gegeben sind.

3. Reihen (unendliche Summen)

Im Schulunterricht wird man die Theorie der unendlichen Reihen nicht allzu weit treiben und sich mit den grundlegenden Definitionen (3.1) sowie einigen wichtigen Beispielen (3.2) begnügen.

3.1. Der Wert einer unendlichen Reihe

Wenn man zum Beispiel für Summanden $a_n = \frac{1}{2^n}$ suggestiv $1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \ldots = 2$ schreibt, so dürfte kein Zweifel bestehen, was damit gemeint ist, nämlich dass sich die sogenannten Partialsummen $s_0 = 1$, $s_1 = 1 + \frac{1}{2} = \frac{3}{2} = 2 - \frac{1}{2}$, $s_2 = 1 + \frac{1}{2} + \frac{1}{4} = \frac{7}{4} = 2 - \frac{1}{4}$, $s_3 = 1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} = \frac{15}{8} = 2 - \frac{1}{8}$ etc. dem Wert 2 annähern. Allgemein und etwas genauer definiert man:

Definition 3.1 *Ist eine Folge mit den Gliedern (Summanden)* a_n, $n \in \mathbb{N}$ *(sinngemäß für* $n \in \mathbb{N}^+$ *etc.), gegeben, dann sind die zugehörigen* Partialsummen s_n *rekursiv durch* $s_0 := a_0$ *und* $s_{n+1} := s_n + a_{n+1}$ *definiert. Die* a_n *heißen* Glieder *der* (unendlichen) Reihe, *für die man symbolisch auch* $\sum_{n\in\mathbb{N}} a_n$ *schreibt. Konvergiert die Folge der Partialsummen* s_n *(im Sinne des dritten Teils von Definition 2.1) gegen ein* $\alpha \in \mathbb{R}$, *so schreibt man auch*

$$\sum_{n=0}^{\infty} a_n := \lim_{n\to\infty} s_n = \alpha,$$

und nennt α *den* Wert der Reihe *und die Reihe* konvergent, *andernfalls* divergent.

In Hinblick auf den Grenzwertbegriff ist beim Übergang von Folgen zu Reihen also überhaupt nichts Neues hinzugekommen. Es wurde lediglich aus der Folge der Summanden a_n (den Gliedern der Reihe) die Folge der Partialsummen s_n gebildet und auf diese der bereits vorhandene Grenzwertbegriff angewendet. Der Unterschied zwischen Folgen und Reihen ist also weniger begrifflicher als psychologischer Natur, weil in der Theorie der Reihen vor allem solche Aussagen von Interesse sind, die unmittelbare Schlüsse von der Folge der a_n auf das Konvergenzverhalten der s_n ermöglichen.

3.2. Einfache Sachverhalte und Beispiele

Satz 6 (Zwei einfache Eigenschaften von Reihen) *Für jede Reihe* $\sum_{n=0}^{\infty} a_n$ *mit Partialsummen* s_n *gilt:*

1. *Ist* $a_n \geq 0$ *für alle n, so konvergiert die Reihe genau dann, wenn die* s_n *beschränkt sind.*
2. *Konvergiert die Reihe, so folgt* $\lim_{n\to\infty} a_n = 0$.

An die zweite Aussage von Satz 6 schließt sich die Frage an, ob es auch divergente Reihen gibt, deren Glieder eine Nullfolge bilden. Ein Beispiel dieser Art erhält man tatsächlich, wenn man die Glieder $a_0 \geq a_1 \geq a_n \geq \ldots$ so definiert, dass $a_0 = 1$, $a_1 = a_2 = \frac{1}{2}$, $a_3 = a_4 = a_5 = \frac{1}{3}$ und allgemein, n aufeinanderfolgende Glieder den Wert $\frac{1}{n}$ haben. Offensichtlich wachsen die Partialsummen dieser Reihe unbeschränkt, folglich ist die Reihe divergent trotz $\lim_{n\to\infty} a_n = 0$. Ein ehrgeizigeres Beispiel ist die harmonische Reihe:

Satz 7 (harmonische Reihe) *Die Reihe mit den Gliedern* $a_n = \frac{1}{n}$, $n \geq 1$, *divergiert.*

Das bei Weitem wichtigste Beispiel einer konvergenten Reihe ist die geometrische Reihe:

Satz 8 (geometrische Reihe) *Die Reihe mit den Gliedern $a_n = q^n$, $n \in \mathbb{N}$, konvergiert für $|q| < 1$ gegen $\frac{1}{1-q}$ und divergiert für $|q| \geq 1$. Die Partialsummen $s_n = \sum_{k=0}^{n} q^n$ haben für $q = 1$ den Wert $s_n = n+1$, in allen anderen Fällen gilt $s_n = \frac{1-q^{n+1}}{1-q}$.*

Darüber hinaus gibt es nicht viel obligatorischen Schulstoff über Reihen. Immerhin erwähnenswert ist das Phänomen, das im Riemannschen Umordnungssatz beschrieben ist: Bei manchen Reihen (nämlich den nicht absolut konvergenten) können Umordnungen der Glieder Einfluss auf das Konvergenzverhalten haben (siehe Beispiel im Anhang der Langversion). Von allgemeinem Interesse könnten auch Reihendarstellungen wie $\pi = 4(1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \ldots)$ für die Kreiszahl π und $e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$ für die Exponentialfunktion sein (siehe auch Winkler (2011/12)).

4. Funktionen und Stetigkeit

Ähnlich wie das Kapitel 2 über Folgen gliedern wir auch jenes über Funktionen und Stetigkeit: Der Abschnitt 4.1 bringt die grundlegenden Definitionen, 4.2 die wichtigsten allgemeinen Sachverhalte, 4.3 Beispiele („fast alle“ in der Schule auftretenden Funktionen sind stetig). Überdies werden im kurzen Abschnitt 4.4 auch noch rekursive Folgen unter dem Gesichtspunkt der Stetigkeit angesprochen.

4.1. Die grundlegenden Definitionen

Wir wollen die begrifflichen Präzisierungen, die uns zum Grenzwert von Folgen geführt haben, zu einer ebenso präzisen Definition des Grenzwertes $\lim_{x \to x_0} f(x)$ einer Funktion f für $x \to x_0$ modifizieren. In engster Verbindung damit steht der Begriff der Stetigkeit, der in der Mathematik einen ähnlichen Rang hat wie jener des Grenzwertes und mit diesem sehr eng verknüpft ist.

Mit der Schreibweise $\lim_{x \to x_0} f(x) = \alpha$ soll anschaulich gemeint sein: Nähert sich x dem Wert x_0 an, so auch $f(x)$ dem Wert α, d.h.: Die Differenz $|f(x) - \alpha|$ wird *beliebig* klein, sofern nur x *hinreichend* nahe bei x_0 ist. Im Vergleich mit dem Folgengrenzwert $\lim_{n \to \infty} a_n$ ist lediglich die Bedingung $n \geq n_0$, also „n hinreichend groß“ durch „$|x - x_0|$ hinreichend klein“ zu ersetzen. Die reelle Zahl x_0 übernimmt also die Rolle von ∞. Als Messgröße, die das „hinreichend klein“ kontrolliert, ist statt einer natürlichen Zahl n_0 als Folgenindex nun eine sehr kleine, aber positive reelle Zahl zu setzen, zu deren Bezeichnung sich der griechische Buchstabe δ eingebürgert hat. Soll $\lim_{x \to x_0} f(x) = \alpha$ gelten, hat also zu jedem $\varepsilon > 0$ ein $\delta > 0$ zu existieren, so dass für alle $x \neq x_0$ aus dem Definitionsbereich von f mit $|x - x_0| < \delta$ die gewünschte Ungleichung $|f(x) - \alpha| < \varepsilon$ gilt. (Dass man $x_0 = x$ nicht einbezieht, entspricht der Tatsache, dass es ja auch bei Folgen kein a_∞ gibt.) Diese Definition ist allerdings nur dann sinnvoll, wenn der Definitionsbereich D von f Punkte $x \neq x_0$ enthält, die beliebig nahe bei x_0 liegen. Ist das der Fall, so heißt x_0 ein *Häufungspunkt* von D, andernfalls ein *isolierter Punkt*. Von einer stetigen Funktion verlangt man, dass der Grenzwert $\lim_{x \to x_0} f(x) = \alpha$ auch mit $f(x_0)$ übereinstimmt. Ist $x_0 \in D$ ein isolierter Punkt von D, dann ist es sinnvoll, f als in x_0 stetig anzusehen. Die genaue Definition lautet also:

Definition 4.1 *Sei $D \subseteq \mathbb{R}$ der Definitionsbereich einer reellen Funktion $f : D \to \mathbb{R}$ und x_0 ein Häufungspunkt von D. Dann heißt $\alpha \in \mathbb{R}$* Grenzwert *oder* Limes *der Funktion f für $x \to x_0$, wenn es für alle reellen $\varepsilon > 0$ eine positive reelle Zahl δ gibt derart, dass für alle $x \neq x_0$ mit $x \in D$ aus $|x - x_0| < \delta$ stets $|f(x) - \alpha| < \varepsilon$ folgt. Als Formel:*

$$\forall \varepsilon > 0 \; \exists \delta > 0 \; \forall x \in D, x \neq x_0 : \; |x - x_0| < \delta \Rightarrow |f(x) - \alpha| < \varepsilon,$$

Kurzschreibweise: $\lim_{x \to x_0} f(x) = \alpha$ oder $f(x) \to \alpha$ für $x \to x_0$. Ist überdies $f(x_0) = \lim_{x \to x_0} f(x)$ für $x_0 \in D$, so heißt f stetig *in x_0. Ebenso heißt f stetig in x_0 für alle $x_0 \in D$, die nicht Häufungspunkt von D sind. Die Funktion f heißt stetig auf D, wenn f stetig in x ist für alle $x \in D$.*

Stetigkeit von f in x_0 lässt sich also auch charakterisieren durch die Formel:

$$\forall \varepsilon > 0 \; \exists \delta > 0 \; \forall x \in D : \; |x - x_0| < \delta \Rightarrow |f(x) - f(x_0)| < \varepsilon$$

Oft sind auch die einseitigen Varianten dieser Begriffe nützlich. Dazu betrachtet man die Einschränkungen $f_{x_0^-}$ und $f_{x_0^+}$ von f auf die Definitionsbereiche $D_{x_0^-} := D\cap\ ]-\infty, x_0[$ links bzw. $D_{x_0^+} := D\cap\ ]x_0,\infty[$ rechts von x_0. Die Grenzwerte $f(x_0^-) := \lim_{x\to x_0} f_{x_0^-}(x)$ und $f(x_0^+) := \lim_{x\to x_0} f_{x_0^+}(x)$ (sofern sie existieren) heißen die *einseitigen Grenzwerte* von f in x_0, jener von $f(x_0^-)$ der *links*-, jener von $f(x_0^+)$ der *rechtsseitige*. Stimmen diese Werte mit $f(x_0)$ überein, so heißt f *links*- bzw. *rechtsseitig* stetig in x_0.

4.2. Eigenschaften stetiger Funktionen

Die Verwandtschaft zwischen Grenzwert von Funktionen mit dem von Folgen drückt sich in folgendem sehr nützlichen Kriterium aus:

Satz 9 (Folgenstetigkeit): *Die reelle Funktion $f : D \to \mathbb{R}$, $D \subseteq \mathbb{R}$, ist genau dann stetig in $x_0 \in D$, wenn für jede Folge $(a_n)_n$ mit $a_n \in D$ aus $\lim_{n\to\infty} a_n = x_0$ die Konvergenz $\lim_{n\to\infty} f(a_n) = f(x_0)$ folgt.*

Stetigkeit lässt sich also auch durch die einprägsame Formel $f(\lim_{n\to\infty} a_n) = \lim_{n\to\infty} f(a_n)$ charakterisieren. Dieses Kriterium erleichtert viele Beweise, insbesondere jene von zwei wichtigen Sätze über stetige Funktionen, die auch im Schulunterricht keinesfalls fehlen dürfen: der Zwischenwertsatz und der Satz vom Maximum:[4]

Satz 10 *Sei $f : [a,b] \to \mathbb{R}$ stetig, $a < b$. Dann gilt:*

1. *(**Zwischenwertsatz**) Die Funktion f nimmt auf $[a,b]$ alle Werte zwischen $f(a)$ und $f(b)$ an.*
2. *(**Satz vom Maximum**) Die Funktion f nimmt auf $[a,b]$ sowohl Minimum als auch Maximum an, d.h. es gibt $x_1, x_2 \in [a,b]$ mit $f(x_1) \le f(x) \le f(x_2)$ für alle $x \in [a,b]$.*[5]

Dass auf die Stetigkeitsvoraussetzung nicht ersatzlos verzichtet werden kann, zeigt für den Zwischenwertsatz das Beispiel der Signumfunktion sgn auf $[a,b] = [-1,1]$. Sie ist durch $\mathrm{sgn}(x) := -1$ für $x < 0$, $\mathrm{sgn}(0) = 0$ und $\mathrm{sgn}(x) = 1$ für $x > 0$ definiert. Für den Satz vom Maximum kann man als Gegenbeispiel etwa die Funktion $f : [0,1] \to \mathbb{R}$ mit $f(0) := 0$ und $f(x) := 1-x$ oder, noch extremer, $f(x) := \frac{1}{x}$ für $x > 0$ nehmen. Die letzten Beispiele dienen, wenn man die Stelle $x = 0$ aus dem Definitionsbereich herausnimmt, auch zur Illustration der Voraussetzung, dass der Definitionsbereich von f abgeschlossen (genauer: sogar kompakt) sein muss. Natürlich gilt der Satz vom Maximum auch nicht für unbeschränkte Definitionsbereiche (z.B. $f(x) := x$ betrachten).

4.3. Viele Beispiele stetiger Funktionen

Die einfachsten Beispiele stetiger Funktionen sind die konstanten, wo an jeder Stelle x_0 zu beliebigem $\varepsilon > 0$ irgendein $\delta > 0$ genommen werden kann, und die identische Funktion $f(x) := x$ (hier setzt man am einfachsten $\delta = \delta(\varepsilon) = \varepsilon$). Wie Satz 11 zeigen wird, überträgt sich die Stetigkeit von diesen speziellen Beispielen auf alle daraus mittels Addition, Subtraktion und Multiplikation aufgebauten Funktionen, also auf alle Polynomfunktionen. Nimmt man auch noch die Division hinzu, so erhält man sämtliche gebrochen rationalen Funktionen, die ebenfalls nach Satz 11 auf ihrem gesamten Definitionsbereich (allfällige Pole, d.h. Nullstellen des Nenners, sind bei gebrochen rationalen Funktionen natürlich auszunehmen) stetig sind. Von vielleicht noch größerer Bedeutung für die Mathematik insgesamt ist die Stetigkeit der Verkettung stetiger Funktionen. Zusammenfassend gilt:

Satz 11 (Verknüpfung von Funktionen und Stetigkeit) *Seien $f, g : D \to \mathbb{R}$ Funktionen und $x_0 \in D$ derart, dass die Grenzwerte $\alpha := \lim_{x\to x_0} f(x)$ und $\beta := \lim_{x\to x_0} g(x)$ existieren. Außerdem möge für die Funktion $h : D_h \to \mathbb{R}$ mit $f(D) \subseteq D_h$ der Grenzwert $\gamma := \lim_{y\to\alpha} h(y)$ existieren. Dann gilt:*

[4] Beide Sätze sind aus Sicht der Topologie von besonderem Interesse. Sie besagen nämlich für den Spezialfall reeller Funktionen, dass die Eigenschaften Zusammenhang bzw. Kompaktheit unter stetigen Abbildungen bewahrt bleiben. Beides gilt allgemein auf beliebigen topologischen Räumen.

[5] Statt des abgeschlossenen Intervalls $[a,b]$ kann auch ein beliebiger kompakter Definitionsbereich $D \subseteq \mathbb{R}$ von f vorausgesetzt werden.

1. *Auch die Grenzwerte* $\lim_{x\to x_0}(f+g)(x)$, $\lim_{x\to x_0}(f-g)(x)$, $\lim_{x\to x_0}(f\cdot g)(x)$ *und, sofern* $g(x_0)\neq 0$, $\lim_{x\to x_0}(\frac{f}{g})(x)$ *(definiert wenigstens in einer Umgebung von* x_0*) existieren und stimmen mit* $\alpha+\beta$, $\alpha-\beta$, $\alpha\beta$ *bzw.* $\frac{\alpha}{\beta}$ *überein.*
2. *Sind* f *und* g *stetig in* x_0*, dann auch die Summe* $f+g$*, die Differenz* $f-g$*, das Produkt* $f\cdot g$ *und, sofern* $g(x_0)\neq 0$*, der Quotient* $\frac{f}{g}$.
3. *Sind* f *und* g *stetig auf ganz* D*, dann auch die Summe* $f+g$*, die Differenz* $f-g$*, das Produkt* $f\cdot g$ *und, sofern* $g(x)\neq 0$ *für alle* $x\in D$*, der Quotient* $\frac{f}{g}$.
4. *Auch der Grenzwert* $\lim_{x\to x_0}(h\circ f)(x)$ *existiert und hat den Wert* γ*. Ist* f *stetig in* x_0 *und* h *in* $\alpha=f(x_0)\in D_h$*, so auch* $h\circ f$ *in* x_0*. Ist* f *stetig auf ganz* D *und* h *auf ganz* D_h*, so auch* $h\circ f: D\to\mathbb{R}$ *auf ganz* D.

Stetig sind außerdem Exponential-, Logarithmus-, Potenz-, trigonometrische und viele andere „elementare" Funktionen. Der Beweis dafür muss allerdings auf die Definition dieser Funktionen zurückgreifen (siehe etwa Winkler (2011/12)), was hier zu weit führte. Das und Satz 11 zeigen, dass man, um überhaupt unstetige Funktionen zu erhalten, andersartige Konstruktionen bemühen muss, typischerweise mittels Fallunterscheidung.

Das einfachste Beispiel ist die sogenannte *Signumfunktion* $\operatorname{sgn}:\mathbb{R}\to\mathbb{R}$, die wir schon als Gegenbeispiel zum Zwischenwertsatz (erste Aussage in Satz 10) kennen gelernt haben. Ein viel extremeres Beispiel, das nicht nur an einer Stelle, sondern überall unstetig ist, ist die berühmte *Dirichletsche Sprungfunktion* $f:\mathbb{R}\to\mathbb{R}$, die an allen rationalen $x\in\mathbb{Q}$ den Wert $f(x)=1$ und an allen irrationalen Stellen $x\in\mathbb{R}\setminus\mathbb{Q}$ den Wert $f(x)=0$ annimmt. Man überzeugt sich leicht (siehe letztes Kapitel der Langversion Winkler (2018/19)), dass diese Funktion tatsächlich an jedem Punkt $x_0\in\mathbb{R}$ unstetig ist. Die Dirichletsche Sprungfunktion dient oft als Gegenbeispiel. Bei der Integration werden auch wir sie als solches verwenden.

4.4. Rekursive Folgen und Stetigkeit

Zum Abschluss des Themenbereichs Stetigkeit kommen wir noch auf rekursive Folgen zu sprechen, d.h. auf solche Folgen $\mathbf{a}=(a_n)_n$, die $a_{n+1}=T(a_n)$ für alle n erfüllen. Dabei wollen wir $T:D\to D$ als eine stetige Funktion voraussetzen, die eine Menge $D\subseteq\mathbb{R}$ auf sich selbst abbildet. Konvergiert so eine Folge gegen ein $x_0\in D$, also $x_0=\lim_{n\to\infty}a_n$, so folgt (Folgenstetigkeit!)

$$T(x_0)=T(\lim_{n\to\infty}a_n)=\lim_{n\to\infty}T(a_n)=\lim_{n\to\infty}a_{n+1}=x_0.$$

Der Grenzwert muss also ein Fixpunkt von T, d.h. Lösung der Gleichung $T(x)=x$ in der Unbekannten x sein. Ist T durch einen Formelausdruck gegeben, so fällt die Lösung dieser Gleichung oft relativ leicht. Ist der Grenzwert einer rekursiv gegebenen Folge gesucht, so ermöglicht diese Vorgangsweise die Beschränkung der Untersuchungen auf wenige Kandidaten für den gesuchten Grenzwert. Um bei gegebenem Anfangswert a_0 die Konvergenz der a_n gegen einen bestimmten Wert nachzuweisen, sind dann aber meist noch andere Überlegungen erforderlich. Ausführlicher ist das beispielsweise in Winkler (2013/14) behandelt.

5. Differentialrechnung

Ähnlich den bisherigen Kapiteln beginnen wir mit den grundlegenden Begriffen in Abschnitt 5.1, lassen in 5.2 die wichtigsten theoretischen Ergebnisse folgen und schließen mit praktischen Aspekten (nämlich den Differentiationsregeln) in Abschnitt 5.3.

5.1. Differentialquotient und Ableitung

Ähnlich wie sich der Wert einer unendlichen Reihe als Grenzwert einer Folge (nämlich der Folge der Partialsummen) deuten lässt, kann der Begriff der Ableitung einer Funktion f in einem Punkt x_0 auf den

Grenzwert einer anderen Funktion für $x \to x_0$ zurückgeführt werden. Und zwar handelt es sich um die sogenannte *Differenzenquotienten-* oder auch *Sekantensteigungsfunktion* $s_{f,x_0} : D \setminus \{x_0\} \to \mathbb{R}$, die durch

$$s_{f,x_0}(x) := \frac{f(x) - f(x_0)}{x - x_0}$$

definiert ist. Die Bezeichnung ergibt sich aus der Interpretation des Wertes $s_{f,x_0}(x)$ als Steigung jener Sekante (= schneidenden Gerade), die durch die beiden auf dem Funktionsgraphen von f liegenden Punkte $(x_0, f(x_0))$ und $(x, f(x))$ geht. Man beachte, dass s_{f,x_0} tatsächlich genau für alle $x \in D \setminus \{x_0\}$ definiert ist. Die als Differentialquotient definierte *Ableitung* von f an der Stelle x_0 stimmt, sofern sie existiert, mit der stetigen Fortsetzung von s_{f,x_0} an der Stelle x_0 überein:

Definition 5.1 *Sei $f : D \to \mathbb{R}$ und $x_0 \in D \subseteq \mathbb{R}$ derart, dass der* Differentialquotient *von f in x_0 genannte Grenzwert*

$$f'(x_0) := \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0} = \lim_{x \to x_0} s_{f,x_0}(x)$$

existiert. Dann heißt dieser Wert die Ableitung *von f an der Stelle x_0, und f heißt an der Stelle x_0* differenzierbar. *Ist f an allen Stellen $x \in D$ differenzierbar, dann heißt die Funktion $f' : D \to \mathbb{R}$, $x \mapsto f'(x)$, die* Ableitung von f, *und f heißt eine* Stammfunktion *von f'.*

Differenzierbarkeit von f in x_0 bedeutet, dass f in einer Umgebung von x_0 „sehr gut“ durch eine lineare Funktion approximiert werden kann, nämlich durch die Funktion $l(x) := f(x_0) + f'(x_0)(x - x_0)$. Dabei heißt „sehr gut“, dass nicht nur $\lim_{x \to x_0}(f(x) - l(x)) = 0$ gilt (das wäre bei stetigem f mit jeder linearen Funktion, die $l(x_0) = f(x_0)$ erfüllt, der Fall), sondern sogar stärker $\lim_{x \to x_0} \frac{f(x) - l(x)}{x - x_0} = 0$. Dieser Gesichtspunkt der linearen Approximierbarkeit lässt sich weitreichend verallgemeinern, insbesondere auf Funktionen in mehreren Variablen, und kann deshalb als paradigmatisch für die Differentialrechnung schlechthin angesehen werden. Die geometrische Interpretation bedeutet in unserem eindimensionalen Fall, dass l die Tangente an den Graphen von f im Punkt $(x_0, f(x_0))$ ist. Erwartungsgemäß hat der Differentialquotient einer linearen Funktion f der Gestalt $f(x) := kx + d$ für alle x_0 den Wert

$$f'(x_0) = \lim_{x \to x_0} \frac{(kx + d) - (kx_0 + d)}{x - x_0} = \lim_{x \to x_0} \frac{k(x - x_0)}{x - x_0} = k.$$

Ähnlich wie bei der Stetigkeit einseitige Grenzwerte, so können nun auch *einseitige*, d.h. *links-* und *rechtsseitige Ableitungen* $f'(x_0^-)$ und $f'(x_0^+)$ von Interesse sein. Sie sind durch $f'(x_0^-) := \lim_{x \to x_0^-} s_{f,x_0}(x)$ und $f'(x_0^+) := \lim_{x \to x_0^+} s_{f,x_0}(x)$ als einseitige Grenzwerte des Differenzenquotienten definiert. Häufig setzt man in der Definition von $f'(x_0)$ voraus, dass x_0 *innerer Punkt* von D ist, was bedeutet, dass es ein $\delta > 0$ gibt derart, dass das gesamte Intervall $U :=]x_0 - \delta, x_0 + \delta[$ (so ein U heißt *Umgebung* von x_0) in D enthalten ist. In diesem Fall wäre zum Beispiel eine Funktion $f : [a, b] \to \mathbb{R}$ in den Randpunkten höchstens einseitig differenzierbar.

5.2. Eigenschaften differenzierbarer Funktionen

Als Veranschaulichung des Unterschieds zwischen stetig und differenzierbar bietet sich die auf ganz $\mathbb{R}$ stetige Betragsfunktion $f : \mathbb{R} \to \mathbb{R}$, $x \mapsto |x|$, an der Stelle $x_0 = 0$ an. An dieser Stelle existieren zwar die einseitigen Ableitungen $f'(0^-) = -1$ und $f'(0^+) = 1$. Diese beiden Werte sind aber verschieden, weshalb kein gemeinsamer Grenzwert $f'(0)$ existiert, f also nicht differenzierbar an der Stelle $x_0 = 0$ ist. Anschaulich spiegelt sich das in dem „Eck“ des Graphen der Betragsfunktion an der Stelle x_0 wider. Stetige Funktionen müssen also nicht differenzierbar sein. Es gibt sogar Funktionen, die überall stetig und nirgends differenzierbar sind. Ihre Konstruktion erfordert allerdings beträchtlichen Aufwand, der den Rahmen des Schulunterrichts sprengen würde. In Übereinstimmung mit der Anschauung, dass ein Funktionsgraph dort, wo eine Tangente existiert, keine Sprünge haben kann, gilt aber umgekehrt:

Satz 12 (differenzierbar impliziert stetig): *Ist $f : D \to \mathbb{R}$ an der Stelle $x_0 \in D \subseteq \mathbb{R}$ differenzierbar, so auch stetig.*

Eine der wichtigsten Anwendungen der Differentialrechnung im Schulunterricht sind Extremwertaufgaben. Die Definition eines Extremums bzw. einer Extremstelle braucht den Begriff der Ableitung nicht und sollte deshalb auch allgemein gefasst werden:

Definition 5.2 *Sei $f: D \to \mathbb{R}$ und x_0, dann heißt $x_0 \in D$* (globale) Extremstelle *und $f(x_0)$* (globales) Extremum *von f in x_0, wenn $f(x_0) \geq f(x)$ für alle $x \in D$ oder $f(x_0) \leq f(x)$ für alle $x \in D$ gilt. Im ersten Fall spricht man von einem* (globalen) Maximum*, im zweiten Fall von einem* (globalen) Minimum. *Man spricht von einem* lokalen *Extremum $f(x_0)$ bzw. einer* lokalen *Extremstelle $x_0 \in D$ von f (Maximum oder Minimum), wenn es ein $\delta > 0$ gibt derart, dass die Einschränkung von f auf die Menge $D \cap]x_0 - \delta, x_0 + \delta[$ (Umgebung von x_0) in x_0 ein Extremum hat.*

Klarerweise ist jedes globale Extremum auch ein lokales, während die Umkehrung nicht gilt. Die prominente Rolle der Ableitung bei der Bestimmung (lokaler und somit auch globaler) Extrema ergibt sich aus folgendem Satz.

Satz 13 (Extremum und Ableitung) *Hat $f: D \to \mathbb{R}$ im inneren Punkt x_0 von D ein Extremum und ist f in x_0 differenzierbar, so folgt $f'(x_0) = 0$.*

So wirkungsvoll dieser Satz auch ist, so dürfen nicht die Voraussetzungen „innerer Punkt" und „differenzierbar" übersehen werden. Außerdem ist, selbst wenn diese Voraussetzungen erfüllt sind, die Bedingung $f'(x_0) = 0$ notwendig, aber nicht hinreichend für das Vorliegen einer Extremstelle in x_0. Man kann untersuchen, ob die zweiten (und eventuell höheren) Ableitungen > 0 oder < 0 sind. Häufig ist aber die Anwendung des Satzes vom Maximum (zweite Aussage in Satz 10) effizienter. Weil das Thema „Extremwertaufgaben" einen eigenen Artikel rechtfertigen würde, begnügen wir uns hier mit der kurzen Erläuterung einer typischen Situation: Ist $f: [a,b] \to \mathbb{R}$ differenzierbar und gibt es z.B. nur zwei Stellen $x_0, x_1 \in D$ mit $f'(x_0) = f'(x_1) = 0$, so kommen als Extrema nur die vier Werte $f(a), f(b), f(x_0)$ und $f(x_1)$ in Frage. Diese Funktionswerte auszuwerten, um sie zu vergleichen, kommt man ohnedies nicht umhin. Der größte von ihnen ist das Maximum, der kleinste das Minimum von f auf $[a,b]$. Es besteht also keine Notwendigkeit, $f''(x_0)$ und $f''(x_1)$ auszurechnen.

Eine für die Theorie generell und insbesondere für die Berechnung von Integralen wichtige Folgerung aus den bisherigen Ergebnissen ist der Mittelwertsatz der Differentialrechnung. Anschaulich besagt er, dass es für eine auf einem Intervall $[a,b]$ differenzierbare Funktion f eine Tangente in einem geeigneten Punkt $(x_0, f(x_0))$ mit $a < x_0 < b$ gibt, die zur Verbindung der beiden Punkte $(a, f(a))$ und $(b, f(b))$ parallel ist.

Satz 14 *(**Mittelwertsatz**) Sei $f: [a,b] \to \mathbb{R}$ $(a < b)$ stetig und auf $]a,b[$ differenzierbar. Dann gibt es ein x_0 mit $a < x_0 < b$ und $f'(x_0) = \frac{f(b)-f(a)}{b-a}$ oder, äquivalent, $f(b) = f(a) + f'(x_0)(b-a)$.*

5.3. Differentiationsregeln

Ein wesentlicher Grund für die Macht der Differentialrechnung liegt darin, dass es einen „vollständigen Werkzeugkasten" an einfachen Regeln gibt, der die Berechnung der Ableitung von allen Funktionen ermöglicht, die in „regulärer" Weise aus „elementaren Funktionen" aufgebaut sind. Dieser Werkzeugkasten enthält zunächst die Ableitungsregeln für Polynome und „elementare" Funktionen wie $\exp' = \exp$, $\ln'(x) = \frac{1}{x}$, $\sin' = \cos$, $\cos' = -\sin$, $(x^\alpha)' = \alpha x^{\alpha-1}$ mit beliebigem $\alpha \in \mathbb{R}$. (Für eine ausführlichere Diskussion solcher Funktionen sei auf Winkler (2011/12) verwiesen.) Vor allem aber gibt es Ableitungsregeln, die klären, wie sich Differentiation mit den vier Grundrechnungsarten, der Verkettung und der Umkehrung von Funktionen verträgt. In weitgehender Analogie zu den entsprechenden Gesetzen sowohl für die Konvergenz von Folgen als auch für die Stetigkeit von Funktionen gilt nämlich:

Satz 15 (Differentiationsregeln) *Seien die Funktionen $f, g: D \to \mathbb{R}$ differenzierbar in $x_0 \in D$. Außerdem sei die Funktion $h: D_h \to \mathbb{R}$ mit $f(D) \subseteq D_h$ differenzierbar in $f(x_0)$. Dann sind auch die Funktionen $f \pm g$, $f \cdot g$, $\frac{f}{g}$ (sofern $g(x) \neq 0$ für alle $x \in D$) und $h \circ f$ in x_0 differenzierbar. Überdies gilt:*

1. $(f \pm g)'(x_0) = f'(x_0) \pm g'(x_0)$ *(*Summenregel*)*

2. $(fg)'(x_0) = f(x_0)g'(x_0) + f'(x_0)g(x_0)$ (Produktregel)
3. $(h \circ f)'(x_0) = h'(f(x_0))f'(x_0)$ (Kettenregel)
4. $\left(\frac{f}{g}\right)'(x_0) = \frac{f'(x_0)g(x_0) - f(x_0)g'(x_0)}{g(x_0)^2}$ (Quotientenregel)
5. *Ist f in einer Umgebung von x_0 stetig und umkehrbar, außerdem $f'(x_0) \neq 0$, so ist die Umkehrfunktion $f^{(-1)}$ in $f(x_0)$ differenzierbar mit $\left(f^{(-1)}\right)'(f(x_0)) = \frac{1}{f'(x_0)}$.* (Ableitungsregel für Umkehrfunktionen)

Insbesondere sind Summe, Differenz, Produkt, Verkettung und Quotient (wo immer definiert) differenzierbarer Funktionen ebenfalls differenzierbar, außerdem die Umkehrfunktion einer umkehrbaren differenzierbaren Funktion mit Ableitung $\neq 0$.

Ganz ähnlich wie bei Stetigkeit lässt sich also auch für Differenzierbarkeit und Ableitung sagen: Alle Funktionen, die sich aus elementaren differenzierbaren Funktionen mittels Grundrechnungsarten und Verkettung aufbauen lassen, sind selbst differenzierbar. Darüber hinaus lassen sich ihre Ableitungen mit Hilfe der Regeln aus Satz 15 berechnen.

6. Integralrechnung

Wir beginnen in Abschnitt 6.1 mit der Definition des Riemann-Integrals mittels Ober- und Untersummen und arbeiten in 6.2 den Grenzwertcharakter dieses Begriffs heraus. Als Schlusspunkt wird in 6.3 die Brücke zwischen Differential- und Integralrechnung geschlagen.

6.1. Definition des Integrals mittels Ober- und Untersummen

Bekanntlich lässt sich das Integral $\int_a^b f(x)\,dx$ einer reellen Funktion $f : D \to \mathbb{R}$ mit $[a,b] \subseteq D$ auf dem Intervall $[a,b]$, $a \leq b$, für $f \geq 0$ als die Fläche der Menge $M := \{(x,y) \in \mathbb{R}^2 : \; a \leq x \leq b, \; 0 \leq y \leq f(x)\}$ interpretieren. Nimmt f auch negative Werte an, so sind Flächenstücke in diesen Bereichen mit negativem Vorzeichen zu versehen. Wir wollen im Folgenden voraussetzen, dass f beschränkt ist.

Zur Präzisierung des Integralbegriffs betrachtet man Zerlegungen $Z = \{a = x_0 \leq x_1 \leq \ldots \leq b = x_n\}$ von $[a,b]$ in Teilintervalle $[x_{i-1}, x_i]$, $i = 1, \ldots, n$.[6] Jedes der Teilintegrale $\int_{x_{i-1}}^{x_i} f(x)\,dx$ sollte einen Wert haben, der zwischen $c_i(x_i - x_{i-1})$ und $C_i(x_i - x_{i-1})$ liegt mit $c_i = \inf\{f(x) : \; x_{i-1} \leq x \leq x_i\}$ und $C_i = \sup\{f(x) : x_{i-1} \leq x \leq x_i\}$.[7] Entsprechend sollte $\int_a^b f(x)\,dx$ zwischen *Untersumme $U(f,Z)$* und *Obersumme $O(f,Z)$* bezüglich der Zerlegung Z liegen:

$$U(f,Z) := \sum_{i=1}^{n} (x_i - x_{i-1})c_i \leq \int_a^b f(x)\,dx \leq O(f,Z) := \sum_{i=1}^{n} (x_i - x_{i-1})C_i$$

Man hofft, dass für hinreichend feine Zerlegungen Unter- und Obersumme einander beliebig nahe kommen und definiert daher:

Definition 6.1 *Die beschränkte Funktion $f : [a,b] \to \mathbb{R}$ heißt* (Riemann-)integrierbar *und $\alpha \in \mathbb{R}$ heißt das* bestimmte Integral *von f (dem sogenannten* Integranden*) auf $[a,b]$ (dem sogenannten* Integrationsbereich*), symbolisch $\alpha = \int_a^b f(x)\,dx$, wenn es zu jedem $\varepsilon > 0$ eine Zerlegung Z von $[a,b]$ gibt mit*

$$\alpha - \varepsilon < U(f,Z) \leq \alpha \leq O(f,Z) < \alpha + \varepsilon.$$

Ist beispielsweise $f(x) = x$ und $[a,b] = [0,1]$, so ergibt das Integral $\int_a^b f(x)\,dx$ als Wert α das Flächenmaß des Dreiecks mit den Eckpunkten $(0,0)$, $(1,0)$ und $(0,1)$, also $\int_0^1 x\,dx = \frac{1}{2}$. Man kann das unschwer nachrechnen (siehe letztes Kapitel der Langversion Winkler (2018/19)).

[6] Manchmal ist es praktisch, auch die Möglichkeit $x_{i-1} = x_i$ zu erfassen. Deshalb wäre es genauer, Z nicht als Menge der x_i, sondern als Vektor zu betrachten. Das schafft jedoch terminologische und notationelle Komplikationen, weshalb wir der einfachen Lesbarkeit halber darauf verzichten und Z schlicht als Menge der Unterteilungspunkte definieren.

[7] Das Symbol $\inf$ steht für *Infimum*, das ist die größte untere Schranke, $\sup$ für *Supremum*, die kleinste obere Schranke.

A priori wäre es denkbar, dass verschiedene Zerlegungen zu verschiedenen Werten des Integrals führen könnten. Dem ist aber nicht so:

Satz 16 (Untersumme stets kleiner als Obersumme) *Sei $f : [a,b] \to \mathbb{R}$ beschränkt und seien Z_1 und Z_2 beliebige Zerlegungen von $[a,b]$. Dann gilt:*

1. *Ist $Z_1 \subseteq Z_2$ (in diesem Fall heißt Z_2 eine* Verfeinerung *von Z_1), so gilt*

$$U(f,Z_1) \leq U(f,Z_2) \leq O(f,Z_2) \leq O(f,Z_1).$$

2. *In jedem Fall gilt $U(f,Z_1) \leq O(f,Z_2)$.*

Also sind alle Untersummen durch jede Obersumme nach oben und alle Obersummen durch jede Untersumme nach unten beschränkt. Im Falle einer Riemann-integrierbaren Funktion, wo Unter- und Obersummen einander beliebig nahe kommen können, bedeutet das: Das Integral stimmt sowohl mit dem Supremum sämtlicher Untersummen als auch mit dem Infimum sämtlicher Obersummen, genommen jeweils über alle Zerlegungen Z von $[a,b]$, überein.

Das klassische Beispiel einer *nicht* Riemann-integrierbaren Funktion ist die Dirichletsche Sprungfunktion f, die wir schon in Abschnitt 4.3 kennengelernt haben. Für jede Zerlegung Z gibt es zwischen zwei Unterteilungspunkten $x_{i-1} < x_i$ sowohl rationale als auch irrationale Stellen mit Werten 1 bzw. 0. Deshalb erhält man als Untersumme immer den Wert $U(f,Z) = 0$, als Obersumme zum Beispiel auf dem Integrationsintervall $[0,1]$ jedoch $O(f,Z) = 1$, unabhängig von Z. Somit ist die Dirichletsche Sprungfunktion auf $[0,1]$ (analog auf jedem anderen Intervall $[a,b]$ mit $a < b$) nicht Riemann-integrierbar.

6.2. Das Integral als Grenzwert

Schreibt man Definition 6.1 quantorenlogisch um, so erhält man die Formel

$$\forall \varepsilon > 0\ \exists Z: \ \alpha - \varepsilon < U(f,Z) \leq \alpha \leq O(f,Z) < \alpha + \varepsilon$$

mit nur zwei Quantoren. Will man analog zu Folgengrenzwert etc. auch das Integral als Grenzwert deuten mit einem dritten Quantor, nämlich einem zusätzlichem Allquantor, so bietet sich an,

$$\forall \varepsilon > 0\ \exists Z\ \forall Z' \supseteq Z: \ \alpha - \varepsilon < U(f,Z') \leq \alpha \leq O(f,Z') < \alpha + \varepsilon$$

zu fordern. Nach den Überlegungen des vorangegangenen Abschnitts sind tatsächlich beide Formeln äquivalent. Noch interessanter aber ist es, den Begriff der *Feinheit* $\|Z\|$ einer Zerlegung Z einzubeziehen. $\|Z\|$ ist für eine Zerlegung $Z = \{x_0, \ldots, x_n\}$ mit $a = x_0 \leq x_1 \leq \ldots \leq x_n = b$ als das Maximum der Längen der Teilintervalle definiert, also $\|Z\| := \max\{x_i - x_{i-1} : \ i = 1, \ldots, n\}$. Dann gilt nämlich:

Satz 17 (Integral als Grenzwert) *Eine beschränkte Funktion $f : [a,b] \to \mathbb{R}$ ist genau dann Riemann-integrierbar, wenn es zu jedem $\varepsilon > 0$ ein $\delta > 0$ gibt mit $O(f,Z) - U(f,Z) < \varepsilon$ für alle Zerlegungen Z von $[a,b]$ mit $\|Z\| < \delta$, als Formel:*

$$\forall \varepsilon > 0\ \exists \delta > 0\ \forall Z: \ \|Z\| < \delta \ \Rightarrow \ O(f,Z) - U(f,Z) < \varepsilon$$

Da das Integral $\int_a^b f(x)\,dx$ sicher zwischen $U(f,Z)$ und $O(f,Z)$ liegt, lässt sich also formulieren: Für *beliebig* vorgegebenes $\varepsilon > 0$ kann bei *hinreichend* kleinem $\delta > 0$ eine Approximationsgüte ε garantiert werden – ganz analog zum Grenzwert von Folgen, Reihen, Funktionen und Ableitung.

Auch die Schreibweise

$$\lim_{\|Z\| \to 0} U(f,Z) = \lim_{\|Z\| \to 0} O(f,Z) = \int_a^b f(x)\,dx$$

bietet sich an. Wir werden im abschließenden Resümee (Kapitel 7) nochmals auf die Analogie zwischen den nunmehr fünf Ausprägungen des Grenzwertbegriffs zurückkommen. Zuvor folgt aber noch die krönenden Verschmelzung von Differential- und Integral- zur Infinitesimalrechnung.

6.3. Die Brücke zwischen Differential- und Integralrechnung

Über die Definition des Integrals hinaus wollen wir wissen, ob das weiter oben erwähnte sehr einfache Beispiel $f(x) = x$ ein glücklicher Sonderfall ist, oder ob die Riemann-integrierbaren Funktionen eine große Klasse bilden. Der wichtigste Satz hierzu lautet:

Satz 18 (Stetige Funktionen sind integrierbar) *Jede stetige Funktion $f : [a,b] \to \mathbb{R}$ ist Riemann-integrierbar.*

Stetigkeit ist für Riemann-Integrierbarkeit also hinreichend, notwendig ist sie nicht. Denn z.B. hat eine Abänderung des Funktionswertes an einer einzigen oder an endlich vielen Stellen keinen Einfluss auf Integrierbarkeit und Wert des Integrals. Auch stückweise stetige Funktionen, also solche mit endlich vielen Unstetigkeitsstellen, sind integrierbar. Das gilt sogar für gewisse Funktionen mit unendlich vielen Unstetigkeitsstellen. Zum Beispiel ist jede monotone Funktion Riemann-integrierbar.[8]

Wir kommen nun zum Hauptsatz der Differential- und Integralrechnung, einem Höhepunkt nicht nur der Analysis, sondern der gesamten Mathematik. Er stellt die Verbindung zwischen Differential- und Integralrechnung her. Indem er Differentiation und Integration als (in einem bestimmten Sinn) zueinander inverse Operationen ausweist, liefert er darüber hinaus auch die wichtigste Methode zur exakten Berechnung von Integralen.

Satz 19 (Hauptsatz der Differential- und Integralrechnung) *Ist $f : [a,b] \to \mathbb{R}$ stetig, so definiert $F(x_0) := \int_a^{x_0} f(x)\,dx$ eine differenzierbare Funktion mit Ableitung $F' = f$.*

Die Berechnung von Integralen mittels Stammfunktionen beruht neben dem Hauptsatz 19 auch auf folgendem Satz, in dessen Beweis der Mittelwertsatz 14 die entscheidende Rolle spielt.

Satz 20 (Stammfunktionen und additive Konstante) *Sind F_1 und F_2 irgendwelche Stammfunktionen der stetigen Funktion $f : [a,b] \to \mathbb{R}$, dann gibt es eine konstante Zahl $c \in \mathbb{R}$ mit $F_1(x) = F_2(x) + c$ für alle $x \in [a,b]$. Die Beziehung $F(b) - F(a) = \int_a^b f(x)\,dx$ gilt für jede beliebige Stammfunktion F von f.*

Für die Berechnung des Integrals $\int_a^b f(x)\,dx$ einer stetigen Funktion f genügt es nach Satz 20, *irgendeine* Stammfunktion F von f zu finden. Deshalb verwendet man das Integralsymbol $\int$ ohne Integrationsgrenzen zur Bezeichnung von Stammfunktionen. Will man ihre Mehrdeutigkeit hervorheben, so schreibt man statt $\int f(x)\,dx$ auch $\int f(x)\,dx + c$ und meint damit die Menge aller Stammfunktionen. Der Formelausdruck $\int f(x)\,dx$ bezeichnet also eine Funktion oder eine Menge von Funktionen, genannt das *unbestimmte Integral* von f. Im Gegensatz dazu bezeichnet das *bestimmte* Integral $\int_a^b f(x)\,dx$ nach Definition 6.1 eine reelle Zahl. Zur Berechnung eines bestimmten Integrals ermittelt man in der Praxis also zunächst ein unbestimmtes Integral, d.h. de facto eine beliebige Stammfunktion F und bildet die Differenz $F(b) - F(a)$ der Funktionswerte an den Integrationsgrenzen.

Zur rechnerischen Ermittlung von Stammfunktionen bietet es sich an, nach Umkehrungen der Differentiationsregeln aus Satz 15 zu suchen. Leider gelingt das nur in gewissen Fällen, wie bei Polynom- und Potenzfunktionen, exp, sin, cos und vereinzelten weiteren Beispielen. Schon die Umkehrung von Produktregel und Kettenregel liefern in Gestalt von partieller Integration und Substitutionsregel Methoden, die nur in manchen Fällen zum Ziel führen. Für manche Funktionen wie $\frac{\sin x}{x}$ und $e^{-\frac{x^2}{2}}$ (Normalverteilung!) kann man sogar zeigen, dass Stammfunktionen wegen des Hauptsatzes 19 zwar existieren, aber nicht elementar dargestellt werden können. Deshalb hat man auch andere, insbesondere numerische Methoden zur Integration entwickelt. Es ist eine Ermessensfrage, wie viel Zeit man im Unterricht für Integrationsmethoden investieren möchte. Im Gegensatz zu den universell wirksamen Ableitungesregeln aus Satz 15 wird jede Liste von Integrationsregeln aber notgedrungen immer nur Stückwerk bleiben.

[8] Das allgemeine Kriterium hierzu, dessen Beweis allerdings den Rahmen des Schulunterrichts deutlich sprengt, besagt, dass eine Funktion $f : [a,b] \to \mathbb{R}$ genau dann Riemann-integrierbar ist, wenn sie beschränkt ist und die Menge U der Unstetigkeitsstellen eine sogenannte *Lebesgue-Nullmenge*, was explizit bedeutet: Für jedes $\varepsilon > 0$ gibt es eine Folge von Intervallen I_n, $n \in \mathbb{N}$, deren Vereinigung V die Menge U enthält und deren Längen λ_n Glieder einer konvergenten Reihe mit $\sum_{n\in\mathbb{N}} \lambda_n < \varepsilon$ sind. Integrierbarkeit hat also tatsächlich viel mit Stetigkeit zu tun.

7. Resümee

Der Begriff des Grenzwertes zieht sich mit geringfügigen Variationen wie ein Rückgrat durch die gesamte Analysis. Immer wieder geht es dabei um die Approximation eines gesuchten Wertes (bei uns: einer reellen Zahl) durch Näherungen, die in gewissem Sinn leichter (z.B. mit endlichen Mitteln) fassbar sind. Begrifflich ist der gesuchte Wert trotzdem exakt definiert, wenn die Einhaltung einer *beliebig* klein vorgegebenen positiven Fehlerschranke (traditionell meist mit ε bezeichnet) garantiert werden kann unter der Bedingung, dass bei der Bestimmung der Näherung ein jeweils angemessener Parameter *hinreichend* gut gewählt wird. Im Falle des Folgengrenzwertes und beim Wert einer unendlichen Reihe ist dieser Parameter der Folgenindex n, der hinreichend groß gesetzt werden muss. Im Falle des Grenzwertes einer Funktion und ihrer Stetigkeit sowie bei ihrer Ableitung an einer Stelle x_0 muss ein (positiver) Abstand $|x - x_0|$ hinreichend klein gewählt werden. Im Fall des Integrals muss einer Zerlegung Z hinreichend fein sein. Die gemeinsame logische Struktur besteht also in der Abfolge dreier logischer Quantoren, verbal: Für alle $\varepsilon > 0$ (Allquantor) gibt es einen gewissen Parameterwert (Existenzquantor), so dass für diesen und alle besseren Wahlen des Parameters (Allquantor) die Fehlerschranke ε eingehalten wird. Auf die drei Quantoren folgt daher eine Implikation der Struktur: „Wenn der Parameter gut genug ist, dann ist der Fehler $< \varepsilon$“. In Formeln für Folge, Reihe, Funktion, Ableitung und Integral:

$$
\begin{array}{rlccl}
\lim\limits_{n\to\infty} a_n = \alpha: & \forall \varepsilon > 0\ \exists n_0 & \forall n: & n \geq n_0 \Rightarrow & |a_n - \alpha| < \varepsilon \\
\sum\limits_{n=0}^{\infty} a_n = \alpha: & \forall \varepsilon > 0\ \exists n_0 & \forall n: & n \geq n_0 \Rightarrow & \left|\sum\limits_{k=0}^{n} a_n - \alpha\right| < \varepsilon \\
\lim\limits_{x\to x_0} f(x) = \alpha: & \forall \varepsilon > 0\ \exists \delta > 0 & \forall x \neq x_0: & |x - x_0| < \delta \Rightarrow & |f(x) - \alpha| < \varepsilon \\
f'(x_0) = \alpha: & \forall \varepsilon > 0\ \exists \delta > 0 & \forall x \neq x_0: & |x - x_0| < \delta \Rightarrow & \left|\dfrac{f(x) - f(x_0)}{x - x_0} - \alpha\right| < \varepsilon \\
\int_a^b f(x)\,dx = \alpha: & \forall \varepsilon > 0\ \exists \delta > 0 & \forall Z: & \|Z\| < \delta \Rightarrow & O(f,Z) - U(f,Z) < \varepsilon
\end{array}
$$

Ein sinnvoller Mathematikunterricht wird diese Analogien nutzen und so mehreren Anliegen gleichzeitig gerecht werden:

- Ein zwar notwendiger, für sich alleine aber unbefriedigender, weil nebuloser „intuitiver Grenzwertbegriff“ wird durch klare Begriffsbildungen ergänzt.
- Durch die Identifikation des begrifflichen Kerns insbesondere mit symbolischer Unterstützung durch logische Quantoren werden fünf Stoffkapitel zum Grenzwertbegriff (Folgen, Reihen, Funktionen und Stetigkeit, Differential- und Integralrechnung) vereinheitlicht und somit besser überschaubar.
- Mit klaren Begriffen entsteht automatisch auch ein klarer Blick dafür, welche über die Definitionen hinausgehenden Inhalte relevant sind.
- Der Zusammenhang zwischen all diesen Inhalten fördert das Bewusstsein für die Mathematik als sinnvollem Ganzen, das weit mehr ist als ein unübersichtliches Chaos von Einzelheiten.
- Es bieten sich unzählige, in diesem Artikel aus Platzgründen gar nicht im Einzelnen hervorgehobene Gelegenheiten, jene vielfältigen Aspekte der Mathematik anzusprechen, die es laut Lehrplan in der AHS-Oberstufe (siehe Lehrplan (2016)) zu vermitteln gilt.

8. Anhang: Beweise

Es folgen nun die Beweise der Sätze und einiger weiterer Behauptungen aus dem Hauptteil. Dieser Anhang ist nur in der längeren Online-, nicht aber in der Druckausgabe des Artikels inkludiert. Im Unterricht wird man wahrscheinlich nicht alle Beweise in der vorliegenden Form behandeln, nur gewisse exemplarisch auswählen und sich bei den übrigen mit anschaulichen Erklärungen begnügen. Unbedingt sollte man an geeigneter Stelle die Grundlagen aus 8.1 thematisieren. Die weiteren Abschnitte entsprechen den fünf Hauptkapiteln des Textes und bringen die Beweise zu Folgen (8.2), Reihen (8.3), Funktionen und Stetigkeit (8.4), zur Differential- (8.5) und zur Integralrechnung (8.6).

8.1. Unverzichtbare Grundlagen

Die entscheidende Eigenschaft des Systems $\mathbb{R}$ der reellen Zahlen, die es von $\mathbb{Q}$, dem der rationalen Zahlen, unterscheidet und worauf die Analysis fußt, ist die **Vollständigkeit**. Wir werden sie bereits im Beweis von zwei Aussagen von Satz 1 verwenden: dass jede monotone und beschränkte Folge konvergiert, und dass jede beschränkte Folge einen Häufungspunkt hat. Viele weitere Beweise werden auf diese beiden Resultate zurückgreifen. Die Vollständigkeit von $\mathbb{R}$ lässt sich auf vielfältige Weise charakterisieren. Am anschaulichsten dürfte die Schnitteigenschaft sein, die sich sehr überzeugend anhand von $\mathbb{R}$ als Zahlengerade beschreiben lässt: Zerlegt man $\mathbb{R}$ in zwei nichtleere und disjunkte Mengen, einen „linken" Teil A und einen „rechten" Teil B, also mit $a < b$ für alle $a \in A$ und $b \in B$, so gibt es einen eindeutigen „Schnittpunkt" auf der Zahlengerade, also eine reelle Zahl x (genannt die *Schnittzahl*) mit $a \leq x \leq b$ für alle $a \in A$ und $b \in B$. Diese Eigenschaft kann als axiomatische Forderung an $\mathbb{R}$ aufgefasst werden. Aber auch als beweisbarer Satz ergibt sie sich recht unmittelbar, wenn man die mengentheoretische Konstruktion von $\mathbb{R}$ nach Dedekind mittels sogenannter Schnitte zugrunde legt. Etwas mehr Arbeit muss man investieren, wenn man die reellen Zahlen nach Cantor mittels Cauchyfolgen oder (weniger elegant) als unendliche dekadische Ziffernfolgen o.ä. einführt. Praktisch ist die Vollständigkeit auch als Supremums- bzw. Infimumseigenschaft zu fassen: Jede nichtleere nach oben/unten beschränkte Teilmenge T von $\mathbb{R}$ hat ein Supremum/Infimum (= eine kleinste obere bzw. größte untere Schranke). Diese Formulierung folgt schnell aus der Schnitteigenschaft: Wenn man $\mathbb{R}$ in die Mengen A und B jener Elemente zerlegt, die obere/untere Schranke von T sind bzw. nicht sind. Die Schnittzahl ist dann das gesuchte Supremum/Infimum. Viele weitere äquivalente Eigenschaften finden sich in Winkler (2008/09).

Eine ähnlich fundamentale Rolle wie die Vollständigkeit für $\mathbb{R}$ spielt das **Induktionsprinzip** für $\mathbb{N}$: Eine Teilmenge von $\mathbb{N}$, in der 0 und mit jedem n auch $n+1$ liegen, enthält bereits alle natürlichen Zahlen. Induktion wirkt auch in die Analysis und den in diesem Artikel behandelten Stoff hinein, besonders wenn Folgen im Spiel sind. Im Zusammenhang mit dem Grenzwertbegriff steht Induktion aber etwas abseits des Hauptinteresses. Für Grundlagenfragen zu $\mathbb{N}$ verweise ich deshalb auf Winkler (2007/08).

Als technisch unerlässliches Hilfsmittel, das wir unzählige Male brauchen werden, sei abschließend auch noch die **Dreiecksungleichung** $|x+y| \leq |x|+|y|$ hervorgehoben. Für reelle x, y ist sie fast trivial: In den beiden Fällen $x, y \geq 0$ und $x, y \leq 0$ gilt sogar Gleichheit, im gemischten Fall offenbar strikte Ungleichung $<$. Wenn wir die Dreiecksungleichung verwenden, werden wir darauf nicht jedes Mal ausdrücklich hinweisen.

8.2. Beweise zu Kapitel 2 (Folgen)

Beweis von Satz 1 (Eigenschaften von Folgen):

1. Jeder Grenzwert einer Folge ist Häufungspunkt dieser Folge:
   Sei $\alpha = \lim_{n\to\infty} a_n$ und $\varepsilon > 0$ beliebig, so gibt es nach der Definition des Grenzwertes ein n_1 mit $|a_n - \alpha| < \varepsilon$ für alle $n \geq n_1$. Insbesondere gibt es dann zu jedem n_0 ein $n \geq n_0$ mit dieser Eigenschaft, z.B. $n := \max\{n_0, n_1\}$.

2. Eine Folge mit mehreren Häufungspunkten ist divergent:

Die Folge mit den Gliedern a_n konvergiere gegen α, habe einen weiteren Häufungspunkt β, und $\varepsilon > 0$ sei beliebig vorgegeben. Unter diesen Annahmen genügt es, die Ungleichung $|\alpha - \beta| < \varepsilon$ zu beweisen. Das gelingt so: Nach der Grenzwertdefinition gibt es ein n_0 mit $|a_n - \alpha| < \frac{\varepsilon}{2}$ für alle $n \geq n_0$. Weil β ein Häufungspunkt ist, gibt es ein $n \geq n_0$ mit $|a_n - \beta| < \frac{\varepsilon}{2}$. Mit diesem n gilt tatsächlich

$$|\alpha - \beta| = |\alpha - a_n + a_n - \beta| \leq |\alpha - a_n| + |a_n - \beta| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

3. Der Grenzwert einer konvergenten Folge ist eindeutig bestimmt:
Gäbe es zwei Grenzwerte $\alpha \neq \beta$, so müssten nach der ersten Aussage beide Häufungspunkte sein, was nach der zweiten Aussage Konvergenz ausschließt, Widerspruch.

4. Jede konvergente Folge ist beschränkt:
Ist $\lim_{n\to\infty} a_n = \alpha \in \mathbb{R}$, so gibt es ein n_0 für $\varepsilon = 1$, also mit $|a_n - \alpha| < 1$ für alle $n \geq n_0$. Für diese n gilt also $|a_n| \leq |\alpha| + 1$. Nehmen wir für S die größte der endlich vielen Zahlen $|a_0|, \ldots, |a_{n_0} - 1|$ und $|\alpha| + 1$, dann gilt $|a_n| \leq S$ sogar für alle n, was die Beschränktheit der Folge beweist.

5. Jede Folge, die monoton und beschränkt ist, konvergiert:
Der Beweis ist von besonderem Interesse, weil er die Vollständigkeit von $\mathbb{R}$ verwendet. Wir führen ihn für eine monoton wachsende Folge $\mathbf{a} = (a_n)_n$; das Argument für monoton fallend ist symmetrisch. Laut Voraussetzung ist die Menge M der oberen Schranken der Folge nicht leer, außerdem nach unten beschränkt, z.B. durch a_0. Die Vollständigkeit von $\mathbb{R}$ garantiert, dass M ein Infimum (eine größte untere Schranke) hat, das wir mit α bezeichnen. Dieses α erfüllt offenbar selbst $a_n \leq \alpha$ für alle $n \in \mathbb{N}$. Wir wollen $\lim_{n\to\infty} a_n = \alpha$ beweisen. Dazu sei $\varepsilon > 0$ vorgegeben. Gäbe es kein n_0 mit $a_{n_0} > \alpha - \varepsilon$, so wäre auch $\alpha' := \alpha - \varepsilon < \alpha$ eine obere Schranke von $\mathbf{a}$, Widerspruch. Also gibt es ein n_0 mit $\alpha - \varepsilon < a_{n_0} \leq \alpha$. Wegen der Monotonie gilt $a_{n_0} \leq a_n$ für alle $n \geq n_0$. Das bedeutet $\alpha - \varepsilon < a_{n_0} \leq a_n \leq \alpha$ und somit $|a_n - \alpha| < \varepsilon$ für alle $n \geq n_0$. Weil $\varepsilon > 0$ beliebig war, beweist das $\lim_{n\to\infty} a_n = \alpha$.

6. Gilt $|a_n| \leq S$ für alle $n \in \mathbb{N}$ und ein $S \in \mathbb{R}$, dann gibt es einen Häufungspunkt x_0 der Folge mit den Gliedern a_n (Satz von Bolzano-Weierstraß 1):
Wir definieren die Folge von Intervallen $I_k = [s_k, t_k]$ rekursiv durch fortgesetzte Halbierung und Auswahl einer jeweils geeigneten Hälfte wie folgt: Sei $I_0 := [s_0, t_0] = [-S, S]$, also $a_n \in I_0$ für alle n. Wir nehmen induktiv an, dass I_k bereits definiert ist und unendlich viele Folgenglieder a_n enthält (genauer: dass $a_n \in I_k$ für unendlich viele n gilt). Wir setzen $I_{k+1} = [s_{k+1}, t_{k+1}] := [s_k, \frac{s_k+t_k}{2}]$, sofern auch dieses Intervall unendlich viele Folgenglieder enthält. Andernfalls enthält $I_{k+1} = [s_{k+1}, t_{k+1}] := [\frac{s_k+t_k}{2}, t_k]$ unendlich viele Folgenglieder. Wegen $s_0 \leq s_1 \leq \ldots \leq t_0$ bilden die s_k eine monotone und beschränkte Folge, konvergieren wegen Aussage 5 daher gegen einen Grenzwert $\alpha := \lim_{n\to\infty} s_n$. Analog konvergieren die t_n (fallend) gegen einen Grenzwert $\beta := \lim_{n\to\infty} t_n$. Wegen $s_n \leq t_n$ für alle n gilt $\alpha \leq \beta$. Wegen der fortgesetzten Halbierung der Intervalle I_k konvergiert ihre Länge $t_k - s_k$ gegen $\lim_{k\to\infty}(t_k - s_k) = \lim_{k\to\infty} \frac{2S}{2^k} = 0$ (Vorgriff auf die geometrische Folge aus Satz 5), was nur für $\alpha = \beta$ möglich ist. Dann ist $x_0 := \alpha = \beta$ der gesuchte Häufungspunkt: Jede Umgebung $]x_0 - \varepsilon, x_0 + \varepsilon[$, $\varepsilon > 0$ beliebig, enthält ein geeignetes I_k (k hinreichend groß wählen) und somit unendlich viele Folgenglieder. Also gibt es zu jedem n_0 ein $n \geq n_0$ mit $|a_n - \alpha| < \varepsilon$.

7. Ist x_0 Häufungspunkt der Folge mit den Gliedern a_n, dann gibt es eine gegen x_0 konvergente Teilfolge der a_n (Satz von Bolzano Weierstraß 2):
Wir müssen eine Folge von Indizes $n_1 < n_2 < \ldots$ angeben mit $\lim_{k\to\infty} a_{n_k} = x_0$. Nach Voraussetzung liegen für jedes $k = 1, 2, \ldots$ unendlich viele Glieder in der Umgebung $]x_0 - \frac{1}{k}, x_0 + \frac{1}{k}[$, erfüllen also $|a_n - x_0| < \frac{1}{k}$. Wir können daher tatsächlich n_1 so wählen, dass $|a_{n_1} - x_0| < 1$ gilt, und außerdem, rekursiv, $n_{k+1} > n_k$ mit $|a_{n_{k+1}} - x_0| < \frac{1}{k+1}$ für alle k, also $\lim_{k\to\infty} a_{n_k} = x_0$. $\square$

Beweis von Satz 2 (Typische Beispiele von Folgen): Es sollten die folgenden Beobachtungen genügen. Zusammen mit Satz 1 ergeben sich daraus alle Behauptungen von Satz 2.

Die Folge mit den Gliedern $a_n = \frac{1}{n}$ ist (wegen $a_n = \frac{1}{n} > \frac{1}{n+1} = a_{n+1}$ für alle n) streng monoton fallend und (wegen $|\frac{1}{n}| \leq 1$ für alle n) beschränkt. Sie konvergiert gegen $\alpha = 0$, weil für beliebiges $\varepsilon > 0$ der Kehrwert $\frac{1}{\varepsilon} \in \mathbb{R}$ gebildet werden kann und jedes $n_0 > \frac{1}{\varepsilon} \in \mathbb{R}$ die in der Definition geforderte Eigenschaft hat, nämlich dass $|a_n - 0| = \left|\frac{1}{n}\right| < \varepsilon$ für alle $n \geq n_0$ gilt. Somit ist 0 auch (der einzige) Häufungspunkt.

Für die Folge mit den Gliedern $a_n = (-1)^n \frac{1}{n}$ erfolgt die Begründung von Beschränktheit, Konvergenz und Häufungspunkt wie für $a_n = \frac{1}{n}$. Wegen $a_1 < a_2 > a_3$ ist diese Folge aber weder monoton wachsend noch fallend.

Die Folge mit den Gliedern $a_n = (-1)^n$ ist klarerweise beschränkt ($|a_n| \leq 1$ für alle n) und nicht monoton ($a_0 > a_1 < a_2$). Offenbar hat sie die beiden Häufungspunkte 1 und -1, ist folglich nicht konvergent.

Die Aussagen betreffend die drei Folgen mit den Gliedern $a_n = n$, $a_n = n + (-1)^n n$ (Häufungspunkt 0) und $a_n = (-1)^n n$ sind offensichtlich. $\square$

Beweis von Satz 3 (Grenzwertsätze): Seien $(a_n)_n$ und $(b_n)_n$ konvergente Folgen mit $\lim_{n\to\infty} a_n = \alpha$ und $\lim_{n\to\infty} b_n = \beta$. Drei Aussagen sind zu beweisen:

1. Für $c_n = a_n \pm b_n$ gilt $\lim_{n\to\infty} c_n = \alpha \pm \beta$ (Summen- und Differenzsatz):
   Für ein beliebig vorgegebenes $\varepsilon > 0$ ist ein Index n_0 zu finden, so dass $|(a_n \pm b_n) - (\alpha \pm \beta)| < \varepsilon$ für alle $n \geq n_0$ gilt. Die folgende Abschätzung bringt die vorausgesetzten Grenzwertbeziehungen ins Spiel, die sicherstellen, dass die Differenzen $|a_n - \alpha|$ und $|b_n - \beta|$ für große n klein werden:
   $$|(a_n \pm b_n) - (\alpha \pm \beta)| = |(a_n - \alpha) \pm (b_n - \beta)| \leq |a_n - \alpha| + |b_n - \beta|.$$
   Die Summe rechts wird sicher dann kleiner ε, wenn z.B. jeder der beiden Summanden $< \frac{\varepsilon}{2}$ wird. Nach Voraussetzung gibt es Indizes n_1 und n_2, so dass das der Fall ist, sofern $n \geq n_1$ (für den ersten) bzw. $n \geq n_2$ (für den zweiten Summanden). Nehmen wir als n_0 das Maximum $\max\{n_1, n_2\}$, so gilt also tatsächlich für alle $n \geq n_0$ die gewünschte Beziehung $|(a_n \pm b_n) - (\alpha \pm \beta)| < \varepsilon$.

2. Für $c_n = a_n b_n$ gilt $\lim_{n\to\infty} c_n = \alpha\beta$ (Produktsatz); ist überdies $\beta = 0$, so genügt Beschränktheit (statt Konvergenz) der a_n, um $\lim_{n\to\infty} a_n b_n = 0$ zu garantieren:
   Für die Multiplikation schätzt man den entsprechenden Term geschickter durch
   $$|a_n b_n - \alpha\beta| = |a_n b_n - a_n\beta + a_n\beta - \alpha\beta| = |a_n(b_n - \beta) + (a_n - \alpha)\beta| \leq |a_n| \cdot |b_n - \beta| + |a_n - \alpha| \cdot |\beta|$$
   ab. Wieder können beide Summanden rechts $< \frac{\varepsilon}{2}$ gemacht werden, sofern $n \geq n_0$ für ein geeignetes n_0 gewährleistet ist. Um so ein n_0 zu finden, verwendet man, dass konvergente Folgen beschränkt sind, folglich ein positives $A \in \mathbb{R}$ existiert mit $|a_n| \leq A$ für alle n. Man wählt n_1 so, dass $|b_n - \beta| < \frac{\varepsilon}{2A}$ für alle $n \geq n_1$ und n_2 so, dass $|a_n - \alpha| < \frac{\varepsilon}{2\beta}$ für alle $n \geq n_2$ (im Fall $\beta = 0$ kann n_2 sogar beliebig gesetzt werden). Wieder folgt für alle $n \geq n_0 := \max\{n_1, n_2\}$ die geforderte Beziehung $|a_n b_n - \alpha\beta| < A\frac{\varepsilon}{2A} + \frac{\varepsilon}{2\beta}\beta = \varepsilon$. Der Beweis zeigt auch: Ist $\beta = 0$, so genügt Beschränktheit (statt Konvergenz) der a_n, um $|a_n b_n - \alpha\beta| \leq |a_n| \cdot |b_n - \beta| < A\frac{\varepsilon}{2A} = \frac{\varepsilon}{2} < \varepsilon$ und somit $\lim_{n\to\infty} a_n b_n = 0$ zu garantieren.

3. Ist $\beta \neq 0$ und $c_n = \frac{a_n}{b_n}$, so gilt $\lim_{n\to\infty} c_n = \frac{\alpha}{\beta}$ (Quotientensatz):
   Wegen $\lim_{n\to\infty} b_n = \beta \neq 0$ kann $|b_n| < \frac{|\beta|}{2}$ und erst recht $b_n = 0$ nur für endlich viele n gelten, die wir für die behauptete Grenzwertaussage ignorieren dürfen. Es genügt daher, diese unter der Annahme $|b_n| \geq \frac{|\beta|}{2}$ für alle n zu beweisen. Wir betrachten zunächst den einfachen Fall $a_n = 1$, also $c_n = \frac{1}{b_n}$ für alle n und beweisen $\lim_{n\to\infty} \frac{1}{b_n} = \frac{1}{\beta}$. Daraus folgt dann wegen $\frac{a_n}{b_n} = a_n \frac{1}{b_n}$ aus dem bereits bewiesenen Multiplikationssatz die allgemeinere Beziehung $\lim_{n\to\infty} a_n b_n = \frac{\alpha}{\beta}$. Im Beweis von $\lim_{n\to\infty} \frac{1}{b_n} = \frac{1}{\beta}$ lautet die entscheidende Umformung
   $$\left|\frac{1}{b_n} - \frac{1}{\beta}\right| = \left|\frac{\beta}{b_n\beta} - \frac{b_n}{b_n\beta}\right| = |\beta - b_n| \cdot \frac{1}{|b_n\beta|}.$$

Wegen $|b_n| \geq \frac{|\beta|}{2}$ sind die $\frac{1}{|b_n\beta|}$ beschränkt. Die Voraussetzung besagt $\lim_{n\to\infty} |\beta - b_n| = 0$. Wegen des Zusatzes am Ende des Beweises zum Multiplikationssatz folgt damit, dass der abzuschätzende Ausdruck $|\beta - b_n| \cdot \frac{1}{|b_n\beta|}$ für hinreichend große n kleiner als ein beliebig vorgegebenes $\varepsilon > 0$ wird, was zu zeigen war. $\square$

Beweis von Satz 4 (gebrochen rationale Folgen):
Eine Folge mit den Gliedern $a_n = \frac{p(n)}{q(n)}$ (wobei $p(n) = \sum_{i=0}^{k} \alpha_i n^i$, $q(n) = \sum_{j=0}^{l} \beta_j n^j$, $\alpha_i, \beta_j \in \mathbb{R}$ und $\alpha_k, \beta_l \neq 0$) ist auf Konvergenz zu untersuchen:
1. Fall, $k < l$: In diesem Fall liefert Division von Zähler und Nenner durch n^l die Darstellung $a_n = \frac{p_0(n)}{q_0(n)}$ mit $p_0(n) = \sum_{i=0}^{k} \alpha_i n^{i-l}$ und $q_0(n) = \sum_{j=0}^{l} \beta_j n^{j-l}$. Die Grenzwertsätze (Satz 3) zeigen $\lim_{n\to\infty} p_0(n) = 0$ und $\lim_{n\to\infty} q_0(n) = \beta_j$, also $\lim_{n\to\infty} a_n = \frac{0}{\beta_l} = 0$.
2. Fall, $k = l$: Wie 1.Fall, mit dem Unterschied, dass $\lim_{n\to\infty} p_0(n) = \alpha_k$, folglich $\lim_{n\to\infty} a_n = \frac{\alpha_k}{\beta_k}$.
3. Fall, $k > l$. Dann fällt die Folge $b_n := a_n^{-1}$ unter den 1.Fall, ist also eine Nullfolge. Folglich divergiert die Folge der a_n. $\square$

Beweis von Satz 5 (geometrische Folge):
Wir beginnen mit dem Fall $q > 1$, also $q = 1 + r$ mit einem Rest $r > 0$, und beweisen die sogenannte *Bernoullische Ungleichung*: $q^n \geq 1 + nr$ für alle $n \in \mathbb{N}$. Für $n = 0, 1$ gilt die Ungleichung offensichtlich. Ist $q^n \geq 1 + nr$, so pflanzt sich das wegen $q^{n+1} = q^n q \geq (1+nr)(1+r) = 1 + nr + r + nr^2 > 1 + (n+1)r$ auf $n + 1$ fort, gilt also (Induktionsprinzip!) für alle $n \in \mathbb{N}$. Weil nr für $n \to \infty$ unbeschränkt wächst (das ist die sogenannte *archimedische Eigenschaft* von $\mathbb{R}$, siehe auch Winkler (2008/09)), tut das auch $q^n \geq 1 + nr$, womit die Divergenz für $q > 1$ und erst recht für $q < -1$ bewiesen ist. Indem man die Kehrwerte betrachtet, erhält man $\lim_{n\to\infty} q^n = 0$ für $|q| < 1$. Die verbleibenden Fälle $q = 1$ und $q = -1$ sind trivial. $\square$

8.3. Beweise zu Kapitel 3 (Reihen)

Beweis von Satz 6 (Zwei einfache Eigenschaften von Reihen):
Die folgenden beiden Aussagen sind zu beweisen:

1. Sind die Glieder a_n einer Reihe nichtnegativ, so konvergiert die Reihe genau dann, wenn die Partialsummen s_n beschränkt sind:
   Gilt $a_n \geq 0$ für alle n, so bilden die Partialsummen wegen $s_{n+1} = s_n + a_{n+1} \geq s_n$ eine monoton wachsende Folge. Für solche Folgen ist nach den entsprechenden Aussagen in Satz 1 Beschränktheit äquivalent mit Konvergenz.

2. Konvergiert die Reihe mit den Gliedern a_n, so folgt $\lim_{n\to\infty} a_n = 0$:
   Konvergieren die Partialsummen s_n gegen ein $\alpha \in \mathbb{R}$, so auch die Folge der s_{n+1}, folglich ist unter Verwendung des Grenzwertsatzes für die Subtraktion (siehe Satz 3)

$$\lim_{n\to\infty} a_n = \lim_{n\to\infty} a_{n+1} = \lim_{n\to\infty} (s_{n+1} - s_n) = \lim_{n\to\infty} s_{n+1} - \lim_{n\to\infty} s_n = \alpha - \alpha = 0. \square$$

Beweis von Satz 7 (Divergenz der harmonischen Reihe):
Da die Glieder $a_n = \frac{1}{n}$ positiv sind, bilden die Partialsummen s_n eine monoton wachsende Folge. Folglich genügt es laut der ersten Aussage von Satz 6, die Unbeschränktheit der s_n nachzuweisen. Dazu betrachten wir die Teilfolge aller s_{2^n}, mit $n = 1, 2, \ldots$. Wir fassen die Summanden a_n in Gruppen von Teilsummen t_n zusammen mit $t_1 := a_1 + a_2 = 1 + \frac{1}{2} \geq \frac{1}{2}$ und

$$t_n := \sum_{i=2^{n-1}+1}^{2^n} a_i = \sum_{i=2^{n-1}+1}^{2^n} \frac{1}{i} \quad \geq \quad 2^{n-1} \frac{1}{2^n} = \frac{1}{2}$$

für alle $n \geq 2$. Es folgt $s_{2^n} = \sum_{k=1}^{n} t_k \geq \frac{n}{2}$, also sind die s_n unbeschränkt. $\square$

Beweis von Satz 8 (geometrische Reihe):
Wir beweisen für $q \neq 1$ die Formel $s_n := \sum_{k=0}^{n} q^n = \frac{1-q^{n+1}}{1-q}$. Ist das bewiesen, so folgt mit Satz 5 für $|q| < 1$ auch die behauptete Formel $\sum_{k=0}^{\infty} q^n = \lim_{n\to\infty} s_n = \frac{1}{1-q}$ für den Grenzwert. Tatsächlich liefert die einfache Rechnung

$$(1-q)s_n = s_n - qs_n = \sum_{i=0}^{n} q^i - \sum_{i=0}^{n} q^{i+1} = \sum_{i=0}^{n} q^i - \sum_{i=1}^{n+1} q^i = 1 - q^{n+1}$$

die behauptete Formel $s_n = \frac{1-q^{n+1}}{1-q}$ für die Partialsummen. Alle anderen Aussagen im Satz sind offensichtlich. □

Beispiel zum Riemannschen Umordnungssatz:
Unter einer absolut konvergenten Reihe versteht man eine Reihe mit Gliedern a_n derart, dass (stärker) auch die Reihe mit den Gliedern $|a_n|$ konvergiert. Eine Reihe die konvergiert, nicht aber absolut konvergiert, nennt man bedingt konvergent. Im reellen Fall bilden die positiven Glieder einer bedingt konvergenten Reihe für sich eine divergente Reihe, ebenso wie die negativen. Ein typisches Beispiel ist die Reihe mit den Gliedern $1, -1, \frac{1}{2}, -\frac{1}{2}, \frac{1}{3}, -\frac{1}{3}, \ldots$, genauer: $a_{2k} = \frac{1}{k+1}$ und $a_{2k+1} = -\frac{1}{k+1}$ für alle $k \in \mathbb{N}$. Dann gilt für die Partialsummen $s_{2k} = a_{2k} = \frac{1}{k+1}$ und $s_{2k+1} = 0$. Folglich konvergiert die Reihe mit den Gliedern a_n gegen den Wert

$$\sum_{n=0}^{\infty} a_n = \lim_{n\to\infty} s_n = 0.$$

Weil die harmonische Reihe divergiert, ist diese Reihe aber sicher nicht absolut konvergent. An Beispiel dieser Reihe wollen wir nun den sogenannten Riemannschen Umordnungssatz illustrieren, wonach für eine bedingt konvergente Reihe die Glieder so umgeordnet werden können, dass die resultierende Reihe gegen einen beliebig vorgegebenen Wert α konvergiert. Die Argumentation lässt sich ohne Schwierigkeiten zu einem Beweis des allgemeinen Satzes modifizieren, Komplikationen betreffen fast nur die Notation.

Wir geben uns also eine beliebige reelle Zahl $\alpha \in \mathbb{R}$ vor und wollen uns überlegen, dass durch geeignete Umordnung der Glieder a_n zu Gliedern b_n eine Reihe $\sum_{n=0}^{\infty} b_n$ entsteht, die gegen α konvergiert. Für $\alpha \geq 0$ kann man wie folgt vorgehen (für $\alpha < 0$ entsprechend). Zunächst wählt man $b_0 = a_0 = 1$, $b_1 = a_2 = \frac{1}{2}, \ldots,$ $b_{n_0} = a_{2n_0} = \frac{1}{n_0+1} > 0$ so lange, bis erstmals $\sum_{i=0}^{n_0} b_i > \alpha$ gilt. Weil die harmonische Reihe divergiert, ist das sicher nach endlich vielen Schritten der Fall. Sodann setzt man $b_{n_0+1} = a_1 = -1$, $b_{n_0+2} = a_3 = -\frac{1}{2}$, $\ldots$, $b_{n_1} = a_{2k_1+1} = -\frac{1}{k_1+1} < 0$ so lange, bis $\sum_{i=0}^{n_1} b_i < \alpha$. Wieder ist das wegen der Konvergenz der harmonischen Reihe irgendwann der Fall. (In unserem Beispiel wird man nur ein oder höchstens zwei Glieder brauchen, es wird also $n_1 = n_0 + 1$ oder $n_1 = n_0 + 2$ sein.) Das Verfahren lässt sich auf diese Weise fortsetzen: Man nimmt der Reihe nach die nächsten noch unverbrauchten positiven Glieder der ursprünglichen Reihe als b_n, bis die resultierende Partialsumme der neuen Reihe den angepeilten Wert α übersteigt, dann wieder negative Glieder, bis die Partialsumme unter α fällt etc. Weil die a_n eine Nullfolge bilden, werden die Schwankungen der Partialsummen der neuen Reihe um den Wert α beliebig klein, also

$$\sum_{n=0}^{\infty} b_n = \alpha.$$

Durch geringfügige Modifikation dieses Verfahrens lässt sich auch Divergenz gegen ∞ (oder auch $-\infty$) erzwingen, indem man die positiven a_n gegenüber den negativen sozusagen privilegiert und früher an die Reihe kommen lässt.

8.4. Beweise zu Kapitel 4 (Funktionen und Stetigkeit)

Beweis von Satz 9 (Folgenstetigkeit): Wir haben die Äquivalenz folgender beiden Aussagen zu beweisen:

1. Die reelle Funktion $f : D \to \mathbb{R}$, $D \subseteq \mathbb{R}$, ist stetig in $x_0 \in D$.

2. Für jede Folge $(a_n)_n$ mit $a_n \in D$ und $\lim_{n\to\infty} = x_0$ folgt $\lim_{n\to\infty} f(a_n) = f(x_0)$.

$1 \Rightarrow 2$: Sei f stetig in x_0 und $(a_n)_{n\in\mathbb{N}}$ eine Folge mit $a_n \in D$ und $\lim_{n\to\infty} a_n = x_0$. Um $\lim_{n\to\infty} f(a_n) = f(x_0)$ zu beweisen, sei $\varepsilon > 0$ beliebig vorgegeben. Wegen der Stetigkeit von f in x_0 gibt es erstens ein $\delta > 0$ derart, dass aus $x \in D$ und $|x - x_0| < \delta$ stets $|f(x) - f(x_0)| < \varepsilon$ folgt. Wegen $\lim_{n\to\infty} a_n = x_0$ gibt es zweitens ein $n_0 \in \mathbb{N}$ mit $|a_n - x_0| < \delta$ für alle $n \geq n_0$. Beide Informationen zusammen zeigen $|f(a_n) - f(x_0)| < \varepsilon$ für alle $n \geq n_0$, was zu zeigen war.
$2 \Rightarrow 1$: Für die umgekehrte Implikation sei Folgenstetigkeit von f in x_0 vorausgesetzt. Wir gehen indirekt vor und nehmen an, f sei nicht stetig in x_0. Dann gibt es ein $\varepsilon > 0$, für das kein entsprechendes $\delta > 0$ existiert, genauer: Zu jedem $\delta > 0$ gibt es wenigstens ein $x_\delta \in D$ mit $|x_\delta - x_0| < \delta$ und $|f(x_\delta) - f(x_0)| \geq \varepsilon$. Wir betrachten die Folge der $a_n := x_{\frac{1}{n+1}}$, $n \in \mathbb{N}$. Nach Konstruktion gilt $\lim_{n\to\infty} a_n = x_0$. Wegen $|f(a_n) - f(x_0)| \geq \varepsilon$ für alle n konvergieren die $f(a_n)$ aber sicher nicht gegen $f(x_0)$, was im Widerspruch zur vorausgesetzten Folgenstetigkeit von f in x_0 steht. □

Beweis von Satz 10 (Zwischenwertsatz und Satz vom Maximum): Sei $f : [a,b] \to \mathbb{R}$ stetig. Wir haben zu beweisen:

1. **Zwischenwertsatz:** Ist y ein beliebiger Wert zwischen $f(a)$ und $f(b)$, dann gibt es ein $x \in [a,b]$ mit $f(x) = y$:
   Ist $f(a) = f(b)$, so auch $f(a) = y = f(b)$, und es kann $x = a$ oder $x = b$ gewählt werden. Wir nehmen nun $f(a) < f(b)$ an, der Fall $f(a) > f(b)$ ist analog zu behandeln. Wir geben uns einen beliebigen Wert y mit $f(a) < y < f(b)$ vor und haben ein $x \in [a,b]$ mit $f(x) = y$ zu finden. Ähnlich wie im Beweis des Satzes von Bolzano-Weierstraß 1 (sechste Aussage in Satz 1) definieren wir mit $I_0 := [a,b] = [a_0, b_0]$ beginnend eine Folge von ineinander enthaltenen Intervallen $I_n = [a_n, b_n]$, von denen jeweils $I_{n+1} = [a_{n+1}, b_{n+1}]$ die linke oder rechte Hälfte von I_n ist, und zwar so, dass stets $f(a_n) \leq y \leq f(b_n)$ gilt. Sowohl die a_n als auch die b_n konvergieren monoton (wachsend bzw. fallend) gegen einen gemeinsamen Wert $x \in [a,b]$ (dem einzigen allen I_n gemeinsamen Punkt):

   $$x = \lim_{n\to\infty} a_n = \lim_{n\to\infty} b_n$$

   Wegen der Stetigkeit von f und dem Kriterium der Folgenstetigkeit 9 folgt daraus

   $$f(x) = \lim_{n\to\infty} f(a_n) = \lim_{n\to\infty} f(b_n).$$

   Nach Konstruktion gilt $f(a_n) \leq y \leq f(b_n)$ für alle n. All das ist nur möglich für $f(x) = y$.

2. **Satz vom Maximum:** Es gibt $x_1, x_2 \in [a,b]$ mit $f(x_1) \leq f(x) \leq f(x_2)$ für alle $x \in [a,b]$:
   Man könnte ganz ähnlich wie beim Zwischenwertsatz mittels Intervallhalbierung vorgehen. Zur Abwechslung sei aber eine andere wichtige Beweismethode vorgestellt, nämlich direkter mit Hilfe des Satzes von Bolzano-Weierstraß (sechste und siebente Aussage in Satz 1). Aus Symmetriegründen genügt es, die Aussage für das Maximum zu beweisen. Wir gehen in zwei Schritten vor. Im ersten beweisen wir die obere Beschränktheit der Wertemenge $W := f([a,b]) = \{f(x) : a \leq x \leq b\}$. Im zweiten Schritt zeigen wir, dass das Supremum sogar ein Maximum ist, d.h. dass es ein $x \in [a,b]$ mit $f(x) = \sup W$ gibt. In beiden Schritten verwenden wir den Satz von Bolzano-Weierstraß (Satz 1) und Folgenstetigkeit (Satz 9).

   Erster Schritt: Wir nehmen indirekt an, W wäre nicht nach oben beschränkt. Dann gäbe es zu jedem $n \in \mathbb{N}$ ein $a_n \in [a,b]$ mit $f(a_n) > n$. Nach Bolzano-Weierstraß hat die Folge der a_n eine Teilfolge, die gegen ein $x \in [a,b]$ konvergiert. Wegen der Folgenstetigkeit müssten die Funktionswerte $f(a_n)$ für a_n aus dieser Teilfolge gegen $f(x)$ konvergieren, was wegen deren Unbeschränktheit der $f(a_n) > n$ aber unmöglich ist, Widerspruch.

   Zweiter Schritt: Nach dem ersten Schritt ist W beschränkt, also gibt es (Vollständigkeit von $\mathbb{R}$) ein Supremum $s := \sup W$. Nach Definition des Supremums (kleinste obere Schranke) gibt es zu jedem $n \geq 1$ ein $a_n \in [a,b]$ mit $f(a_n) > s - \frac{1}{n}$. Wieder liefert Bolzano Weierstraß eine gegen ein $x \in [a,b]$

konvergente Teilfolge, deren Funktionswerte wegen der Stetigkeit von f gegen $f(x)$ konvergieren. Es folgt $f(x) \geq s - \frac{1}{n}$ für alle n, gleichzeitig $f(x) \leq s$, also gilt $f(x) = s = \sup W$, und $x_2 := x$ erfüllt die Behauptung. □

Beweis von Satz 11 (Verknüpfung von Funktionen und Stetigkeit): Zu beweisen ist, dass sich Grenzwerte von zwei Funktionen in natürlicher Weise auf deren Summe, Differenz, Produkt, Quotient und Verkettung übertragen. Der Beweis gelingt durchwegs sehr schnell mittels Folgenstetigkeit (Satz 9) und den entsprechenden Aussagen für den Grenzwert von Folgen aus Satz 3. Nun zu den vier Aussagen:

1. Für die vier Operationen $+, -, \cdot$ und : schreiben wir vereinheitlichend $*$. Wir haben

   $$\lim_{x\to x_0} (f * g)(x) = \lim_{x\to x_0} f(x) * \lim_{x\to x_0} g(x) = \alpha * \beta$$

   zu beweisen. Sei daher $(a_n)_n$ irgendeine Folge mit $a_n \in D$ und $\lim_{n\to\infty} a_n = x_0$. Nach Voraussetzung existieren die Grenzwerte $\lim_{n\to\infty} f(a_n) = \alpha$ und $\lim_{n\to\infty} g(a_n) = \beta$, wegen Satz 3 also auch

   $$\lim_{n\to\infty} (f * g)(a_n) = \lim_{n\to\infty} f(a_n) * g(a_n) = \left(\lim_{n\to\infty} f(a_n)\right) * \left(\lim_{n\to\infty} g(a_n)\right) = \alpha * \beta,$$

   was zu zeigen war.

2. Sind f und g stetig in x_0, so auch $f * g$ (Notation wie im ersten Teil):
   Stetigkeit in x_0 bedeutet $f(x_0) = \lim_{x\to x_0} f(x) = \alpha$ und $g(x_0) = \lim_{x\to x_0} g(x) = \beta$. Damit folgt aus dem ersten Teil

   $$\lim_{x\to x_0} (f * g)(x) = \alpha * \beta = f(x_0) * g(x_0) = (f * g)(x_0),$$

   also die Stetigkeit von $f * g$ in x_0.

3. Summe, Differenz, Produkt und Quotient (sofern definiert) von zwei stetigen Funktionen ist wieder stetig:
   Sind f und g auf ganz D stetig, so gilt die Argumentation im zweiten Teil nicht nur für ein bestimmtes, sondern für alle $x_0 \in D$.

4. Wir wenden die Überlegungen aus den ersten drei Teilen auf $h \circ f$ statt auf $f * g$ an:
   Aus $\lim_{n\to\infty} a_n = x_0$ mit $a_n \in D$ folgt wegen der Stetigkeit von f in x_0 zunächst $\lim_{n\to\infty} f(a_n) = \lim_{x\to x_0} f(x) = f(x_0) = \alpha$, folglich wegen der Stetigkeit von h in $\alpha = f(x_0)$ auch $\lim_{n\to\infty} h(f(a_n)) = \lim_{y\to\alpha} h(y) = \gamma$, die erste Behauptung des vierten Teils. Sind f und h stetig in x_0 bzw. α, so gilt $f(x_0) = \alpha$ und $h(\alpha) = \gamma$, also auch $(h \circ f)(x_0) = h(f(x_0)) = h(\alpha) = \gamma$. Somit ist $h \circ f$ stetig in x_0, die zweite Behauptung. Sind f und h auf ganz D bzw. ganz D_h stetig, so ist wegen der zweiten Behauptung auch $h \circ f$ auf ganz D stetig, die dritte und letzte Behauptung. □

Beweis, dass die Dirichletsche Sprungfunktion überall unstetig ist: Sei also $f(x) = 1$ für $x \in \mathbb{Q}$ und $f(x) = 0$ für $x \in \mathbb{R} \setminus \mathbb{Q}$ und $x_0 \in \mathbb{R}$ beliebig. Man setze in der Stetigkeitsdefinition $\varepsilon = 1$. Zu beliebig vorgegebenem $\delta > 0$ gibt es im offenen Intervall $]x_0 - \delta, x_0 + \delta[$ sowohl rationale als auch irrationale Punkte x. Ist x_0 rational, so wähle man x irrational und umgekehrt. Dann ist $|x - x_0| < \delta$, jedoch $|f(x) - f(x_0)| = 1 \geq \varepsilon$, also nicht $< \varepsilon$. Das zeigt die Unstetigkeit von f in x_0. □

8.5. Beweise zu Kapitel 5 (Differentialrechnung)

Beweis von Satz 12 (differenzierbar impliziert stetig): Sei $f : D \to \mathbb{R}$ an der Stelle $x_0 \in D \subseteq \mathbb{R}$ differenzierbar. Wir müssen zeigen, dass f in x_0 auch stetig ist.
Für $x \in D$ mit $x \neq x_0$ wenden wir auf

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

den Grenzübergang $x \to x_0$ an. Weil f an der Stelle x_0 differenzierbar ist, konvergiert der erste Term rechts gegen $f'(x_0)$, der zweite gegen 0. Wegen der Aussage über die Multiplikation in Satz 11 gilt also

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

bzw., gleichwertig, $\lim_{x\to x_0} f(x) = f(x_0)$, was die Stetigkeit von f in x_0 zeigt. $\square$

Beweis von Satz 13 (Extremum und Ableitung): Sei $f: D \to \mathbb{R}$ im inneren Punkt x_0 von D differenzierbar und x_0 eine lokale Extremstelle von f. Zu zeigen ist $f'(x_0) = 0$.
Wir nehmen indirekt zunächst $f'(x_0) = \lim_{x\to x_0}\frac{f(x)-f(x_0)}{x-x_0} > 0$ an. Weil x_0 innerer Punkt von D ist, müsste es dann beliebig nahe rechts von x_0 Werte von x geben mit $f(x) > f(x_0)$, analog beliebig nahe links von x_0 Werte von x mit $f(x) < f(x_0)$. Dann wäre x_0 aber weder lokales Minimum, noch lokales Maximum von x. Genauso führt die Annahme $f'(x_0) < 0$ zu einem Widerspruch. Also folgt $f'(x_0) = 0$. $\square$

Beweis von Satz 14 (Mittelwertsatz): Sei $f: [a,b] \to \mathbb{R}$ $(a < b)$ stetig und auf $]a,b[$ differenzierbar. Wir haben ein x_0 mit $a < x_0 < b$ und $f'(x_0) = \frac{f(b)-f(a)}{b-a}$ zu finden.
Im ersten Schritt beweisen wir die Behauptung für den Spezialfall $f(a) = f(b)$ (den sogenannten *Satz von Rolle*): Als stetige Funktion nimmt f Maximum und Minimum auf $[a,b]$ an. Stimmen diese beiden überein, so ist f konstant, also gilt $f'(x_0) = 0$ sogar für alle $x_0 \in]a,b[$, und wir können irgendein $x_0 \in]a,b[$ nehmen. Andernfalls muss an mindestens einer der beiden Extremstellen x_0 (Maximum oder Minimum) $f(x_0) \neq f(a) = f(b)$ gelten. Also ist $x_0 \neq a,b$ innerer Punkt, und es gilt nach Satz 13 $f'(x_0) = 0 = \frac{f(b)-f(a)}{b-a}$.

Im allgemeinen Fall betrachtet man die Funktion $g(x) := f(x) - \frac{f(b)-f(a)}{b-a}(x-a)$, also $g(x) = f(x) - (kx+d)$ mit $k = \frac{f(b)-f(a)}{b-a}$. Mit f ist auch g auf $[a,b]$ stetig und auf $]a,b[$ differenzierbar. Überdies gilt $g(b) = f(b) - \frac{f(b)-f(a)}{b-a}(b-a) = f(a) = g(a)$. Somit ist der bereits behandelte Spezialfall auf g anwendbar und liefert ein x_0 mit $a < x_0 < b$ mit $g'(x_0) = 0$. Wegen $g'(x) = f'(x) - \frac{f(b)-f(a)}{b-a}$ folgt daraus (siehe Summenregel im folgenden Satz 15 und die am Ende von Abschnitt 5.1 bewiesene Regel $(kx+d)' = k$) $f'(x_0) = \frac{f(b)-f(a)}{b-a}$. $\square$

Beweis von Satz 15 (Differentiationsregeln): Unter den im Satz angegebenen Differenzierbarkeitsvoraussetzungen an f und g in x_0 bzw. an h in $f(x_0)$ berechnet man die Differentialquotienten für die Funktionen $f \pm g$, $f \cdot g$, $h \circ f$, $\frac{f}{g}$ (sofern $g(x) \neq 0$ für alle $x \in D$) und in x_0 bzw. der Umkehrfunktion $f^{(-1)}$ in $f(x_0)$ und beweist damit unter Verwendung der entsprechenden Regeln für Grenzwerte von Funktionen (Satz 11) die einzelnen Differentiationsregeln.

1. Summenregel:

$$(f \pm g)'(x_0) = \lim_{x\to x_0}\frac{(f(x)\pm g(x)) - (f(x_0)\pm g(x_0))}{x-x_0} = \lim_{x\to x_0}\frac{(f(x)-f(x_0)) \pm (g(x)-g(x_0))}{x-x_0} =$$
$$= \lim_{x\to x_0}\frac{f(x)-f(x_0)}{x-x_0} \pm \lim_{x\to x_0}\frac{g(x)-g(x_0)}{x-x_0} = f'(x_0) \pm g'(x_0)$$

2. Produktregel: Den Zähler im Differentialquotienten für $(fg)'(x_0)$ formt man ähnlich um, wie schon für den Produktsatz (siehe Satz 3), nämlich zu

$$f(x)g(x) - f(x_0)g(x_0) = f(x)g(x) - f(x_0)g(x) + f(x_0)g(x) - f(x_0)g(x_0) =$$
$$= (f(x)-f(x_0))g(x) + f(x_0)(g(x)-g(x_0)).$$

Dividiert man diesen Ausdruck entsprechend der Definition des Differentialquotienten durch $x - x_0$ und geht man zum Grenzwert für $x \to x_0$ über, wird der erste Summand zu $f'(x_0)g(x_0)$ (hier verwendet man u.a. Satz 12, wonach g stetig in x_0 ist, also $\lim_{x\to x_0} g(x) = g(x_0)$ erfüllt), der zweite zu $f(x_0)g'(x_0)$, also folgt $(fg)'(x_0) = f(x_0)g'(x_0) + f'(x_0)g(x_0)$.

3. Kettenregel: Die Rechnung

$$
\begin{aligned}
(h\circ f)'(x_0) &= \lim_{x\to x_0}\frac{h(f(x))-h(f(x_0))}{x-x_0} = \lim_{x\to x_0}\left(\frac{h(f(x))-h(f(x_0))}{f(x)-f(x_0)}\cdot\frac{f(x)-f(x_0)}{x-x_0}\right) = \\
&= \lim_{x\to x_0}\frac{h(f(x))-h(f(x_0))}{f(x)-f(x_0)}\cdot\lim_{x\to x_0}\frac{f(x)-f(x_0)}{x-x_0} = h'(f(x_0))\cdot f'(x_0)
\end{aligned}
$$

bedarf folgender Erklärungen: Weil f als differenzierbare Funktion stetig ist (Satz 12), folgt aus $x\to x_0$ auch $f(x)\to f(x_0)$. Also lässt sich der erste Faktor als Differentialquotient von h an der Stelle $f(x_0)$ deuten und ist nach Voraussetzung $= h'(f(x_0))$.

Die einzige denkbare Komplikationen, die obige Rechnung ungültig machen könnte, liegt vor, wenn es $x\neq x_0$ beliebig nahe bei x_0 mit $f(x)=f(x_0)$ gibt. Das bedeutet $f'(x_0)=0$. Dann ist $(h\circ f)'(x_0)=0$ zu zeigen: Aus der Definition des Differentialquotienten für h bei $y_0 := f(x_0)$ liest man

$$|h(y)-h(y_0)| \leq (|h'(y_0)|+1)\cdot|y-y_0|$$

ab, sofern nur $y := f(x)$ hinreichend nahe bei y_0 liegt, was wegen der Stetigkeit von f bei x_0 für x hinreichend nahe bei x_0 garantiert werden kann. Für x genügend nahe bei x_0 kann wegen $f'(x_0)=0$ auch $|y-y_0| = |f(x)-f(x_0)| \leq \frac{|\varepsilon|}{|h'(y_0)|+1}\cdot|x-x_0|$ für beliebig vorgegebenes $\varepsilon>0$ erreicht werden. Insgesamt zeigt dies

$$
\begin{aligned}
|h(f(x))-h(f(x_0))| &= |h(y)-h(y_0)| \leq (|h'(y_0)|+1)\cdot|y-y_0| \leq \\
&\leq (|h'(y_0)|+1)\cdot\frac{|\varepsilon|}{|h'(y_0)|+1}\cdot|x-x_0| \leq |\varepsilon|\cdot|x-x_0|.
\end{aligned}
$$

Für den Differenzenquotienten folgt daraus

$$\left|\frac{h(f(x))-h(f(x_0))}{x-x_0}\right| \leq \varepsilon,$$

wegen der Beliebigkeit von $\varepsilon>0$ daher

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

4. Quotientenregel: Zunächst zeigen wir (Spezialfall $f(x)=1$ und $g(x)=x$) für die Funktion $q(x) := \frac{1}{x}$ die Differentiationsregel $q'(x_0) = -\frac{1}{x_0^2}$:

$$q'(x_0) = \lim_{x\to x_0}\frac{\frac{1}{x}-\frac{1}{x_0}}{x-x_0} = \lim_{x\to x_0}\frac{x_0-x}{x_0x(x-x_0)} = \lim_{x\to x_0}\frac{-1}{x_0x} = -\frac{1}{x_0^2}.$$

Damit liefert die Kettenregel allgemeiner

$$\left(\frac{1}{g}\right)'(x_0) = (q\circ g)'(x_0) = q'(g(x_0))g'(x_0) = -\frac{g'(x_0)}{g(x_0)^2}.$$

Die Quotientenregel ergibt sich nunmehr aus einer Anwendung der Produktregel auf $\frac{f}{g} = f\cdot\frac{1}{g}$:

$$
\begin{aligned}
\left(\frac{f}{g}\right)'(x_0) &= f(x_0)\left(\frac{1}{g}\right)'(x_0) + f'(x_0)\frac{1}{g(x_0)} = f(x_0)\cdot\left(-\frac{g'(x_0)}{g(x_0)^2}\right) + \frac{f'(x_0)}{g(x_0)} = \\
&= \frac{f'(x_0)g(x_0)-f(x_0)g'(x_0)}{g(x_0)^2}
\end{aligned}
$$

5. Umkehrfunktion: Weil f in einer Umgebung von x_0 stetig und umkehrbar ist, muss f auf dieser Umgebung streng monoton sein; andernfalls ergäbe der Zwischenwertsatz, siehe Satz 10, einen Widerspruch. In dieser Umgebung gilt folglich nicht nur $\lim_{x\to x_0} f(x) = f(x_0)$, sondern umgekehrt auch $\lim_{y\to y_0:=f(x_0)} f^{(-1)}(y) = f^{(-1)}(y_0) = x_0$. Deshalb lässt sich der gesuchte Differentialquotient wie folgt berechnen:

$$\left(f^{(-1)}\right)'(f(x_0)) = \lim_{f(x)=y\to f(x_0)=y_0} \frac{f^{(-1)}(y) - f^{(-1)}(y_0)}{y-y_0} = \lim_{x\to x_0} \frac{x-x_0}{f(x)-f(x_0)} = \left(f'(x_0)\right)^{-1}. \ \square$$

8.6. Beweise zu Kapitel 6 (Integralrechnung)

Beweis von $\int_0^1 x\,dx = \frac{1}{2}$ anhand von Definition 6.1: Wir verwenden die Formel $s_n := \sum_{i=1}^n i = \frac{n(n+1)}{2}$, die sich mittels Induktion oder – der anekdotisch tradierten Idee des kleinen Carl Friedrich Gauß folgend – aus

$$2s_n = \sum_{i=1}^n i + \sum_{i=1}^n i = \sum_{i=1}^n i + \sum_{i=1}^n (n+1-i) = \sum_{i=1}^n (i+n+1-i) = \sum_{i=1}^n (n+1) = n(n+1)$$

ergibt. Für die Zerlegungen $Z_n = \{0, \frac{1}{n}, \dots, 1\}$, $n = 1, 2, \dots$, ergeben sich daraus die Untersummen

$$U(f,Z_n) = \sum_{i=1}^n \frac{1}{n}\cdot\frac{i-1}{n} = \frac{1}{n^2}\sum_{i=1}^n (i-1) = \frac{(n-1)n}{2n^2} = \frac{1}{2}(1-\frac{1}{n})$$

und die Obersummen

$$O(f,Z_n) = \sum_{i=1}^n \frac{1}{n}\cdot\frac{i}{n} = \frac{1}{n^2}\sum_{i=1}^n i = \frac{(n+1)n}{2n^2} = \frac{1}{2}(1+\frac{1}{n}).$$

Zu gegebenem $\varepsilon > 0$ gilt für $\alpha = \frac{1}{2}$ und hinreichend großes n daher die geforderte Ungleichung

$$\frac{1}{2} - \varepsilon < U(f,Z_n) < \frac{1}{2} < O(f,Z_n) < \frac{1}{2} + \varepsilon.$$

Folglich ist $f(x) = x$ auf $[0,1]$ gemäß Definition 6.1 Riemann-integrierbar mit Wert $\int_0^1 x\,dx = \frac{1}{2}$. $\square$

Beweis von Satz 16 (Untersumme stets kleiner als Obersumme) Für beschränktes $f : [a,b] \to \mathbb{R}$ und Zerlegungen Z_1 und Z_2 von $[a,b]$ sind Ungleichungen zu beweisen:

1. Ist $Z_1 \subseteq Z_2$, so folgt $U(f,Z_1) \le U(f,Z_2) \le O(f,Z_2) \le O(f,Z_1)$:
Seien zunächst $Z_0 = \{a,b\}$ die triviale und $Z = \{a = x_0 \le \dots \le x_n = b\}$ irgendeine Zerlegung von $[a,b]$, außerdem

$$c := \inf\{f(x) :\ a \le x \le b\} \quad \le \quad c_i := \inf\{f(x) :\ x_{i-1} \le x \le x_i\}$$

für $i = 1, \dots, n$. Dann gilt

$$U(f,Z_0) = c(b-a) = c\sum_{i=1}^n (x_i - x_{i-1}) = \sum_{i=1}^n c(x_i - x_{i-1}) \le \sum_{i=1}^n c_i(x_i - x_{i-1}) = U(f,Z).$$

Wir wenden diese Ungleichung statt auf $[a,b]$ auf jedes der durch die Zerlegung Z_1 gegebenen Teilintervalle sowie auf die Zerlegungen Z_1 (statt Z_0) und Z_2 (statt Z) an. Summation über diese Teilintervalle liefert $U(f,Z_1) \le U(f,Z_2)$. Aus Symmetriegründen gilt für die Obersummen entsprechend $O(f,Z_2) \le O(f,Z_1)$. Trivialerweise gilt $U(f,Z_2) \le O(f,Z_2)$. Damit ist die zu beweisende Ungleichungskette komplett.

2. In jedem Fall gilt $U(f,Z_1) \le O(f,Z_2)$:
Die gemeinsame Verfeinerung $Z := Z_1 \cup Z_2$ erfüllt nach dem im ersten Teil bereits Bewiesenen $U(f,Z_1) \le U(f,Z) \le O(f,Z) \le O(f,Z_2)$. $\square$

Beweis von Satz 17 (Integral als Grenzwert):
Zu zeigen ist, dass die Riemann-Integrierbarkeit einer beschränkten Funktion $f : [a,b] \to \mathbb{R}$ zu folgender Bedingung äquivalent ist: Zu jedem $\varepsilon > 0$ gibt es ein $\delta > 0$ mit $O(f,Z) - U(f,Z) < \varepsilon$ für alle Zerlegungen Z von $[a,b]$ der Feinheit $\|Z\| < \delta$. Klarerweise ist die Bedingung hinreichend, weil es genügt, irgendein solches Z zu nehmen, um Definition 6.1 zu erfüllen. Die Bedingung ist aber auch notwendig:

Gehen wir also davon aus, dass f Definition 6.1 erfüllt, und geben wir uns ein $\varepsilon > 0$ vor, zu dem wir ein $\delta > 0$ finden wollen. Es gibt eine Zerlegung $Z_0 = \{a = x_0 \leq x_1 \leq \ldots \leq x_n = b\}$ mit $O(f,Z_0) - U(f,Z_0) < \frac{\varepsilon}{2}$. Weil f beschränkt ist, existieren $m := \inf\{f(x) :\ a \leq x \leq b\}$ und $M := \sup\{f(x) :\ a \leq x \leq b\}$. Wir wählen δ so klein, dass $\delta n(M-m) < \frac{\varepsilon}{2}$ und $\delta < \|Z_0\|$, und behaupten, dass alle Zerlegungen Z mit $\|Z\| < \delta$ die Ungleichung $O(f,Z) - U(f,Z) < \varepsilon$ erfüllen. Sei also ein solches Z gegeben. Die Verfeinerung $Z' := Z_0 \cup Z$ von Z_0 erfüllt wegen der ersten Aussage in Satz 16 die Ungleichung $O(f,Z') - U(f,Z') \leq O(f,Z_0) - U(f,Z_0) < \frac{\varepsilon}{2}$. Die Differenz $d := (O(f,Z) - U(f,Z)) - (O(f,Z') - U(f,Z'))$ setzt sich aus den Beiträgen zusammen, wo sich Z und Z' unterscheiden. Das sind jene Teilintervalle von Z, die in ihrem Inneren ein $x_i \in Z_0$ enthalten. Die Länge dieser Intervalle ist $< \delta$, ihre Anzahl ist $< n$ und die maximale Differenz zwischen Supremum und Infimum ist stets $< M - m$, also ist $d < \delta n(M-m) < \frac{\varepsilon}{2}$. Daraus folgt

$$O(f,Z) - U(f,Z) = O(f,Z') - U(f,Z') + d < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon,$$

was zu zeigen war. □

Beweis von Satz 18 (Stetige Funktionen sind integrierbar)
Sei $f : [a,b] \to \mathbb{R}$ stetig. Zu zeigen ist, dass das Integral $\int_a^b f(x)\,dx$ existiert. Zunächst ein kurzer Diskurs, der das Hauptproblem erläutert, sodann der Beweis einer Hilfsbehauptung (gleichmäßige Stetigkeit von f) und erst dann der Beweis des Satzes.

Stetigkeit an einer Stelle $x_0 \in [a,b]$ bedeutet, dass zu beliebig vorgegebenem $\varepsilon > 0$ ein $\delta > 0$ existiert, so dass alle $x \in [a,b]$ mit $|x - x_0| < \delta$ die Beziehung $|f(x) - f(x_0)| < \varepsilon$ erfüllen. Unser Problem besteht darin, dass dieses δ nicht nur von ε, sondern a priori auch von x_0 abhängt, eventuell also kein „gleichmäßiges" δ existiert. Formelschreibweise mit Quantoren macht deutlich, worin der feine Unterschied besteht. Die vorausgesetzte Stetigkeit bedeutet $\forall \varepsilon > 0\ \forall x_0\ \exists \delta\ \forall x :\ |x - x_0| < \delta \Rightarrow |f(x) - f(x_0)| < \varepsilon$. Die zu beweisende sogenannte *gleichmäßige Stetigkeit* hingegen bedeutet $\forall \varepsilon > 0\ \exists \delta\ \forall x_0\ \forall x :\ |x - x_0| < \delta \Rightarrow |f(x) - f(x_0)| < \varepsilon$. Der Unterschied besteht also in der Vertauschung der Quantoren $\forall x_0$ und $\exists \delta$. Die Abhängigkeit von x_0, die Probleme bereiten könnte, besteht in Wahrheit aber nicht, weil ganz allgemein jede stetige Funktion auf $[a,b]$ sogar gleichmäßig stetig ist (Hilfsbehauptung).

Beweis der Hilfsbehauptung, d.h. der gleichmäßigen Stetigkeit von f: Angenommen, es gibt zu irgendeinem $\varepsilon > 0$ kein gemeinsames δ. Dann gibt es zu jedem $n = 1, 2, \ldots$ zwei Punkte $a_n, b_n \in [a,b]$ mit $|a_n - b_n| < \frac{1}{n}$ aber $|f(a_n) - f(b_n)| \geq \varepsilon$. Nach dem Satz von Bolzano-Weierstraß gibt es eine Teilfolge der a_n, die gegen einen Grenzwert $x_0 \in [a,b]$ konvergiert. Weil auch die entsprechenden b_n gegen dasselbe x_0 konvergieren, kann wegen $|f(a_n) - f(b_n)| \geq \varepsilon$ kein gemeinsamer Grenzwert $\lim_{x \to x_0} f(x) = \lim_{n \to \infty} f(a_n) = \lim_{n \to \infty} b_n$ existieren, was der Stetigkeit von f widerspricht. Also gibt es doch ein $\delta > 0$ mit $|f(x) - f(x')| < \varepsilon$ für alle $x, x' \in [a,b]$, die $|x - x'| < \delta$ erfüllen.

Wir wollen nun die in Satz 18 behauptete Riemann-Integrierbarkeit von f beweisen. Dafür müssen wir uns ein $\varepsilon > 0$ wie in Definition 6.1 vorgeben und eine Zerlegung mit $O(f,Z) - U(f,Z) < \varepsilon$ finden. Nach dem eben bewiesenen gibt es ein δ derart, dass für beliebige $x, x' \in [a,b]$ mit $|x - x'| < \delta$ die Ungleichung $|f(x) - f(x')| < \frac{\varepsilon}{b-a}$ gilt. Wir wählen irgendeine Zerlegung $Z = \{a = x_0 \leq \ldots \leq x_k = b\}$ mit $\|Z\| < \delta$, z.B. mit $x_i = a + \frac{i}{k}(b-a)$, $i = 0, \ldots, k$, und irgendeinem $k > \frac{b-a}{\delta}$. Auf jedem Teilintervall liegen Infimum und Supremum (die nach dem Satz vom Maximum, der zweiten Aussage in Satz 10 sogar als Minimum bzw. Maximum von f angenommen werden) weniger als $\frac{\varepsilon}{b-a}$ auseinander, was sich nach Summation über alle Teilintervalle zur Abschätzung $O(f,Z) - U(f,Z) < \varepsilon$ aufaddiert. □

Beweis von Satz 19 (Hauptsatz der Differential- und Integralrechnung): Für stetiges $f : [a,b] \to \mathbb{R}$ und $F(x_0) := \int_a^{x_0} f(x)\,dx$ ist $F' = f$ zu zeigen.

Wegen Satz 18 ist F tatsächlich auf ganz $[a,b]$ definiert. Der interessierende Differenzenquotient (Sekantensteigung, siehe Abschnitt 5.1) hat die Gestalt

$$s_{F,x_0}(x) = \frac{F(x) - F(x_0)}{x - x_0} = \frac{\int_a^x f(t)\,dt - \int_a^{x_0} f(t)\,dt}{x - x_0} = \frac{\int_{x_0}^x f(t)\,dt}{x - x_0}.$$

(Hier haben wir die offensichtliche Formel $\int_a^x f(t)\,dt = \int_a^{x_0} f(t)\,dt + \int_{x_0}^x = f(t)\,dt$ verwendet, auf deren formalen Beweis wir verzichten dürfen.) Bezeichnen $c(x)$ und $C(x)$ Infimum bzw. Supremum von f im Intervall $[x_0,x]$, so liegt der Wert des Zählers auf der rechten Seite zwischen $c(x)(x-x_0)$ und $C(x)(x-x_0)$. Nach Division durch $x-x_0$ erhalten wir daher

$$c(x) \leq s_{F,x_0}(x) \leq C(x).$$

Wegen der Stetigkeit von f ist $\lim_{x\to x_0} c(x) = \lim_{x\to x_0} C(x) = f(x_0)$, also

$$F'(x_0) = \lim_{x\to x_0} s_{F,x_0}(x) = f(x_0),$$

was zu beweisen war. □

Beweis von Satz 20 (Stammfunktionen und additive Konstante): Für zwei Stammfunktionen F_1 und F_2 von $f : [a,b] \to \mathbb{R}$ ist eine konstante Zahl $c \in \mathbb{R}$ mit $F_1(x) = F_2(x) + c$ für alle $x \in [a,b]$ zu finden und daraus $F(b) - F(a) = \int_a^b f(x)\,dx$ für jede Stammfunktion F von f zu folgern.
Die Ableitung der Funktion $F_1 - F_2$ ist wegen $(F_1 - F_2)' = F_1' - F_2' = f - f = 0$ die Nullfunktion. Der Mittelwertsatz 14 liefert für ein beliebiges $x \in]a,b]$ ein $x_0 \in]a,x[$ mit

$$F_1(x) - F_2(x) = (F_1 - F_2)(x) = (F_1 - F_2)(a) + (F_1 - F_2)'(x_0)(x-a) = (F_1 - F_2)(a) = F_1(a) - F_2(a).$$

Mit $c := F_1(a) - F_2(a)$ gilt folglich tatsächlich $F_1(x) = F_2(x) + c$ für alle $x \in [a,b]$. Setzen wir speziell $F_1 := F$ mit der Funktion F aus der ersten Aussage des Hauptsatzes 19, so gilt $F_2(b) - F_2(a) = (F(b) - c) - (F(a) - c) = F(b) - F(a) = \int_a^b f(x)\,dx$. Da F_2 eine beliebige Stammfunktion von f war, ist damit alles bewiesen. □

Literatur

Lehrplan: *AHS-Lehrpläne Oberstufe neu: Mathematik.* Am 10.9.2018 verfügbar unter: `https://bildung.bmbwf.gv.at/schulen/unterricht/lp/lp_neu_ahs_07_11859.pdf`

Reinhard Winkler. *Wir zählen bis drei – und sogar darüber hinaus.* DH[9] 40 (2007/08), 129-141.

Reinhard Winkler. *Die reellen Zahlen sind anders.* DH[9] 41 (2008/09), 140-153.

Reinhard Winkler. *Im Anfang war die Exponentialfunktion.* DH[9] 44 (2011/12), 98-109.

Reinhard Winkler. *Dynamische Systeme als Chance für den Schulunterricht*. Druckversion: DH[9] 46 (2013/14), 108-122. Langversion nur online verfügbar.

Reinhard Winkler. *Der Grenzwert – Zentralbegriff der Analysis* Druckversion: DH[9] 51 (2018/19). Langversion nur online verfügbar.

[9] Schriftenreihe zur Didaktik der Mathematik der Österreichischen Mathematischen Gesellschaft (ÖMG) – ehemals Didaktikhefte der ÖMG, online verfügbar unter `https://www.oemg.ac.at/DK/Didaktikhefte/index.html`, meine eigenen Artikel auch unter `http://dmg.tuwien.ac.at/winkler/pub/`.