Kapitel II
Konvergenz und Stetigkeit

Tout va par degrés dans la nature
et rien par saut, et cette règle à
l'égard des changements est une
partie de ma loi de la continuité.
Leibniz

Wir erklären den Begriff der Konvergenz und des Grenzwertes von Folgen, Reihen und Funktionen und den Begriff der Stetigkeit von Funktionen. Der Begriff der Konvergenz ist der eigentliche Grundbegriff der Analysis.

§ 1. Folgen und Reihen reeller Zahlen

Definition. *Eine* **Folge** *reeller Zahlen ist eine Abbildung* $\mathbb{N} \to \mathbb{R}$, $n \mapsto a_n$. *Wir bezeichnen sie durch* $(a_n \mid n \in \mathbb{N})$ *oder* $(a_n)_{n\in\mathbb{N}}$ *oder kurz* (a_n) *und oft auch durch*

$$a_1, a_2, a_3, \dots .$$

Auch Abbildungen $\{n \in \mathbb{Z} \mid n \geq k\} \to \mathbb{R}$ *bezeichnen wir als Folgen*

$$a_k, a_{k+1}, \dots .$$

(1.1) Beispiele

(i) $\left(\frac{1}{n}\right) = 1,\ \frac{1}{2},\ \frac{1}{3},\ \frac{1}{4},\ \dots$

(ii) $\left(\frac{n}{n+1}\right) = \frac{1}{2},\ \frac{2}{3},\ \frac{3}{4},\ \frac{4}{5},\ \dots$

(iii) $(x^n) = x,\ x^2,\ x^3,\ \dots$.
Dies ist für jedes $x \in \mathbb{R}$ eine reelle Folge.

(iv) $\left(\dfrac{n}{2^n}\right) = \frac{1}{2},\ \frac{2}{4},\ \frac{3}{8},\ \dots$

(v) Man kann auch Folgen rekursiv definieren, z.B. **die Fibonacci-Folge**, definiert durch $a_0 = 0$, $a_1 = 1$, $a_n = a_{n-1} + a_{n-2}$, also

$$(a_n) = 0, 1, 1, 2, 3, 5, 8, 13, 21, \dots \ .$$

Der entscheidende Begriff der Analysis, mit dessen Entwicklung auch eigentlich die Mathematik der Neuzeit beginnt, ist der der Konvergenz und des Grenzwerts.

Definition. *Eine Folge (a_n) heißt* **konvergent** *gegen $a \in \mathbb{R}$, und a ist* **Grenzwert** *(***Limes***) von (a_n), geschrieben*

$$lim_{n\to\infty}(a_n) = a, \quad \textit{oder} \quad (a_n) \to a,$$

falls folgendes gilt:

Zu jeder reellen Zahl $\varepsilon > 0$ existiert eine Zahl $N \in \mathbb{N}$, sodaß

$$|a_n - a| < \varepsilon$$

für alle $n > N$ aus $\mathbb{N}$.

Weil man schließlich auf eine Ungleichung $|a_n - a| < \varepsilon$ hinaus will, kommt es ersichtlich nur darauf an, beliebig kleine $\varepsilon > 0$ zu betrachten: Für jedes auch noch so kleine $\varepsilon > 0$ existiert $N \in \mathbb{N}$, sodaß $|a_n - a| < \varepsilon$, falls $n > N$.

Bezeichnen wir das Intervall

$$(a - \varepsilon, a + \varepsilon) =: U_\varepsilon(a)$$

als ε-**Umgebung** von a, so sagt die Bedingung der Konvergenz: $a_n \in U_\varepsilon(a)$ für alle n bis auf endlich viele, nämlich bis auf allenfalls $n \in \{1, 2, \dots, N\}$. Man sagt dafür auch

$$a_n \in U_\varepsilon(a) \quad \text{für } \textbf{fast alle } n,$$

oder für **genügend große** n oder **schließlich**, und benutzt die Sprechweise: "fast alle" heißt "alle bis auf endlich viele". Unsere Definition lautet dann: $(a_n) \to a$, falls gilt:

Ist $\varepsilon > 0$, so ist $a_n \in U_\varepsilon(a)$ für fast alle $n \in \mathbb{N}$.

Oder: Ist $\varepsilon > 0$, so gilt schließlich $a_n \in U_\varepsilon(a)$.

Eine nicht konvergente Folge heißt **divergent**. Die Bedingung der Divergenz lautet also: Zu jedem $a \in \mathbb{R}$ gibt es ein $\varepsilon > 0$, sodaß zu jedem N ein $n > N$ existiert, mit $|a_n - a| \geq \varepsilon$. Oder anders ausgedrückt: Jedes $a \in \mathbb{R}$ besitzt eine ε-Umgebung $U_\varepsilon(a)$, außerhalb von welcher unendlich viele Folgenglieder liegen. (a_n) konvergiert nicht gegen a, falls a eine ε-Umgebung U besitzt, außerhalb der unendlich viele Folgenglieder liegen. Merke also: Die Negation von "fast alle" ist: Für unendlich viele nicht.

Prüfen wir zur Einübung des Begriffs sogleich die Beispiele (1.1):

(i) $(1/n) \to 0$.

Beweis: Sei $\varepsilon > 0$ gegeben. Nach Archimedes gibt es eine Zahl N, sodaß $N > \frac{1}{\varepsilon}$, und ist $n > N$, so ist erst recht $n > N > \frac{1}{\varepsilon}$. Also wenn $n > N$ ist, gilt nach unserer Kenntnis über die Anordnung:

$$0 < \tfrac{1}{n} < \varepsilon, \quad \text{das heißt} \quad |\tfrac{1}{n} - 0| < \varepsilon. \qquad \square$$

(ii) $\left(\frac{n}{n+1}\right) \to 1$.

Beweis: $|1 - \frac{n}{n+1}| = \frac{1}{n+1}$, und ist $\varepsilon > 0$ gegeben, so haben wir eben schon gesehen, daß $0 < \frac{1}{n+1} < \varepsilon$ für fast alle n. $\square$

(iii) Bei der Folge (x^n) hängt das Verhalten von x ab.

Fall 1. $x = 1$, also $x^n = 1$, die Folge $1, 1, 1, \ldots$ konvergiert offenbar gegen 1.

Fall 2. $x = -1$; wir erhalten die Folge $-1, 1, -1, 1, \ldots$, die offenbar divergiert. Eine ganzzahlige Folge konvergiert überhaupt nur, wenn sie schließlich konstant ist. Wäre nämlich a der Grenzwert, so wäre schließlich $|a_n - a| < \frac{1}{2}$ und $|a - a_{n+1}| < \frac{1}{2}$, also nach der Dreiecksungleichung $|a_n - a_{n+1}| < 1$, also $a_n = a_{n+1}$, wenn beide ganz sind. □

Hier ist ein kleiner Schluß verborgen: Gilt für fast alle n die Aussage $A(n)$ und auch für fast alle n die Aussage $B(n)$, so gilt für fast alle n alles beides: $A(n)$ und $B(n)$, denn gilt etwa $A(n)$ für $n > N_1$, $B(n)$ für $n > N_2$, so $A(n)$ und $B(n)$ für $n > \max\{N_1, N_2\}$.

Fall 3. $|x| > 1$. Sei $a = |x|$, dann ist $a = 1 + \delta$ für ein $\delta > 0$, also nach Bernoulli

$$a^n = (1+\delta)^n \geq 1 + n\delta \geq n\delta,$$

und nach Archimedes gilt: Ist $R \in \mathbb{R}$ beliebig vorgegeben, so ist für fast alle n

$$a^n \geq n\delta > R.$$

Also ist (x^n) divergent, denn aus $(x^n) \to g$ würde folgen $|x^n - g| < 1$ für fast alle n, also $a^n = |x^n| < |g| + 1$. In der Tat wird x^n für $x > 1$ schließlich beliebig groß:

Definition. *Wir sagen:* $(a_n) \to \infty$ *oder* $\lim_{n\to\infty} a_n = \infty$, *falls gilt: Für jedes* $R > 0$ *ist* $a_n > R$ *für fast alle* $n \in \mathbb{N}$. *Entsprechend* $(a_n) \to -\infty$, *falls* $(-a_n) \to \infty$. *Die Folge* (a_n) *heißt in diesen Fällen* **bestimmt divergent**.

Als nächstes haben wir die Folge (x^n) für $|x| < 1$ zu betrachten. Ist wieder $a = |x|$, so ist $0 \leq a < 1$. Für $a = 0$ ist die Folge konstant 0 und konvergiert gegen 0. Für $0 < a < 1$ wissen wir $1/a > 1$, also wenn $\varepsilon > 0$ gegeben ist, so ist für fast alle n, wie eben gesehen,

$$(1/a)^n > 1/\varepsilon, \quad \text{also} \quad a^n < \varepsilon,$$

und das heißt $|x^n - 0| < \varepsilon$ für fast alle n, die Folge konvergiert also gegen 0. □

(iv) Die Folge $(\frac{n}{2^n})$ konvergiert gegen 0. Es ist nämlich nach dem binomischen Lehrsatz für $n \geq 2$:

$$2^n = (1+1)^n = 1 + n + \frac{n(n-1)}{2} + \cdots \geq \frac{n(n+1)}{2} \geq \frac{n^2}{2},$$

also $n/2^n \leq \frac{2}{n}$, und $|\frac{2}{n}| < \varepsilon$ für fast alle n. □

(v) Die Fibonacci-Folge ist bestimmt divergent, denn:

$$\text{für } n \geq 5 \text{ ist } a_n \geq n.$$

Man prüft diese Behauptung für $n = 5$ und 6. Beim Induktionsschritt setzt man die Behauptung für $n+1$ und n voraus und erhält:

$$a_{n+2} = a_{n+1} + a_n \geq n + 1 + n > n + 2.$$ □

Hier haben wir ein Beispiel, wo die Induktion nicht bei 1 beginnt und wo man beim Induktionsschritt die Behauptung über zwei vorhergehende Zahlen braucht. Will man das Induktionsschema formal anwenden, so muß man die Behauptung zugleich für n und $n+1$ aufstellen. Dies ist dann induktiv gezeigt.

Mit der Schreibweise $\lim (a_n) = a$ ist schon unterstellt, daß eine Folge, wenn überhaupt, nur *einen* Grenzwert haben kann. Das zeigen wir jetzt.

(1.2) Satz *(Eindeutigkeit des Grenzwertes). Konvergiert die Folge (a_n) gegen a und gegen b, so ist $a = b$.*

Beweis: Angenommen etwa $a < b$, so wähle $\varepsilon = \frac{b-a}{2}$, dann ist für fast alle $n \in \mathbb{N}$ erfüllt:

$$|a_n - a| < \frac{b-a}{2}, \quad |b - a_n| < \frac{b-a}{2},$$

also

$$b - a = |b-a| \leq |b - a_n| + |a_n - a| < b - a.$$ ✠ □

(1.3) Satz. *Eine konvergente Folge ist beschränkt, d.h. die Menge* $\{a_n \mid n \in \mathbb{N}\}$ *der Folgenglieder ist beschränkt.*

Beweis: Angenommen $(a_n) \to a$, dann gilt für alle $n > N$:

$|a_n - a| < 1$, also $|a_n| - |a| \leq |a_n - a| < 1$, also $|a_n| \leq |a| + 1$.

Also gilt für alle n:

$$|a_n| \leq \max\{|a_1|, |a_2|, \ldots, |a_N|, |a| + 1\}. \qquad \square$$

Folgen kann man addieren, multiplizieren, manchmal auch dividieren, nämlich man definiert:

$$\lambda(a_n) + \mu(b_n) := (\lambda a_n + \mu b_n) \quad \text{für} \quad \lambda, \mu \in \mathbb{R}.$$
$$(a_n) \cdot (b_n) := (a_n \cdot b_n).$$

Ist $a_n \neq 0$ für alle n, so ist

$$1/(a_n) := (1/a_n).$$

Die erste Operation macht die Menge aller Folgen zu einem reellen Vektorraum; die Multiplikation ist assoziativ, kommutativ und distributiv:

$$(a_n)\big((b_n) + (c_n)\big) = (a_n)(b_n) + (a_n)(c_n).$$

Die konstante Folge (1) ist ein neutrales Element der Multiplikation. Die Grenzwertbildung ist mit den algebraischen Operationen **verträglich**:

(1.4) Satz. *Es konvergiere* $(a_n) \to a$ *und* $(b_n) \to b$, *dann gilt:*

(i) lim *ist linear, das heißt:* $\lambda(a_n) + \mu(b_n) \to \lambda a + \mu b$.
(ii) lim *ist multiplikativ, das heißt:* $(a_n) \cdot (b_n) \to a \cdot b$.
(iii) *Ist* $b \neq 0$, *so ist* $b_n \neq 0$ *für fast alle* n, *also etwa für alle* $n > N$. *Es gilt:*

$$(1/b_n)_{n>N} \to 1/b.$$

Beweis: (i) Sei $\varepsilon > 0$ gegeben. Wir müssen zeigen, daß für fast alle n gilt

$$|\lambda a_n + \mu b_n - \lambda a - \mu b| < \varepsilon.$$

Die linke Seite schätzen wir ab:

$$|\lambda(a_n - a) + \mu(b_n - b)\,| \le |\lambda| \cdot |a_n - a| + |\mu| \cdot |b_n - b|.$$

Sind nun $\eta > 0$, $\zeta > 0$ vorgegeben, so ist für fast alle n nach Voraussetzung

$$|a_n - a| < \eta, \quad |b_n - b| < \zeta,$$

also ist die rechte Seite der Ungleichung für fast alle n

$$< |\lambda|\eta + |\mu|\zeta < B(\eta + \zeta), \quad \text{mit} \quad B = |\lambda| + |\mu| + 1.$$

Zu dem gegebenen ε wähle jetzt $\eta = \zeta = \frac{\varepsilon}{2B}$, dann ist $B(\eta + \zeta) \le \varepsilon$, also die behauptete Ungleichung für fast alle n gezeigt.

(ii) Zu gegebenem $\varepsilon > 0$ müssen wir für fast alle n zeigen:

$$|a_n b_n - ab| < \varepsilon.$$

Die linke Seite schätzen wir folgendermaßen ab:

$$|\cdots| = |a_n b_n - a b_n + a b_n - ab| \le |b_n| \cdot |a_n - a| + |a| \cdot |b_n - b|.$$

Jetzt wähle B so groß, daß $|b_n| < B$ für alle n, und $|a| < B$ (Satz 1.3), dann setzt sich die Abschätzung fort:

$$|\cdots| < B \cdot (|a_n - a| + |b_n - b|).$$

Für fast alle n ist nun $|a_n - a| < \frac{\varepsilon}{2B}$, $|b_n - b| < \frac{\varepsilon}{2B}$, also $|\cdots| < \varepsilon$.

(iii) Weil $|b| > 0$ ist, können wir die Definition der Konvergenz insbesondere für $\varepsilon = \frac{1}{2}|b|$ anwenden und finden, daß für fast alle n gilt: $|b_n - b| < \frac{1}{2}|b|$, und daher

$$|b_n| = |b - (b - b_n)| \ge |b| - |b - b_n| > \tfrac{1}{2}|b|.$$

Also ist $|b_n|$ fast immer ungleich 0, und es gilt fast immer:

$$\left|\frac{1}{b_n} - \frac{1}{b}\right| = \left|\frac{b_n - b}{b b_n}\right| = \frac{1}{|b||b_n|}|b_n - b| \le \frac{2}{|b|^2} \cdot |b_n - b|.$$

Aber auch fast immer ist $|b_n - b| < \varepsilon \cdot \frac{|b|^2}{2}$, also $\left|\frac{1}{b_n} - \frac{1}{b}\right| < \varepsilon$ für fast alle n. □

Ist (a_n) eine Folge, und $(n_k)_{k\in\mathbb{N}}$ eine Folge natürlicher Zahlen, sodaß $n_1 < n_2 < \cdots$, so heißt die Folge

$$(a_{n_k})_{k\in\mathbb{N}} = a_{n_1}, a_{n_2}, a_{n_3} \ldots$$

eine **Teilfolge** von (a_n). Ist (a_n) konvergent, so auch jede Teilfolge mit dem selben Grenzwert. Allgemeiner gilt:

(1.5) Satz. *Sei* $\lim(a_n) = a$ *und* $\rho : \mathbb{N} \to \mathbb{N}$ *injektiv, dann ist*

$$\lim(a_{\rho(n)} \mid n \in \mathbb{N}) = a.$$

Den Fall der Teilfolge erhält man, wenn man $\rho(k) = n_k$ setzt. Hier wird aber ausgesagt, daß man Folgenglieder auch beliebig umordnen darf, ohne daß sich das Grenzverhalten ändert.

Beweis: $(a_n) \to a$ heißt: Für jedes $\varepsilon > 0$ ist $a_n \in U_\varepsilon(a)$ für fast alle n, also für alle bis auf etwa $n \in \{1, 2, \ldots, K\}$. Weil ρ injektiv ist, gibt es höchstens K Zahlen n, sodaß $\rho(n) \in \{1, \ldots, K\}$, also für fast alle n gilt $\rho(n) \notin \{1, \ldots, K\}$. Also für fast alle n ist $a_{\rho(n)} \in U_\varepsilon(a)$, und das heißt $(a_{\rho(n)}) \to a$. □

Das Grenzverhalten von Folgen ist auch im gewissen Maße mit der Anordnung der reellen Zahlen verträglich:

(1.6) Satz. *Sind* (a_n) *und* (b_n) *konvergente Folgen und ist* $a_n \leq b_n$ *für fast alle* n*, so ist* $\lim(a_n) \leq \lim(b_n)$.

Warnung: Man darf jedoch **nicht** aus $a_n < b_n$ für alle n auf $\lim(a_n) < \lim(b_n)$ schließen! Zur Warnung dient folgendes Beispiel:

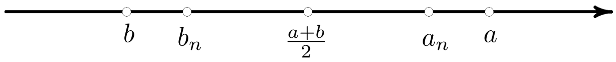
Es sei $a_n = 0$ für alle n, und $b_n = \frac{1}{n}$. Offenbar $a_n = 0 < \frac{1}{n} = b_n$, aber $0 = \lim(a_n) = \lim(b_n)$.

Beweis des Satzes: Angenommen $(a_n) \to a$, $(b_n) \to b$, und $a > b$, so wähle $\varepsilon = \frac{a-b}{2}$, dann ist für fast alle n:

$$|a_n - a| < \frac{a-b}{2}, \quad |b_n - b| < \frac{a-b}{2},$$

also

$$b_n < b + \frac{a-b}{2} = \frac{a+b}{2} = a - \frac{a-b}{2} < a_n. \qquad ✠ \qquad \square$$



Eng verwandt mit dem Begriff der Folge ist der einer Reihe, ja eigentlich handelt es sich recht besehen nur um eine andere Sprechweise für diesselbe Sache, die aber oft angemessener ist. Man kann sich eine Reihe als unendliche Summe vorstellen.

Definition *(Reihe). Sei (a_n) eine reelle Folge. Die Folge (A_n) der Summen*

$$A_n := \sum_{k=1}^{n} a_k$$

heißt **Reihe** *mit* **Gliedern** a_k *und wird mit* $\sum_{k=1}^{\infty} a_k$ *bezeichnet. Man nennt A_n die n-te* **Partialsumme** *der Reihe. Oft beginnt die Summation auch bei einer andern ganzen Zahl, die Reihe $\sum_{k=n_0}^{\infty} a_k$ ist die Folge der Partialsummen $\sum_{n_0 \le k \le n} a_k$, die dann auch für $n \ge n_0$ erklärt sind. Konvergiert eine Reihe (d.h. die Folge der Partialsummen), so bezeichnet man auch den Grenzwert mit $\sum_{k=1}^{\infty} a_k$ und nennt ihn die* **Summe** *der Reihe.*

Also beachte: Das Symbol $\sum_{k=1}^{\infty} a_k$ *bezeichnet:*

1) Die Folge $(A_n \mid n \in \mathbb{N})$ *der Partialsummen.*
2) Den Grenzwert $\lim_{n\to\infty}(\sum_{k=1}^{n} a_k)$, *falls er existiert.*

Offenbar läßt sich jede reelle Folge (A_n) als Reihe $\sum_{k=1}^{\infty} a_k$ schreiben, man wähle $a_1 = A_1$, $a_n = A_n - A_{n-1}$ für $n > 1$. Die Betrachtung von Folgen und Reihen ist also eigentlich dasselbe, und alles was über Folgen gesagt ist, läßt sich für Reihen umformulieren. Nur treten eben viele Reihen von Natur als Reihen auf und haben als Reihen eine durchsichtige Form, die beim Übergang zur Folge der Partialsummen verlorengeht. Ein klassisches Beispiel einer Reihe:

(1.7) Geometrische Reihe.

$$\sum_{k=0}^{\infty} x^k = \frac{1}{1-x} \quad \textit{für } |x| < 1,$$

$$\sum_{k=0}^{\infty} x^k \quad \textit{ist divergent für } |x| \geq 1.$$

Hier hat man beide Bedeutungen des Zeichens $\sum_{k=0}^{\infty} x^k$ beieinander.

Beweis: Wir berechnen die n-te Partialsumme:

$$\Big(\sum_{k=0}^{n} x^k\Big)(1-x) = \sum_{k=0}^{n} x^k - \sum_{k=0}^{n} x^{k+1} = \sum_{k=0}^{n} x^k - \sum_{k=1}^{n+1} x^k = 1 - x^{n+1},$$

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

falls $x \neq 1$. Dies schreiben wir in der Form:

$$X_n = (1-x)^{-1} \cdot (1-x^{n+1}) \;\to\; (1-x)^{-1}$$

für $|x| < 1$, weil $(x^n) \to 0$. □

Für $|x| \geq 1$ ist die Reihe sicher divergent, denn

(1.8) Bemerkung. *Ist die Reihe* $\sum_{k=1}^{\infty} a_k$ *konvergent, so gilt* $(a_k) \to 0$. *Die Umkehrung ist, wie wir bald sehen werden, nicht richtig, die Bedingung* $(a_k) \to 0$ *ist* **notwendig**, *nicht* **hinreichend** *für Konvergenz.*

Beweis: Ist $A_n = \sum_{k=1}^{n} a_k$ und $a = \lim_n(A_n)$, so folgt:
$(a_k) = (A_k - A_{k-1}) \to a - a = 0$. □

Wir beschließen diesen Abschnitt mit dem Beispiel einer Folge, wo man die Charakterisierung der Divergenz zu einem Konvergenzbeweis benutzen kann:

$$\lim(\sqrt[n]{n}) = 1.$$

Beweis: Offenbar ist $\sqrt[n]{n} \geq 1$, denn wäre $\sqrt[n]{n} < 1$, so $n < 1^n = 1$. Angenommen $\sqrt[n]{n}$ geht nicht gegen 1, so gäbe es ein $\varepsilon > 0$ und dazu unendlich viele n, sodaß $\sqrt[n]{n} \geq (1+\varepsilon)$, also

$$n \geq (1+\varepsilon)^n = 1 + n\varepsilon + \frac{n(n-1)}{2}\varepsilon^2 + \cdots \geq \frac{n(n-1)}{2}\varepsilon^2,$$

also $1 \geq (n-1)\frac{\varepsilon^2}{2}$ für unendlich viele n, im Widerspruch zu Archimedes. □

§ 2. Konvergenzsätze

Sätze über Folgen lassen sich als Sätze über Reihen schreiben und umgekehrt. Der Satz zum Beispiel, daß die Grenzwertbildung mit Linearkombinationen verträglich, also *lim* linear ist, nimmt für Reihen folgende Gestalt an:

(2.1) Satz. *Sind die Reihen* $\sum_{k=1}^{\infty} c_k$ *und* $\sum_{k=1}^{\infty} d_k$ *konvergent und* $\lambda, \mu \in \mathbb{R}$, *so ist*

$$\sum_{k=1}^{\infty}(\lambda c_k + \mu d_k) \;=\; \lambda \sum_{k=1}^{\infty} c_k \;+\; \mu \sum_{k=1}^{\infty} d_k. \qquad \square$$

Für nicht konvergente Reihen definieren wir die rechte Seite der Formel durch die linke; dann handelt es sich wieder um eine Gleichung zwischen Folgen, nicht wie hier zwischen reellen Zahlen. Auch Definitionen übertragen sich: Eine Reihe heißt **beschränkt**, falls die Folge der Partialsummen beschränkt ist. Eine konvergente Reihe ist beschränkt.

Bisher ging es einfach zu: Wir haben die Konvergenz einer Folge oder Reihe bewiesen, indem wir den Grenzwert angegeben und die in der Definition der Konvergenz geforderte Abschätzung nachgewiesen haben. Jetzt aber kommen wir zu einer ganz neuen Art von Sätzen: sie behaupten die Existenz eines Grenzwertes, ohne ihn vorzuweisen. Ja, wir werden oft reelle Zahlen $e, \pi, \sqrt{2}, \ldots$ eben als solche Grenzwerte erst bestimmen und benennen.

Definition *(Monotonie). Eine Folge* (a_n) *heißt:*

$$\left.\begin{array}{l}\textbf{monoton wachsend}\\ \textbf{monoton fallend}\\ \textbf{streng monoton wachsend}\\ \textbf{streng monoton fallend}\end{array}\right\} \quad \begin{array}{c}\textit{wenn für}\\ \textit{alle } n \textit{ gilt:}\end{array} \quad \left\{\begin{array}{l} a_n \le a_{n+1},\\ a_n \ge a_{n+1},\\ a_n < a_{n+1},\\ a_n > a_{n+1}\,.\end{array}\right.$$

Für die zugeordnete Reihe bedeutet das, daß ihre Glieder jeweils ≥ 0, ≤ 0, > 0, < 0 sind. Reihen mit nicht negativen Gliedern nennen wir oft kurz **positiv**. Der Sprachgebrauch ist hier wie bei den natürlichen Zahlen etwas schwankend, und Konsequenz führt leicht zur Unbequemlichkeit. Die Nullen mögen gern unsere Aufmerksamkeit über Gebühr beanspruchen.

(2.2) Satz *(Monotone Konvergenz). Eine monoton wachsende nach oben beschränkte Folge konvergiert. Eine beschränkte Reihe mit nicht negativen Gliedern konvergiert. (Entsprechendes gilt für fallende Folgen und für Reihen mit Gliedern* ≤ 0 *).*

Beweis: Die zweite Behauptung ist eine Umformulierung der ersten, und die erste braucht man nur für wachsende Folgen zu zeigen (Übergang zum Negativen). Sei also $a_n \leq a_{n+1} \leq K$ für alle $n \in \mathbb{N}$. Sei $a = \sup\{a_n \mid n \in \mathbb{N}\}$. Dann ist $a_n \leq a$ für alle n, aber ist $\varepsilon > 0$, so ist $a_N \geq a - \varepsilon$ für ein $N \in \mathbb{N}$, und daher auch $a_n \geq a - \varepsilon$ für alle $n > N$ wegen der Monotonie. Folglich ist $a_n \in (a - \varepsilon, a]$ für fast alle n. □

Betrachten wir als Anwendung die Folge (a_n), die durch die Rekursionsformel

$$a_0 = a + 1, \quad a_{n+1} = a_n \left(1 + \frac{a - a_n^k}{k a_n^k}\right)$$

für ein festes $a > 0$ und ein festes $k \in \mathbb{N}$ definiert ist. Man sieht:

(i) $a_n > 0$, (ii) $a_n < a_{n-1}$, (iii) $a_n^k > a$.

Beweis durch Induktion nach n: Die Behauptungen, soweit sinnvoll, gelten für $n = 0$ und seien für ein n nun angenommen. Dann folgt $a_{n+1} < a_n$, weil $a - a_n^k < 0$, also (ii) für $n + 1$; $a_{n+1} > 0$,weil $k a_n^k + a - a_n^k > 0$, also (i) für $n + 1$; schließlich nach Bernoulli

$$a_{n+1}^k = a_n^k \left(1 + \frac{a - a_n^k}{k a_n^k}\right)^k \geq a_n^k \left(1 + \frac{k(a - a_n^k)}{k a_n^k}\right) = a,$$

also (iii) für $n + 1$. □

Demnach ist die Folge monoton fallend und durch 0 nach unten beschränkt, konvergiert also gegen eine Zahl $s \geq 0$, ihr Infimum.

Was wissen wir über diesen Grenzwert s? Es ist ja nach der Rekursionsformel

$$k\, a_{n+1} a_n^{k-1} = (k-1)a_n^k + a,$$

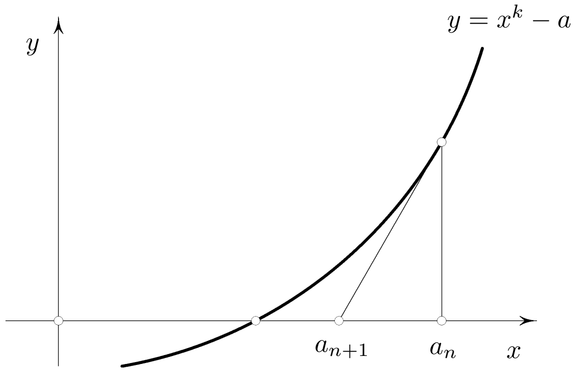
und gehen wir beidseits in allen Summanden und Faktoren zum Grenzwert über, was wir nach (II, 1.4) dürfen, so folgt $ks^k = (k-1)s^k + a$, also $s^k = a$; die Folge konvergiert demnach gegen eine Zahl s mit $s^k = a$. Daraus entnehmen wir insbesondere die

(2.3) Bemerkung. *Jede positive reelle Zahl besitzt zu jedem $k \in \mathbb{N}$ genau eine positive k-te Wurzel.*

Beweis: Die Existenz haben wir gerade gesehen, die Eindeutigkeit sieht man leicht: Ist $0 \le s < t$, so $0 \le s^k < t^k$. Ist also $0 \le s \le t$ und $s^k = t^k = a$, so $s = t$. □

Interessanter als der Beweis ist freilich, wie man auf die Folge kommt: Man sucht ja eine Nullstelle der Funktion $f(x) = x^k - a$. Ist a_n eine noch zu große Schätzung, so bestimmt man eine verbesserte Schätzung a_{n+1} so, daß

$$\frac{f(a_n)}{a_n - a_{n+1}} = f'(a_n).$$



In unserem Fall hat f die Steigung $f'(a_n) = k\,a_n^{k-1}$ am Punkt a_n. Der Beweis zeigt explizit, daß man bei Iteration des Verfahrens — welches nach Newton heißt — die Nullstelle als Grenzwert erhält. Hier ist, wie man sieht, noch viel zu lernen.

Die Beschränktheit einer positiven Reihe prüft man meist so nach, daß man die Reihe mit einer andern vergleicht, über deren Konvergenz man Bescheid weiß.

(2.4) Majorantenkriterium. *Sei $c_n \geq 0$ für alle n. Eine Reihe $\sum_{n=1}^{\infty} a_n$ heißt* **Majorante** *von $\sum_{n=1}^{\infty} c_n$, falls $c_n \leq a_n$ für alle n.*

Besitzt eine Reihe mit nicht negativen Gliedern eine konvergente Majorante, so konvergiert sie.

Beweis: Mit Bezeichnungen des Satzes ist

$$\sum_{n=1}^{k} c_n \leq \sum_{n=1}^{k} a_n \leq \sup\Bigl\{\sum_{n=1}^{k} a_n \mid k \in \mathbb{N}\Bigr\} = \sum_{n=1}^{\infty} a_n.$$

Also ist $\sum_{n=1}^{\infty} c_n$ beschränkt und folglich konvergent. □

Das Konvergenzverhalten ändert sich übrigens nicht, wenn man endlich viele Glieder der Reihe ändert, denn die Partialsummen von $\sum_{k=1}^{\infty} c_k$ und $\sum_{k=K}^{\infty} c_k$ unterscheiden sich für $n \geq K$ um die Konstante $\sum_{k=1}^{K-1} c_k$, also $\sum_{k=1}^{\infty} c_k = \sum_{k=1}^{K-1} c_k + \sum_{k=K}^{\infty} c_k$, die rechte Seite konvergiert genau wenn die linke konvergiert.

Das Majorantenkriterium hat viele Anwendungen, z.B.:

(2.5) Verdichtungslemma von Cauchy. *Sei (c_n) eine nicht negative reelle monoton fallende Folge, dann konvergiert $\sum_{n=1}^{\infty} c_n$ genau dann, wenn $\sum_{k=1}^{\infty} 2^k \cdot c_{2^k}$ konvergiert.*

Beweis: Wir schätzen die 2^k Summanden c_n mit $2^k \le n \le 2^{k+1}-1$ alle durch den ersten c_{2^k} von oben ab und erhalten

$$\sum_{n=2^k}^{2^{k+1}-1} c_n \le 2^k c_{2^k}.$$

Also wenn $\sum_k 2^k c_{2^k}$ beschränkt ist, so auch $\sum_n c_n$. Umgekehrt schätzen wir die 2^{k-1} Summanden c_n mit $2^{k-1}+1 \le n \le 2^k$ alle von unten durch den letzten c_{2^k} ab und erhalten

$$2 \cdot \sum_{n=2^{k-1}+1}^{2^k} c_n \ge 2 \cdot 2^{k-1} c_{2^k} = 2^k c_{2^k}.$$

Also wenn $\sum_n c_n$ beschränkt ist, so auch $\sum_k 2^k c_{2^k}$. □

(2.6) Anwendung. *Die Reihe* $\sum_{n=1}^{\infty} \frac{1}{n^\alpha}$ *ist genau dann konvergent, wenn* $\alpha > 1$ *ist.*

Beweis: Die Reihe konvergiert nicht für $\alpha \le 0$, weil $(\frac{1}{n^\alpha})$ dann keine **Nullfolge** ist (d.h. nicht gegen Null konvergiert). Für $\alpha > 0$ ist das Lemma anwendbar, und wir haben die geometrische Reihe $\sum 2^k/2^{k\alpha} = \sum (2^{1-\alpha})^k$ zu untersuchen. Diese konvergiert genau wenn $2^{1-\alpha} < 1$, also wenn $2 < 2^\alpha$, und dies bedeutet $\alpha > 1$. Für nicht rationale α ist zwar vorläufig x^α jedenfalls nicht definiert, aber das werden wir schon so nachholen, daß dies ein allgemein gültiger Beweis ist. □

Insbesondere **divergiert** die **Harmonische Reihe**

$$\sum_{n=1}^{\infty} \frac{1}{n},$$

obwohl die Folge $(1/n)$ ihrer Glieder eine Nullfolge ist.

Die geometrische Reihe ist die wichtigste Vergleichsreihe bei Anwendungen des Majorantenkriteriums. Klassische Anwendungen sind die folgenden:

(2.7) Quotientenkriterium. *Sei $c_n \geq 0$, und es existiere eine Zahl $\vartheta < 1$, sodaß $c_{n+1} \leq \vartheta c_n$ für fast alle n. Dann konvergiert $\sum_{n=0}^{\infty} c_n$.*

Beweis: Weil es auf endlich viele Glieder nicht ankommt, sei oBdA $c_{n+1} \leq \vartheta c_n$ für alle n, dann folgt induktiv $c_n \leq \vartheta^n c_0$, also hat $\sum c_n$ die Majorante $c_0 \sum \vartheta^n$, die wegen $\vartheta < 1$ konvergiert. □

Anwendung. *Die* **Exponentialreihe**

$$e^x := \sum_{n=0}^{\infty} \frac{x^n}{n!}$$

konvergiert für alle $x \geq 0$. Sie wird uns noch oft begegnen, sie konvergiert tatsächlich für alle x, siehe unten (2.14).

Beweis: Das Quotientenkriterium ist schlüssig: $c_{n+1}/c_n = \frac{x}{n+1} \to 0$ für $n \to \infty$, also $c_{n+1} < \frac{1}{2} c_n$ für fast alle n. □

Bei der harmonischen Reihe erhalten wir $c_{n+1}/c_n = \frac{n}{n+1} < 1$, aber wie wir wissen gilt $\left(\frac{n}{n+1}\right) \to 1$, es gibt keine Zahl $\vartheta < 1$ **unabhängig** von n, sodaß $c_{n+1}/c_n < \vartheta$. Dasselbe finden wir aber auch bei der Reihe $\sum \frac{1}{n^2}$. Hier ist $c_{n+1}/c_n = \frac{n^2}{(n+1)^2} = \left(\frac{n}{n+1}\right)^2 \to 1$, und diese Reihe konvergiert. In diesen Fällen lehrt das Quotientenkriterium nichts.

(2.8) Wurzelkriterium. *Sei $c_n \geq 0$, und es existiere eine Zahl $\vartheta < 1$, sodaß $\sqrt[n]{c_n} \leq \vartheta$ für fast alle n, dann konvergiert $\sum_{n=0}^{\infty} c_n$. Ist $\sqrt[n]{c_n} \geq 1$ für unendlich viele n, so divergiert $\sum c_n$.*

Beweis: Die zweite Behauptung ist trivial, es folgt ja $c_n \geq 1$. Ist oBdA $\sqrt[n]{c_n} \leq \vartheta$, so ist $c_n \leq \vartheta^n$ und $\sum \vartheta^n$ eine Majorante. □

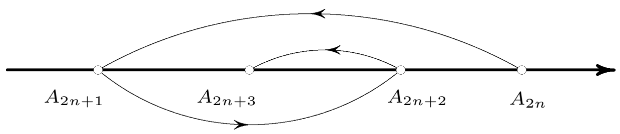
Auch dieses Kriterium ist für die Reihen $\sum_{n=1}^{\infty} \frac{1}{n^\alpha}$ nicht schlüssig. Ist das Quotientenkriterium schlüssig, so auch das Wurzelkriterium, denn ist $c_{n+1} \leq \vartheta c_n$, $\vartheta < 1$, also $c_n \leq \vartheta^n c_0$, so ist $\sqrt[n]{c_n} \leq \vartheta \cdot \sqrt[n]{c_0}$, und $\sqrt[n]{c_0} \to 1$ für $c_0 > 0$. Das Quotientenkriterium ist jedoch oft leichter anzuwenden, weil man Quotienten leichter berechnet als n-te Wurzeln.

Bisher haben wir Reihen mit nicht negativen Gliedern betrachtet. Ihr Konvergenzverhalten lehrt auch viel über beliebige Reihen, wie wir bald sehen werden. Zuvor jedoch bemerken wir das klassische

(2.9) Leibniz-Kriterium. *Ist* (a_k) *eine positive monoton fallende Folge, so heißt die Reihe* $\sum_{k=0}^{\infty}(-)^k a_k$ **alternierend**. *Sie konvergiert genau wenn* (a_k) *eine Nullfolge ist.*

Beweis: Sei $A_n = \sum_{k=0}^{n}(-)^k a_k$ die n-te Partialsumme. Wir schreiben kurz $(-)^k := (-1)^k$. Es gilt:

(i) $A_{2n} \geq A_{2n} - a_{2n+1} + a_{2n+2} = A_{2n+2}$,
(ii) $A_{2n+1} \leq A_{2n+1} + a_{2n+2} - a_{2n+3} = A_{2n+3}$,
(iii) $A_{2n} \geq A_{2n} - a_{2n+1} = A_{2n+1}$,
(iv) $A_{2n} - A_{2n+1} = a_{2n+1} \to 0$.



Die Folge (A_{2n}) ist nach (i) monoton fallend, und nach (iii), (ii) ist $A_{2n} \geq A_{2n+1} \geq A_{2n-1} \ldots \geq A_1$, also ist (A_{2n}) nach unten beschränkt, daher $(A_{2n}) \to a$. Nach (iv) gilt auch $(A_{2n+1}) = (A_{2n}) - (a_{2n+1}) \to a - 0 = a$, und aus beidem zusammen folgt $(A_n) \to a$, denn in der Tat, schließlich mal ist $|A_{2n} - a| < \varepsilon$ und $|A_{2n+1} - a| < \varepsilon$. □

Wir bemerken unter der Hand: Wenn (a_{2n}) und (a_{2n+1}) gegen die gleiche Zahl konvergieren, so konvergiert (a_n) gegen dieselbe.

Anwendung. *Die folgenden Reihen sind konvergent (die Angabe des Grenzwertes beweisen wir freilich noch nicht):*

$$1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + - \cdots = \log 2,$$
$$1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + - \cdots = \frac{\pi}{4}.$$

Die bisher betrachteten Kriterien sind oft nützlich, aber oft auch nicht schlüssig. Es gibt aber eine wichtige allgemeine Charakterisierung konvergenter Folgen, die nicht benutzt, daß man etwa den Grenzwert kennt — das ist ja das Handicap bei der Definition der Konvergenz. Zuvor ein technisches

(2.10) Lemma. *Jede Folge besitzt eine monotone Teilfolge.*

Beweis: Sei (a_n) die Folge. Eine Stelle $n \in \mathbb{N}$ heiße "niedrig", wenn $a_{n+k} \geq a_n$ für alle $k \in \mathbb{N}$ ist. Zwei Fälle sind möglich:

1. Fall: Es gibt unendlich viele niedrige Stellen (nach jedem $N \in \mathbb{N}$ noch ein niedriges n). Dann bilden die Folgenglieder a_n an den niedrigen Stellen n eine monoton wachsende Teilfolge.

2. Fall: Nach einem $N \in \mathbb{N}$ kommt kein niedriges n mehr. Zu jedem $n > N$ gibt es dann ein $k \in \mathbb{N}$ mit $a_{n+k} < a_n$, weil n ja nicht "niedrig" ist. Daher findet man induktiv eine monoton fallende Teilfolge $(a_{n_\ell} \mid \ell \in \mathbb{N})$. □

(2.11) Satz (Bolzano-Weierstraß). *Jede beschränkte Folge besitzt eine konvergente Teilfolge.*

Beweis: Wähle eine monotone Teilfolge nach dem Lemma; sie konvergiert. □

Ein Punkt $a \in \mathbb{R}$ heißt **Häufungspunkt** der Folge (a_n), wenn es eine Teilfolge von (a_n) gibt, die gegen a konvergiert. Eine beschränkte Folge besitzt also stets einen Häufungspunkt. Eine konvergente Folge besitzt genau einen Häufungspunkt, ihren Grenzwert. Eine beschränkte Folge ist genau dann konvergent, wenn sie nur einen Häufungspunkt besitzt. Ist nämlich a der Häufungspunkt und $\varepsilon > 0$, und sind nicht fast alle a_n in $(a-\varepsilon, a+\varepsilon)$, so gibt es eine Teilfolge $a_{n_k} \notin (a-\varepsilon, a+\varepsilon)$, die einen anderen Häufungspunkt als a hat. Dieser wäre auch Häufungspunkt von (a_n).

Bemerkung. *Die Zahl a ist Häufungspunkt von (a_n) genau wenn jede Umgebung von a unendlich viele a_n enthält.*

Beweis:
$\Rightarrow$: eine ε-Umgebung U von a enthält a_{n_k} für fast alle k, falls $(a_{n_k} \mid k \in \mathbb{N}) \to a$.
$\Leftarrow$: Wähle $\varepsilon = \frac{1}{k}$, und $n_k > n_{k-1}$, so daß $a_{n_k} \in U_\varepsilon(a)$, dann folgt $(a_{n_k}) \to a$, also a ist Häufungspunkt von (a_n). □

Nimmt man Punkte ∞, $-\infty$ zu $\mathbb{R}$ hinzu, so besitzt jede Folge eine konvergente Teilfolge, also einen Häufungspunkt, und eine Folge ist genau dann konvergent, wenn sie genau einen Häufungspunkt besitzt, wenn man auch Konvergenz gegen ∞ bzw. $-\infty$ zuläßt. Die ε-Umgebungen von ∞ bzw. $-\infty$ sind die Mengen $(\varepsilon, \infty] = \{x \mid x > \varepsilon\}$ bzw. $[-\infty, -\varepsilon) = \{x \mid x < -\varepsilon\}$ für beliebig große ε.

Für eine Folge (a_n) heißt die Zahl

$$c = \overline{\lim}\,(a_n) := \inf\{x \mid a_n \leq x \text{ für fast alle } n\} =: \limsup\,(a_n)$$

der **Limes superior** oder **obere Häufungspunkt** von (a_n), und in der Tat ist c der größte Häufungspunkt von (a_n) — wenn man $\infty, -\infty$ zu den Punkten hinzunimmt. Offenbar gibt es keinen grösseren, größer als $c+\varepsilon$, man muß also nur sehen, daß c Häufungspunkt ist. Nun sind nicht fast alle $a_n \leq c-\varepsilon$, also ∞ viele $a_n > c-\varepsilon$, und fast alle $a_n \leq c+\varepsilon$, also ∞ viele $a_n \in U_\varepsilon(c)$, was zu zeigen war.

Man findet entsprechend, daß der **Limes inferior** oder auch **untere Häufungspunkt**

$$\underline{\lim}\,(a_n) := \sup\{x \mid a_n \geq x \text{ für fast alle } n\} =: \liminf\,(a_n)$$

der minimale Häufungspunkt von (a_n) ist. Ist $\varepsilon > 0$ beliebig klein, so ist

$$\underline{\lim}\,(a_n) - \varepsilon < a_n < \overline{\lim}\,(a_n) + \varepsilon$$

für fast alle n, und (a_n) ist genau dann konvergent, wenn $\underline{\lim}\,(a_n) = \overline{\lim}\,(a_n)$. Beachte aber, daß es üblich ist, für $\overline{\lim}$, $\underline{\lim}$ alle Punkte aus

$$\bar{\mathbb{R}} := \mathbb{R} \cup \{\infty, -\infty\}$$

zuzulassen, während man eine Folge nur **konvergent** nennt, wenn sie gegen *eine Zahl aus* $\mathbb{R}$ konvergiert, also auch Häufungspunkte nur in $\mathbb{R}$ betrachtet, es sei denn daß man ausdrücklich etwas anderes sagt.

Soweit war das angekündigte Allgemeine nur eine Anhäufung von neuen Wörtern, Vokabelnlernen, aber jetzt betritt eine überaus wichtige Beschreibung konvergenter Folgen die Bühne (mit ihrem vertrauten Pseudonym, Cauchy war nicht der erste, der den Satz formuliert hat, und beweisen konnte er ihn auch nicht wirklich zureichend).

(2.12) Cauchy-Kriterium für Folgen. *Eine Folge* (a_n) *ist genau dann konvergent, wenn gilt: Zu jedem* $\varepsilon > 0$ *existiert ein* $m \in \mathbb{N}$, *sodaß für alle* $k \in \mathbb{N}$

$$|a_{m+k} - a_m| < \varepsilon.$$

Beweis: Ist $\lim (a_n) = a$, so existiert ein $m \in \mathbb{N}$, sodaß für alle $k \in \mathbb{N}_0$ gilt $|a_{m+k} - a| < \varepsilon/2$, also

$$|a_m - a_{m+k}| \ \leq \ |a_{m+0} - a| + |a - a_{m+k}| < \varepsilon.$$

Ist umgekehrt $|a_{m+k} - a_m| < \varepsilon$ für alle k zu einem festen m, so ist die Folge $(a_{m+k} \mid k \in \mathbb{N})$ beschränkt, und daher auch die Folge (a_n). Eine Teilfolge (a_{n_ℓ}) konvergiert nach Bolzano-Weierstraß gegen ein a, und wir zeigen $a = \lim (a_n)$.
Sei also $\varepsilon > 0$, dann ist für ein geeignetes m

(i) $|a_m - a_{m+k}| < \varepsilon/3$ für alle k. Insbesondere ist also

(ii) $|a_m - a_{n_\ell}| < \varepsilon/3$ für fast alle ℓ, nämlich alle, für die $n_\ell > m$.
Aber auch für fast alle ℓ ist

(iii) $|a_{n_\ell} - a| < \varepsilon/3$.

Aus (ii), (iii) also $|a_m - a| < \frac{2\varepsilon}{3}$, und hieraus mit (i)

$$|a - a_{m+k}| < \varepsilon \quad \text{für alle } k. \qquad \square$$

Aus der Ungleichung im Cauchy-Kriterium folgt übrigens

$$|a_{m+k} - a_{m+\ell}| < 2\varepsilon.$$

Daher hätte man, statt $|a_{m+k} - a_m| < \varepsilon$, auch $|a_k - a_\ell| < \varepsilon$ für alle $k, \ell > m$ setzen können, was man auch oft findet.

(2.13) Cauchy-Kriterium für Reihen. *Eine Reihe* $\sum_{n=0}^{\infty} c_n$ *ist genau dann konvergent, wenn es zu jedem* $\varepsilon > 0$ *ein* $m \in \mathbb{N}$ *gibt, sodaß für alle* $k \in \mathbb{N}$

$$\Big| \sum_{n=m}^{m+k} c_n \Big| < \varepsilon. \qquad \square$$

(2.14) Absolute Konvergenz. *Eine Reihe* $\sum_{k=0}^{\infty} c_k$ *heißt absolut konvergent, wenn die Reihe der Beträge* $\sum_{k=0}^{\infty} |c_k|$ *konvergiert. Eine absolut konvergente Reihe ist konvergent.*

Beweis: aus dem Cauchy-Kriterium und der Dreiecksungleichung:

$$\left|\sum_{n=m}^{m+k} c_n\right| \leq \sum_{n=m}^{m+k} |c_n| \, .$$

Die Reihe konvergiert absolut, wenn die rechte Seite fr̈ große m stets $< \varepsilon$ ist, und sie konvergiert, wenn dasselbe für die linke Seite gilt.

$\square$

Ist (c_n) beliebig, so heißt $\sum a_n$ eine **Majorante** von $\sum c_n$, wenn $\sum a_n$ eine Majorante der positiven Reihe $\sum |c_n|$ ist. Die Konvergenzkriterien für positive Reihen liefern unmittelbar Kriterien für absolute Konvergenz, man muß sie nur auf die Reihe $\sum |c_n|$ anwenden.

Beispiel. *Ist* (a_n) *beschränkt, so konvergiert die Reihe*

$$\sum_{k=0}^{\infty} a_k x^k$$

für alle $x \in \mathbb{R}$ *mit* $|x| < 1$ *absolut.*

Beweis: $\sum |a_k x^k|$ hat die Majorante $A \cdot \sum |x|^k$, die für $|x| < 1$ konvergiert. $\square$

Insbesondere konvergieren alle unendlichen Dezimalentwicklungen

$$\pm \sum_{k=-m}^{\infty} a_k 10^{-k}, \quad a_k \in \{0, 1, \ldots, 9\},$$

und allgemeiner gilt:

(2.15) Satz (*b-al-Zahl-Entwicklung*). *Sei* $b \geq 2$ *eine natürliche Zahl, dann läßt sich jede reelle Zahl* x *in der Form*

$$x = \pm \sum_{k=-m}^{\infty} a_k b^{-k}, \quad a_k \in \{0, \ldots, b-1\},$$

darstellen. Die Darstellung ist eindeutig, wenn man ausschließt, daß $a_k = b-1$ *für fast alle* k.

Beweis: OBdA ist $x \geq 0$. Man bestimmt a_ℓ induktiv so, daß $a_\ell = 0$ falls $b^{-\ell} > x$ (also für kleine ℓ), und dann a_ℓ maximal, sodaß $\sum_{k=-m}^{\ell} a_k b^{-k} \leq x$. Mit andern Worten: die ℓ-te Partialsumme der b-alen Entwicklung von x ist die größte b-ale Zahl mit ℓ Stellen hinter dem Komma, die noch nicht größer als x ist. Dann folgt induktiv sofort $|x - \sum_{k=-m}^{\ell} a_k b^{-k}| < b^{-\ell}$, also $x = \sum_{k=-m}^{\infty} a_k b^{-k}$.

Hat man zwei b-ale Entwicklungen

$$x = \sum_{-m}^{\infty} a_k b^{-k} = \sum_{-m}^{\infty} c_k b^{-k},$$

sodaß beidemal nicht fast immer $a_k = b-1$ bzw. $c_k = b-1$, so sei ℓ die erste abweichende Stelle, und oBdA $a_\ell > c_\ell$, dann ist $\sum_{-m}^{\infty} a_k b^{-k} \geq \sum_{-m}^{\ell} a_k b^{-k}$, und

$$\sum_{-m}^{\infty} c_k b^{-k} < \sum_{-m}^{\ell-1} a_k b^{-k} + (a_\ell - 1) b^{-\ell} + \sum_{k=\ell+1}^{\infty} (b-1) b^{-k} = \sum_{-m}^{\ell} a_k b^{-k}. \quad ✠$$

$\square$

Wir rechnen mit $b = 10$, weil wir 10 Finger haben. Manchen Vorzug hätte $b = 12$, weil 12 viele Teiler hat. Rechenmaschinen rechnen mit $b = 2$, Dualzahlen, weil alles in Zellen kodiert ist, die nur zwei Zustände haben. Wer viel mit Computern arbeitet, rechnet daher gerne mit $b = 2^3 = 8$ oder $b = 2^4 = 16$, weil dann die Umrechnung in Dualzahlen leichtfällt, man aber auch nahe bei $b = 10$ bleibt.

Für die Begründung der Axiome des reellen Zahlkörpers sind b-ale Zahlentwicklungen nicht geeignet, weil zum Beispiel die Summe zweier solcher Entwicklungen nicht wieder die vorgeschriebene Gestalt hat. Gangbar ist jedoch folgender Weg: Man sagt, eine reelle Zahl ist durch eine Cauchy-Folge rationaler Zahlen gegeben, wobei man das ε im Cauchy-Kriterium auch auf rationale Zahlen beschränkt — das ist ja kein Verlust. Und zwei rationale Cauchy-Folgen sollen dieselbe reelle Zahl bedeuten, wenn ihre Differenz eine Nullfolge, d.h. eine gegen 0 konvergente Folge ist. Wie man dann die algebraischen Verknüpfungen erklärt, ist klar, und man nennt eine rationale Cauchy-Folge (a_n) **positiv**, wenn ein $\varepsilon > 0$ in $\mathbb{Q}$ existiert, sodaß $a_n > \varepsilon$ für fast alle n. Positive Folgen definieren die positiven reellen Zahlen. Mit diesen Erklärungen und Vertrauen in die Logik ist es nicht schwer, die Axiome zu bestätigen. Auch zeigt man nacheinander, daß durch die Axiome $\mathbb{N}$ und damit der Ring $\mathbb{Z}$, der Körper $\mathbb{Q}$ und schließlich $\mathbb{R}$ mit algebraischer Struktur und dadurch bestimmter Anordnung bis auf Isomorphie eindeutig bestimmt sind. Es gibt bis auf Umbenennung einen und nur einen Körper $\mathbb{R}$, der den mitgeteilten Axiomen genügt. Das wollen wir nicht weiter verfolgen.

Statt "b-al" sagt man vielfach auch "b-adisch", aber das scheint mir eher verwirrend, weil das Wort in der Zahlentheorie etwas äußerlich Ähnliches, tatsächlich aber sehr Verschiedenes bedeutet.

Für absolut konvergente Reihen gelten sehr starke Versionen des kommutativen Gesetzes:

(2.16) Umordnungssatz. *Sei $\sum_{k=1}^{\infty} c_k$ absolut konvergent, und sei $\rho : \mathbb{N} \to \mathbb{N}$ bijektiv, dann konvergiert $\sum_{k=1}^{\infty} c_{\rho(k)}$ absolut gegen denselben Grenzwert.*

Beweis: Sei $a_n = \sum_{k=1}^{n} c_k$, $b_n = \sum_{k=1}^{n} c_{\rho(k)}$, $\varepsilon > 0$, und m so groß gewählt, daß $\sum_{k=m}^{m+\ell} |c_k| < \varepsilon$ für alle $\ell \in \mathbb{N}$. Für genügend große $r \in \mathbb{N}$ ist dann zugleich $r > m$ und $\rho(r) > m$.

Wählen wir nun n so groß, daß diese Bedingung für alle $r \geq n$ erfüllt ist, so ist b_n eine Summe von Gliedern c_j, worunter jedenfalls alle $c_1, \dots, c_m$ auftreten. Dasselbe gilt für a_n, und die Differenz ist folglich eine Summe von Gliedern $\pm c_j$, mit $j > m$. Daher

$$|a_n - b_n| \leq \sum_{j>m} |c_j| < \varepsilon.$$

Das zeigt schon, daß (b_n) gegen denselben Grenzwert wie (a_n) konvergiert. Wendet man das schon Gezeigte auf die Reihe $\sum |c_n|$ an, so folgt die absolute Konvergenz der umgeordneten Reihe. □

Für nicht absolut konvergente Reihen gilt das gerade Gegenteil:

(2.17) Satz. *Ist $\sum_{k=1}^{\infty} c_k$ konvergent aber nicht absolut konvergent und $x \in \mathbb{R}$ beliebig, so existiert eine Umordnung $\rho : \mathbb{N} \to \mathbb{N}$, sodaß $\sum_{k=1}^{\infty} c_{\rho(k)} = x$.*

Beweis: Sei (a_k) die Folge der positiven und (b_k) die Folge der negativen Glieder c_k der betrachteten Reihe. Weil $\sum c_k$ konvergiert, aber nicht absolut, gilt offenbar

$$(a_k) \to 0, \qquad (b_k) \to 0,$$
$$\Big(\sum_{k=1}^{n} a_k\Big) \to \infty, \quad \Big(\sum_{k=1}^{n} b_k\Big) \to -\infty.$$

Wir wählen jetzt als Folge $c_{\rho(\ell)}$ rekursiv jeweils das nächste noch nicht gewählte Glied der Folge (a_k) bzw. (b_k), je nach dem

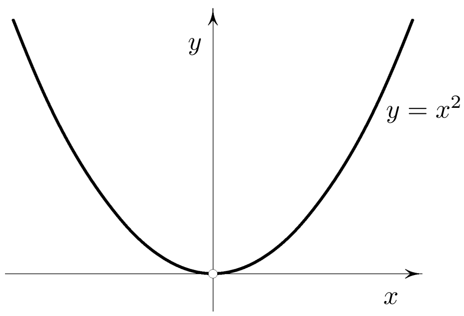
$$d_{\ell-1} := \sum_{k=1}^{\ell-1} c_{\rho(k)} < x \quad \text{oder} \quad \geq x.$$

Es ist dann schließlich $|c_{\rho(\ell)}| < \varepsilon$ für $\ell \geq L$, und ist etwa $d_L < x$, $d_{L+k} \geq x$, so ist $|d_\ell - x| < \varepsilon$ für $\ell > L + k$. □

Man könnte mit ähnlichem Argument auch $x \in \{\pm\infty\}$ wählen. Später in der allgemeinen Maßtheorie wird sich mit ergeben, daß für absolut konvergente Reihen ein noch stärkerer Umordnungssatz als (2.16) gilt (siehe Bd. 2, Aufg. 12 zu Kap. III).

§ 3. Stetige Funktionen

Eine Funktion $f : \mathbb{R} \to \mathbb{R}$ heißt stetig, wenn sie keine Sprünge macht. Mit dieser Erklrung hat man sich lange begnügt, und wir wollen das auch nicht schlechtmachen. Zum Beispiel die Funktion $f(x) = x^2$ ist demnach stetig.

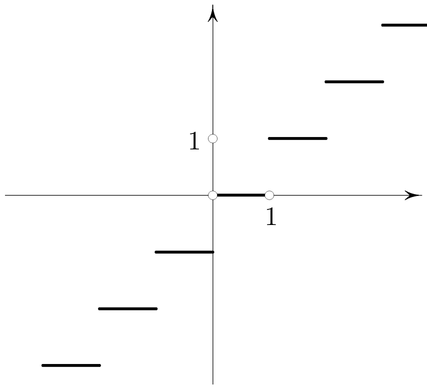


Ursprünglich hat man überhaupt nur solche Funktionen betrachtet, die sich in einer Formel hinschreiben lassen, doch dann hat die Entwicklung der Analysis selbst die Möglichkeiten, Formeln zu erzeugen, ins Ungewisse ausgeweitet, und jetzt lassen wir ganz Beliebiges zu.

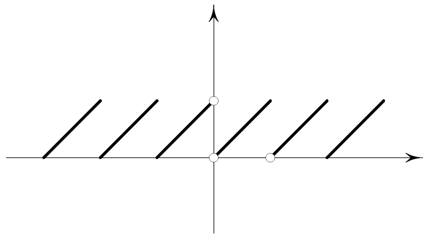
Die Funktion

$$x \mapsto [x] := \max\{k \in \mathbb{Z} \mid k \leq x\},$$

der **ganzzahlige Anteil** von x, ist unstetig an den Punkten $k \in \mathbb{Z}$.



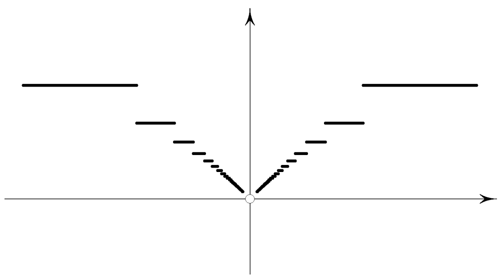
Ebenso die Funktion $f(x) = x - [x]$, die Sägezahnfunktion.



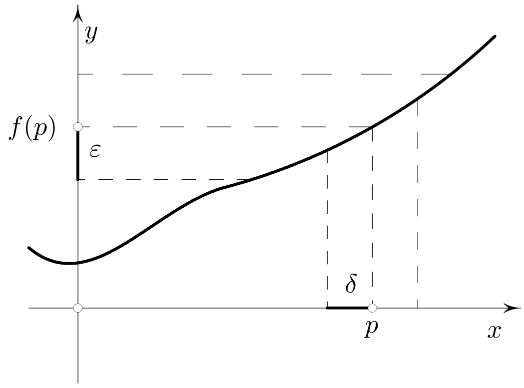
Wie aber steht es mit der Funktion

$$f(x) = \tfrac{1}{n} \quad \text{für} \quad \tfrac{1}{n} \leq |x| < \tfrac{1}{n-1}, \quad f(0) = 0,$$

im Nullpunkt?



Hier ist schon ein genauer und formaler Begriff der Stetigkeit nötig. Die Vorstellung, die uns leitet, ist: Wenn sich x nur wenig ändert, so ändert sich auch $f(x)$ nur wenig, und zwar ändert sich $f(x)$ *beliebig wenig*, wenn sich x *genügend wenig* ändert. "Beliebig wenig" heißt: Wenn $\varepsilon > 0$ beliebig vorgegeben ist, so ändert sich f um weniger als ε, falls sich dafür x "genügend wenig" ändert, also weniger als ein geeignetes $\delta > 0$. So kommen wir zu folgender



Definition *(Stetigkeit). Sei $D \subset \mathbb{R}$ und $f : D \to \mathbb{R}$ eine Funktion mit* **Definitionsgebiet** *D, und sei $p \in D$. Dann heißt f* **stetig am Punkt** *p, wenn es zu jedem $\varepsilon > 0$ ein $\delta > 0$ gibt, sodaß für alle $x \in D$ gilt:*

$$|x - p| < \delta \Longrightarrow |f(x) - f(p)| < \varepsilon.$$

Die Funktion f heißt **stetig**, *falls sie an jedem Punkt $p \in D$ stetig ist. Statt "am Punkt p" sagt man auch "an p" oder "bei p".*

Wir können uns diese Definition noch etwas mundlicher zurichten. Dazu erinnern wir an folgende Sprechweise der Mengenlehre: Gegeben sei eine Abbildung $f : M \to N$.

Ist $A \subset M$, so ist $f(A) := \{f(x) \mid x \in A\} =$ **Bild** der Menge A.

Ist $B \subset N$, so ist $f^{-1}(B) := \{x \in M \mid f(x) \in B\} =$ **Urbild** der Menge B.

Die letzte Implikation in der Definition der Stetigkeit können wir dann mit Umgebungen so aussprechen:

$$x \in U_\delta(p) \Longrightarrow f(x) \in U_\varepsilon\big(f(p)\big),$$

und das heißt

$$fU_\delta(p) \subset U_\varepsilon\big(f(p)\big) \quad \text{oder} \quad U_\delta(p) \subset f^{-1}U_\varepsilon\big(f(p)\big).$$

Hierbei meinen wir mit $U_\delta(p)$ die δ-Umgebung von p im Definitionsgebiet D, also den Durchschnitt einer δ-Umgebung in $\mathbb{R}$ mit D. Demnach also ist f stetig bei p, wenn das Urbild einer jeden Umgebung von $f(p)$ stets eine Umgebung von p in D enthält. Etwas salopper sagt man wohl auch: Wenn x gegen p geht, geht $f(x)$ gegen $f(p)$. Dies rechtfertigt sich durch den

(3.1) Satz *(Folgenstetigkeit). Genau dann ist $f : D \to \mathbb{R}$ stetig bei $p \in D$, wenn folgendes gilt:*
Für jede Folge $(x_n) \to p$, $x_n \in D$, gilt $\big(f(x_n)\big) \to f(p)$.

Beweis: Zwei Richtungen sind zu zeigen. Sei also f stetig bei p und (x_n) eine Folge in D mit $(x_n) \to p$. Wir müssen $\big(f(x_n)\big) \to f(p)$ zeigen. Ist $\varepsilon > 0$ gegeben, so wähle dazu δ nach der Definition der Stetigkeit. Dann ist schließlich $x_n \in U_\delta(p)$, weil $(x_n) \to p$. Also $f(x_n) \in U_\varepsilon\big(f(p)\big)$. Das zeigt $\big(f(x_n)\big) \to f(p)$. Nun sei umgekehrt die Bedingung über Folgen im Satz erfüllt. Angenommen, f ist unstetig bei p. Dann gilt: Es gibt ein $\varepsilon > 0$, sodaß für alle $\delta > 0$ ein $x \in D$ existiert, mit $|x - p| < \delta$ und $|f(x) - f(p)| \geq \varepsilon$. So nämlich lautet die Negation der Stetigkeitsdefinition.

Nun, was für alle δ gilt, gilt insbesondere für $\delta = \frac{1}{n}$, $n \in \mathbb{N}$. Also finden wir ein $\varepsilon > 0$ und dazu eine Folge (x_n) in D, mit

$$|x_n - p| < 1/n \quad \text{und} \quad |f(x_n) - f(p)| \geq \varepsilon$$

für alle $n \in \mathbb{N}$. Die erste Ungleichung sagt $(x_n) \to p$ nach Archimedes, nach der zweiten aber konvergiert $\big(f(x_n)\big)$ nicht gegen $f(p)$, im Widerspruch zur Folgenbedingung. Damit ist die Annahme, f sei unstetig bei p, widerlegt. ✠ $\square$

Was wir über Konvergenz wissen, läßt sich mit diesem Satz oft auch als Aussage über Stetigkeit deuten. So zum Beispiel (II, 1.4):

(3.2) Satz *(über rationale Operationen). Sind die beiden Funktionen $f, g : D \to \mathbb{R}$ stetig am Punkt $p \in D$, und sind $\lambda, \mu \in \mathbb{R}$, dann sind auch die Funktionen $\lambda f + \mu g$ und $f \cdot g$ stetig am Punkt p. Ist $f(x) \neq 0$ für alle $x \in D$, so ist auch $1/f$ stetig bei p.* $\square$

Dabei sind die rationalen Operationen im Satz durch das Entsprechende für die Werte definiert, also

$$\begin{aligned}(\lambda f + \mu g)(x) &:= \lambda f(x) + \mu g(x),\\ (f \cdot g)(x) &:= f(x) \cdot g(x),\\ (1/f)(x) &:= 1/f(x).\end{aligned}$$

Ist übrigens $f(p) \neq 0$ und f stetig bei p, so ist $f(x) \neq 0$ in einer Umgebung um p, oder wie man sagt: **lokal** um p. Wählt man nämlich $\varepsilon = |f(p)|$ und dazu δ nach der Stetigkeitsdefinition, und ist $x \in U_\delta(p)$, so ist $|f(p)| - |f(x)| \leq |f(p) - f(x)| < \varepsilon = |f(p)|$, also $|f(x)| > 0$.

Offenbar ist die **identische Abbildung**

$$\mathrm{id} : \mathbb{R} \to \mathbb{R}, \quad x \mapsto x$$

stetig, wähle $\delta = \varepsilon$. Man bezeichnet diese Funktion kurzerhand mit x, wie man ja überhaupt eine Funktion durch den Term bezeichnet, der sagt, wo x hingeht. Aus (3.2) folgt dann, daß jedes **Polynom**

$$f(x) = \sum_{k=0}^{n} a_k\, x^k$$

überall stetig ist. Ist $a_n \neq 0$, so heißt a_n der **Leitkoeffizient** und n der **Grad** des Polynoms f, und alle Polynome aller Grade bilden den reellen **Polynomring** $\mathbb{R}[x]$. Sind f, g zwei Polynome und $g \neq 0$, womit man meint, daß nicht alle Koeffizienten von g verschwinden, so hat man die

rationale Funktion: $\dfrac{f}{g}\,, \qquad f, g \in \mathbb{R}[x], \qquad g \neq 0\,.$

Diese definiert auf der Menge $D = \{x \mid g(x) \neq 0\}$ eine stetige Funktion. Es ist nicht schwer zu sehen, daß ein Polynom vom Grad n nur höchstens n Nullstellen haben kann.

Stetigkeit ist eine **lokale Eigenschaft** von Funktionen, d.h. man braucht f nur in einer beliebig kleinen Umgebung von p zu kennen, um zu entscheiden, ob f am Punkt p stetig ist.

(3.3) Satz. *Eine Zusammensetzung stetiger Funktionen ist stetig.*

Beweis: Seien $C, D \subset \mathbb{R}$ und seien Funktionen

$$C \xrightarrow{f} D \xrightarrow{g} \mathbb{R}, \quad g \circ f\ (x) := g\big(f(x)\big), \qquad f(p) = q,$$

gegeben. Die Behauptung meint: Ist f stetig bei p und g stetig bei $q = f(p)$, so ist $g \circ f$ stetig bei p. Nun, angenommen $(x_n) \to p$, dann folgt $\big(f(x_n)\big) \to f(p) = q$, weil f stetig ist, und dann $\big(gf(x_n)\big) \to g(q)$, weil g stetig ist. Zusammen $(x_n) \to p \implies \big(g \circ f\ (x_n)\big) \to g \circ f(p)$. $\square$

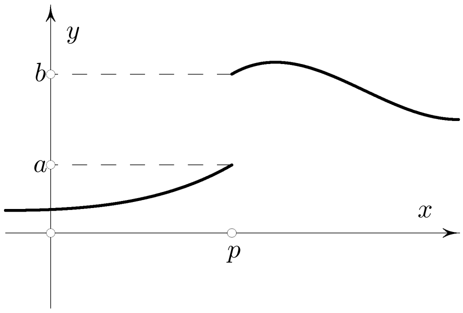
Eng verwandt mit dem Begriff der Stetigkeit ist der allgemeine Konvergenzbegriff für Funktionen statt Folgen:

Definition. *Sei $f : D \to \mathbb{R}$ eine Funktion, $p \in \bar{\mathbb{R}}$, und es gebe mindestens eine Folge (x_n) in D, die gegen p konvergiert. Dann sei*

$$\lim_{x \to p} f(x) = a,$$

falls für jede Folge (x_n) *in* D, *die gegen* p *konvergiert,* $\big(f(x_n)\big)$ *gegen* a *geht.*

Ist $p \in D$, so bedeutet dies: $f(p) = a$, und f ist stetig am Punkt p. Jedoch bleibt die Definition sinnvoll für $p = \pm\infty$, und man kann erklären: Ist D nach oben unbeschränkt, so heißt $f : D \to \mathbb{R}$ **stetig bei** ∞, wenn $\lim_{x\to\infty} f(x)$ existiert. Schränkt man f auf die $x < p$ bzw. die $x > p$ aus D ein, so schreibt man auch $\lim_{x\nearrow p} f(x)$ bzw. $\lim_{x\searrow p} f(x)$.



$\lim_{x\nearrow p} f(x) = a$

$\lim_{x\searrow p} f(x) = b$

Natürlich kann man diese Konvergenzbegriffe auch durch ε-δ-Sprüche erklären — wie?

Wir lassen als Intervallgrenzen im allgemeinen auch $\pm\infty$ zu, also z.B. $[a, \infty)$ ist die Menge der reellen Zahlen $\geq a$. Zur Unterscheidung nennen wir ein abgeschlossenes Intervall $[a, b]$ mit endlichen a, b **kompakt**, also kompakt heißt beschränkt und abgeschlossen. Der Satz von Bolzano-Weierstraß lehrt, daß eine Folge in einem kompakten Intervall eine konvergente Teilfolge hat, und weil aus $a \leq x_n \leq b$ für alle n die entsprechende Ungleichung für den Grenzwert folgt, liegt der Grenzwert der Folge wieder in dem kompakten Intervall.

Wir wollen jetzt unsere Aufmerksamkeit auf stetige Funktionen richten, die auf kompakten Intervallen $[a,b]$ definiert sind, und die naheliegenden Folgerungen und Umformulierungen für den Konvergenzbegriff nicht weiter ausbreiten.

(3.4) Satz. *Sei K ein kompaktes Intervall und $f : K \to \mathbb{R}$ stetig. Dann ist f beschränkt und nimmt auf K ein Maximum und ein Minimum an.*

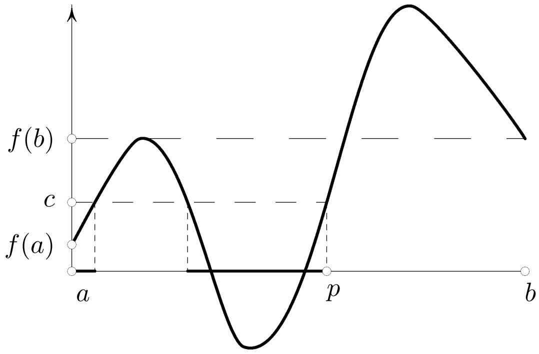
Beweis: Angenommen, f wäre nach oben unbeschränkt. Dann wähle eine Folge von Punkten $x_n \in K$ mit $f(x_n) > n$. Nach Bolzano-Weierstraß konvergiert eine Teilfolge, also oBdA die Folge (x_n) selbst, gegen ein $p \in K$. Dann aber konvergiert $\big(f(x_n)\big)$ gegen $f(p)$, weil f stetig ist, und das widerspricht $f(x_n) > n$. Folglich ist f beschränkt. Sei

$$a := \sup\{f(x) \mid x \in K\}.$$

Angenommen $f(x) \neq a$ für alle $x \in K$. Dann ist $a - f(x) \neq 0$, also $\big(a - f(x)\big)^{-1}$ stetig auf K und nicht beschränkt, weil ja die Werte $f(x)$ ihrem Supremum a beliebig nahe kommen. Das widerspricht der ersten Aussage. Die Beschränktheit nach unten und die Aussage über das Minimum folgen analog. □

Die Funktion $x \mapsto 1/x$ ist auf dem Intervall $(0,1]$ unbeschränkt; die Kompaktheit ist wesentlich. Wir werden bald lernen, daß es sich hier um eine topologische Eigenschaft handelt (VI, 7.10).

(3.5) Zwischenwertsatz. *Eine stetige Funktion $f : [a,b] \to \mathbb{R}$ nimmt jeden Wert zwischen $f(a)$ und $f(b)$ an.*



Beweis: Sei etwa $f(a) < c < f(b)$, und wir wollen zeigen, daß f den Wert c annimmt. Sei $p \in [a, b]$ die obere Grenze der $x \in [a, b]$ mit $f(x) \leq c$, dann behaupten wir: $f(p) = c$. In der Tat, wäre $f(p) < c - \varepsilon$ für ein $\varepsilon > 0$, so wähle dazu δ nach der Stetigkeitsdefinition, dann ist $|f(p + x) - f(p)| < \varepsilon$ für $|x| < \delta$, also $f(p + x)$ $< c$ für $|x| < \delta$, im Widerspruch zur Definition von p. Ganz analog, wenn $f(p) > c + \varepsilon$, so wäre $f(p - x) > c$ für $|x| < \delta$, was auch der Definition von p widerspricht. Weil also $c - \varepsilon \leq f(p) \leq c + \varepsilon$ für alle $\varepsilon > 0$, folgt $f(p) = c$. □

Dies ist ein sehr wichtiger und tausendfach benutzter Satz. Er sagt eigentlich, daß die reelle Gerade und daher ein Intervall keine Löcher hat und nirgends auseinanderfällt. Er ist auch sehr anschaulich, aber beachte, daß die Behauptung gleich falsch wird, wenn man nur einen Punkt aus dem Intervall wegläßt — ein Unterschied, der physikalisch gar nicht faßbar ist. So einfach kann man eben die mathematische Einsicht nicht physikalisch deuten.

Eine leichte Folgerung aus dem Satz kennen wir schon: Jede positive reelle Zahl a hat eine positive k-te Wurzel. Das Polynom $f(x) = x^k - a$ hat nämlich den Wert $f(0) = -a < 0$ und $f(a+1) = (a+1)^k - a$ $\geq 1 + (k-1)a > 0$, also muß es zwischen 0 und $a + 1$ eine Nullstelle haben. Man kann damit übrigens die nicht negativen reellen Zahlen rein algebraisch charakterisieren: Es sind genau die Quadrate. Eine ähnlich wichtige algebraische Folgerung besagt:

(3.6) Satz. *Jedes reelle Polynom ungeraden Grades hat eine reelle Nullstelle. (Eine Nullstelle eines Polynoms nennt man auch eine* **Wurzel***).*

Beweis: Wir dividieren das Polynom durch den Leitkoeffizienten, dann haben wir ein Polynom der Gestalt

$$f(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_0 = x^n(1 + a_{n-1}x^{-1} + \cdots + a_0x^{-n})$$

für $x \neq 0$. Für $|x| \to \infty$ konvergiert der Faktor $(1 + \cdots)$ gegen 1, und x^n hat das Vorzeichen von x, weil n ungerade ist. Also ist $f(x) > 0$ für genügend große x, und $f(x) < 0$ für genügend kleine. Dazwischen muß es einmal verschwinden. □

Das Polynom $x^2 + 1$ hat, wie wir wissen, keine reelle Nullstelle.

(3.7) Satz. *Das Bild eines Intervalls unter einer stetigen reellen Funktion ist wieder ein Intervall. Dabei lassen wir auch* $\pm\infty$ *als Intervallgrenzen zu.*

Beweis: Sei $f : D \to \mathbb{R}$ die betrachtete Abbildung eines Intervalls und

$$a = \inf\bigl(f(D)\bigr), \quad b = \sup\bigl(f(D)\bigr).$$

Weil es Funktionswerte beliebig nahe an a und b gibt und auch alles dazwischen getroffen wird, liegt jedenfalls das ganze offene Intervall (a, b) im Bild von f, und nach Definition von a und b kein kleinerer Punkt als a, kein größerer als b. Es kommen also allenfalls a oder b selbst hinzu, und in jedem Fall ist $f(D)$ ein Intervall mit Grenzen a und b. □

(3.8) Satz. *Eine stetige reelle Funktion auf einem Intervall ist genau dann injektiv, wenn sie streng monoton ist.*

Mit **streng monoton wachsend** ist hier natürlich gemeint $x < y \implies f(x) < f(y)$, und entsprechend für fallende Funktionen. Man kann den Satz durch allerlei Fallunterscheidungen zeigen, aber wir gehen den Weg übers Zweidimensionale.

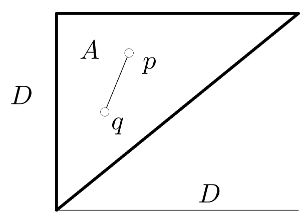
Beweis: Offenbar ist eine streng monotone Funktion injektiv. Sei also $f : D \to \mathbb{R}$ injektiv auf dem Intervall D. Wir müssen zeigen, daß f streng monoton ist. Betrachte das Dreieck

$$A := \{(x, y) \in D \times D \mid x < y\}.$$

Die Voraussetzung sagt, daß die Funktion

$$\varphi : A \to \mathbb{R}, \quad (x, y) \mapsto f(x) - f(y)$$

keine Nullstelle hat, und die Behauptung sagt, daß entweder gilt: $\varphi(p) > 0$ für alle $p = (x, y) \in A$, oder: $\varphi(p) < 0$ für alle p.



Nun, *angenommen* $\varphi(p) > 0$, $\varphi(q) < 0$ *für zwei Punkte* $p, q \in A$. Dann verbinden wir sie durch die Strecke $(1 - t)p + tq$, $0 \le t \le 1$. Wir erhalten die Funktion

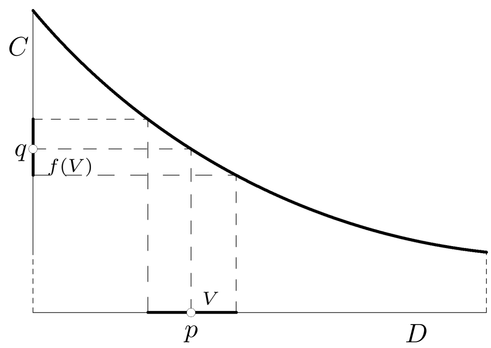
$$[0, 1] \to \mathbb{R}, \quad t \mapsto \varphi\big((1 - t)p + tq\big).$$

Sie ist stetig und hat für $t = 0$ den Wert $\varphi(p) > 0$, für $t = 1$ den Wert $\varphi(q) < 0$, also an einem Zwischenpunkt τ hat man den Wert $\varphi\big((1-\tau)p + \tau q\big) = 0$, im Widerspruch zur Voraussetzung. $\square$

Wir kommen zum Ziel dieser Überlegungen.

(3.9) Satz über die Umkehrfunktion. *Sei D ein Intervall und $f : D \to \mathbb{R}$ stetig und injektiv. Dann ist $f(D) =: C$ ein Intervall, die Funktion f ist streng monoton, und f besitzt eine stetige und streng monotone Umkehrabbildung.*

Beweis: Nur die Stetigkeit von f^{-1} ist noch zu zeigen. Sei etwa D offen — die anderen Fälle muß man für die Randpunkte analog behandeln — dann ist C auch offen, weil f streng monoton ist. Sei $f(p) = q$. Daß f stetig bei p ist heißt: Ist U ein offenes Intervall um q, so enthält $f^{-1}U$ ein offenes Intervall um p. Daß also f^{-1} bei q stetig ist heißt nach dem selben Muster: Ist V ein offenes Intervall um p, so enthält $(f^{-1})^{-1}V$ ein offenes Intervall um q. Aber $(f^{-1})^{-1}V = f(V)$ ist ja ein offenes Intervall um q, weil f stetig und streng monoton ist. $\square$



Anwendung: Ist n ungerade, oder beschränken wir uns auf $x \geq 0$, so ist die Funktion $x \mapsto x^n$ stetig und streng monoton. Die Umkehrfunktion

$$x \mapsto x^{\frac{1}{n}}$$

ist also ebenfalls stetig und streng monoton. Auch die Betragsfunktion $x \mapsto |x|$ ist stetig, denn sie ist die Zusammensetzung

$$x \mapsto x^2 \mapsto \sqrt{x^2}.$$

Sind f und g stetig, so auch die Funktion $\max(f,g)$, die durch $x \mapsto \max\{f(x), g(x)\}$ definiert ist, denn

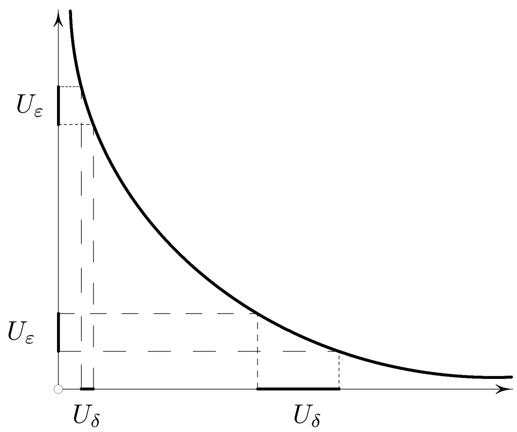
$$\max(f,g) = \tfrac{1}{2}(f+g) + \tfrac{1}{2}|f-g|.$$

Entsprechend fürs Minimum.

Schließlich müssen wir einen etwas delikaten aber wichtigen technischen Punkt berühren, und dafür wollen wir die Stetigkeitsdefinition noch einmal sorgfältig betrachten. Eine Funktion $f : D \to \mathbb{R}$ ist danach stetig auf ganz D, wenn gilt:

Zu jedem $\varepsilon > 0$ und jedem $p \in D$ gibt es ein $\delta > 0$, sodaß
$$|x-p| < \delta \Longrightarrow |f(x) - f(p)| < \varepsilon.$$

Das δ hängt also nicht nur von ε, sondern auch vom Punkt p ab. Zum Beispiel bei der Funktion $y = x^{-1}$ auf $\mathbb{R}_+$ muß man δ um so kleiner wählen, je näher p an 0 liegt.



Die Funktion f heißt gleichmäßig stetig, wenn man δ unabhängig von p wählen kann.

Definition. *Eine Funktion* $f : D \to \mathbb{R}$ *heißt* **gleichmäßig stetig**, *falls gilt: Zu jedem* $\varepsilon > 0$ *gibt es ein* $\delta > 0$, *sodaß für alle* $x, p \in D$ *gilt:*

$$|x - p| < \delta \Longrightarrow |f(x) - f(p)| < \varepsilon.$$

Hier ist also $\delta = \delta(\varepsilon)$, während bei bloßer Stetigkeit $\delta = \delta(\varepsilon, p)$ ist. Hier gehen auch x und p gleichberechtigt und symmetrisch in die Definition ein, was bei der Stetigkeitsdefinition nicht der Fall ist.

(3.10) Satz. *Eine stetige Funktion auf einem kompakten Intervall ist dort gleichmäßig stetig.*

Beweis: Sei $f : D \to \mathbb{R}$ die Funktion, $\varepsilon > 0$, und wir nehmen an, daß dazu kein δ existiert, wie es in der Definition gefordert ist. Dann ist auch $\delta = \frac{1}{n}$ für kein n geeignet. Es gibt also ein x_n und p_n in D mit

$$|x_n - p_n| < \tfrac{1}{n} \quad \text{und} \quad |f(x_n) - f(p_n)| \geq \varepsilon.$$

Nach Bolzano-Weierstraß , mit einem Übergang zu Teilfolgen, dürfen wir annehmen, daß (x_n) gegen ein $q \in D$ konvergiert, und wegen $|x_n - p_n| < \frac{1}{n}$ konvergiert dann (p_n) gegen dasselbe q. Weil f stetig ist, konvergieren $\big(f(x_n)\big)$ und $\big(f(p_n)\big)$ beide gegen $f(q)$, und daher $\big(f(x_n) - f(p_n)\big) \to 0$. Das widerspricht der Ungleichung $|f(x_n) - f(p_n)| \geq \varepsilon$. □

Hier wird sich der Student von der Anschauung verlassen und unangenehm auf das Formale der Definition verwiesen fühlen. Er mag sich damit trösten, daß auch Cauchy an dieser Stelle gestolpert ist.

§ 4. Folgen und Reihen von Funktionen

Eine Folge von Funktionen auf D ordnet jeder ganzen Zahl $n \geq k$ eine Funktion $f_n : D \to \mathbb{R}$ zu. Wir bezeichnen sie wieder durch $(f_n \mid n \geq k)$ oder $(f_n)_{n \geq k}$ oder einfach durch (f_n), und manchmal lassen wir auch noch die Klammern weg.

Definition. *Die Folge von Funktionen* $\big(f_n\big)$ **konvergiert punktweise** *gegen* $f : D \to \mathbb{R}$, *wenn für jedes* $x \in D$ *die reelle Folge* $\big(f_n(x)\big)$ *gegen* $f(x)$ *konvergiert.*

Das heißt also: Zu jedem $\varepsilon > 0$ und $x \in D$ existiert eine natürliche Zahl N, sodaß für $n > N$ gilt:

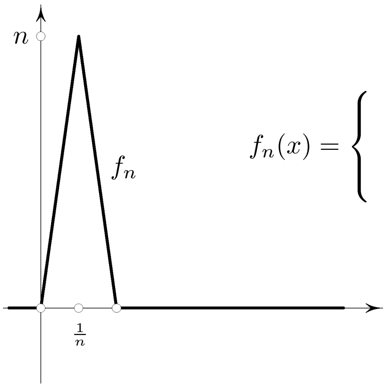
$$|f_n(x) - f(x)| < \varepsilon.$$

Das ist eine naheliegende Definition, aber sie sagt nicht, wie man auf den ersten Blick glauben könnte, daß f_n schließlich nahe an f liegt, also f gut approximiert. Für Funktionenfolgen sind mehrere wesentlich verschiedene Begriffe der Konvergenz sinnvoll. Wir werden darauf in Kapitel VI systematischer eingehen und außer den hier erklärten Konvergenzbegriffen noch andere kennenlernen, die durch Integrale erklärt sind.

Die formale Umschreibung mit ε und N weist darauf hin, wo die Schwierigkeit liegt: das N hängt nicht nur von ε, sondern auch vom Punkt x ab. Für jedes auch noch so große n kann es immer noch viele Punkte x geben, wo $f_n(x)$ sich weit von $f(x)$ entfernt, auch wenn der Definitionsbereich ein kompaktes Intervall ist und alle beteiligten Funktionen stetig sind.

Folgendes Beispiel wird uns auch in der Integrationstheorie wieder vor allzu voreiligen Schlüssen warnen:

(4.1) Beispiel.



$$f_n(x) = \begin{cases} n^2 x \text{ für } 0 \le x \le \frac{1}{n}, \\ 2n - n^2 x \text{ für } \frac{1}{n} \le x \le \frac{2}{n}, \\ 0 \text{ für } x \le 0 \text{ oder } x \ge \frac{2}{n}. \end{cases}$$

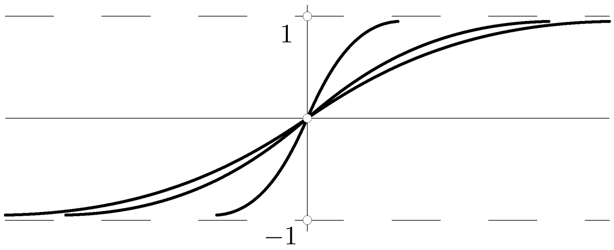
Die letzte Bedingung für f_n bewirkt $f_n(x) = 0$ für $n \ge \frac{2}{x}$ oder $x \le 0$, also für jedes gegebene x verschwindet $f_n(x)$ schließlich. Daher konvergiert $(f_n) \to 0$ punktweise. Trotzdem ist f_n für kein n eine gute Approximation der Nullfunktion.

Man sieht, daß punktweise Konvergenz oft zu wenig ist. Wenn alle f_n stetig sind und (f_n) punktweise gegen f geht, braucht der Limes f auch nicht stetig zu sein. Stetigkeit ist ja selbst durch eine Grenzwertbildung definiert, und wenn man es so betrachtet, geht es hier um die Frage, ob man zwei Grenzwertbildungen miteinander vertauschen darf — man darf im allgemeinen nicht!

Beispiel.

$$f_n(x) = \frac{nx}{1 + |nx|} \to \begin{cases} 1 & \textit{für } x > 0, \\ 0 & \textit{für } x = 0, \\ -1 & \textit{für } x < 0. \end{cases}$$

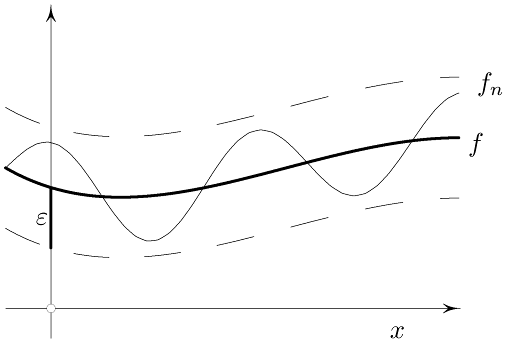
Beweis: Für $x \neq 0$ gilt $f_n(x) = \dfrac{1}{\frac{1}{nx} + \frac{|x|}{x}} \to \dfrac{x}{|x|} = \operatorname{sign}(x)$. □



Das pathologische Verhalten liegt daran, daß die Funktionenfolge nicht gleichmäßig konvergiert.

Definition *(gleichmäßige Konvergenz): Die Folge von Funktionen* (f_n) *auf* D *konvergiert* **gleichmäßig** *gegen* $f : D \to \mathbb{R}$*, wenn gilt: Zu jedem* $\varepsilon > 0$ *gibt es ein* $N \in \mathbb{N}$ *, sodaß für alle* $n > N$ *und alle* $x \in D$ *zugleich*

$$|f_n(x) - f(x)| < \varepsilon.$$



Wir schreiben kurz

$$f \le g \quad :\Longleftrightarrow \quad f(x) \le g(x) \quad \text{für alle } x \in D,$$

und entsprechend fr $f < g$. Dann besagt die Definition: Für große n ist

$$|f_n - f| < \varepsilon.$$

Hier fassen wir $|f_n - f|$ als Funktion $x \mapsto |f_n(x) - f(x)|$ auf. Man kann dieselbe Tatsache auch als Ungleichung zwischen reellen Zahlen fassen. Man erklärt nämlich für Funktionen $f : D \to \mathbb{R}$ mit nicht leerem Definitionsgebiet die

Supremumsnorm $\quad \|f\|_D := \sup\{|f(x)| \mid x \in D\}$.

Damit ist $\|f\|_D = \min\{a \mid |f| \leq a\}$ nach Definition der oberen Grenze. Wenn kein Zweifel über das Definitionsgebiet besteht, schreiben wir kurz $\|f\|$. Im allgemeinen ist $\|f\| \in [0, \infty]$, und $\|f\| \in \mathbb{R}$ genau wenn f beschränkt ist. Die Supremumsnorm hat die

(4.2) Normeigenschaften.

(i) $\|f\| \geq 0$ *und* $\|f\| = 0$ *genau wenn* $f = 0$.

(ii) **Positive Homogenitt:**

$$\|\lambda f\| = |\lambda| \cdot \|f\| \text{ für konstante } \lambda \in \mathbb{R}.$$

(iii) **Dreiecksungleichung:** $\|f + g\| \leq \|f\| + \|g\|$.

Beweis: Nur die Dreiecksungleichung ist nicht ganz selbstverständlich, aber

$$\|f + g\| = \big\||f + g|\big\| \leq \big\||f| + |g|\big\| \leq \|f\| + \|g\|,$$

letztere Ungleichung weil $|f| \leq \|f\|$, $|g| \leq \|g\|$. □

Wenn man nun den reellen Vektorraum aller Funktionen $D \to \mathbb{R}$ mit dieser Norm versieht, so bedeutet gleichmäßige Konvergenz das Natürliche, nämlich (f_n) konvergiert gleichmäßig gegen f, wenn für jedes $\varepsilon > 0$ schließlich $\|f_n - f\| < \varepsilon$ ist. Daher kann es nicht verwundern, daß sich gewohnte Konvergenzsätze auf gleichmäßige Konvergenz von Funktionenfolgen übertragen.

(4.3) Cauchy-Kriterium. *Genau dann ist die Folge von Funktionen (f_n) auf D gleichmäßig konvergent, wenn zu jedem $\varepsilon > 0$ ein $m \in \mathbb{N}$ existiert, sodaß für alle $k \in \mathbb{N}$*

$$\|f_{m+k} - f_m\| < \varepsilon.$$

Beweis: Ist (f_n) gleichmäßig konvergent gegen f, so gilt, wenn m genügend groß ist, $\|f_{m+k} - f\| < \frac{\varepsilon}{2}$ für alle $k \in \mathbb{N}_0$, also

$$\|f_m - f\| < \frac{\varepsilon}{2}, \quad \|f - f_{m+k}\| < \frac{\varepsilon}{2}, \quad \text{also} \quad \|f_m - f_{m+k}\| < \varepsilon.$$

Umgekehrt sei die Cauchy-Bedingung für (f_n) erfüllt. Dann ist insbesondere $\big(f_n(x)\big)$ für jedes $x \in D$ eine Cauchy-Folge reeller Zahlen, bestimmt also eindeutig $f(x) := \lim_{n\to\infty} f_n(x)$. Ist nun $\varepsilon > 0$ gegeben und m genügend groß, so ist $\|f_m - f_{m+k}\| < \frac{\varepsilon}{2}$ für alle k. Aber für jede Stelle $x \in D$ kann man ein $k = k(x)$ so groß wählen, daß $|f_{m+k}(x) - f(x)| < \frac{\varepsilon}{2}$. Folglich ist $|f_m(x) - f(x)| < \varepsilon$ für alle x, also $\|f_m - f\| \le \varepsilon$. □

Wie bei Zahlenfolgen überträgt sich alles Gesagte auf Reihen von Funktionen. Wir sprechen also von punktweiser und gleichmäßiger Konvergenz einer Reihe $\sum_k f_k$ von Funktionen, und von absoluter Konvergenz wenn die Reihe $\sum_k |f_k|$ konvergiert. Konvergiert letztere Reihe gleichmäßig, so nennen wir die Reihe $\sum_k f_k$ gleichmäßig absolut konvergent. Aber die Supremumsnorm führt zu einem neuen und für uns bald sehr wichtigen Begriff: Eine Reihe $\sum_{k=0}^{\infty} f_k$ heißt **normal konvergent**, wenn die Reihe

$$\sum_{k=0}^{\infty} \|f_k\|$$

der Supremumsnormen konvergiert.

(4.4) Konvergenzsatz von Weierstraß. *Eine normal konvergente Reihe $\sum_{k=0}^{\infty} f_k$ von Funktionen $f_k : D \to \mathbb{R}$ konvergiert gleichmäßig absolut gegen eine Funktion $f : D \to \mathbb{R}$.*

Beweis: Dies folgt unmittelbar aus dem Cauchy-Kriterium und der Dreiecksungleichung

$$\Big\| \sum_{k=m}^{m+\ell} |f_k| \Big\| \le \sum_{k=m}^{m+\ell} \|f_k\|.$$

Normale Konvergenz bedeutet, daß die rechte Seite für große m kleiner ε wird, gleichmäßig absolute Konvergenz bedeutet dasselbe für die linke Seite. □

Die wichtigste Bemerkung jedoch, für die der ganze Begriffsapparat aufgefahren ist, besagt:

(4.5) Satz. *Ein gleichmäßiger Limes stetiger Funktionen ist stetig.*

Beweis: Sei (f_n) auf D gleichmäßig konvergent gegen f, die f_n seien stetig, sei ein Punkt $p \in D$ betrachtet und $\varepsilon > 0$ gegeben. Dann wähle $n \in \mathbb{N}$ so groß, daß $|f_n - f| < \frac{\varepsilon}{3}$ und zu diesem n und p wähle $\delta > 0$ so, daß $|f_n(p) - f_n(x)| < \frac{\varepsilon}{3}$ für $|x - p| < \delta$. Das geht, weil f_n stetig ist. Für $|x - p| < \delta$ ist dann

$$\begin{aligned} |f(x) - f(p)| &\le |f(x) - f_n(x)| + |f_n(x) - f_n(p)| + |f_n(p) - f(p)| \\ &< \tfrac{\varepsilon}{3} + \tfrac{\varepsilon}{3} + \tfrac{\varepsilon}{3} = \varepsilon. \end{aligned}$$ □

Für ein mit Pomp angekündigtes Hauptergebnis kann es enttäuschend scheinen, daß der Beweis auf einen $\varepsilon/3$-Schluß von einer Zeile hinausläuft. Darin liegt auch nicht der Witz: das Wichtigste ist eben der richtige Begriff, der den Satz richtig macht: *gleichmäßig konvergent*.

Einen reellen Vektorraum mit einer reellwertigen Norm mit den Eigenschaften (4.2), in dem jede Cauchy-Folge konvergiert, nennt man einen **Banachraum**. Die beschränkten stetigen Funktionen auf einer Menge $D \neq \varnothing$ mit der Supremumsnorm bilden einen Banachraum.

§ 5. Treppenfunktionen

Stetige Funktionen verdienen wohl, daß man sie um ihrer selbst willen mit Liebe betrachtet. Aber in diesem Abschnitt müssen wir kurz unsere Aufmerksamkeit einer Funktionenklasse zuwenden, die für sich weniger anziehend ist, aber nützliche technische Hilfe in der Integrationstheorie leistet, von der dann im nächsten Kapitel gleich die Rede sein wird.

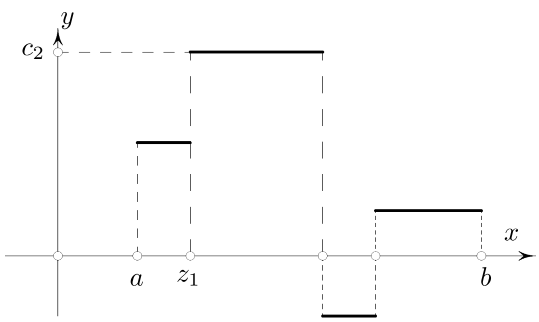
Eine **Zerlegung** Z eines kompakten Intervalls $[a, b]$ ist ein $(n+1)$-Tupel $(z_0, \ldots, z_n)$ von Zahlen, sodaß

$$a = z_0 \leq z_1 \leq \cdots \leq z_n = b.$$

Eine Zerlegung Z_1 ist **feiner** als Z oder eine **Verfeinerung** von Z, wenn Z_1 aus Z durch Hinzunahme weiterer Punkte entsteht. Durch Vereinigung erhält man zu zwei Zerlegungen Z, Z_1 eine gemeinsame Verfeinerung $Z \cup Z_1$. Eine Funktion $\varphi : [a, b] \to \mathbb{R}$ heißt eine **Treppenfunktion**, wenn es eine Zerlegung Z von $[a, b]$ wie oben und dazu Konstanten $c_1, \ldots, c_n$ in $\mathbb{R}$ gibt, mit

$$\varphi(t) = c_k \quad \text{für} \quad z_{k-1} < t < z_k.$$

Die Werte auf den Zerlegungspunkten z_k selbst sind gleichgültig.



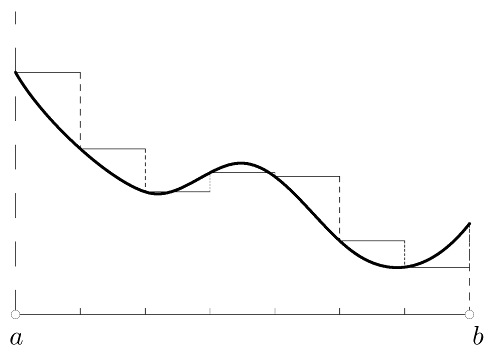
Die sämtlichen Treppenfunktionen auf $[a,b]$ bilden offenbar einen reellen Vektorraum. Zwei Treppenfunktionen φ, ψ auf $[a,b]$ lassen sich (gemeinsame Verfeinerung) durch dieselbe Zerlegung definieren, sodaß also φ und ψ beide auf den Intervallen $z_k < t < z_{k+1}$ konstant sind, und dann ist auch $\lambda\varphi + \mu\psi$ dort konstant, also eine Treppenfunktion.

(5.1) Satz. *Eine stetige Funktion auf einem kompakten Intervall ist ein gleichmäßiger Limes von Treppenfunktionen.*

Beweis: Eine stetige Funktion $f : [a,b] \to \mathbb{R}$ ist gleichmäßig stetig. Zu gegebenem $\varepsilon > 0$ wähle danach $\delta > 0$ so, daß

$$|x - p| < \delta \Longrightarrow |f(x) - f(p)| < \varepsilon$$

für alle $x, p \in D$. Dann wähle eine Zerlegung $a = z_0 \leq \cdots \leq z_n = b$ mit Maschenweite $z_{k+1} - z_k < \delta$, und bestimme dazu die Treppenfunktion φ durch $\varphi(t) = f(z_k)$ für $z_k \leq t < z_{k+1}$.



Dann ist $|f(t) - \varphi(t)| = |f(t) - f(z_k)| < \varepsilon$ für alle $z_k \leq t < z_{k+1}$, und das heißt $|f - \varphi| \leq \varepsilon$.

Mit einem gleichmäßigen Limes ist natürlich die Grenzfunktion einer gleichmäßig konvergenten Folge (von Treppenfunktionen) gemeint. Eine solche Folge findet man wie immer mit dem ($\varepsilon = \frac{1}{n}$)-Trick: Man findet nach dem Gesagten Treppenfunktionen φ_n mit $|f - \varphi_n| < 1/n$. Die Folge (φ_n) konvergiert dann gleichmäßig gegen die Funktion f. □

Die Zerlegung kann man hier äquidistant wählen, also $z_k = a + k(b-a)/n$ mit einem so großen n, daß $n\delta > b - a$. Ein gleichmäßiger Limes von Treppenfunktionen heißt **Regelfunktion**. Stetige Funktionen sind also Regelfunktionen, aber auch z.B. monotone Funktionen, Treppenfunktionen, ...

(5.2) Bemerkung. *Zu jeder Regelfunktion f auf dem kompakten Intervall $[a,b]$ und jedem $\varepsilon > 0$ gibt es Treppenfunktionen φ, ψ mit*

$$\varphi < f < \psi \quad \textit{und} \quad \psi - \varphi = \varepsilon.$$

Beweis: Wähle eine Treppenfunktion τ mit $|f - \tau| < \frac{\varepsilon}{2}$ und setze $\varphi = \tau - \frac{\varepsilon}{2}, \quad \psi = \tau + \frac{\varepsilon}{2}$. □