Kapitel 2
Folgen und Funktionen

Im ersten Kapitel haben wir, ausgehend von den natürlichen Zahlen, den Zahlbegriff so erweitert, daß wir möglichst viele Rechen- und Vergleichsoperationen durchführen können; Ziel war vor allem die Konstruktion der reellen und der komplexen Zahlen. Wesentlicher Gegenstand der Analysis sind allerdings nicht Zahlen, sondern deren Veränderung: Wir interessieren uns für die zeitliche Entwicklung einer Größe oder für die Abhängigkeiten zwischen verschiedenen Größen, um so Entwicklungen und Zusammenhänge zu beschreiben und eventuell sogar vorhersagen zu können.

Da sich viele Größen höchstens mit einer begrenzten Genauigkeit ermitteln lassen und wir bei komplizierten Rechnungen meist auf numerische Näherungslösungen angewiesen sind, müssen wir uns dabei auch überlegen, wie unsere Ansätze auf kleine oder auch größere Variationen der Eingangsgrößen reagieren; wenn ein Modell praktisch nützlich sein soll, dürfen kleine Veränderungen der Ausgangsdaten nicht zu dramatischen Veränderungen bei den Endergebnissen führen.

Um dies quantitativ fassen zu können, wollen wir zunächst den Begriff des *Abstands* abstrakt definieren. Das mag auf den ersten Blick überflüssig erscheinen, da wir schließlich alle aus der Schule wissen, was Abstände sind und wie man sie berechnet. Wir werden aber gleich sehen, daß je nach Anwendung durchaus verschiedene Definitionen des Abstands zweier Punkte sinnvoll sein können: Gerade bei wirtschaftlichen Anwendungen hängen die Kosten eines Transports beispielsweise selten vom geometrischen Abstand ab, sondern von komplizierter zu berechnenden Abstandsmaßen.

Auch wenn es in dieser Vorlesung fast ausschließlich um reelle Zahlen geht, möchte ich die entsprechenden Definitionen gleich so bringen, daß sie auch für mehrdimensionale Räume anwendbar sind. Da wir in einer dreidimensionalen Welt leben, uns meist auf zweidimensionalen Flächen bewegen und auch auf solchen Flächen schreiben, haben wir deutlich mehr Erfahrung mit zwei- oder dreidimensionalen Konstruktionen als mit eindimensionalen, so daß zweidimensionale Beispiele wahrscheinlich anschaulicher sind als eindimensionale.

§ 1: Metrische Räume

Von einem vernünftigen Abstandsbegriff erwarten wir, daß er mindestens die folgenden drei Forderungen erfüllt:

- Entfernungen hängen nicht von der Richtung ab: Von Mannheim nach Heidelberg ist es genauso weit wie von Heidelberg nach Mannheim.
- Entfernungen können nicht negativ sein und sind genau dann gleich Null, wenn Anfangs- und Endpunkt übereinstimmen.
- Durch Umwege lassen sich Entfernungen nicht verkürzen: Die Entfernung von Mannheim nach Heidelberg ist nicht größer als die Summe der Entfernungen von Mannheim nach Karlsruhe und von Karlsruhe nach Heidelberg.

Dementsprechend definieren wir

Definition: Eine *Metrik* auf einer Menge X ist eine Abbildung, die je zwei Elementen $x, y \in X$ eine reelle Zahl $d(x, y)$ zuordnet, so daß gilt:

1.) Für alle $x, y \in X$ ist $d(x, y) = d(y, x)$ *(Symmetrie)*

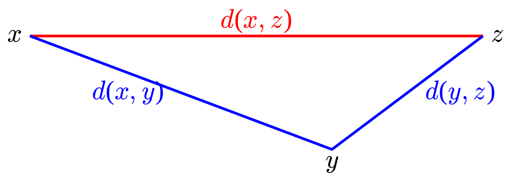
2.) $d(x, y) \geq 0$ für alle $x, y \in X$ mit Gleichheit genau dann, wenn $x = y$ ist *(positive Definitheit)*

3.) $d(x, z) \leq d(x, y)+d(y, z)$ für alle $x, y, z \in X$ *(Dreiecksungleichung)*

Die Menge X zusammen mit dieser Metrik bezeichnen wir als einen *metrischen Raum.*

Der Name *Dreiecksungleichung* kommt natürlich daher, daß drei Punkte x, y, z ein Dreieck definieren; die Ungleichung besagt, daß dabei die

Summe zweier Seiten nicht kürzer sein kann als die dritte:

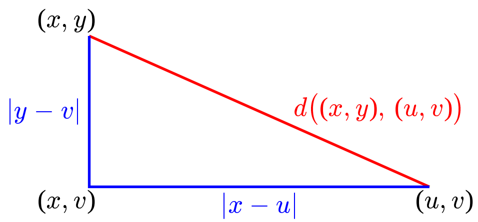


Als Beispiel könnten wir etwa die Menge X aller deutscher Städte betrachten und für zwei Städte x,y die Entfernung zwischen deren irgendwie festzulegenden Zentren als $d(x,y)$ definieren.

In dieser Vorlesung wird X meist die Menge der reellen Zahlen sein, anschaulicher wird der Begriff der Metrik aber, wenn wir zunächst die reelle Zahlen*ebene* betrachten. Wie aus der Schule gewohnt, identifizieren wir deren Punkte über ein kartesisches Koordinatensystem mit der Menge aller Paare (x,y) von reellen Zahlen, also mit

$$\mathbb{R}^2 \underset{\text{def}}{=} \left\{(x,y) \mid x,y \in \mathbb{R}\right\} .$$

Für die Definition einer Metrik haben wir verschiedene Möglichkeiten. Die bekannteste davon ist natürlich der „übliche“ Abstand aus der EUKLIDischen Geometrie. Für zwei Punkte (x,y) und (u,v) aus $\mathbb{R}^2$ kann er leicht über den Satz des PYTHAGORAS bestimmt werden:



Die Verbindungsstrecke der beiden Punkte ist die Hypothenuse des rechtwinkligen Dreiecks mit Ecken (x,y), (x,v) und (u,v); die Katheten sind parallel zu den Koordinatenachsen und haben die Längen $|x-u|$ beziehungsweise $|y-v|$. Die Länge der Verbindungsstrecke ist daher wieder

$$d_E\big((x,y),(u,v)\big) = \sqrt{(x-u)^2+(y-v)^2} .$$

Die so definierte Abbildung d_E bezeichnen wir als die EUKLID*ische Metrik* auf $\mathbb{R}^2$.

Um dies zu rechtfertigen, müssen wir natürlich nachweisen, daß d_E wirklich die drei Forderungen an eine Metrik erfüllt. Mit der ersten, der Symmetrie, gibt es keine Probleme: Wenn wir (x,y) und (u,v) vertauschen, wird $(x-u)^2$ ersetzt durch $(u-x)^2$ und $(y-v)^2$ durch $(v-y)^2$; dadurch ändert sich natürlich nicht das geringste.

Auch die zweite Forderung, die positive Definitheit, ist problemlos: Unter der Wurzel steht eine Summe zweier Quadrate; wie wir wissen, sind diese stets größer oder gleich Null; insbesondere ist also die Quadratwurzel wirklich eine reelle Zahl – was wir streng genommen eigentlich gleich zu Beginn als Rechtfertigung der Definition hätten nachweisen müssen. Beim gegenwärtigen Stand der Vorlesung sollten wir aber schon so weit sein, daß ein so einfaches Argument stillschweigend übergangen werden kann. Auch das, worum es bei der zweiten Forderung eigentlich geht, ist klar: Nach Definition der Quadratwurzel einer reellen Zahl ist diese stets nichtnegativ, und wenn sie Null ist, muß auch unter der Wurzel eine Null stehen. Dort steht aber eine Summe zweier Quadrate; die kann nur dann verschwinden, wenn beide Quadrate verschwinden, und das wiederum ist nur dann möglich, wenn beidesmal die Null quadriert wird. Also ist $x-u = y-v = 0$, d.h. $(x,u) = (y,v)$. Umgekehrt ist in diesem Fall natürlich $d_E\big((x,y),(u,v)\big) = 0$.

Bleibt noch die dritte Forderung, die Dreiecksungleichung. Für diese können wir uns hier auf die klassische EUKLIDische Geometrie berufen, die uns in der Tat lehrt, daß in jedem echten Dreieck die Summe zweier Seitenlängen größer ist als die Länge der dritten Seite; wenn das Dreieck zu Strecke entartet, haben wir Gleichheit.

Natürlich kann man die Dreiecksungleichung auch explizit nachrechnen; ohne den Gebrauch von Vektoren ist diese (an einer Stelle auch etwas trickreiche) Rechnung allerdings ziemlich unübersichtlich. Da die Dreiecksungleichung in der Linearen Algebra ohnehin in einem allgemeineren Zusammenhang bewiesen wird, sei hier auf einen rechnerischen Beweis verzichtet.

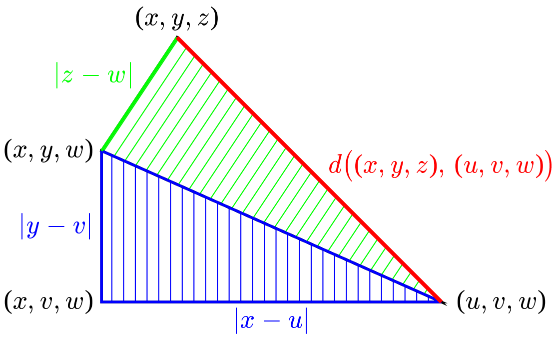
Nicht nur in der Ebene, sondern auch im Raum können wir über den Satz des PYTHAGORAS einen EUKLIDischen Abstand berechnen: Dort haben wir drei Koordinaten, wir rechnen also in

$$\mathbb{R}^3 \underset{\text{def}}{=} \left\{(x,y,z) \;\middle|\; x,y,z \in \mathbb{R}\right\} .$$

Um den klassischen EUKLIDischen Abstand zwischen zwei Punkten (x,y,z) und (u,v,w) zu berechnen, beginnen wir mit dem Abstand zwischen den beiden Punkten (x,y,w) und (u,v,w). Diese liegen beide in der Ebene, in der die dritte Koordinate konstant gleich w ist. Die Punkte (x,y,w), (x,v,w) und (u,v,w) aus dieser Ebene sind die Ecken eines rechtwinkliges Dreiecks, und wir sind in derselben Situation wie oben beim Dreieck mit Ecken (x,y), (x,v) und (u,v): Der Abstand zwischen (x,y,w) und (u,v,w) ist nach PYTHAGORAS

$$\sqrt{(x-u)^2+(y-v)^2} .$$

Nun gehen wir in die Ebene, die auf der gerade betrachteten senkrecht steht und die Gerade durch die Punkte (x,y,w) und (u,v,w) enthält.



Dort liegt auch der Punkt (x,y,z) und bildet zusammen mit diesen beiden Punkten ein rechtwinkliges Dreieck, dessen Hypothenuse die Strecke zwischen (x,y,z) und (u,v,w) ist. Das Quadrat ihre Länge ist nach PYTHAGORAS die Summe der Quadrate aus der Länge der gerade berechneten Strecke sowie dem Abstand der Punkte (x,y,z) und (x,y,w). Diese beiden Punkte liegen auf einer Geraden parallel zur z-Achse; ihr Abstand ist also $|z-w|$. Der gesuchte Abstand ist somit

$$d_E\big((x,y,z),(u,v,w)\big) \underset{\text{def}}{=} \sqrt{(x-u)^2+(y-v)^2+(z-w)^2} .$$

Wenn wir die Dimension noch weiter erhöhen wollen, verläßt uns die

Anschauung, aber zumindest im Prinzip ist alles klar: Bei Punkten aus dem n-dimensionalen Raum

$$\mathbb{R}^n \underset{\text{def}}{=} \left\{(x_1,\ldots,x_n) \;\middle|\; x_1,\ldots,x_n \in \mathbb{R}^n\right\}$$

können wir uns schrittweise hocharbeiten von der Ebenen zum n-dimensionalen Raum, und indem wir den Satz des PYTHAGORAS $(n-1)$-mal anwenden, erhalten wir für den Abstand zweier Punkte die Formel

$$d_E\big((x_1,\ldots,x_n),\,(y_1,\ldots,y_n)\big) \underset{\text{def}}{=} \sqrt{\sum_{i=1}^{n}(y_i-x_i)^2} .$$

Auch hier müssen wir uns wieder überlegen, daß dies wirklich eine Metrik definiert. Bei der Symmetrie und der positiven Definitheit können wir genauso argumentieren wie im zweidimensionalen Fall, und die Dreiecksungleichung folgt auch hier aus der klassischen EUKLIDischen Geometrie: Zwar war dort nie von einem n-dimensionalen Raum die Rede, aber auch für $n \geq 3$ spannen drei Punkte höchstens eine Ebene auf, so daß wir es auch hier mit gewöhnlichen Dreiecken zu tun haben. Im übrigen wird die Dreiecksungleichung wie bereits erwähnt in der Linearen Algebra in einer Form bewiesen werden, die insbesondere auch den hier betrachteten Fall einschließt.

Damit haben wir die EUKLIDische Metrik für beliebige Dimensionen erklärt; was uns in dieser Vorlesung wirklich interessiert, ist allerdings nur der Fall $n=1$. Dort haben wir die Formel

$$d_E(x,y) = \sqrt{(y-x)^2} = |y-x| ,$$

und das ist natürlich die offensichtliche Weise, wie man im Eindimensionalen einen Abstand definiert. Wir bezeichnen diese Metrik daher als die *Standardmetrik* auf $\mathbb{R} = \mathbb{R}^1$ und werden zumindest in der Vorlesung nie eine andere benutzen.

In höheren Dimensionen freilich kann die EUKLIDische Metrik durch die Quadratwurzel in ihrer Definition gelegentlich recht unangenehm werden. Außerdem liefert sie auch nicht immer das, was für eine gegebene Anwendung den richtigen Begriff von „Nähe“ formalisiert. Wenn wir beispielsweise ein nichtlineares Gleichungssystem mit n Unbekannten lösen wollen, sind die Lösungen Punkte aus $\mathbb{R}^n$. Nur in seltenen Fällen haben wir die Möglichkeit, diese exakt zu berechnen; üblicherweise müssen wir uns mit Näherungslösungen zufrieden geben. Oft sind die einzelnen Komponenten der Lösung für sich selbst interessante Größen, von denen wir gerne wissen möchten, mit welcher Genauigkeit wir sie kennen. Bei der EUKLIDischen Metrik könnten, gerade für große Dimensionen n, hohe Genauigkeiten einiger Komponenten größere Fehler bei anderen kaschieren. In solchen Fällen arbeitet man besser mit der sogenannten Maximums-Metrik

$$d_{\max}\big((x_1,\dots,x_n),\,(y_1,\dots,y_n)\big) \underset{\text{def}}{=} \max\big\{|y_i-x_i| \bigm| i=1,\dots,n\big\}\,.$$

Auch dies ist eine Metrik auf $\mathbb{R}^n$: Die Symmetrie ist wieder klar, da $|y_i-x_i| = |x_i-y_i|$ für alle i, und auch die positive Definitheit folgt, analog zum Fall der EUKLIDischen Metrik, daraus, daß Beträge nie negativ und abgesehen vom Fall der Null stets positiv sind. Bleibt die Dreiecksungleichung, die wir hier wirklich nachrechnen müssen – was zum Glück aber deutlich weniger Aufwand erfordert, als der entsprechende Nachweis für die EUKLIDische Metrik: Wir gehen aus von drei Punkten $(x_1,\dots,x_n)$, $(y_1,\dots,y_n)$ und $(z_1,\dots,z_n)$ aus $\mathbb{R}^n$. Für jeden Index i ist dann

$$\begin{aligned}|z_i-x_i| &= |(z_i-y_i)+(y_i-x_i)| \le |(z_i-y_i| + |y_i-x_i|\\ &\le d\big((z_1,\dots,z_n),\,(y_1,\dots,y_n)\big) + d\big((y_1,\dots,y_n),\,(x_1,\dots,x_n)\big)\,,\end{aligned}$$

denn die Metrik ist ja jeweils das Maximum der Betragsdifferenzen. Da diese Abschätzung für alle i gilt, gilt sie insbesondere für das (nicht notwendigerweise eindeutig bestimmte) i, für das $|z_i-x_i|$ maximal wird, und dieses Maximum ist der Abstand zwischen $(x_1,\dots,x_n)$ und $(z_1,\dots,z_n)$ in der Maximums-Metrik. Damit ist die Dreiecksungleichung bewiesen.

Im Spezialfall $n=1$ wird die Maximums-Metrik offensichtlich zur Standardmetrik, und das gilt auch für die dritte Metrik auf $\mathbb{R}^n$, die hier noch kurz erwähnt werden soll:

Die Taxi-Metrik (oder Manhattan-Metrik) auf $\mathbb{R}^n$ ordnet zwei Punkten $(x_1,\dots,x_n)\in\mathbb{R}^n$ und $(y_1,\dots,y_n)\in\mathbb{R}^n$ den Abstand

$$d_{\text{tax}}\big((x_1,\dots,x_n),\,(y_1,\dots,y_n)\big) \underset{\text{def}}{=} \sum_{i=1}^n |y_i-x_i|$$

zu. Symmetrie und positive Definitheit zeigt man genauso wie bei der Maximums-Metrik, und für die Dreiecksungleichung müssen wir einfach die Ungleichungen $|z_i-x_i| \le |z_i-y_i| + |y_i-x_i|$ aufaddieren:

$$\begin{aligned}d_{\text{tax}}\big((x_1,\dots,x_n),\,(z_1,\dots,z_n)\big) &= \sum_{i=1}^n |z_i-x_i|\\ &\le \sum_{i=1}^n |z_i-y_i| + \sum_{i=1}^n |y_i-x_i|\\ &= d_{\text{tax}}\big((x_1,\dots,x_n),\,(y_1,\dots,y_n)\big) + d_{\text{tax}}\big((y_1,\dots,y_n),\,(z_1,\dots,z_n)\big)\end{aligned}$$

Die Taxi-Metrik hat ihren Namen daher, daß Entfernungen in Städten, z.B. in den Mannheimer Quadraten, aus der Sicht eines taxifahrenden Passanten nichts mit der Luftlinie, dem EUKLIDischen Abstand, zu tun haben, sondern mit dem Straßenverlauf. Im Falle eines rechtwinklig organisierten Straßennetzes sind dies bei jeder entfernungsoptimalen Route genau die durch die Taxi-Metrik berechneten. Der Name *Manhattan-Metrik* kommt natürlich daher, daß das rechtwinklige Straßennetz von Manhattan noch bekannter ist als das der Mannheimer Innenstadt.

Gelegentlich werden in dieser Vorlesung auch die komplexen Zahlen ein Gastspiel geben; in Analogie zur Standardmetrik auf $\mathbb{R}$ definieren wir eine Metrik auf $\mathbb{C}$ durch die Vorschrift

$$d(z,w) \underset{\text{def}}{=} |z-w|\ .$$

Schreiben wir $z=x+iy$ und $w=u+iv$, so ist

$$d(z,w) = |(x-u)+i(y-v)| = \sqrt{(x-u)^2+(y-v)^2}\,;$$

identifizieren wir die komplexe Zahl $x+iy$ also mit dem Punkt (x,y) der reellen Zahlenebene, so ist dies gerade der EUKLIDische Abstand der beiden Punkte. Insbesondere ist damit klar, daß die obige Definition wirklich auf eine Metrik führt.

Dies soll uns an Beispielen genügen; wie bereits mehrfach erwähnt, wird es in dieser Vorlesung im wesentlichen ohnehin nur um $\mathbb{R}$ mit der Standardmetrik gehen. Allgemeine Aussagen über metrische Räume interessieren uns deshalb hier noch nicht besonders; nur einige wenige Aussagen, bei denen dies gegenüber einem direkten Beweis nur für $\mathbb{R}$ zu keinem größeren Aufwand führt, sollen allgemein bewiesen werden.

Eine solche Aussage ist eine Verschärfung der Dreiecksungleichung, die $d(x, z)$ auch nach unten abschätzt:

Verschärfte Dreiecksungleichung: Für drei Elemente x, y, z eines metrischen Raums X mit Metrik d gilt stets

$$|d(x, y) - d(y, z)| \le d(x, z) \le d(x, y) + d(y, z) .$$

Beweis: Die rechte Ungleichung ist die in der Definition einer Metrik postulierte Dreiecksungleichung; wir müssen also nur die linke beweisen. Da der Betrag einer reellen Zahl entweder die Zahl selbst ist oder ihr Negatives, müssen wir zeigen, daß

$$d(x, y) - d(y, z) \le d(x, z) \quad \text{und} \quad d(y, z) - d(x, y) \le d(x, z)$$

ist. Dazu wenden wir die Dreiecksungleichung zunächst an auf $d(x, y)$ und erhalten mit z als Zwischenpunkt

$$d(x, y) \le d(x, z) + d(z, y) \quad \text{oder} \quad d(x, y) - d(y, z) \le d(x, z) ,$$

wobei wir für die zweite Formel auch die Symmetrie von d ausgenutzt haben. Entsprechend folgt

$$d(y, z) \le d(y, x) + d(x, z) \quad \text{oder} \quad d(y, z) - d(x, y) \le d(x, z) .$$

Damit ist die verschärfte Dreiecksungleichung bewiesen. ■

Speziell für $X = \mathbb{R}$ oder $X = \mathbb{C}$ wird die verschärfte Dreiecksungleichung zur Ungleichung

$$\big||x - y| - |y - z|\big| \le |x - z| \le |x - y| + |y - z| ,$$

die uns im weiteren Verlauf der Vorlesung gelegentlich nützlich sein wird. Insbesondere folgt für $y = 0$ die Ungleichung $\big||x| - |z|\big| \le |x - z|$.

Wir interessieren uns für Metriken, weil wir präzise definieren wollen, was es bedeutet, daß uns ein Verfahren zur Produktion von Näherungslösungen, wie etwa das von HERON, bei hinreichend häufiger Anwendung beliebig genaue Näherungen liefern kann oder auch weil wir wissen wollen, ob eine Funktion uns auch dann noch zumindest näherungsweise gute Werte liefert, wenn wir als Argument nicht einen exakten Wert wie $\sqrt{2}$ einsetzen, sondern nur eine taschenrechner- oder computergenaue Näherung. Dies führt auf die folgenden Definitionen:

Definition: Die Folge $(x_n)_{n\in\mathbb{N}}$ von Punkten eines metrischen Raums X *konvergiert* bezüglich der Metrik d gegen den Punkt $x \in X$, wenn die Folge $\big(d(x, x_n)\big)_{n\in\mathbb{N}}$ der Abstände zwischen x und x_n eine Nullfolge ist, wenn es also für jedes $\varepsilon > 0$ ein $n_0 \in \mathbb{N}$ gibt, so daß $d(x, x_n) \le \varepsilon$ für alle $n \ge n_0$, Wir bezeichnen dann x als den *Grenzwert* oder *Limes* der Folge und schreiben $x = \lim\limits_{n\to\infty} x_n$. Eine Folge heißt *konvergent*, wenn es einen Punkt $x \in X$ gibt, gegen den sie konvergiert.

Natürlich gibt es für eine beliebige Folge keinerlei Grund, warum sie konvergieren sollte; um einen ersten Eindruck davon zu bekommen, was hier passieren kann, wollen wir einige Folgen in der reellen Zahlenebene betrachten.

Fangen wir an mit der Folge aus den Paaren $(x_n, y_n) \in \mathbb{R}^2$ mit

$$x_n = 1 + \frac{3}{n} \quad \text{und} \quad y_n = 2 - \frac{4}{n} .$$

Wenn unsere Definition sinnvoll ist, sollte diese Folge unabhängig von der Metrik gegen den Punkt $(1, 2) \in \mathbb{R}^2$ konvergieren. Bezüglich der EUKLIDischen Metrik auf $\mathbb{R}^2$ ist der Abstand

$$d\big((x_n, y_n), (1, 2)\big) = \sqrt{\left(\frac{3}{n}\right)^2 + \left(-\frac{4}{n}\right)^2} = \sqrt{\frac{25}{n^2}} = \frac{5}{n} ,$$

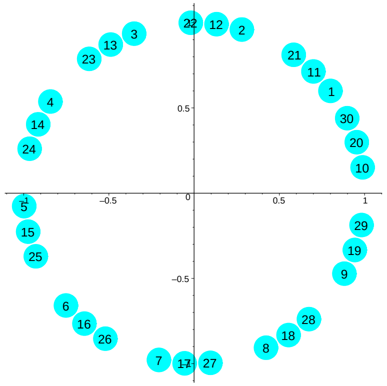
und die Zahlen $5/n$ bilden natürlich eine Nullfolge: Für jedes vorgegebene $\varepsilon > 0$ ist $5/n < \varepsilon$ für alle n ab der kleinsten natürlichen Zahl n_0 größer oder gleich $5/\varepsilon$. Geometrisch betrachtet liegen die Punkte (x_n, y_n) alle auf der Geraden durch den Punkt $(1, 2)$ mit Steigung $4/3$, und gehen auf dieser schnurstracks gegen den Grenzwert $(1, 2)$.

Als nächstes betrachten wir die Folge der Punkte $(x_n, y_n)_{n\in\mathbb{N}}$ mit $x_n = n$ und $y_n = n^2$. Diese Punkte liegen offensichtlich alle auf der Parabel $y = x^2$ und entfernen sich immer weiter vom Nullpunkt sowie auch von jedem anderen Punkte der Ebenen. Diese Folge ist also offensichtlich in keiner der oben definierten Metriken konvergent.

Beim HERON-Verfahren zur Berechnung der Quadratwurzel haben wir rekursiv definierte Folgen kennen gelernt; natürlich können wir auch Folgen im $\mathbb{R}^2$ rekursiv definieren. Wir können etwa ausgehen von einem festen Punkt $(x_0, y_0) = (1, 0)$ und dann definieren

$$(x_n, y_n) = \left(\tfrac{4}{5}x_{n-1} - \tfrac{3}{5}y_{n-1},\ \tfrac{3}{5}x_{n-1} + \tfrac{4}{5}y_{n-1}\right) \quad \text{für alle } n \in \mathbb{N}\,.$$

Die folgende Abbildung zeigt die ersten dreißig Punkte dieser Folge, wobei das Paar (x_n, y_n) dargestellt ist durch eine Scheibe, in der die Zahl n steht:



Laut Bild liegen alle Punkte auf einem Kreis um den Nullpunkt, und in der Tat können wir das auch leicht durch Nachrechnen bestätigen: Für $n \geq 1$ ist der EUKLIDische Abstand des Punkts (x_n, y_n) vom Nullpunkt die Wurzel aus

$$x_n^2 + y_n^2 = \left(\tfrac{4}{5}x_{n-1} - \tfrac{3}{5}y_{n-1}\right)^2 + \left(\tfrac{3}{5}x_{n-1} + \tfrac{4}{5}y_{n-1}\right)^2$$
$$= \tfrac{16}{25}x_{n-1}^2 - \tfrac{24}{25}x_{n-1}y_{n-1} + \tfrac{9}{25}y_{n-1}^2 + \tfrac{9}{25}x_{n-1}^2 + \tfrac{24}{25}x_{n-1}y_{n-1} + \tfrac{16}{25}y_{n-1}^2$$
$$= x_{n-1}^2 + y_{n-1}^2\,,$$

der Punkt (x_n, y_n) hat also den gleichen Abstand vom Nullpunkt wie sein Vorgänger. Induktiv folgt daraus, daß alle Punkte den gleichen Abstand vom Nullpunkt haben wie der Punkt (x_0, y_0), und für den ist der Abstand offensichtlich gleich eins.

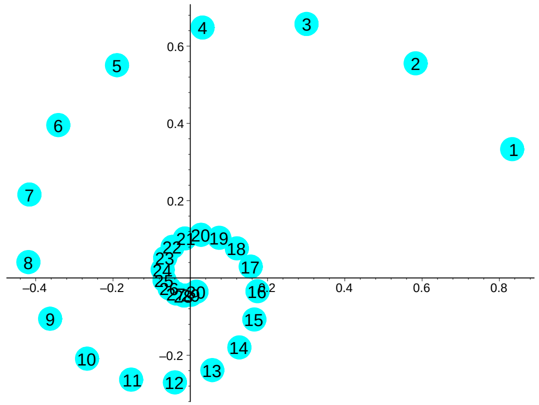
Die Punkte (x_n, y_n) liegen also alle auf dem Kreis mit Radius eins um den Nullpunkt, allerdings ist nichts davon zu sehen, daß sie sich dort auf einen festen Punkt zubewegen. Wenn man die entsprechenden Begriffe aus der Linearen Algebra kennt, kann man leicht zeigen, daß der Punkt (x_n, y_n) aus dem Punkt (x_{n-1}, y_{n-1}) entsteht durch Drehung um den

Nullpunkt mit einem festen Winkel α; hier ist dieser Winkel der Arkuskosinus von $\frac{4}{5}$; das sind ungefähr 36,86989765°. Damit ist klar, daß die Folge dieser Punkte unmöglich gegen einen festen Punkt konvergieren kann: Wenn ein Punkt (x_n, y_n) aus der Folge in der Nähe eines festen Punkts (x, y) liegt, kann unmöglich auch sein um α weitergedrehter Nachfolger dort liegen. Diese Folge ist somit nicht konvergent.

Eine leichte Modifikation macht sie konvergent: Wir starten wieder mit $(x_0, y_0) = (1, 0)$, setzen jetzt aber

$$(x_n, y_n) = \left(\tfrac{5}{6}x_{n-1} - \tfrac{1}{3}y_{n-1},\ \tfrac{1}{3}x_{n-1} + \tfrac{5}{6}y_{n-1}\right) \quad \text{für alle } n \in \mathbb{N}\,.$$

Jetzt legt die Abbildung der ersten dreißig Punkte die Vermutung nahe, daß sich die Folgenglieder wohl spiralförmig an den Nullpunkt annähern:



Wir vermuten daher, daß die Folge gegen den Nullpunkt konvergiert. Um diese Vermutung zu beweisen, vergleichen wir wieder den Abstand eines Punkts (x_n, y_n) vom Nullpunkt mit dem seines Vorgängers. Das Quadrat dieses Abstands ist

$$x_n^2 + y_n^2 = \left(\tfrac{5}{6}x_{n-1} - \tfrac{1}{3}y_{n-1}\right)^2 + \left(\tfrac{1}{3}x_{n-1} + \tfrac{5}{6}y_{n-1}\right)^2$$
$$= \tfrac{25}{36}x_{n-1}^2 - \tfrac{10}{18}x_{n-1}y_{n-1} + \tfrac{1}{9}y_{n-1}^2 + \tfrac{1}{9}x_{n-1}^2 + \tfrac{10}{18}x_{n-1}y_{n-1} + \tfrac{25}{36}y_{n-1}^2$$
$$= \tfrac{29}{36}x_{n-1}^2 + \tfrac{29}{36}y_{n-1}^2\,,$$

also ist

$$d\big((x_n, y_n),\,(0,0)\big) = \frac{\sqrt{29}}{6}\,d\big((x_{n-1}, y_{n-1}),\,(0,0)\big) \quad \text{für alle } n \geq 1\,.$$

Da der Punkt $(x_0, y_0) = (1,0)$ Abstand eins vom Nullpunkt hat, folgt induktiv:

$$d\big((x_n, y_n),\,(0,0)\big) = \left(\frac{\sqrt{29}}{6}\right)^n .$$

Wir müssen zeigen, daß dies eine Nullfolge ist, daß es also zu jedem $\varepsilon > 0$ ein n_0 gibt, so daß $(\sqrt{29}/6)^n$ für jedes $n \geq n_0$ kleiner als ε ist.

Dazu schätzen wir zunächst die Basis $\sqrt{29}/6$ nach oben ab: 29 liegt zwischen 5^2 und 6^2; zwischen diesen beiden Zahlen liegt $5\frac{1}{2} = \frac{11}{2}$ und

$$(5 + \tfrac{1}{2})^2 = 5^2 + 2 \cdot 5 \cdot \tfrac{1}{2} + \tfrac{1}{4} = 30\tfrac{1}{4} > 29\,.$$

Daher ist

$$\frac{\sqrt{29}}{6} < \frac{11}{2 \cdot 6} = \frac{11}{12} \quad \text{und} \quad \left(\frac{\sqrt{29}}{6}\right)^n < \left(\frac{11}{12}\right)^n .$$

Für ein festes n ist offenbar $(\frac{11}{12})^n < \varepsilon$ genau dann, wenn $(\frac{12}{11})^n > 1/\varepsilon$ ist. Die linke Seite können wir abschätzen nach der Ungleichung von BERNOULLI, die wir am Ende von Kap I, §3 bewiesen haben: Danach ist

$$\left(\frac{12}{11}\right)^n = \left(1 + \frac{1}{11}\right)^n > 1 + \frac{n}{11} \quad \text{für alle } n \geq 2\,.$$

Die Ungleichung $1 + n/11 > 1/\varepsilon$ ist insbesondere dann erfüllt, wenn $n > 11/\varepsilon$ ist; für jede natürliche Zahl $n_0 > 11/\varepsilon$ und (wegen der Ungleichung von BERNOULLI) $n_0 \geq 2$ ist damit auch für jedes $n \geq n_0$

$$d\big((x_n, y_n),\,(0,0)\big) = \left(\frac{\sqrt{29}}{6}\right)^n < \left(\frac{12}{11}\right)^n < \varepsilon\,.$$

Damit haben wir gezeigt, daß die Folge der (x_n, y_n) gegen den Nullpunkt konvergiert.

Um noch eine weitere Art von dynamischem Verhalten kennenzulernen, modifizieren wir die Rekursionsvorschrift noch einmal: Wir starten wieder mit $(x_0, y_0) = (1,0)$, setzen jetzt aber

$$(x_n, y_n) = \left(\tfrac{5}{7}x_{n-1} - \tfrac{6}{7}y_{n-1},\ \tfrac{6}{7}x_{n-1} + \tfrac{5}{7}y_{n-1}\right) \quad \text{für alle } n \in \mathbb{N}\,.$$

Betrachten wir wieder die Abstände vom Nullpunkt, so gilt hier für deren Quadrate die Beziehung

$$x_n^2 + y_n^2 = \left(\tfrac{5}{7}x_{n-1} - \tfrac{6}{7}y_{n-1}\right)^2 + \left(\tfrac{6}{7}x_{n-1} + \tfrac{5}{7}y_{n-1}\right)^2$$
$$= \tfrac{25}{49}x_{n-1}^2 - \tfrac{60}{49}x_{n-1}y_{n-1} + \tfrac{36}{49}y_{n-1}^2 + \tfrac{36}{49}x_{n-1}^2 + \tfrac{60}{49}x_{n-1}y_{n-1} + \tfrac{25}{49}y_{n-1}^2$$
$$= \tfrac{61}{49}x_{n-1}^2 + \tfrac{61}{49}y_{n-1}^2\,,$$

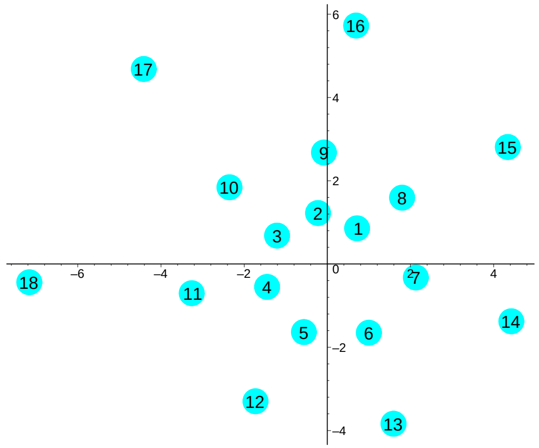
also ist

$$d\big((x_n, y_n),\,(0,0)\big) = \frac{\sqrt{61}}{7}\,d\big((x_{n-1}, y_{n-1}),\,(0,0)\big) \quad \text{für alle } n \geq 1\,.$$

Da der Punkt $(x_0, y_0) = (1,0)$ Abstand eins vom Nullpunkt hat, folgt wieder induktiv:

$$d\big((x_n, y_n),\,(0,0)\big) = \left(\frac{\sqrt{61}}{7}\right)^n .$$

Da $7^2 = 49 < 61$ ist, wird hier der Abstand vom Nullpunkt immer größer, zumindest gegen den Nullpunkt konvergiert diese Folge also sicher nicht.



Um zu sehen, daß sie auch gegen keinen anderen Punkt konvergieren kann, überlegen wir uns zunächst, daß der Abstand von (x_n, y_n) zum Nullpunkt beliebig groß werden kann, daß es also für jede Schranke

$M \in \mathbb{R}$ ein $n_0 \in \mathbb{N}$ gibt, so daß der Abstand zwischen den beiden Punkten größer als M ist. Dazu verwenden wieder die Ungleichung von BERNOULLI: Da das Quadrat von $7\frac{1}{2}$ mit $56\frac{1}{4}$ kleiner ist als 61, ist

$$d\big((x_n, y_n), (0,0)\big) = \left(\frac{\sqrt{61}}{7}\right)^n > \left(\frac{7\frac{1}{2}}{7}\right)^n = \left(1 + \frac{1}{14}\right)^n > 1 + \frac{n}{14} > \frac{n}{14}\,,$$

und das ist größer als M, wenn n größer ist als die kleinste natürliche Zahl n_0 größer oder gleich $14M$.

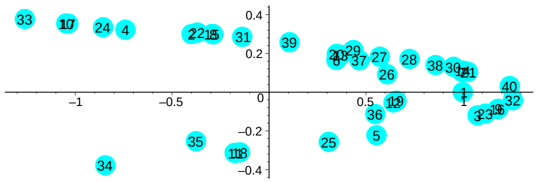
Wenn es nun einen Punkt $(x, y) \in \mathbb{R}^2$ gäbe, gegen den die Folge der $(x_n, y_n)_{n\in\mathbb{N}}$ konvergiert, so gäbe es insbesondere für $\varepsilon = 1$ ein $n_1 \in \mathbb{N}$, so daß der Abstand zwischen (x, y) und (x_n, y_n) für alle $n \ge n_1$ kleiner als eins wäre. Somit wäre für alle $n \ge n_1$ nach der Dreiecksungleichung

$$d\big((x_n, y_n), (0,0)\big) \le d\big((x_n, y_n), (x,y)\big) + d\big((x,y), (0,0)\big) < 1 + d\big((x,y), (0,0)\big)\,.$$

Andererseits haben wir bei unserer Folge für jedes M, also auch für $M = 1 + d\big((x,y), (0,0)\big)$, ein n_0, so daß $d\big((x_n, y_n), (0,0)\big) > M$ für alle $n \ge n_0$. Ist $n \ge n_0$ und $n \ge n_1$, so haben wir hier zwei einander widersprechende Ungleichungen für den Abstand zwischen (x_n, y_n) und dem Nullpunkt, also kann die Folge nicht konvergieren.

Als letztes Beispiel betrachten wir die Folge $(x_n, y_n)_{n\in\mathbb{N}}$, die vom Startwert $(x_0, y_0) = (0,0)$ ausgeht und für $n \in \mathbb{N}$ das Folgenglied (x_n, y_n) rekursiv definiert durch

$$x_n = y_{n-1} + 1 - 1{,}4x_{n-1}^2 \quad \text{und} \quad y_n = 0{,}3x_{n-1}\,.$$



Hier gibt uns das Bild keinerlei Hinweis auf das Konvergenzverhalten; die Punkte (x_n, y_n) bewegen sich anscheinend ziemlich planlos und

chaotisch auf oder zumindest in der Nähe einer etwas seltsamen Kurve. In der Tat haben wir es hier mit einem sogenannten chaotischen System zu tun; es wurde eingeführt von dem französischen Mathematiker und Astronomen MICHEL HÉNON (*1931) vom Observatoire de la Côte d'Azur in Nizza in seiner Arbeit

M. HÉNON: A Two-Dimensional Map with a Strange Attractor, *Commun. Math. Phys.* **50** (1976), S. 69–77

(Tatsächlich betrachtete er an Stelle der Konstanten 1,4 und 0,3 beliebige Parameter a und b und legte sich auch nicht auf einen bestimmten Startwert fest; seine numerischen Experimente (mit einem programmierbaren Taschenrechner!) führte er aber mit den hier verwendeten Werten durch.) Mit Methoden, die jenseits einer Vorlesung *Analysis I* liegen, konnte HÉNON diese Folge genauer untersuchen; insbesondere folgt aus seinen Ergebnissen, daß sie tatsächlich nicht konvergiert.

Diese Beispiele sollen uns fürs erste genügen; sie zeigen, daß sich Folgen in sehr unterschiedlicher Weise verhalten können und daß Konvergenz alles andere als selbstverständlich ist.

Trotzdem sind gerade die konvergenten Folgen wichtig für uns, denn da wir reelle Zahlen meist nur näherungsweise in den Griff bekommen, können wir die Lösung eines Problems oft nur angeben durch eine gegen diese Lösung konvergente Folge. Sinnvoll ist das natürlich nur, wenn diese Folge nur gegen einen Grenzwert konvergiert; wir wollen uns daher zum Abschluß dieses Paragraphen noch überlegen, daß dies in metrischen Räumen stets der Fall ist:

Lemma: Eine konvergente Folge $(x_n)_{n\in\mathbb{N}}$ in einem metrischen Raum X mit Metrik d konvergiert gegen *genau* einen Grenzwert.

Beweis: Wir nehmen an, die Folge würde sowohl gegen x als auch gegen y konvergieren. Zu jedem $\varepsilon > 0$ gibt es dann ein $n_1 \in \mathbb{N}$, so daß $d(x, x_n) < \frac{1}{2}\varepsilon$ ist für alle $n \ge n_1$ und ein $n_2 \in \mathbb{N}$, so daß $d(y, x_n) < \frac{1}{2}\varepsilon$ ist für alle $n \ge n_2$. Für alle n ab dem Maximum $n_0 = \max(n_1, n_2)$ der beiden Zahlen ist dann nach der Dreiecksungleichung

$$d(x,y) \le d(x, x_n) + d(x_n, y) = d(x, x_n) + d(y, x_n) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon\,.$$

Wenn der Abstand zwischen x und y aber für jedes $\varepsilon > 0$ kleiner als ε ist, muß er gleich Null sein, und das ist nach Definition einer Metrik nur möglich für $x = y$. Somit ist der Grenzwert einer konvergenten Folge eindeutig bestimmt.

■

Erst damit ist die Bezeichnung $\lim_{n\to\infty} x_n$ wirklich gerechtfertigt; bislang war noch nicht klar, daß dieser Wert für jede konvergente Folge eindeutig bestimmt ist.

§2: Folgen reeller und komplexer Zahlen

Wir spezialisieren uns nun auf die für die *Analysis I* wichtigen Folgen, also auf Folgen reeller Zahlen. Da sich auch komplexe Zahlen selbst bei der Lösung rein reeller Probleme immer wieder als nützlich erweisen werden, wollen wir, soweit dies ohne nennenswerten Zusatzaufwand möglich ist, unsere Resultate gleich auch für Folgen komplexer Zahlen beweisen.

Die einzige Metrik, die wir hier betrachten, ist sowohl im Reellen als auch im Komplexen die Metrik mit $d(x,y) = |x-y|$; eine Folge $(x_n)_{n\in\mathbb{N}}$ konvergiert also genau dann gegen eine reelle oder komplexe Zahl x, wenn es zu jedem $\varepsilon > 0$ ein $n_0 \in \mathbb{N}$ gibt, so daß $|x - x_n| < \varepsilon$ für alle $n \geq n_0$. Das ist natürlich gleichbedeutend damit, daß die Folge der Differenzen $(x - x_n)_{n\in\mathbb{N}}$ eine Nullfolge ist. Insbesondere sind die uns bereits bekannten Nullfolgen genau die Folgen, die gegen Null konvergieren.

Beginnen wir zunächst mit einigen einfachen Folgen, die uns zeigen, daß es selbst hier im Eindimensionalen schon eine Vielzahl von Möglichkeiten für das Langzeitverhalten einer Folge gibt.

Für zwei natürliche Zahlen $a, b \in \mathbb{N}$ kennen wir aus der Schule die Division mit Rest: Es gibt zu a und b stets zwei eindeutig bestimmte Zahlen $q, r \in \mathbb{N}_0$, so daß $a = qb + r$ ist mit $0 \leq r < b$. Man schreibt dann

$$a : b = q \text{ Rest } r ,$$

und wir bezeichnen den Divisionsrest r als a mod b, gesprochen a modulo b. Mit dieser Bezeichnung definieren wir für eine feste natürliche

Zahl m eine Folge $(x_n)_{n\in\mathbb{N}}$ durch die Vorschrift

$$x_n \underset{\text{def}}{=} n \bmod m .$$

Für $m = 1$ erhalten wir die konstante Folge, deren sämtliche Glieder gleich Null sind; schließlich ist jede natürliche Zahl ohne Rest durch eins teilbar. Diese Folge ist natürlich konvergent mit Grenzwert Null.

Für $m = 2$ ist $x_n = 0$ für gerade n und $x_n = 1$ für ungerade; die Folge wechselt also ständig zwischen 0 und 1 und kann somit nicht konvergent sein: Würde sie nämlich gegen eine reelle oder komplexe Zahl x konvergieren, so gäbe es für jedes $\varepsilon > 0$ ein $n_0 \in \mathbb{N}$, so daß $|x - x_n| < \varepsilon$ wäre für alle $n \geq n_0$. Unter den $n \geq n_0$ gibt es aber sowohl gerade als auch ungerade, also wäre

$$|x - 0| = |x| < \varepsilon \quad \text{und} \quad |x - 1| < \varepsilon .$$

Nach der Dreiecksungleichung wäre daher

$$1 = |1 - 0| \leq |1 - x| + |x - 0| = |x - 1| + |x| < \varepsilon + \varepsilon = 2\varepsilon ,$$

was für $\varepsilon < \frac{1}{2}$ offensichtlich nicht richtig sein kann.

Für $m = 3$ haben wir eine Folge, die ihre drei möglichen Werte $0, 1, 2$ ständig zyklisch reproduziert; ähnlich ist es für $m > 3$, wo die m Zahlen zwischen 0 und $m - 1$ immer wieder aufeinander folgen. In so einem Fall sagen wir, die Folge sei *zyklisch* mit Periode m; wie oben bei $m = 2$ sieht man leicht, daß sie für $m > 1$ nicht konvergent sein kann.

Als nächstes triviales Beispiel betrachten wir die Folge $(x_n)_{n\in\mathbb{N}}$ mit $x_n = n$, also einfach die Folge der natürlichen Zahlen. Selbstverständlich ist sie nicht konvergent; die Folgenglieder x_n werden schließlich immer größer. Trotzdem gibt es einen wichtigen Unterschied zum vorigen Beispiel: Während dort die Folgenglieder für $m > 1$ endlos einen immer gleichen Zyklus wiederholten, werden sie hier zielgerichtet immer größer und wachsen schließlich über jede vorgegebene Grenze. Entsprechend wird die Folge mit $x_n = -n$ immer kleiner und fällt schließlich unter jede vorgegebene Grenze. Um solche Folgen zu beschreiben, führen wir ein neues Symbol ∞ für „unendlich“ ein und definieren

Definition: Eine Folge $(x_n)_{n\in\mathbb{N}}$ reeller Zahlen heißt *bestimmt divergent* gegen ∞, wenn es für jede Schranke $M \in \mathbb{R}$ ein $n_0 \in \mathbb{N}$ gibt, so daß

$x_n > M$ für alle $n \geq n_0$. Sie heißt bestimmt divergent gegen $-\infty$, wenn es für jede Schranke $M \in \mathbb{R}$ ein $n_0 \in \mathbb{N}$ gibt, so daß $x_n < M$ für alle $n \geq n_0$. Sie heißt *unbestimmt divergent,* wenn sie weder konvergiert noch bestimmt divergiert gegen ∞ oder $-\infty$.

Damit ist also die Folge $(x_n)_{n\in\mathbb{N}}$ mit $x_n = n$ bestimmt divergent gegen ∞, die mit $x_n = -n$ gegen $-\infty$. Die Folgen mit $x_n = n \bmod m$ dagegen sind für $m \geq 2$ unbestimmt divergent.

Eine Folge ist sicherlich dann nicht bestimmt divergent gegen ∞, wenn es eine Schranke gibt, unter der alle Folgenglieder liegen; entsprechend kann sie nicht gegen $-\infty$ divergieren, wenn es eine untere Schranke gibt. Wir definieren

Definition: *a)* Eine Folge $(x_n)_{n\in\mathbb{N}}$ reeller Zahl heißt *nach oben beschränkt,* wenn es eine reelle Zahl M gibt, so daß $x_n \leq M$ für alle $n \in \mathbb{N}$. Sie heißt *nach unten beschränkt,* wenn es eine reelle Zahl N gibt, so daß $x_n \geq N$ für alle $n \in \mathbb{N}$. Die Zahlen M und N bezeichnen wir als obere *bzw.* untere Schranken.
b) Eine Folge $(x_n)_{n\in\mathbb{N}}$ reeller oder komplexer Zahlen heißt *beschränkt,* wenn es eine reelle Zahl M gibt, so daß $|x_n| \leq M$ für alle $n \in \mathbb{N}$.

Im reellen Fall ist die Folge natürlich genau dann beschränkt, wenn sie sowohl nach oben als auch nach unten beschränkt ist: Falls die Ungleichung $N \leq x_n \leq M$ gilt, ist $|x_n| \leq \max(|N|, |M|)$, und ist $|x_n| \leq M$, so ist $-M \leq x_n \leq M$.

Die Folgen $(n \bmod m)_{n\in\mathbb{N}}$ sind Beispiele beschränkter, aber nicht konvergenter Folgen.

Da fast alle Hörer dieser Vorlesung etwas studieren, das *Wirtschaft* im Namen hat, wollen wir als nächstes Beispiel das einfachste Beispiel einer Kapitalanlage betrachten: Wir nehmen an, es gäbe heute noch irgendeine Bank, die anbietet, ein angelegtes Kapitel für alle Zeiten zu einem festen, im Voraus vereinbarten Satz zu verzinsen. Wir legen also ein festes Kapital $x_0 = a > 0$ an, und jedes Jahr gibt uns die Bank darauf p% Zinsen, multipliziert das bis dahin vorhandene Kapital also mit einen Faktor $q = 1 + r$, wobei r als Abkürzung für $p/100$ stehen soll. Bezeichnen wir das nach n Jahren vorhandene Kapital mit x_n, so ist also $x_n = x_{n-1}q = x_{n-1}(1 + r)$, woraus wir sofort induktiv folgern,

können, daß $x_n = x_0 q^n = aq^n$ ist. Sinn der Kapitalanlage ist natürlich, daß unser Kapital wächst und wächst und wächst; in der Mathematik sagen wir, daß es *streng monoton* wachsen soll:

Definition: Eine Folge $(x_n)_{n\in\mathbb{N}}$ von reellen Zahlen heißt *monoton wachsend,* wenn für alle $n \in \mathbb{N}$ gilt: $x_{n+1} \geq x_n$. Falls sogar $x_{n+1} > x_n$, reden wir von einer *streng monoton wachsenden* Folge. Entsprechend heißt die Folge *monoton fallend,* wenn $x_{n+1} \leq x_n$ für alle $n \in \mathbb{N}$ und *streng monoton fallend,* wenn sogar $x_{n+1} < x_n$ ist für alle $n \in \mathbb{N}$.

So versprechen uns beispielsweise viele Politiker, daß die Staatsverschuldung der nächsten Jahre eine streng monoton fallende Folge bilden werden, sind aber leider durch widrige Umstände gezwungen, daraus eine sehr streng monoton wachsende Folge zu machen.

Bei unserer Kapitalanlage haben wir mit der Bank natürlich einen Zinssatz $p > 0$ vereinbart; damit ist auch $r = p/100$ positiv, und nach der Ungleichung von BERNOULLI ist

$$q^n = (1 + r)^n > 1 + nr \quad \text{für alle } n \geq 2\,.$$

In diesem Zusammenhang sagt uns die BERNOULLIsche Ungleichung also einfach, daß Zinseszins mehr einbringt als bloßer Zins.

Sie sagt uns allerdings auch, daß unser Kapital im Laufe der Zeit beliebig groß wird: $aq^n > a(1 + nr)$, und offensichtlich können wir auch noch zu einer beliebig großen Zahl M ein n_0 finden, so daß $a(1 + nr) > M$ ist für alle $n \geq n_0$. (Die BERNOULLIsche Ungleichung kann uns freilich nicht garantieren, daß wir in n_0 Jahren noch leben.)

Dieses Ergebnis wollen wir gleich etwas allgemeiner festhalten:

Lemma: Für eine komplexe Zahl q ist die Folge $(q^n)_{n\in\mathbb{N}}$ eine Nullfolge, falls $|q| < 1$; sie divergiert und wächst unbeschränkt, falls $|q| > 1$. Für $q = 1$ ist die Folge konstant gleich eins, für $|q| = 1$, aber $q \neq 1$ divergiert sie.

Beweis: Sei zunächst $|q| > 1$. Dann ist $r = |q| - 1 > 0$ und für $n \geq 2$ ist nach der Ungleichung von BERNOULLI

$$|q^n| = |q|^n = (1 + r)^n > 1 + rn > rn\,.$$

Für eine vorgegebene Schranke $M \in \mathbb{R}$ finden wir daher eine natürliche Zahl n_0, z.B. jedes $n_0 \ge M/r$, so daß $|q|^n > M$ ist für alle $n \ge n_0$.

Ist $|q| < 1$, so müssen wir zeigen, daß $(q^n)_{n\in\mathbb{N}}$ eine Nullfolge ist. Sei also $\varepsilon > 0$ vorgegeben. Da $|1/q| > 1$ ist, können wir das gerade bewiesene Resultat anwenden, wonach es ein $n_0 \in \mathbb{N}$ gibt, so daß $|1/q^n| > 1/\varepsilon$ ist für alle $n \ge n_0$. Für diese n ist dann aber auch $|q^n| < \varepsilon$, womit die Nullfolgeneigenschaft bewiesen ist.

Für $q = 1$ schließlich ist die Behauptung trivial; ist $q \ne 1$, aber $|q| = 1$, so ist natürlich auch $|q^n| = 1$ für alle n, aber die Folge der q^n divergiert: Andernfalls müßte sie gegen einen Grenzwert x konvergieren, es gäbe also zu jedem $\varepsilon > 0$ ein n_0, so daß $|x - q^n| < \varepsilon$ für alle $n \ge n_0$. Ist $n \ge n_0$; so ist auch $n + 1 \ge n_0$; nach der Dreiecksungleichung wäre daher

$$\left|q^n - q^{n+1}\right| \le |x - q^n| + \left|x - q^{n+1}\right| < 2\varepsilon\,.$$

Nun ist aber $\left|q^n - q^{n+1}\right| = |q^n| \cdot |1-q| = |1-q|$ unabhängig von n, und für $\varepsilon \le \frac{1}{2}|1-q|$ kann dies unmöglich echt kleiner 2ε sein. Somit divergiert die Folge.

■

Bislang haben wir vor allem Aussagen über Konvergenz, Divergenz und Grenzwerte bewiesen, die eigentlich auch so klar waren; um auch nicht offensichtliche Grenzwerte ausrechnen zu können, benötigen wir als erstes Beziehungen zwischen den Grenzwerten verschiedener Folgen. Dazu diesen die folgenden

Rechenregeln für Grenzwerte: Sind $(x_n)_{n\in\mathbb{N}}$ und $(y_n)_{n\in\mathbb{N}}$ konvergente Folgen reeller oder komplexer Zahlen mit Grenzwerten x und y, so gilt:

a) Die Folge $(x_n \pm y_n)_{n\in\mathbb{N}}$ konvergiert gegen $x \pm y$.

b) Die Folge $(x_n y_n)_{n\in\mathbb{N}}$ konvergiert gegen xy.

c) Ist $y \ne 0$, so gibt es eine natürliche Zahl n_1 derart, daß $y_n \ne 0$ für alle $n \ge n_1$. Betrachtet man nur Folgenglieder mit Index $n \ge n_1$, so konvergiert die Folge $(x_n/y_n)_{n\ge n_1}$ gegen x/y.

d) Die Folge $(|x_n|)_{n\in\mathbb{N}}$ konvergiert gegen $|x|$.

e) Für reelle Folgen gilt: Ist $x_n \le y_n$ für alle n, so ist auch $x \le y$.

Beweis: a) Wir müssen zeigen, daß es zu jedem $\varepsilon > 0$ ein $n_0 \in \mathbb{N}$ gibt, so daß für alle $n \ge n_0$ gilt

$$|(x \pm y) - (x_n \pm y_n)| \le \varepsilon\,.$$

Nach der Dreiecksungleichung ist

$$|(x \pm y) - (x_n \pm y_n)| = |(x - x_n) \pm (y - y_n)| \le |x - x_n| + |y - y_n|\,.$$

Da die Folge $(x_n)_{n\in\mathbb{N}}$ gegen x konvergiert, gibt es ein $n_1 \in \mathbb{N}$, so daß $|x - x_n| < \frac{1}{2}\varepsilon$ ist für alle $n \ge n_1$; genauso gibt wegen der Konvergenz der y_n gegen y ein $n_2 \in \mathbb{N}$, so daß $|y - y_n| < \frac{1}{2}\varepsilon$ für alle $n \ge n_2$. Für $n \ge n_0 = \max(n_1, n_2)$ gelten beide Ungleichungen, also auch die zu beweisende.

b) Wir müssen zu vorgegebenem $\varepsilon > 0$ ein $n_0 \in \mathbb{N}$ finden, so daß $|xy - x_n y_n| < \varepsilon$ ist. Dazu schreiben wir

$$\begin{aligned}|xy - x_n y_n| &= |x(y - y_n) + y_n(x - x_n)|\\ &\le |x| \cdot |y - y_n| + |y_n| \cdot |x - x_n|\,.\end{aligned}$$

$|x|$ ist eine Konstante, $|y_n|$ jedoch nicht. Um trotzdem eine Aussage über $|y_n|$ zu bekommen, beweisen wir zunächst die folgende

Zwischenbehauptung: Ist $(y_n)_{n\in\mathbb{N}}$ keine Nullfolge, so gibt es einen Index $n_1 \in \mathbb{N}$, so daß

$$\frac{|y|}{2} < |y_n| < \frac{3\,|y|}{2} \quad \text{für alle } n \ge n_1\,.$$

Zum *Beweis* müssen wir einfach in der Definition einer konvergenten Folge $\varepsilon = \frac{1}{2}|y|$ setzen; dann erhalten wir ein n_1, so daß für alle $n \ge n_1$ gilt $|y_n - y| \le \frac{1}{2}|y|$. Nach der verschärften Dreiecksungleichung ist dann auch $\big||y_n| - |y|\big| \le \frac{1}{2}|y|$ und

$$|y| - \frac{|y|}{2} < |y_n| < |y| + \frac{|y|}{2} \quad \text{für alle } n \ge n_1\,.$$

Das ist genau die behauptete Ungleichung.

■

Falls $(y_n)_{n\in\mathbb{N}}$ keine Nullfolge ist, haben wir demnach die Abschätzung

$$|xy - x_n y_n| \le |x| \cdot |y - y_n| + \frac{3}{2}|y| \cdot |x - x_n| \quad \text{für alle } n \ge n_1\,.$$

Um die rechte Seite unter eine vorgegebene Schranke ε zu drücken, betrachten wir ein $n_2 \in \mathbb{N}$, so daß für alle $n \geq n_2$ gilt $|x - x_n| < \varepsilon/3\,|y|$ und, falls $x \neq 0$, ein $n_3 \in \mathbb{N}$, so daß $|y - y_n| < \varepsilon/2\,|x|$ ist. Im Falle $x \neq 0$ haben wir dann die Abschätzung

$$|xy - x_n y_n| \leq |x| \cdot |y - y_n| + \frac{3}{2}\,|y| \cdot |x - x_n|$$
$$\leq |x| \cdot \frac{\varepsilon}{2\,|x|} + \frac{3}{2}\,|y| \cdot \frac{\varepsilon}{3\,|y|} = \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$$

für alle $n \geq n_0 = \max(n_1, n_2, n_3)$. Für $x = 0$ ist

$$|xy - x_n y_n| \leq |x| \cdot |y - y_n| + \frac{3}{2}\,|y| \cdot |x - x_n|$$
$$= \frac{3}{2}\,|y| \cdot |x - x_n| \leq \frac{3}{2}\,|y| \cdot \frac{\varepsilon}{3\,|y|} = \frac{\varepsilon}{2} < \varepsilon$$

für alle $n \geq n_0 = \max(n_1, n_2)$.

Bleibt noch der Fall, daß $(y_n)_{n\in\mathbb{N}}$ eine Nullfolge ist. In diesem Fall wählen wir unser n_1 so, daß $|y_n| < 1$ ist für alle $n \geq n_1$ und n_2 so, daß $|x - x_n| < \frac{1}{2}\varepsilon$ für alle $n \geq n_2$. Damit ist auch hier der Summand $\frac{3}{2}\,|y| \cdot |x - x_n|$ kleiner als $\frac{1}{2}\varepsilon$ für $n \geq \max(n_1, n_2)$. Im Fall $x = 0$ reicht das; für $x \neq 0$ schätzen wir den Summanden $|x| \cdot |y - y_n|$ wie oben ab und haben für $n \geq \max(n_1, n_2, n_3)$ die gleiche Ungleichung wie dort.

c) Da $(y_n)_{n\in\mathbb{N}}$ keine Nullfolge ist, gibt es nach der in *b)* bewiesenen Zwischenbehauptung ein $n_1 \in \mathbb{N}$, so daß

$$\frac{|y|}{2} < |y_n| < \frac{3\,|y|}{2} \quad \text{für alle } n \geq n_1\,.$$

Insbesondere ist also $y_n \neq 0$ für alle $n \geq n_1$.

Wegen *b)* genügt es, wenn wir zeigen, daß die Folge $(1/y_n)_{n\geq n_1}$ gegen $1/y$ konvergiert.

$$\left|\frac{1}{y} - \frac{1}{y_n}\right| = \left|\frac{y_n - y}{y y_n}\right| \leq \left|\frac{y_n - y}{\frac{1}{2}y^2}\right| = \frac{2}{|y|^2}\,|y - y_n|\,.$$

Zu jedem $\varepsilon > 0$ gibt es wegen der Konververgenz der Folge $(y_n)_{n\in\mathbb{N}}$ ein $n_2 \in \mathbb{N}$, so daß $|y - y_n| < \frac{1}{2}\,|y|^2\,\varepsilon$ ist. Für $n \geq n_0 = \max(n_1, n_2)$ ist dann nach der gerade bewiesenen Abschätzung auch $|1/y - 1/y_n| < \varepsilon$, die Folge der Kehrwerte konvergiert also gegen $1/y$.

d) Wir müssen zeigen, daß es zu jedem $\varepsilon > 0$ ein $n_0 \in \mathbb{N}$ gilt, so daß

$$\big||x| - |x_n|\big| < \varepsilon \quad \text{für alle } n \geq n_0\,.$$

Wir wissen, daß es ein n_0 gibt, so daß

$$|x - x_n| < \varepsilon \quad \text{für alle } n \geq n_0\,,$$

und wenn wir in der verschärften Dreiecksungleichung

$$\big||x - y| - |y - z|\big| \leq |x - z| \leq |x - y| + |y - z|$$

$y = 0$ und $z = x_n$ setzen, erhalten wir die gewünschte Ungleichung

$$\big||x| - |x_n|\big| \leq |x - x_n| < \varepsilon \quad \text{für alle } n \geq n_0\,.$$

e) Wäre $x > y$, so gäbe es für $\varepsilon = \frac{1}{2}(x - y)$ natürliche Zahlen n_1 und n_2, so daß $|x - x_n| < \varepsilon$ wäre für alle $n \geq n_1$ und $|y - y_n| < \varepsilon$ für alle $n \geq n_2$. Für $n \geq \max(n_1, n_2)$ wäre dann

$$x_n > x - \varepsilon = \frac{x + y}{2} = y + \varepsilon > y_n\,,$$

im Widerspruch zur Voraussetzung $x_n \leq y_n$. Also ist $x \leq y$. ■

Damit haben wir erste Instrumente in der Hand, um mit Grenzwerten zu rechnen; wir können diese Rechenregeln auch so formulieren, daß die Berechnung von Grenzwerten konvergenter Folgen vertauschbar ist mit den Grundrechenarten, der Betragsfunktion und, im reellen Fall, der Ordnungsrelation $\leq$:

$$\lim_{n\to\infty}(x_n \pm y_n) = \lim_{n\to\infty} x_n \pm \lim_{n\to\infty} y_n\,,$$
$$\lim_{n\to\infty}(x_n \cdot y_n) = \lim_{n\to\infty} x_n \cdot \lim_{n\to\infty} y_n\,,$$
$$\lim_{n\to\infty}\frac{x_n}{y_n} = \frac{\lim_{n\to\infty} x_n}{\lim_{n\to\infty} y_n} \quad \text{falls } (y_n)_{n\in\mathbb{N}} \text{ keine Nullfolge ist}\,,$$
$$\lim_{n\to\infty}|x_n| = \left|\lim_{n\to\infty} x_n\right|\,,$$
$$\lim_{n\to\infty} x_n \leq \lim_{n\to\infty} y_n \quad \text{für reelle Folgen mit } x_n \leq y_n \text{ für alle } n\,.$$

Man beachte, daß die letzte Zeile falsch wird, wenn wir $\leq$ durch $<$ ersetzen: Ist $x_n = 1/2n$ und $y_n = 1/n$ so ist $x_n < y_n$ für alle n, aber die Grenzwerte sind beide gleich Null.

Bei der Aussage über die Quotientenfolge darf man n natürlich erst ab einem Wert laufen lassen, jenseits dessen kein y_n mehr verschwinden kann.

§3: Die Vollständigkeit der reellen Zahlen

Bisher haben wir die Konvergenz einer Folge immer über die Definition bewiesen, d.h. wir haben bewiesen, daß sie gegen eine uns bekannte, konkrete Zahl konvergiert. Eine der Hauptanwendungen von Folgen besteht aber darin, daß wir damit *unbekannte* Zahlen berechnen wollen. Wir brauchen daher unbedingt Kriterien, mit denen wir die Existenz eines Grenzwerts beweisen können, ohne diesen bereits zu kennen. Damit soll sich dieser Paragraph beschäftigen.

Definition: k sei ein angeordneter Körper und A eine Teilmenge von k. Eine Zahl $M \in k$ heißt $\left\{\begin{smallmatrix}\textit{obere}\\ \textit{untere}\end{smallmatrix}\right\}$ *Schranke* von A, wenn $\left\{\begin{smallmatrix}x\le M\\ x\ge M\end{smallmatrix}\right\}$ für alle $x \in A$. Sie heißt $\left\{\begin{smallmatrix}\textit{Supremum}\\ \textit{Infimum}\end{smallmatrix}\right\}$ von A, wenn sie eine $\left\{\begin{smallmatrix}\textit{obere}\\ \textit{untere}\end{smallmatrix}\right\}$ Schranke von A ist und wenn für jede andere $\left\{\begin{smallmatrix}\textit{obere}\\ \textit{untere}\end{smallmatrix}\right\}$ Schranke N von A gilt: $\left\{\begin{smallmatrix}M\le N\\ M\ge N\end{smallmatrix}\right\}$. Die Menge A heißt nach $\left\{\begin{smallmatrix}\textit{oben}\\ \textit{unten}\end{smallmatrix}\right\}$ beschränkt, wenn sie eine $\left\{\begin{smallmatrix}\textit{obere}\\ \textit{untere}\end{smallmatrix}\right\}$ Schranke hat. Wenn es sowohl eine obere als auch eine untere Schranke gibt, bezeichnen wir A als *beschränkt.*

Das Supremum einer Teilmenge $A \subseteq k$, falls es existiert, ist die *kleinste* obere Schranke. Sie muß nicht existieren: Für $k = \mathbb{Q}$ beispielsweise hat die Teilmenge $A = \left\{x \in \mathbb{Q} \mid x^2 < 2\right\}$ zwar viele obere Schranken, aber keine kleinste: Ist nämlich M irgendeine obere Schranke, so gilt in $\mathbb{R}$ die Ungleichung $M > \sqrt{2}$, denn da M eine rationale Zahl sein muß, ist $M \neq \sqrt{2}$. Damit ist $x = M - \sqrt{2}$ eine positive reelle Zahl, und wir können ein $\varepsilon > 0$ finden, die kleiner ist, z.B. eine untere Grenze eines Intervalls aus einer Intervallschachtelung die x definiert. Dann ist aber auch $M - \varepsilon$ eine obere Schranke von M.

Wenn allerdings ein Supremum existiert, dann ist es eindeutig bestimmt: Sind nämlich M und N zwei Suprema, so sind beide insbesondere obere Schranken; da M Supremum ist, muß daher$M \le N$ sein, und da N Supremum ist auch $N \le M$. Somit ist $M = N$.

Ein wesentlicher Unterschied zwischen reellen und rationalen Zahlen besteht darin, daß jede beschränkte Teilmenge von $\mathbb{R}$ ein Infimum und ein Supremum hat:

Satz: Jede nichtleere nach oben beschränkte Teilmenge $A \subseteq \mathbb{R}$ hat ein Supremum.

Beweis: Wir konstruieren das Supremum über eine Intervallschachtelung mit rationalen Grenzen. Dazu gehen wir aus von einer oberen Schranke M sowie einem Element $x \in A$. Beides sind reelle Zahlen; für unser erstes Intervall $[a_1, b_1]$ wählen wir eine rationale Zahl $a_1 \le x$ und eine rationale Zahl $b_1 \ge M$. Die obere Intervallgrenze ist somit eine obere Schranke von A, und das Intervall enthält (mindestens) ein Element $x \in A$. Diese beiden Eigenschaften werden auch die rekursiv konstruierten weiteren Intervalle haben: Wenn wir die Intervalle bis $[a_n, b_n]$ konstruiert haben, betrachten wir den Intervallmittelpunkt $c_n = \frac{1}{2}(a_n + b_n)$. Falls c_n eine obere Schranke von A ist, setzen wir $a_{n+1} = a_n$ und $b_{n+1} = c_n$. Alle $x \in A$, die in $[a_n, b_n]$ liegen, sind auch im neuen Intervall, und $b_{n+1} = c_n$ ist eine obere Schranke von A.

Falls c_n keine obere Schranke ist, gibt es ein $x \in A$ mit $c_n \le y$. Wir setzen $a_{n+1} = c_n$ und $b_{n+1} = b_n$. Wieder ist b_{n+1} obere Schranke und das Intervall $[a_{n+1}, b_{n+1}]$ enthält mindestens ein $x \in A$.

Das neue Intervall $[a_{n+1}, b_{n+1}]$ ist die linke oder die rechte Hälfte seines Vorgängers $[a_n, b_n]$ und ist somit auch nur halb so lang; damit haben wir eine Intervallschachtelung, und die definiert eine reelle Zahl S. Wir wollen uns überlegen, daß S das Supremum von A ist.

Zunächst ist S eine obere Schranke von A, denn sonst gäbe es ein Element $y \in A$ mit $y > S$. Damit gäbe es nach Definition der Größerbeziehung auf $\mathbb{R}$ auch ein n, so daß $y > b_n$ wäre, aber das ist nicht möglich, denn alle b_n sind obere Schranken von A.

Nun sei N irgendeine obere Schranke von A; wir müssen zeigen, daß $S \le N$ ist. Angenommen, S wäre größer als N. Dann gäbe es ein n mit $a_n > N$. Nach Konstruktion der Intervalle gibt es aber mindestens ein Element $y \in A$ mit $y \ge a_n$, also wäre auch $y \ge N$, im Widerspruch zur Schrankeneigenschaft. Somit ist $S \le N$, d.h. S ist das Supremum von A. ■

Definition: Ein angeordneter Körper k heißt *vollständig,* wenn jede nach oben beschränkte Teilmenge $A \subseteq k$ ein Supremum hat.

Somit ist also der Körper der reellen Zahlen vollständig, nicht aber der der rationalen Zahlen.

Wir können uns fragen, ob es außer den reellen Zahlen noch andere vollständige angeordnete Körper gibt, oder ob $\mathbb{R}$ durch diese Forderung bereits eindeutig charakterisiert ist. Diese Frage ist für den Rest der Vorlesung zwar nicht weiter wichtig; da es aber eine ganze Reihe von Lehrbüchern der Analysis gibt, die den Körper der reellen Zahlen axiomatisch einführen, möchte ich doch kurz darauf eingehen.

Man kann vollständige angeordnete Körper konstruieren, die größer sind als der Körper der reellen Zahlen, indem man z.B. ein neues Element T dazunimmt, das größer sein soll als alle reellen Zahlen. (Das Element $1/T$ ist dann zwar positiv, aber kleiner als jedes reelle $\varepsilon > 0$.) Die Potenzen T^n dieses Elements müssen dann alle verschieden sein, denn $T < T^2 < T^3 < \cdots$.

Wegen der Vollständigkeit enthält der neue Körper nicht nur Polynome in T, sondern auch sogenannte Potenzreihen (mit denen wir uns später in dieser Vorlesung noch ausführlich beschäftigen werden), die wegen der Körpereigenschaft auch zumindest endlich viele negative Potenzen von T enthalten dürfen. Man kann zeigen, daß man so einen vollständigen angeordneten Körper erhält.

Um derartige Körper auszuschließen, müssen wir eine bislang nie explizit erwähnte weitere Eigenschaft reeller Zahlen betrachten: Für jede reelle Zahl x gibt es eine natürliche Zahl $N > x$, denn wegen der Konstruktion der reellen Zahlen via Intervallschachtelungen gibt es rationale Zahlen $b > x$, und zu jeder rationalen Zahl kann man mit elementarer Bruchrechnung eine größere natürliche Zahl finden.

Allgemein bezeichnet man einen angeordneten Körper k als *archimedisch,* wenn es zu jedem $x \in k$ eine natürliche Zahl N gibt mit $N > x$ und kann dann zeigen, daß es außer den reellen Zahlen keinen weiteren archimedischen vollständigen angeordneten Körper gibt. Die Idee ist einfach, wenn auch die detaillierte Ausführung etwas länglich und mühsam wäre:

In einem angeordneten Körper ist $1 > 0$; die Summe von n Einsen ist daher stets größer als die von $n-1$ Einsen; insbesondere sind diese Summen alle verschieden. Identifizieren wir die Summe von n Einsen mit der natürlichen Zahl n, haben wir also $\mathbb{N}$ mit einer Teilmenge von k identifiziert; wegen der Körperaxiome folgt leicht, daß wir dann auch eine Teilmenge finden, die wir mit $\mathbb{Q}$ identifizieren können. Wie wir gleich sehen werden, muß sich in jedem vollständigen angeordneten Körper jede Intervallschachtelung auf ein Element des Körpers zusammenziehen; somit können wir auch $\mathbb{R}$ mit einer Teilmenge von k identifizieren. Wegen der Archimedizität schließlich lassen sich zu jedem $x \in k$ ganze Zahlen N, M finden mit $N \le x \le M$; durch Intervallhalbierung läßt sich daraus eine Intervallschachtelung mit rationalen Grenzen finden, sie sich auf x zusammenzieht. Somit stimmt k (modulo der obigen Identifizierungen) mit $\mathbb{R}$ überein.

Die im Rest dieses Paragraphen für $\mathbb{R}$ formulierten Sätze gelten, wie man sich leicht überlegt, für beliebige vollständige angeordnete Körper; da wir außer $\mathbb{R}$ keinen brauchen, werde ich sie aber der Übersichtlichkeit halber nur für $\mathbb{R}$ formulieren.

Aus dem gerade bewiesenen Satz über die Existenz von Suprema können wir als Folgerung sofort eine Aussage über die Existenz von Infima ableiten:

Korollar: Jede nach unten beschränkte nichtleere Teilmenge $A \subseteq \mathbb{R}$ hat ein Infimum.

Beweis: Ist M untere Schranke von A, so ist $-M$ obere Schranke der Menge $B = \{x \in k \mid -x \in A\}$. Also hat B ein Supremum S, und offensichtlich ist $-S$ das Infimum von A. ■

Nach diesen Vorbereitungen können wir nun endlich eine Aussage über die Existenz von Grenzwerten beweisen:

Satz: Jede monoton wachsende und nach oben beschränkte Folge $(x_n)_{n\in\mathbb{N}}$ reeller Zahlen konvergiert.

Beweis: $A = \{x_n \mid n \in \mathbb{N}\}$ sei die Menge aller Folgenglieder. Nach Voraussetzung hat sie eine obere Schranke, also hat sie auch ein Supremum x. Wir wollen uns überlegen, daß die Folge gegen x konvergiert. Wir betrachten also ein vorgegebenes $\varepsilon > 0$. Dann kann $x - \varepsilon$ keine obere Schranke von A sein, denn jede obere Schranke ist größer oder gleich x. Deshalb gibt es ein $n_0 \in \mathbb{N}$, so daß $x_{n_0} > x - \varepsilon$ ist. Wegen der Monotonie der Folge ist $x_{n_0} \le x_n$ für alle $n \ge n_0$, und da x insbesondere eine obere Schranke ist, muß $x_n \le x$ sein für alle n. Für $n \ge n_0$ ist somit

$$x - \varepsilon < x_{n_0} \le x_n \le x\,.$$

Damit muß $|x - x_n| < \varepsilon$ sein, die Folge konvergiert also gegen x. ■

Korollar: Jede monoton fallende und nach unten beschränkte Folge $(x_n)_{n\in\mathbb{N}}$ reeller Zahlen konvergiert.

Beweis: $(-x_n)_{n\in\mathbb{N}}$ ist eine monoton wachsende und nach oben beschränkte Folge, hat also nach dem Satz einen Grenzwert y, gegen den sie konvergiert. Damit konvergiert die Folge der x_n gegen $x = -y$. ■

Lemma: Ist $([a_n, b_n])_{n\in\mathbb{N}}$ eine Intervallschachtelung mit rationalen oder reellen Intervallgrenzen, so konvergieren die Folgen $(a_n)_{n\in\mathbb{N}}$ und $(b_n)_{n\in\mathbb{N}}$ beide gegen dieselbe reelle Zahl x.

Beweis: Nach Definition einer Intervallschachtelung gelten für alle $n \in \mathbb{N}$ die Ungleichungen $a_1 \le a_n \le b_n \le b_1$; somit ist b_1 eine obere Schranke für $(a_n)_{n\in\mathbb{N}}$ und a_1 eine untere Schranke für $(b_n)_{n\in\mathbb{N}}$. Da $[a_{n+1}, b_{n+1}]$ stets in $[a_n, b_n]$ liegen muß, ist die Folge der unteren Grenzen monoton wachsend, konvergiert also nach dem gerade bewiesenen Satz gegen eine reelle Zahl x. Genauso folgt aus dem Korollar, daß die oberen Grenzen gegen eine reelle Zahl y konvergieren. Für jedes n ist $a_n \le x \le y \le b_n$; da die Differenzen $(b_n - a_n)_{n\in\mathbb{N}}$ eine Nullfolge bilden, muß $x = y$ sein. ■

Im Falle einer Intervallschachtelung mit rationalen Grenzen ist x gerade die durch die Intervallschachtelung definierte reelle Zahl; für Intervallschachtelungen mit reellen Grenzen zeigt uns das Lemma, daß wir die reellen Zahlen nicht wie die rationalen Zahlen via Intervallschachtelungen erweitern können zu irgendwelchen hyperreellen Zahlen; aus diesem Grund reden wir auch von einem *vollständigen* Körper. Wir wollen sehen, daß wir dort nicht nur bei monotonen, sondern auch bei beliebigen reellen Folgen nützliche Konvergenzkriterien haben.

Ausgangspunkt dafür ist die Beobachtung, daß eine beliebige Folge reeller Zahlen mindestens eine monotone Teilfolge hat. Bevor wir das zeigen können, müssen wir zunächst definieren, was eine Teilfolge sein soll; anschaulich entsteht sie aus einer Folge dadurch, daß gewisse Folgenglieder gestrichen werden und wir das, was übrig bleibt, als neue Folge nehmen.

Definition: Eine Folge $(y_n)_{n\in\mathbb{N}}$ heißt *Teilfolge* der Folge $(x_n)_{n\in\mathbb{N}}$, wenn es eine streng monoton wachsende Folge $(\nu_n)_{n\in\mathbb{N}}$ von natürlichen Zahlen gibt, so daß $y_n = x_{\nu_n}$ für alle $n \in \mathbb{N}$.

So ist beispielsweise die Folge $(y_n)_{n\in\mathbb{N}}$ mit $y_n = 1/(2n+1)$ eine Teilfolge der Folge $(x_n)_{n\in\mathbb{N}}$ mit $x_n = 1/n$, denn setzen wir $\nu_n = 2n+1$, so definiert dies eine streng monoton wachsende Folge natürlicher Zahlen, und $y_n = x_{\nu_n}$ für alle $n \in \mathbb{N}$. Genauso ist auch die Folge $(z_n)_{n\in\mathbb{N}}$ mit $z_n = 1/n^2$ eine Teilfolge von $(x_n)_{n\in\mathbb{N}}$; hier ist $z_n = x_{\mu_n}$ mit $\mu_n = n^2$, was ebenfalls eine streng monoton wachsende Folge definiert. Dagegen ist die Folge $(w_n)_{n\in\mathbb{N}}$ mit $w_n = 1/(1 + |n^2 - 11|)$ *keine* Teilfolge, denn zwar können wir auch hier schreiben $w_n = x_{\lambda_n}$ mit $\lambda_n = 1 + |n^2 - 11|$, aber die Folge $(\lambda_n)_{n\in\mathbb{N}}$ ist nicht streng monoton wachsend: Ihre ersten Folgenglieder sind $11, 8, 3, 6, 15$.

Lemma: Jede reelle Zahlenfolge $(x_n)_{n\in\mathbb{N}}$ hat eine monotone Teilfolge.

Beweis: Als erstes müssen wir entscheiden, ob wir nach einer monoton wachsenden oder einer monoton fallenden Teilfolge suchen wollen. Dazu betrachten wir die Menge $J \subseteq \mathbb{N}$ aller Indizes n, für die gilt: $x_m \le x_n$ für alle $m \ge n$. Für ein $n \in J$ sind die Folgenglieder, die *hinter* x_n stehen also allesamt kleiner oder gleich x_n. Im Falle einer monoton fallenden Folge $(x_n)_{n\in\mathbb{N}}$ wäre offensichtlich $J = \mathbb{N}$, bei einer streng monoton wachsenden Folge wäre $J = \emptyset$ die leere Menge. Wir haben es mit einer beliebigen Folge zu tun; hier ist für J grundsätzlich alles möglich.

Wenn J eine unendliche Teilmenge der natürlichen Zahlen ist, ordnen wir ihre Elemente der Größe nach und bezeichnen das n-te mit ν_n. Dann ist $(\nu_n)_{n\in\mathbb{N}}$ eine streng monoton wachsende Folge natürlicher Zahlen, und die Folge $(y_n)_{n\in\mathbb{N}}$ mit $y_n = x_{\nu_n}$ ist offensichtlich monoton fallend: Da ν_{n+1} größer als ν_n ist, muß nach Definition von J insbesondere $y_{n+1} = x_{\nu_{n+1}} \le x_{\nu_n} = y_n$ sein für alle $n \in \mathbb{N}$.

Falls J endlich ist, können wir eine natürliche Zahl ν_1 finden, die größer ist als alle Elemente von J: Im Falle $J = \emptyset$ setzen wir einfach $\nu_1 = 1$, ansonsten nehmen wir beispielsweise das um eins vergrößerte Maximum von J.

Beginnend mit ν_1 konstruieren wir rekursiv eine streng monoton wachsende Folge $(\nu_n)_{n\in\mathbb{N}}$ von natürlichen Zahlen: Angenommen, wir haben die Folge bis zum n-ten Folgenglied ν_n konstruiert. Dann ist sicher $\nu_n \notin J$, denn bereits ν_1 ist größer als alle Elemente von J, und die folgenden ν_n werden immer größer. Daher gibt es nach Definition von J mindestens einen Index $m > \nu_n$, so daß $x_m > x_{\nu_n}$; andernfalls wäre $\nu_n \in J$. Wir wählen einen solchen Index m aus und definieren $\nu_{n+1} = m$. Dann ist $\nu_{n+1} > \nu_n$ und $x_{\nu_{n+1}} > x_{\nu_n}$.

Setzen wir nun noch $y_n = x_{\nu_n}$, so haben wir offenbar eine sogar streng monoton wachsende Teilfolge von $(x_n)_{n\in\mathbb{N}}$ gefunden. ■

Als unmittelbare Folgerung erhalten wir den

Satz von Bolzano-Weierstraß: Jede beschränkte Folge reeller Zahlen hat eine konvergente Teilfolge.

Beweis: Wir wir gerade gesehen haben, hat *jede* Folge $(x_n)_{n\in\mathbb{N}}$ reeller Zahlen eine monotone Teilfolge; falls $(x_n)_{n\in\mathbb{N}}$ eine beschränkte Folge ist, gilt das erst recht für die Teilfolge, und damit konvergiert diese. ■

BERNARD PLACIDUS JOHANN NEPOMUK BOLZANO (1781–1848) wurde in Prag geboren als Sohn eines aus Italien eingewanderten Kunsthändlers und einer deutschsprachigen Kaufmannstochter. An der Prager Karls-Universität studierte er ab 1876 Philosophie, Physik und Mathematik; ab Herbst 1800 schrieb er sich zusätzlich noch für Theologie ein, arbeitete aber neben seinem Theologiestudium auch an einer Dissertation über Geometrie. 1804 wurde er damit promoviert; zwei Tage später folgte die Priesterweihe. Er bewarb sich sowohl um einen Lehrstuhl für Mathematik als auch um einen für Religionsphilosophie und bekam den letzteren. Wegen aufklärerischer Tendenzen wurde er 1819 suspendiert und unter Hausarrest gestellt. Die meisten seiner mathematischen Arbeiten entstanden zwischen 1810 und 1817; ab 1837 folgten Arbeiten zur Wissenschaftstheorie und zur Philosophie der Mathematik.

KARL THEODOR WILHELM WEIERSTRASS (1815–1897) wurde im westfälischen Ostenfelde geboren. Schon während seiner Gymnasialzeit in Paderborn las er regelmäßig mathematische Fachzeitschriften; trotzdem studierte er auf Wunsch seines Vaters Rechts- und Wirtschaftswissenschaften an der Universität Bonn. Seine mathematischen Kenntnisse erwarb er im Selbststudium. Da er darüber sein eigentliches Studium vernachlässigte, mußte er nach acht Semestern die Universität ohne Abschluß verlassen und ließ sich in Münster zum Gymnasiallehrer ausbilden. Auch während dieser Ausbildung und später an der Schule setzte er seine mathematischen Forschungen fort. Mit einer 1854 erschienenen Arbeit über Abelsche Funktionen wurde er erstmals einer breiteren mathematischen Öffentlichkeit bekannt, bekam die Ehrendoktorwürde der Universität Königsberg und verschiedene Rufe auf Lehrstühle. Im Oktober 1856 entschied er sich für eine Professur an der Universität Berlin, wo er Studenten aus aller Welt anzog. In einigen seiner Vorlesungen beschäftigte er sich mit der präzisen Grundlegung der Analysis und entwickelte das Schema, dem im wesentlich auch heute noch alle Anfängervorlesungen über Analysis folgen. Auch viele Resultate der komplexen Analysis (Funktionentheorie) gehen auf ihn zurück.

Die bloße Tatsache, daß eine *Teilfolge* konvergiert, mag nicht sonderlich aufregend erscheinen; sie kann aber helfen, unter geeigneten Zusatzbedingungen die Konvergenz der gesamten Folge zu zeigen, und das auch in sehr allgemeinen Zusammenhängen, die erst in späteren Semestern auftauchen werden.

Wir betrachten daher wieder einmal zumindest kurzfristig einen beliebigen metrischen Raum X mit Metrik d. Angenommen, die Folge $(x_n)_{n\in\mathbb{N}}$ von Elementen aus X konvergiert gegen einen Punkt $x \in X$. Dann können wir für jedes $\varepsilon > 0$ eine natürliche Zahl n_0 finden, so daß $d(x, x_n) < \varepsilon/2$ für alle $n \geq n_0$. Für zwei beliebige natürliche Zahlen $n, m \geq n_0$ ist somit nach der Dreiecksungleichung

$$d(x_n, x_m) \leq d(x, x_n) + d(x, x_m) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon\,.$$

Definition: Eine Folge $(x_n)_{n\in\mathbb{N}}$ von Elementen aus einem metrischen Raum X mit Metrik d heißt CAUCHY-Folge, wenn es zu jedem $\varepsilon > 0$ ein $n_0 \in \mathbb{N}$ gibt, so daß $d(x_n, x_m) < \varepsilon$ für alle $n, m \geq n_0$.

Wie wir gerade gesehen haben, ist jede konvergente Folge eine CAUCHY-Folge. Die Umkehrung kann allerdings falsch sein: Ist etwa $X = \mathbb{Q}$ mit der üblichen Metrik $d(x, y) = |x - y|$, so liefert uns das Verfahren von HERON eine Folge $(x_n)_{n\in\mathbb{N}}$ von Approximationen an $\sqrt{2}$. Da diese Folge in $\mathbb{R}$ gegen $\sqrt{2}$ konvergiert, ist sie insbesondere eine CAUCHY-Folge, aber als Folge von Elementen aus $\mathbb{Q}$ ist sie nicht konvergent: Schließlich ist $\sqrt{2}$ keine rationale Zahl. Aus eben diesem Grund hatten wir die rationalen Zahlen erweitert zu den reellen, und hier sagt uns CAUCHY, daß *jede* CAUCHY-Folge konvergiert:

Cauchysches Konvergenzkriterium: Jede CAUCHY-Folge $(x_n)_{n\in\mathbb{N}}$ von reellen oder komplexen Zahlen konvergiert.

Zum *Beweis* betrachten wir zunächst nur reelle Folgen. Hier überlegen wir uns als erstes, daß jede CAUCHY-Folge $(x_n)_{n\in\mathbb{N}}$ beschränkt ist:

Setzen wir in der Definition einer CAUCHY-Folge $\varepsilon = 1$, so erhalten wir eine natürliche Zahl n_0 mit der Eigenschaft, daß $|x_n - x_m| < 1$ ist für alle $n, m \geq n_0$. Insbesondere ist für alle $n \geq n_0$

$$|x_n| = \left|x_{n_0} + (x_n - x_{n_0})\right| \leq \left|x_{n_0}\right| + \left|x_n - x_{n_0}\right| \leq \left|x_{n_0}\right| + 1\,.$$

Bezeichnet M das Maximum der n_0 Zahlen $|x_1|, \ldots, \left|x_{n_0-1}\right|$ und $\left|x_{n_0}\right| + 1$, ist daher $|x_n| \leq M$ für alle $n \in \mathbb{N}$.

Damit können wir den Satz von BOLZANO-WEIERSTRASS anwenden, der uns die Existenz einer konvergenten Teilfolge garantiert. Der Grenzwert dieser Teilfolge sei x; wir wollen uns überlegen, daß auch die Gesamtfolge gegen x konvergiert.

Sei also $\varepsilon > 0$ vorgegeben. Da $(x_n)_{n\in\mathbb{N}}$ eine CAUCHY-Folge ist, gibt es ein $n_1 \in \mathbb{N}$, so daß $|x_n - x_m| < \frac{1}{2}\varepsilon$ ist für alle $n, m \geq n_1$. Da wir eine gegen x konvergente Teilfolge haben, finden wir auch ein $n_2 \in \mathbb{N}$, so daß $|x - x_n| < \frac{1}{2}\varepsilon$ ist für alle jene $n \geq n_2$, *für die x_n zur Teilfolge gehört*. Wie üblich setzen wir $n_0 = \max(n_1, n_2)$; außerdem wählen wir uns ein $m \geq n_0$, für das x_m zur Teilfolge gehört. Für $n \geq n_0$ ist dann

$$\begin{aligned}|x - x_n| &= |(x - x_m) + (x_m - x_n)| \leq |x - x_m| + |x_m - x_n| \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon\,;\end{aligned}$$

die Folge $(x_n)_{n\in\mathbb{N}}$ konvergiert also gegen x.

Damit ist der reelle Fall erledigt; wir müssen noch zeigen, daß die Aussage auch im Komplexen gilt. Dazu sei $(z_n)_{n\in\mathbb{N}}$ eine CAUCHY-Folge von komplexen Zahlen. Wir schreiben $z_n = x_n + iy_n$ mit reellen Zahlen x_n, y_n und wollen uns zunächst überlegen, daß auch $(x_n)_{n\in\mathbb{N}}$ und $(y_n)_{n\in\mathbb{N}}$ CAUCHY-Folgen sind: Zu jedem $\varepsilon > 0$ gibt es ein $n_0 \in \mathbb{N}$, so daß

$$\begin{aligned}|z_n - z_m| &= |(x_n - x_m) + i(y_n - y_m)| \\ &= \sqrt{(x_n - x_m)^2 + (y_n - y_m)^2} < \varepsilon\end{aligned}$$

ist für alle $n, m \geq n_0$. Da $(x_n - x_m)^2$ und $(y_n - y_m)^2$ beide größer oder gleich Null sind, ist dann auch

$$|x_n - x_m| < \varepsilon \quad \text{und} \quad |y_n - y_m| < \varepsilon \quad \text{für alle } n, m \geq n_0\,.$$

Damit wissen wir, daß die Folgen $(x_n)_{n\in\mathbb{N}}$ und $(y_n)_{n\in\mathbb{N}}$ konvergieren; ihre Grenzwerte seien x und y. Natürlich erwarten wir, daß die Folge $(z_n)_{n\in\mathbb{N}}$ gegen $z = x + iy$ konvergiert, und das läßt sich auch leicht nachrechnen: Zu vorgegebenem $\varepsilon > 0$ gibt es natürliche Zahlen n_1, n_2, so daß $|x - x_n| < \frac{1}{2}\varepsilon$ für alle $n \geq n_1$ und $|y - y_n| < \frac{1}{2}\varepsilon$ für alle $n \geq n_2$. Für $n \geq n_0 = \max(n_1, n_2)$ ist daher

$$\begin{aligned}|z - z_n| &= |(x - x_n) + i(y - y_n)| = \sqrt{(x - x_n)^2 + (y - y_n)^2} \\ &< \sqrt{\left(\frac{\varepsilon}{2}\right)^2 + \left(\frac{\varepsilon}{2}\right)^2} = \sqrt{\frac{\varepsilon^2}{2}} < \varepsilon\,,\end{aligned}$$

womit auch der komplexe Fall erledigt wäre. ■

Definition: Ein metrischer Raum X heißt *vollständig*, wenn jede CAUCHY-Folge $(x_n)_{n\in\mathbb{N}}$ in X gegen einen Punkt $x \in X$ konvergiert.

Wie wir gerade gezeigt haben, sind also $\mathbb{R}$ und $\mathbb{C}$ mit ihren Standardmetriken vollständig, $\mathbb{Q}$ aber nicht. Da alle Schritte, die zum Beweis des CAUCHYschen Konvergenzkriteriums für reelle Zahlen führten, in einem beliebigen vollständigen angeordneten Körper k gültig sind; ist auch jeder solche Körper bezüglich der durch den Betrag gegebenen Metrik vollständig. Umgekehrt überlegt man sich leicht, daß ein angeordneter Körper, der in diesem Sinne ein vollständiger metrischer Raum ist, auch als angeordneter Körper vollständig ist; für angeordnete Körper stimmen die beiden Vollständigkeitsbegriffe also überein.

§4: Die Exponentialfunktion

Kehren wir zurück zum Beispiel einer Kapitalanlage mit konstantem Zinssatz von p%. Falls es während eines gesamten Jahres weder Ein- noch Auszahlungen gibt, wird das am Jahresanfang vorhandene Kapital zum Jahresende mit $1 + p/100$ multipliziert.

Nun wird aber erstens kaum ein Kapital am ersten Januar einbezahlt (da sind die Banken schließlich geschlossen), und in vielen Fällen gibt es auch Kontobewegungen während des Jahres. Damit sind die Banken natürlich wohlvertraut, und sie haben auch Rechenregeln entwickelt, um damit umzugehen. Aus der Zeit, als noch alle Berechnungen von Hand oder mit einfachen mechanischen Rechenmaschinen durchgeführt werden mußten, stammt die Regel, daß banktechnisch gesehen jeder Monat 30 Tage und das Jahr damit 360 Tage hat. Inzwischen werden Zinsen natürlich schon längst per Computer berechnet; da zumindest der traditionelle Bankenbereich nicht sehr veränderungsfreudig ist, sind die heutigen Regeln, auch die auf den Euro-Kapitalmärkten angewendeten, immer noch eine Mischung aus Jahren mit 360 Tagen und tatsächlicher Zeitrechnung. So kann eine Anlage von n Tagen je nach Anfangsdatum bei gleichem Zinssatz verschieden hohe Zinsen abwerfen.

Wir wollen hier nicht auf solche banktechnische Feinheiten eingehen; wir nehmen einfach an, daß eine Einlage über x Jahre für *jede reelle Zahl* $x \leq 1$ bei einem Zinssatz von p % mit $1+px/100$ multipliziert wird. Das ist zwar eine Idealisierung, liefert aber Ergebnisse, die nicht sehr von den nach tatsächlich praktizierten Bankregeln berechneten abweichen.

Angenommen, wir legen unser Kapital zu Jahresbeginn bei einer Bank A an und lassen es dort ein Jahr lang stehen. Bei einem Zinssatz von p % wird es dann mit $1 + p/100$ multipliziert. Wenn wir es allerdings zur Jahresmitte wieder abziehen, wird es nur mit $1 + p/200$ multipliziert. Legen wir es anschließend wieder bei einer Bank B an, wird es zum Jahresende ebenfalls nur mit $1+p/200$ multipliziert, insgesamt also mit

$$\left(1+\frac{p}{200}\right)^2 = 1+\frac{p}{100}+\frac{p^2}{40\,000}\,,$$

was mehr ist, als wir innerhalb eines Jahres bei Bank A erzielt hätten.

In einer Zeit, in der an den großen Börsen Computerprogramme Kauf- und Verkaufsentscheidungen im Sekundenrhythmus treffen und in der Transaktionsgebühren zumindest für große Kapitalbeträge für alle praktischen Zwecke vernachlässigt werden können, zwingt uns natürlich niemand, unser Geld gleich sechs Monate lang bei einer festen Bank zu lassen; wir können auch schon nach drei Monaten wechseln. Bei dieser Strategie brauchen wir vier Banken pro Jahr, dafür hat sich unser Kapital im ersten Halbjahr schon mit

$$\left(1+\frac{p}{400}\right)^2 = 1+\frac{p}{200}+\frac{p^2}{160\,000}$$

multipliziert, was mehr ist, als wenn wir es ein halbes Jahr lang bei einer Bank gelassen hätten. Insgesamt wird das Kapital am Ende des Jahres multipliziert mit

$$\left(1+\frac{p}{400}\right)^4 = 1+\frac{p}{100}+\frac{3p^2}{80\,000}+\frac{p^3}{16\,000\,000}+\frac{p^4}{25\,600\,000\,000}\,.$$

Zur weiteren Optimierung des Kapitalzuwachs können wir auch jeden Monat wechseln, oder jede Woche, jeden Tag, und – sofern uns nicht die Banken ausgehen – jede Stunde, jede Minute, jede Sekunde und so weiter. Wenn wir von einem (derzeit sicherlich übertriebenen) marktüblichen Zinssatz von 5% ausgehen, ergibt dies folgende effektive Verzinsung pro Jahr:

Anlageperiode	*effektiver Zinssatz*
1 Jahr	5 %
$\frac{1}{2}$ Jahr	5,0625 %
$\frac{1}{4}$ Jahr	5,0945337 %
1 Monat = $\frac{1}{12}$ Jahr	5,1161898 %
1 Woche = $\frac{1}{52}$ Jahr	5,1245842 %
1 Tag = $\frac{1}{365}$ Jahr	5,1267450 %
1 Stunde	5,1270946 %
1 Minute	5,1271094 %
1 Sekunde	5,1271096 %

(Ab der stündlichen Verzinsung habe ich mit einem Kalenderjahr aus 365,2422 Tagen gerechnet.)

Aus Sicht der Bank ist es, zumindest wenn wir ein wirklich großes Kapital anzulegen haben, nicht attraktiv, uns nur einen Monat, eine Woche, einen Tag, eine Stunde, eine Minute oder gar nur eine Sekunde als Kunden zu haben. Sie könnte uns natürlich länger halten, wenn sie uns beispielsweise das Doppelte des marktüblichen Zinssatzes anböte; es ist aber fraglich, ob sie mit einem solchen Angebot lange überleben könnte.

Die Bank muß somit eine Strategie entwickeln, die es zwar für uns unattraktiv macht, ständig die Bank zu wechseln, die es aber auch der Bank ermöglicht, im Rahmen des marktüblichen Zinssatzes zu bleiben.

Sie muß uns daher für die Vermehrung unseres Kapitals einen Multiplikator anbieten, den wir nicht durch Unterteilung verbessern können. Ist $f(x)$ der Multiplikator, den sie uns für eine Anlage von x Jahren anbietet, wobei x eine beliebige positive reelle Zahl ist, muß also gelten

$$f(x+y) \geq f(x)f(y)\,.$$

Aus Sicht der Bank wäre hier ein echtes Größerzeichen ein überflüssiger und deshalb kostentreibender Luxus, denn um uns vom Wechseln abzuhalten, reicht es, daß wir dabei nichts gewinnen können; ein Größerzeichen würde bedeuten, daß wir bei Unterteilung sogar verlieren.

Ist $p\,\%$ der marktübliche Zinssatz und $r = p/100$, so muß sie auch konkurrieren mit Banken, die selbst für beliebig kurze Zeiträume $x > 0$ eine Multiplikation mit $1 + rx$ anbieten. Daher muß gelten

$$f(x) \ge 1 + rx \quad \text{für alle } x \ge 0\,.$$

Beide Bedingungen sind leicht erfüllbar mit einem Zinssatz, der hinreichend deutlich über $p\%$ liegt, aber auch das liegt natürlich nicht im Interesse der Bank.

Um $f(x)$ nach oben zu begrenzen, könnte die Bank folgendermaßen argumentieren: Bei Geldanlagen zu einem festen Zinssatz $p\,\%$ können wir deshalb durch Wechseln mehr Zins realisieren, weil der Zins nach jeder Neuanlage auf das bis dahin akkumulierte Kapital bezahlt wird statt nur auf das Ausgangskapital. Wir können also auf keinen Fall mehr bekommen als rx mal das Endkapital. Die Zinsen, die wir bekommen, sind $f(x)-1$ mal das Startkapital; würde das Endkapital mit rx multipliziert, hätten wir das $rx\cdot f(x)$-fache des Startkapitals als Zinsen. Mehr braucht uns die Bank nicht zu bieten; also kann sie auch die Ungleichung

$$f(x) - 1 \le rx \cdot f(x) \quad \text{oder} \quad f(x)\cdot(1 - rx) \le 1$$

fordern. Insgesamt sucht also die Bank eine Funktion f mit den Eigenschaften

$$f(x + y) = f(x) \cdot f(y) \quad \text{für alle } x, y \ge 0\,,$$

$$f(x) \ge 1 + rx \quad \text{und} \quad f(x)\cdot(1 - rx) \le 1 \quad \text{für alle } x \ge 0\,.$$

Diese Funktion interessiert zunächst nur für Werte $x \ge 0$, es könnte aber trotzdem sinnvoll sein, sie auch für negative Werte zu definieren: Um aus dem gegenwärtigen Kontostand a den Kontostand zu einem künftigen Zeitpunkt in x Jahren zu berechnen, müssen wir einfach mit $f(x)$ multiplizieren. Entsprechend sollte der Kontostand *vor* x Jahren – falls das Konto damals schon bestand – einfach als $a \cdot f(-x)$ berechnet werden können. Da der damalige Kontostand mit $f(x)$ multipliziert wurde um seinen jetzigen Wert x zu erreichen, ist $a \cdot f(-x) \cdot f(x) = a$, die einzig sinnvolle Weise, eine derartige Funktion f ins Negative

fortzusetzen, besteht also darin, sie für $x < 0$ als $f(x) = 1/f(-x)$ zu definieren. Schwierigkeiten mit einer Division durch Null gibt es dabei nicht, denn $f(x) \ge 1 + rx$ für alle positiven x, und r ist sogar größer als eins.

Wenn f die obigen Forderungen erfüllt, ist für die so fortgesetzte Funktion $f(x+y) = f(x)\cdot f(y)$ für *alle* $x, y \in \mathbb{R}$, und auch die Ungleichungen in der zweiten Zeile sind erfüllt: Für $x < 0$ ist

$$f(x) = \frac{1}{f(-x)} \le \frac{1}{1 - rx}\,. \quad \text{also} \quad f(x)(1 - rx) \le 1 \quad \text{und}$$

$$f(-x)(1 + rx) = \frac{1 + rx}{f(x)} \le 1\,, \quad \text{also} \quad f(x) \ge 1 + rx\,.$$

Wir suchen also zu einem Parameter $r > 0$ eine Funktion $f\colon \mathbb{R} \to \mathbb{R}$, die folgenden Bedingungen genügt:

1.) $f(x + y) = f(x) \cdot f(y)$ für alle $x, y \in \mathbb{R}$
2.) $f(x) \ge 1 + rx$ für alle $x \in \mathbb{R}$
3.) $f(x)(1 - rx) \le 1$ für alle $x \in \mathbb{R}$.

Wie die gerade durchgeführte Rechnung zeigt, ist die zweite Bedingung für positive x äquivalent dazu, daß die dritte für negative r gilt und umgekehrt. Es genügt daher, eine der beiden Bedingungen nachzuweisen; die andere folgt automatisch.

Als nächstes wollen wir uns überlegen, daß wir uns auch über den Parameter r keine großen Gedanken machen müssen:

Behauptung: Erfüllt die Funktion f die obigen Bedingungen mit $r = 1$, so erfüllt für irgendein vorgegebenes $r \in \mathbb{R}$ die Funktion $g(x) = f(rx)$ dieselben Bedingungen mit diesem r.

Beweis: Wie wir gerade gesehen haben, reicht es, wenn wir die ersten beiden Bedingungen nachrechnen. Für $x, y \in \mathbb{R}$ ist

$$g(x + y) = f\bigl(r(x + y)\bigr) = f(rx + ry) = f(rx) \cdot f(ry) = g(x) \cdot g(y)$$
$$\text{und} \qquad g(x) = f(rx) \ge 1 + rx\,,$$

wie gewünscht. ■

Damit gibt uns der folgende Satz eine allgemeine Übersicht über die Existenz und Eindeutigkeit der gerade betrachteten Funktionen:

Satz: Es gibt genau eine Funktion $f\colon \mathbb{R} \to \mathbb{R}$ mit den Eigenschaften
1.) $f(x+y) = f(x) \cdot f(y)$ für alle $x, y \in \mathbb{R}$
2.) $f(x) \geq 1 + x$ für alle $x \in \mathbb{R}$

Beweis: Wir beginnen damit, aus den drei Forderungen weitere Konsequenzen zu ziehen, um so hoffentlich auf eine mögliche Definition eines Kandidaten für f zu kommen. Wie wir bereits wissen, erfüllt f automatisch auch die Bedingung
3.) $f(x)(1-x) \leq 1$ für alle $x \in \mathbb{R}$.

Offensichtlich ist $f(0) = 1$ für jede Funktion f, die diese drei Bedingungen erfüllt: Nach der zweiten Bedingung ist $f(0) \geq 1$, und nach der ersten ist $f(0) = f(0+0) = f(0) \cdot f(0)$. Außerdem haben wir uns bereits überlegt, daß stets $f(-x) = 1/f(x)$ sein muß.

Die erste Bedingung läßt sich leicht verallgemeinern auf beliebig viele Summanden:
$$f(x_1 + \cdots + x_n) = f(x_1) \cdot \cdots \cdot f(x_n) \quad \text{für alle } x_1, \ldots, x_n \in \mathbb{R}\,.$$
(Wir haben inzwischen schon so viel Erfahrung mit Beweisen, daß jedem klar sein sollte, wie er das nötigenfalls formal durch vollständige Induktion beweisen kann.)

Wenden wir dies an auf den Fall, daß alle $x_i = x/n$ sind, erhalten wir die Formel
$$f(x) = f\left(n \cdot \frac{x}{n}\right) = f\left(\frac{x}{n}\right)^n \quad \text{für alle } x \in \mathbb{R} \text{ und } n \in \mathbb{N}\,.$$
Damit können wir die Bedingungen 2.) und 3.) verschärfen: Für $x < 1$ können wir bei 3.) durch $1-x$ dividieren und erhalten die Ungleichung
$$f(x) \leq \frac{1}{1-x} \quad \text{für alle } x < 1\,.$$
Für eine beliebige vorgegebene reelle Zahl x wählen wir ein $n \in \mathbb{N}$ mit $n > |x|$ und haben dann die Ungleichung
$$1 + \frac{x}{n} \leq f\left(\frac{x}{n}\right) \leq \frac{1}{1 - \frac{x}{n}} = \left(1 - \frac{x}{n}\right)^{-1}\,.$$
Hier stehen überall positive Zahlen; da wir Ungleichungen mit positiven Zahlen multiplizieren dürfen, bleibt dies also richtig, wenn wir bei allen drei Termen zur n-ten Potenz übergehen:
$$\left(1 + \frac{x}{n}\right)^n \leq f\left(\frac{x}{n}\right)^n = f(x) \leq \frac{1}{\left(1 - \frac{x}{n}\right)^n}$$
für alle $n > |x|$.

Die gerade bewiesene Ungleichung erinnert sehr an Intervallschachtelungen; wenn wir zeigen können, daß wir hier tatsächlich eine Intervallschachtelung haben, definiert sie eine reelle Zahl y, und wenn es eine Funktion f mit den behaupteten Eigenschaften gibt, muß $f(x) = y$ sein.

Wir wollen also zeigen, daß die Folge der Intervalle
$$\left[\left(1 + \frac{x}{n}\right)^n,\ \left(1 - \frac{x}{n}\right)^{-n}\right] \quad \text{mit} \quad n > |x|$$
eine Intervallschachtelung bilden.

Als erstes müssen wir zeigen, daß wir es tatsächlich mit Intervallen zu tun haben, daß also die untere Intervallgrenze stets kleiner oder gleich der oberen ist. Angesichts der gerade bewiesenen Ungleichungen für $f(x)$ mag das auf den ersten Blick überflüssig erscheinen, aber wir dürfen nicht vergessen, daß $f(x)$ bislang erst ein fiktiver Wert ist; wir wissen noch nicht, ob es eine reelle Zahl mit diesen Eigenschaften gibt.

Da wir nur Indizes $n > |x|$ betrachten, sind $(1 + x/n)^n$ und $(1 - x/n)^{-n}$ positive Zahlen; daher ist die Ungleichung $(1 + x/n)^n < (1 - x/n)^{-n}$ äquivalent zur Ungleichung
$$\left(1 + \frac{x}{n}\right)^n \left(1 - \frac{x}{n}\right)^n = \left(1 - \frac{x^2}{n^2}\right)^n < 1\,,$$
und da x^2/n^2 wegen $n > |x|$ zwischen null und eins liegt, ist letztere Ungleichung offensichtlich richtig.

Nachdem wir Intervalle haben, müssen wir zeigen, daß jedes davon in seinem Vorgänger enthalten ist. Für die unteren Grenzen bedeutet dies, daß für alle $n \in \mathbb{N}$ gelten muß
$$\left(1 + \frac{x}{n}\right)^n \leq \left(1 + \frac{x}{n+1}\right)^{n+1}\,.$$
Das Problem dabei ist, daß selbstverständlich
$$1 + \frac{x}{n} \geq 1 + \frac{x}{n+1}$$

ist; für $x \neq 0$ gilt sogar das echte Größerzeichen. Der um eins größere Exponent auf der rechten Seite muß also zu einer Umkehr der Relation führen.

Um das zu beweisen, könnten wir versuchen, die Differenz der beiden Zahlen zu berechnen. Da wir dazu alles ausmultiplizieren müßten, wäre das allerdings sehr aufwendig, und das bei eher ungewissen Erfolgsaussichten. Wir wollen stattdessen versuchen, den Quotienten

$$\left(1+\frac{x}{n+1}\right)^{n+1} \Big/ \left(1+\frac{x}{n}\right)^{n}$$

mit Hilfe der BERNOULLIschen Ungleichung

$$(1+x)^n > 1+nx \quad \text{für } n \geq 2 \text{ und } x \geq -1 \text{ aber } x \neq 0$$

abzuschätzen. Da diese nur eine Abschätzung in einer Richtung liefert, können wir sie nicht sowohl auf den Zähler als auch auf den Nenner anwenden, sondern höchstens auf die n-te Potenz des Quotienten von $1+x/(n+1)$ und $1+x/n$. Dazu müssen wir diesen Quotienten umformen zu einem Ausdruck der Form $1+z$, was am einfachsten dadurch geschieht, daß wir – wie schon so oft bei Übungsaufgaben zur Konvergenz – den Zähler schreiben als Nenner plus weiteren Term, konkret also

$$1+\frac{x}{n+1} = 1+\frac{x}{n}+\left(\frac{x}{n+1}-\frac{x}{n}\right) = 1+\frac{x}{n}-\frac{x}{n(n+1)}\,;$$

Dann erhalten wir den Quotienten als

$$\frac{1+\dfrac{x}{n+1}}{1+\dfrac{x}{n}} = 1-\frac{\dfrac{x}{n(n+1)}}{1+\dfrac{x}{n}} = 1-\frac{\dfrac{x}{n+1}}{n+x} = 1-\frac{x}{(n+1)(n+x)}\,.$$

Da wir davon ausgehen, daß $n > |x|$ ist, können wir für $x \neq 0$ die Ungleichung von BERNOULLI anwenden mit dem Ergebnis, daß die n-te Potenz des Quotienten größer ist als

$$1-\frac{nx}{(n+1)(n+x)}\,.$$

Für $x = 0$ ist beides gleich eins; also haben wir für $n \geq |x|$ die

Abschätzung

$$\begin{aligned}\frac{\left(1+\dfrac{x}{n+1}\right)^{n+1}}{\left(1+\dfrac{x}{n}\right)^{n}} &\geq \left(1-\frac{n}{n+x}\,\frac{x}{n+1}\right)\left(1+\frac{x}{n+1}\right)\\ &= 1+\left(1-\frac{n}{n+x}\right)\frac{x}{n+1}-\frac{n}{n+x}\left(\frac{x}{n+1}\right)^2\\ &= 1+\frac{x}{n+x}\,\frac{x}{n+1}-\frac{n}{n+x}\left(\frac{x}{n+1}\right)^2\\ &= 1+\frac{(n+1)x^2)-nx^2}{(n+x)(n+1)^2} = 1+\frac{x^2}{(n+x)(n+1)^2}\,.\end{aligned}$$

Für $n > |x|$ ist dies offensichtlich größer oder gleich eins, d.h.

$$\left(1+\frac{x}{n+1}\right)^{n+1} \geq \left(1+\frac{x}{n}\right)^{n}\,.$$

Zumindest für die unteren Schranken ist die Inklusionsbedingung also erfüllt.

Für die oberen Schranken müssen wir nachweisen, daß

$$\left(1-\frac{x}{n+1}\right)^{-(n+1)} \leq \left(1-\frac{x}{n}\right)^{-n}$$

gilt für alle $n > |x|$. Durch Übergang zu den Kehrwerten wird dies zur äquivalenten Ungleichung

$$\left(1-\frac{x}{n+1}\right)^{n+1} \geq \left(1-\frac{x}{n}\right)^{n} \quad \text{oder}$$

$$\left(1+\frac{(-x)}{n+1}\right)^{n+1} \geq \left(1+\frac{(-x)}{n}\right)^{n}\,,$$

die wir gerade bewiesen haben: Wir hatten schließlich keine Annahmen über das Vorzeichen von x gemacht.

Damit ist die erste Bedingung an eine Intervallschachtelung nachgewiesen; die zweite Bedingung war, daß die Intervallängen, also die Differenzen

$$\left(1-\frac{x}{n}\right)^{-n} - \left(1+\frac{x}{n}\right)^{n}\,,$$

eine Nullfolge bilden. Nach der dritten binomischen Formel können wir diese umformen zu

$$\left(1-\frac{x}{n}\right)^{-n}\left(1-\left(1-\frac{x^2}{n^2}\right)^n\right).$$

Von den linken Faktoren haben wir gerade bewiesen, daß sie eine monoton fallende Folge bilden; sie sind also nach oben beschränkt. Da wir bereits gezeigt haben, daß die behauptete Intervallschachtelung wirklich aus Intervallen besteht, sind sie auch nach unten beschränkt, zum Beispiel durch jedes beliebige untere Ende eines Intervalls. Um zu zeigen, daß wir insgesamt eine Nullfolge haben, reicht es deshalb, wenn wir nachweisen, daß die zweiten Faktoren eine Nullfolge bilden. Nach den Rechenregeln für Grenzwerte aus dem vorigen Paragraphen reicht dazu wiederum, daß

$$\lim_{n\to\infty}\left(1-\frac{x^2}{n^2}\right)^n = 1$$

ist. Da $(x^2/n)_{n\in\mathbb{N}}$ offensichtlich eine Nullfolge ist, folgt dies (und damit die Intervallschachtelungseigenschaft) aus der folgenden etwas allgemeineren

Restbehauptung: Ist $(y_n)_{n\in\mathbb{N}}$ eine Nullfolge, so ist

$$\lim_{n\to\infty}\left(1+\frac{y_n}{n}\right)^n = 1\,.$$

Beweis: Mit der Ungleichung von BERNOULLI können wir die Folgenglieder nach unten abschätzen:

$$\left(1+\frac{y_n}{n}\right)^n > 1+n\cdot\frac{y_n}{n} = 1+y_n\,,$$

falls $1+y_n/n > 0$ und $y_n \neq 0$. Da die y_n eine Nullfolge bilden, gibt es ein $n_1\in\mathbb{N}$, so daß $|y_n|<1$ für alle $n\geq n_1$; erst recht ist dann $1+y_n>0$. Für etwaige Folgenglieder $y_n=0$ müssen wir in der Ungleichung noch das Gleichheitszeichen zulassen; für $n\geq n_1$ ist also

$$\left(1+\frac{y_n}{n}\right)^n \geq 1+y_n\,.$$

Um auch eine Abschätzung nach oben zu bekommen, drehen wir die Ungleichung einfach um: Wenn wir für eine reelle Zahl u die Gleichung

$1+u = 1/(1-v)$ nach v auflösen wollen, sehen wir, daß das für $u=-1$ nicht funktionieren kann; ansonsten aber erhalten wir sofort die Gleichung $v = u/(1+u)$. Da wir von einer Nullfolge $(y_n)_{n\in\mathbb{N}}$ ausgehen, gibt es sicherlich ein $n_2\in\mathbb{N}$, so daß $y_n/n\neq -1$ für alle $n\geq n_2$. Für $n\geq n_2$ gibt es daher reelle Zahlen z_n, so daß

$$1+\frac{y_n}{n} = \frac{1}{1-z_n}$$

ist, nämlich $z_n = (y_n/n)\big/(1+y_n/n)$. Da $(y_n)_{n\in\mathbb{N}}$ eine Nullfolge ist, konvergiert nach den Rechenregeln für Grenzwerte auch die Folge der z_n gegen Null, es gibt also ein $n_3\in\mathbb{N}$, so daß $z_n<1$ für alle $n\geq n_3$. Für $n\geq n_0=\max(n_1,n_2,n_3)$ ist daher nach der BERNOULLIschen Ungleichung

$$\left(1+\frac{y_n}{n}\right)^n = \left(\frac{1}{1-z_n}\right)^n \leq \frac{1}{1-nz_n}\,.$$

Für $n\geq n_0$ ist daher

$$1+y_n \leq \left(1+\frac{y_n}{n}\right)^n \leq \frac{1}{1-nz_n}\,.$$

Nach den Rechenregeln für Grenzwerte ist damit

$$1 = \lim_{n\to\infty}(1+y_n) \leq \lim_{n\to\infty}\left(1+\frac{y_n}{n}\right)^n \leq \lim_{n\to\infty}\frac{1}{1-nz_n} = 1\,,$$

also ist der betrachtete Grenzwert tatsächlich gleich eins.

Damit sind alle Forderungen an eine Intervallschachtelung nachgewiesen; *falls* es eine Funktion f mit den behaupteten Eigenschaften gibt, ist der Funktionswert $f(x)$ an der Stelle x also durch die Intervallschachtelung

$$\left[\left(1+\frac{x}{n}\right)^n,\ \left(1-\frac{x}{n}\right)^{-n}\right] \quad \text{mit} \quad n>|x|$$

definiert. Insbesondere ist er dann, wie wir im vorigen Paragraphen gesehen haben, Limes der Folge der unteren Intervallgrenzen, d.h.

$$f(x) = \lim_{n\to\infty}\left(1+\frac{x}{n}\right)^n\,.$$

Dies können wir nun als *Definition* einer Funktion f betrachten und müssen beweisen, daß sie die beiden im Satz geforderten Eigenschaften hat. Bei der zweiten folgt das sofort aus der Ungleichung von BERNOULLI, denn danach ist für jedes $n > |x|$

$$\left(1 + \frac{x}{n}\right)^n \geq 1 + n \cdot \frac{x}{n} = 1 + x\,,$$

also gilt dies auch für den Limes.

Für die Gleichung $f(x + y) = f(x) \cdot f(y)$ berechnen wir zunächst für $n > |x + y|$ den Quotienten

$$\frac{\left(1 + \frac{x}{n}\right)\left(1 + \frac{y}{n}\right)}{\left(1 + \frac{x+y}{n}\right)} = \frac{1 + \frac{x+y}{n} + \frac{yx}{n^2}}{\left(1 + \frac{x+y}{n}\right)} = 1 + \frac{\frac{xy}{n^2}}{1 + \frac{x+y}{n}} = 1 + \frac{z_n}{n}$$

mit $z_n = \dfrac{xy}{n + x + y}$. Die Folge der z_n mit $n > |x + y|$ ist offensichtlich eine Nullfolge; nach der obigen *Restbehauptung* ist also

$$\lim_{n\to\infty} \frac{\left(1 + \frac{x}{n}\right)^n \left(1 + \frac{y}{n}\right)^n}{\left(1 + \frac{x+y}{n}\right)^n} = \lim_{n\to\infty} \left(1 + \frac{z_n}{n}\right)^n = 1\,.$$

Andererseits ist der linksstehende Limes nach den Rechenregeln für Grenzwerte gleich

$$\frac{f(x)f(y)}{f(x + y)}\,,$$

also ist dieser Quotient gleich eins und damit $f(x+y) = f(x)f(y)$, womit auch diese Gleichung bewiesen wäre.

■

Damit haben wir also gezeigt, daß es genau eine Funktion $f\colon \mathbb{R} \to \mathbb{R}$ gibt, für die gilt

$$f(x + y) = f(x) \cdot f(y) \quad \text{und} \quad f(x) \geq 1 + x$$

für alle $x, y \in \mathbb{R}$. Die erste Gleichung erinnert an eines der Rechengesetze für Potenzen: Für jede reelle Zahl a und alle natürlichen Zahlen m, n gilt: $a^{m+n} = a^m \cdot a^n$.

Bezeichnen wir die reelle Zahl $f(1)$ mit dem Buchstaben e (wie EULER; man redet auch von der EULERschen Zahl), so ist $f(2) = f(1 + 1) = e^2$ und allgemein $f(n + 1) = f(n) \cdot e$; also folgt induktiv, daß $f(n) = e^n$

für alle $n \in \mathbb{N}$. Für natürliche Zahlen stimmt unsere Funktion f also überein mit den Potenzen der festen reellen Zahl

$$e = f(1) = \lim_{n\to\infty} \left(1 + \frac{1}{n}\right)^n \approx 2{,}71828\,18284\,59045\,23536\ldots\,.$$

Für eine beliebige reelle Zahl x sind Potenzen mit Exponent x nicht definiert; wir können sie zumindest für die Basis e *definieren* über die gerade konstruierte Funktion f:

Definition: Für eine reelle Zahl x schreiben wir e^x oder $\exp(x)$ für den Funktionswert an der Stelle x der einzigen Funktion $f\colon \mathbb{R} \to \mathbb{R}$, die den Bedingungen

$$f(x + y) = f(x) \cdot f(y) \quad \text{und} \quad f(x) \geq 1 + x$$

für alle $x, y \in \mathbb{R}$ genügt. Die Funktion f bezeichnen wir als Exponentialfunktion.

(Die Schreibweise $\exp(x)$ wird vor allem dann verwendet, wenn das Funktionsargument relativ kompliziert ist:

$$\exp\left(\sqrt[5]{\frac{x^2 + x + 1}{x^2 - x + 1}}\right) \quad \text{ist besser lesbar als} \quad e^{\sqrt[5]{\frac{x^2+x+1}{x^2-x+1}}}\,.$$

Da $f(-x) = 1/f(x)$ gibt es hier keine Konflikte mit der üblichen Konvention $a^{-n} = 1/a^n$; auch Schreibweisen wie $a^{1/n}$ für die n-te Wurzel aus a sind mit dieser Definition verträglich, denn $f(1/n)^n = f(1) = e$.

Die Lösung unseres ursprünglichen Problems läßt sich mit diesen Bezeichnungen dann so zusammenfassen:

Satz: Zu jeder reellen Zahl r gibt genau es eine Funktion $f_r\colon \mathbb{R} \to \mathbb{R}$ mit den Eigenschaften

$$f_r(x + y) = f_r(x) \cdot f_r(y) \quad \text{und} \quad f_r(x) \geq 1 + rx$$

für alle $x, y \in \mathbb{R}$, nämlich die Funktion $f_r(x) = e^{rx}$.

■

Eine Bank, die uns bei einem marktüblichen Zinssatz von p % davon abhalten will, ständig Banken zu wechseln, bietet uns also an, das Kapital bei einer Anlage von $x \in \mathbb{R}$ Jahren mit dem Faktor e^{rx} zu multiplizieren, wobei $r = p/100$ ist. Bei einem Zinssatz von 5% wäre dies für ein Jahr der Faktor $e^{0,05} \approx 1{,}051271096$, wir bekämen also statt fünf Prozent 5,1271096%, was sich bei der angegebenen Genauigkeit nicht vom Wert bei sekündlicher Verzinsung unterscheidet. Wenn man mit drei Nachkommastellen mehr rechnet, sieht man, daß es dann bei sekündlicher Verzinsung mit 336 weitergeht, bei der Exponentialfunktion aber mit 376; bezogen auf eine Anlagesumme von einer Milliarde Euro entspricht dies einem Zinsgewinn von vier Cent.

Im Bankenalltag spielt diese sogenannte *kontinuierliche Verzinsung* keine große praktische Rolle, sie ist allerdings beliebt bei finanzmathematischen Modellrechnungen, da man mit der Exponentialfunktion einfacher und besser rechnen kann als mit tatsächlichen Zinsformeln.

Auch die Funktion e^{rx} verhält sich wie eine Potenzfunktion: Für eine natürliche Zahl n ist $e^{rn} = \left(e^r\right)^n$. Dies legt es nahe, über die Exponentialfunktion Potenzen a^x mit *beliebigen* reellen Basen a und Exponenten x zu definieren: Ist $a = e^r$, so soll $a^x = e^{rx}$ sein.

Für *beliebige* reelle Basen a ist das allerdings nicht möglich: Schließlich ist $e^x \geq 1 + x$ für $x \geq 0$ insbesondere positiv, und da $e^{-x} = 1/e^x$ ist, muß e^x sogar für alle $x \in \mathbb{R}$ positiv sein. Wir müssen uns also zumindest auf positive Basen a beschränken.

Unser nächstes Ziel ist es, zu zeigen, daß es für jedes positive $a \in \mathbb{R}$ in der Tat *genau* eine reelle Zahl x gibt mit $e^x = a$. Dazu sammeln wir zunächst einige Eigenschaften der Exponentialfunktion:

Lemma: *a)* $e^x > 0$ für alle $x \in \mathbb{R}$
b) $e^{x+y} = e^x \cdot e^y$ für alle $x, y \in \mathbb{R}$ und $e^0 = 1$
c) $e^{x-y} = e^x/e^y$ für alle $x, y \in \mathbb{R}$.
d) Für $x > y$ ist auch $e^x > e^y$.

Beweis: a) haben wir uns bereits überlegt, und die erste Gleichung aus *b)* ist einfach eine der beiden Forderungen an die Exponentialfunktion. Nach dieser Gleichung ist insbesondere $e^0 = e^{0+0} = e^0 \cdot e^0$, was wegen der Positivität von e^0 nur für $e^0 = 1$ möglich ist. Weiter ist

$$e^{x-y} \cdot e^y = e^{(x-y)+y} = e^x ,$$

also $e^{x-y} = e^x/e^y$, wie in *c)* behauptet. *d)* ist wegen der Positivität aller Funktionswerte gleichbedeutend mit der Ungleichung $e^x/e^y > 1$.

Nach *c)* ist $e^x/e^y = e^{x-y}$, und das ist wegen der zweiten Forderung an die Exponentialfunktion größer oder gleich $1 + (x - y)$, also echt größer als eins. Damit ist alles bewiesen. ■

Satz: Zu jeder positiven Zahl $x \in \mathbb{R}$ gibt es genau eine reelle Zahl y mit, so daß $e^y = x$ ist.

Beweis: Als erstes wollen wir uns überlegen, daß es eine reelle Zahl b_1 gibt mit $e^{b_1} \geq x$: Nach der zweiten Forderung an die Exponentialfunktion ist $e^y \geq 1 + y$ für alle $y \in \mathbb{R}$, insbesondere ist also $e^{x-1} \geq 1 + (x - 1) = x$, so daß wir zum Beispiel $b_1 = x - 1$ setzen können.

Genauso finden wir leicht ein $a_1 \in \mathbb{R}$ mit $e^{a_1} \leq x$: Wir können uns nämlich wieder zunächst ein $c \in \mathbb{R}$ verschaffen mit $e^c \geq 1/x$, und wegen der Positivität beider Seiten ist dann $e^{-c} \leq x$, so daß wir $a_1 = -c$ setzen können.

Ausgehend vom Intervall $[a_1, b_1]$ konstruieren wir wie beim Beweis der Vollständigkeit von $\mathbb{R}$ eine Intervallschachtelung für die gesuchte Zahl y: Wenn wir ein Intervall $[a_n, b_n]$ konstruiert haben, für das $e^{a_n} \leq x \leq e^{b_n}$ ist, betrachten wir die Intervallmitte $c_n = \frac{1}{2}(a_n + b_n)$. Falls $e^{c_n} \geq x$ ist, nehmen wir als nächstes Intervall die linke Hälfte von $[a_n, b_n]$, setzen also $a_{n+1} = a_n$ und $b_{n+1} = c_n$; andernfalls nehmen wir die rechte Hälfte, setzen also $a_{n+1} = c_n$ und $b_{n+1} = b_n$. Die Bedingungen an eine Intervallschachtelung sind offensichtlich erfüllt, also definiert dies eine reelle Zahl y. Für jeden Index n ist $e^{a_n} \leq e^y \leq e^{b_n}$; wir müssen uns überlegen, daß mit den Differenzen $b_n - a_n$ auch die Differenzen $e^{b_n} - e^{a_n}$ beliebig klein werden. Für jedes n ist

$$e^{b_n} - e^{a_n} = a^{a_n}\left(e^{b_n - a_n} - 1\right)$$

Nach der Eigenschaft 3) aus dem Beweis des Satzes über die Existenz der Exponentialfunktion gilt für jedes $x \in \mathbb{R}$ die Ungleichung

$$e^x(1 - x) \leq 1 \quad \text{oder} \quad e^x - 1 \leq xe^x .$$

Somit ist

$$\begin{aligned} e^{b_n} - e^{a_n} &= e^{a_n}\left(e^{b_n - a_n} - 1\right) \leq e^{a_n}\left((b_n - a_n)e^{b_n - a_n}\right) \\ &= e^{b_n}(b_n - a_n) \leq e^{b_1}(b_n - a_n) , \end{aligned}$$

denn $b_n \le b_1$ für alle n. Die Zahl e^{b_1} ist eine Konstante; da $(b_n - a_n)_{n\in\mathbb{N}}$ eine Nullfolge ist, gilt somit dasselbe auch für die Folge der Differenzen $e^{b_n} - e^{a_n}$. Da die Folge $\left(e^{a_n}\right)_{n\in\mathbb{N}}$ monoton wachsend und durch x nach oben beschränkt ist, $\left(e^{b_n}\right)_{n\in\mathbb{N}}$ monoton fallend und durch x nach unten beschränkt, konvergieren beide; da die Differenzen eine Nullfolge bilden, haben auch beide denselben Grenzwert, der sowohl größer oder gleich als auch kleiner oder gleich x sein muß, also gleich x ist. Somit ist $e^y = x$, es gibt also eine reelle Zahl y mit $e^x = y$. Diese ist eindeutig bestimmt, denn jedes $z \neq y$ ist entweder echt größer oder echt kleiner als y, und damit muß nach obigem Lemma auch e^z echt größer oder echt kleiner als $e^y = x$ sein.

■

Definition: Ist $x > 0$ eine positive reelle Zahl, so bezeichnen wir die eindeutig bestimmte reelle Zahl y mit $e^y = x$ als den *natürlichen Logarithmus* $y = \log x$ von x. Einige Autoren schreiben auch $y = \ln x$ um speziell darauf hinzuweisen, daß der *natürliche* Logarithmus gemeint ist und nicht einer der weiter unten definierten anderen.

Bis zum Aufkommen erschwinglicher Taschenrechner spielten Logarithmen eine wichtige Rolle für praktische Rechnungen. Der Grund dafür liegt im folgenden

Lemma: Für zwei positive reelle Zahlen x, y ist $\log(xy) = \log x + \log y$.

Beweis: Nach der ersten definierenden Eigenschaft der Exponentialfunktion ist $e^{\log x + \log y} = e^{\log x} \cdot e^{\log y} = x \cdot y$. Somit ist $\log x + \log y$ die eindeutig bestimmte reelle Zahl, die von der Exponentialfunktion auf xy abgebildet wird. Das ist aber nach Definition des Logarithmus gerade $\log(xy)$.

■

Für jemanden, der seine Rechnungen nur mit Bleistift und Papier durchführen muß, sind Additionen erheblich weniger aufwendig als Multiplikationen. Daher gab es früher dicke *Logarithmentafeln,* in denen die Logarithmusfunktion tabelliert war, daneben auch beispielsweise die Logarithmen der trigonometrischen Funktionen. Zur Multiplikation zweier Zahlen konnte man deren Logarithmen aus den Tafeln ablesen, diese addieren und dann nachschauen, welche Zahl (ungefähr) diese Summe als Logarithmus hat.

Für uns ist die Logarithmusfunktion beispielsweise interessant zur Definition allgemeiner Potenzen: Für eine positive reelle Zahl a und eine beliebige reelle Zahl x definieren wir

$$a^x \underset{\text{def}}{=} e^{x \cdot \log a} .$$

Für eine natürliche Zahl n ist $e^{n \log a} = \left(e^{\log a}\right)^n = a^n$, die Definition führt also für natürliche Exponenten zum gewohnten Ergebnis.

Lemma: Zu zwei reellen Zahlen $a > 1$ und $x > 0$ gibt es stets genau eine reelle Zahl y, so daß $a^y = x$ ist.

Beweis: Die Gleichung $a^y = e^{y \cdot \log a} = x$ ist gleichbedeutend damit, daß $y \cdot \log a = \log x$ ist. Da $a > 1$ vorausgesetzt war, ist $\log a > 0$, wir können also durch $\log a$ dividieren und erhalten $y = \dfrac{\log x}{\log a}$.

■

Definition: Diese eindeutig bestimmte Zahl y wird als *Logarithmus von x zur Basis a* bezeichnet: $y = \log_a x$.

Gebräuchlich sind vor allem der früher zum Rechnen mit Logarithmentafeln benutzte *dekadische* oder BRIGGSsche Logarithmus $\lg x = \log_{10} x$ sowie der *binäre* Logarithmus $\operatorname{lb} x = \log_2 x$. Logarithmen zu einer Basis $0 < a < 1$ könnte man zwar problemlos definieren, sie spielen aber weder in der Mathematik noch in deren Anwendungen eine nennenswerte Rolle.

Logarithmen sind auch heute noch nützlich unter anderem zur graphischen Darstellung schnell wachsender Größen: Trägt man diese auf einer logarithmischen statt einer linearen Skala auf, erhält man im allgemeinen aussagekräftigere Darstellungen.

§5: Stetige Funktionen

Wir können die meisten reellen Zahlen, mit denen wir es in Theorie oder Anwendungen zu tun haben, nur näherungsweise angeben. Trotzdem wollen wir diese Näherungswerte in Funktionen einsetzen und hoffen dann, daß sich der so berechnete Funktionswert nicht wesentlich von dem unterscheidet, der beim Einsetzen des exakten Werts entstanden wäre.

Das muß nicht immer so sein: Definieren wir eine Funktion $f\colon \mathbb{R} \to \mathbb{R}$ durch die Vorschrift

$$f(x) = \begin{cases} 1 & \text{falls } x \geq e \\ 0 & \text{sonst} \end{cases},$$

so ist $f(e) = 1$. Verwenden wir aber den bei einer Genauigkeit von fünf Nachkommastellen korrekten Näherungswert $e \approx x = 2{,}71828$, so ist $f(x) = 0$, denn mit zehn Nachkommastellen ist $e \approx 2{,}7182818285$, d.h. $x < e$. Schlimmer noch: Wir können e mit beliebiger Genauigkeit annähern durch die Zahlen $\left(1+\frac{1}{n}\right)^n$. Da die Folge dieser Zahlen monoton wächst, sind alle Folgenglieder echt kleiner als e, werden von f also auf die Null abgebildet. Trotzdem ist $f(e) = 1$.

Setzen wir x dagegen ein in die Funktion $g(x) = x^2$, erhalten wir $g(x) = 7{,}389046158$ was sich nur geringfügig vom auf 15 Nachkommastellen korrekten Näherungswert $g(e) = e^2 \approx 7{,}389056098930650$ unterscheidet. Funktionen wie g werden wir in diesem Paragraphen als *stetig* bezeichnen, während die gerade betrachtete Funktion f zumindest bei e *unstetig* ist.

Funktionen müssen nicht auf ganz $\mathbb{R}$ definiert sein: So ist beispielsweise die Funktion $f(x) = 1/(x^2 - 1)$ nur für Argumente $x \neq \pm 1$ definiert, der Logarithmus aus dem vorigen Paragraphen nur für $x > 0$. Wir sollten den Stetigkeitsbegriff daher so formulieren, daß er auch auf Funktionen anwendbar ist, die nur auf einer Teilmenge von $\mathbb{R}$ definiert sind.

Tatsächlich gibt es keinen Grund, daß wir uns auf $\mathbb{R}$ beziehungsweise Teilmengen von $\mathbb{R}$ beschränken; da Stetigkeit etwas mit Nähe zu tun haben soll, können wir in einem beliebigen metrischen Raum arbeiten. Wir können allerdings weder dort noch in $\mathbb{R}$ mit beliebigen Teilmengen arbeiten: Wenn wir möchten, daß eine Funktion robust auf kleinere Schwankungen beim Argument reagiert, müssen wir zumindest verlangen, daß sie definiert bleibt, wenn wir das Argument durch einen Näherungswert ersetzen. Diese Forderung formalisiert der Begriff der *offenen Menge:*

Definition: Eine Teilmenge $D \subseteq X$ eines metrischen Raums X mit Metrik d heißt *offen*, wenn es für jeden Punkt $x \in D$ eine reelle Zahl $\varepsilon > 0$ gibt, so daß jedes $y \in X$ mit $d(x, y) < \varepsilon$ in D liegt.

Anschaulich bedeutet dies, daß D mit jedem Punkt auch noch eine gewisse Kreisscheibe/Kugel/Intervall um diesen herum enthalten soll, so daß man sich in jeder Richtung zumindest ein kleines Stück weit bewegen kann, ohne die Menge D zu verlassen.

In $\mathbb{R}$ ist beispielsweise jedes offene Intervall

$$(a, b) = \left\{x \in \mathbb{R} \;\middle|\; a < x < b\right\}$$

eine offene Teilmenge, denn für $x \in (a, b)$ sind die Differenzen $x - a$ und $b - x$ beide positiv; bezeichnen wir mit ε die kleinere dieser beiden Zahlen, so ist für jede reelle Zahl y mit $|x - y| < \varepsilon$

$$y - a = x - a + (y - x) > x - a - |x - y| \geq 0 \quad \text{und}$$

$$b - y = b - x + (x - y) > b - x - |x - y| \geq 0\,,$$

so daß auch y im Intervall (a, b) liegt. Ein abgeschlossenes Intervall

$$[a, b] = \left\{x \in \mathbb{R} \;\middle|\; a \leq x \leq b\right\}$$

dagegen ist nie offen: Gäbe es beispielsweise für die untere Intervallgrenze a ein $\varepsilon > 0$, so daß $y \in [a, b]$ wann immer $|a - y| < \varepsilon$, so läge insbesondere $a - \varepsilon/2$ in $[a, b]$, was natürlich nicht der Fall ist.

In offenen Mengen können wir uns somit bewegen und sinnvoll definieren was es heißt, daß die Funktion f, ausgewertet an einem Näherungswert für x, einen Näherungswert für $f(x)$ liefern soll. Wir beginnen mit der allgemeinen Situation:

Definition: Eine Abbildung $f\colon D \to Y$ von der offenen Teilmenge $D \subseteq X$ des metrischen Raums X mit Metrik d in den metrischen Raum Y mit Metrik e heißt *stetig* im Punkt $x \in D$, wenn es zu jedem $\varepsilon > 0$ ein $\delta > 0$ gibt, so daß gilt: Erfüllt ein Element $x' \in D$ die Ungleichung $d(x, x') < \delta$, so ist $e\big(f(x), f(x')\big) < \varepsilon$. Die Abbildung f heißt *stetig* (auf D), wenn sie in jedem Punkt $x \in D$ stetig ist.

Eine Funktion $f\colon D \to \mathbb{R}$ auf einer offenen Teilmenge $D \subseteq \mathbb{R}$ ist also genau dann stetig im Punkt $x \in D$. wenn es zu jedem $\varepsilon > 0$ ein $\delta > 0$ gibt, so $|f(x) - f(y)| < \varepsilon$ ist wann immer $|x - y| < \delta$ ist.

Wörtlich dasselbe gilt auch für komplexe Funktionen: Eine Funktion $f\colon D \to \mathbb{C}$ auf einer offenen Teilmenge $D \subseteq \mathbb{C}$ ist also genau dann stetig im Punkt $x \in D$. wenn es zu jedem $\varepsilon > 0$ ein $\delta > 0$ gibt, so $|f(x) - f(y)| < \varepsilon$ ist wann immer $|x - y| < \delta$ ist.

Beginnen wir mit zwei trivialen Beispielen stetiger Funktionen auf $\mathbb{R}$:

Die konstante Abbildung $f: \mathbb{R} \to \mathbb{R}$, die jedem $x \in \mathbb{R}$ denselben Wert $f(x) = a \in \mathbb{R}$ zuordnet, ist stetig in jedem Punkt $x \in \mathbb{R}$ und damit auf ganz $\mathbb{R}$, denn für jedes $y \in \mathbb{R}$ und jedes $\varepsilon > 0$ ist $|f(x) - f(y)| = |a - a| = 0 < \varepsilon$; hier kann δ also völlig beliebig gewählt werden.

Auch die identische Abbildung $f: \mathbb{R} \to \mathbb{R}$ mit $f(x) = x$ für alle $x \in \mathbb{R}$ ist stetig in jedem Punkt $x \in \mathbb{R}$ und damit auf ganz $\mathbb{R}$, denn wählen wir zu vorgegebenem $x \in \mathbb{R}$ und $\varepsilon > 0$ einfach $\delta = \varepsilon$, so ist $|f(x) - f(y)| = |x - y| < \varepsilon$, wenn $|x - y| < \delta = \varepsilon$ ist.

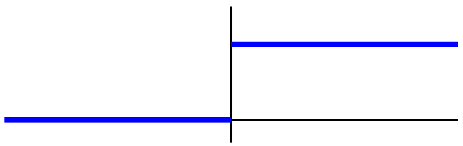
So uninteressant diese beiden Beispiele erscheinen mögen, sind sie doch die beiden Bausteine, mit denen wir zeigen können, daß alle Polynomfunktionen stetig sind.

Bevor wir das beweisen, wollen wir aber zunächst noch zwei Beispiele unstetiger Funktionen betrachten. Das erste ist genauso konstruiert wie das Eingangsbeispiel dieses Paragraphen: Wir setzen

$$f(x) = \begin{cases} 1 & \text{falls x > 0} \\ 0 & \text{sonst} \end{cases} .$$

Diese Funktion ist an der Stelle $x = 0$ unstetig, denn nehmen wir etwa $\varepsilon = \frac{1}{2}$, so kann es kein $\delta > 0$ geben, für das $|f(y) - f(0)| = |f(y)| < \varepsilon$ ist wann immer $|y - 0| = |y| < \delta$ ist: Nehmen wir etwa $y = \delta/2$. so ist $f(y) = 1 > \varepsilon$.

Für $x \neq 0$ ist die Funktion f in einer gewissen Umgebung von x konstant und damit stetig in x; die Sprungstelle bei $x = 0$ ist also die einzige Unstetigkeitsstelle von f.



Die DIRICHLETsche Sprungfunktion

$$f(x) = \begin{cases} 1 & \text{falls } x \in \mathbb{Q} \\ 0 & \text{sonst} \end{cases}$$

ist offenbar in *jedem* Punkt $x \in \mathbb{R}$ unstetig, denn für jedes $\delta > 0$ gibt es sowohl rationale als auch irrationale Zahlen y mit $|x - y| < \delta$; diese Funktion ist also in keinem Punkt stetig.

Um möglichst schnell möglichst viele Beispiele stetiger Funktionen zu bekommen, wollen wir uns überlegen, daß auch Summen, Produkte und Quotienten stetiger Funktionen wieder stetig sind. Ähnliche Aussagen hatten wir bereits über die Konvergenz von Folgen bewiesen; um dies anwenden zu können, wollen wir Stetigkeit via Folgen charakterisieren:

Lemma: Eine Abbildung $f: D \to Y$ von der offenen Teilmenge $D \subseteq X$ des metrischen Raums X mit Metrik d in den metrischen Raum Y mit Metrik e ist genau dann stetig im Punkt $x \in D$, wenn für jede Folge $(x_n)_{n\in\mathbb{N}}$ von Elementen aus D mit Grenzwert x gilt:

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

Zum *Beweis* nehmen wir zunächst an, die Funktion f sei stetig in x und die Folge $(x_n)_{n\in\mathbb{N}}$ konvergiere gegen x. Zu jedem $\varepsilon > 0$ gibt es dann ein $\delta > 0$, so daß $d\big(f(y), f(x)\big) < \varepsilon$ ist, falls $d(x, y) < \delta$. Wegen der Konvergenz der Folge gegen x gibt es zu diesem δ ein $n_0 \in \mathbb{N}$, so daß $d(x, x_n) < \delta$ für alle $n \geq n_0$. Für diese n ist dann auch $d\big(f(x), f(x_n)\big) < \varepsilon$, die Folge der $f(x_n)$ konvergiert also gegen $f(x)$.

Umgekehrt konvergiere für jede gegen x konvergente Folge $(x_n)_{n\in\mathbb{N}}$ von Elementen aus D die Folge der $f(x_n)$ gegen $f(x)$; wir müssen zeigen, daß f stetig ist in x. Wäre dies nicht der Fall, gäbe es ein $\varepsilon > 0$, so daß es zu jedem $\delta > 0$ ein $y \in D$ gäbe mit $d(x, y) < \delta$, aber $d\big(f(x), f(y)\big) \geq \varepsilon$. Insbesondere gäbe es für jede natürliche Zahl n ein $x_n \in D$, so daß $d(x, x_n) < 1/n$ wäre, aber $d\big(f(x), f(x_n)\big) > \varepsilon$. Natürlich konvergiert die Folge der x_n gegen x, aber die Folge der $f(x_n)$ kann nicht gegen $f(x)$ konvergieren, da alle Abstände der $f(x_n)$ zu $f(x)$ mindestens ε sind. Dies widerspricht der Voraussetzung; also ist f stetig in x.

■

Damit können wir nun ziemlich leicht aus bekannten stetigen Funktionen neue konstruieren:

Lemma: Sind die Funktionen $f, g\colon D \to \mathbb{R}$ auf der offenen Teilmenge $D \subseteq \mathbb{R}$ stetig im Punkt $z \in D$, so sind auch die Funktionen

$$f \pm g\colon \begin{cases} D \to \mathbb{R} \\ x \mapsto f(x) \pm g(x) \end{cases} \quad \text{und} \quad f \cdot g\colon \begin{cases} D \to \mathbb{R} \\ x \mapsto f(x) \cdot g(x) \end{cases}$$

stetig in z. Ist $g(x) \neq 0$ für alle $x \in D$, so ist auch

$$\frac{f}{g}\colon \begin{cases} D \to \mathbb{R} \\ x \mapsto \dfrac{f(x)}{g(x)} \end{cases}$$

stetig in z.

Beweis: Wir benutzen das gerade bewiesene Lemma: Wegen der vorausgesetzten Stetigkeit der Funktionen f und g in z ist für jede gegen z konvergente Folge $(z_n)_{n\in\mathbb{N}}$ von Elementen aus D

$$\lim_{n\to\infty} f(z_n) = f(z) \quad \text{und} \quad \lim_{n\to\infty} g(z_n) = g(z)\,.$$

Das Lemma folgt damit sofort aus den im vorigen Paragraphen bewiesenen Rechenregeln für Grenzwerte. ■

Daraus folgt nun fast sofort, daß alle Polynomfunktionen stetig sind: Wenden wir dieses Lemma nämlich an auf das Produkt der Funktion $f(x) = x$ mit sich selbst, folgt, daß auch die Funktion g mit $g(x) = x^2$ stetig auf ganz $\mathbb{R}$ ist und induktiv weiter, daß für jedes $n \in \mathbb{N}$ die Funktion, die x auf x^n abbildet, stetig ist. Multiplizieren wir noch mit einer konstanten Funktion, folgt als nächstes, daß auch alle Funktionen der Form $f(x) = ax^n$ mit $a \in \mathbb{R}$ und $n \in \mathbb{N}$ stetig sind, und indem wir mehrere solcher Funktionen und gegebenenfalls noch eine konstante Funktion aufaddieren folgt dann

Lemma: Für alle $n \in \mathbb{N}$ und $a_0, \dots, a_n \in \mathbb{R}$ ist die Funktion

$$f\colon \begin{cases} \mathbb{R} \to \mathbb{R} \\ x \mapsto a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 \end{cases}$$

stetig auf ganz $\mathbb{R}$. ■

Bei Quotienten von Polynomen müssen wir etwas vorsichtiger sein, da deren Nenner nirgends verschwinden darf. Deshalb gilt hier nur

Lemma: $f(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$ und $g(x) = b_m x^m + b_{m-1}x^{m-1} + \cdots + b_1 x + b_0$ seien zwei Polynomfunktionen und $D \subseteq \mathbb{R}$ sei eine offene Teilmenge von $\mathbb{R}$, so daß $g(x) \neq 0$ für alle $x \in D$. Dann ist der Quotient $f(x)/g(x)$ eine auf ganz D stetige Funktion. ■

Außer durch Grundrechenarten können wir Funktionen auch durch Hintereinanderausführen miteinander verknüpfen. Hier gilt:

Lemma: Sind $f\colon D \to \mathbb{R}$ und $g\colon D' \to \mathbb{R}$ stetige Funktionen derart, daß $f(x) \in D'$ für alle $x \in D$, so ist auch die Hintereinanderausführung $g \circ f\colon D \to R$ mit $(g \circ f)(x) = g\big(f(x)\big)$ stetig.

Beweis: Zu jedem $x \in D$ und $\varepsilon > 0$ gibt es zunächst wegen der Stetigkeit von g im Punkt $f(x)$ ein $\delta > 0$, so daß $\big|g\big(f(x)\big) - g(y)\big| < \varepsilon$ ist wann immer $|f(x) - y| < \delta$ ist. Wegen der Stetigkeit von f im Punkt x gibt es zu diesem δ wiederum ein η, so daß $|f(x) - f(y)| < \delta$ ist für alle $x \in D$ mit $|x - y| < \eta$. Für diese y ist daher auch $\big|g\big(f(x)\big) - g\big(f(y)\big)\big| < \varepsilon$. Damit ist die Stetigkeit von $g \circ f$ in x gezeigt. ■

Lemma: Die Exponentialfunktion ist stetig auf ganz $\mathbb{R}$.

Beweis: Für jedes $u \in \mathbb{R}$ ist nach Eigenschaft 3) der Exponentialfunktion $e^u(1-u) \le 1$, also $e^u - 1 \le ue^u$. Damit ist für zwei reelle Zahlen $u < v$

$$e^v - e^u = e^u(e^{v-u} - 1) \le e^u \cdot (v-u)e^{v-u} = (v-u)e^v\,.$$

Ist $w > v$, so gilt wegen der Monotonie der Exponentialfunktion erst recht die Ungleichung

$$e^v - e^u \le (v-u)e^w \quad \text{für alle } u < v < w\,.$$

Daraus folgt nun sofort die Stetigkeit der Exponentialfunktion in einem vorgegebenen Punkt x: Zu gegebenem $\varepsilon > 0$ wählen wir eine Zahl w derart, daß $e^w > e^x + \varepsilon$. Mit $\delta = e^{-w}\varepsilon$ ist dann für y mit $|x - y| < \delta$

$$|e^x - e^y| \le e^w \cdot |y - x| < e^w\delta = \varepsilon\,.$$

Dies zeigt die Stetigkeit der Exponentialfunktion. ■

Korollar: Für jedes $a > 0$ ist die Funktion $f(x) = a^x$ stetig auf ganz $\mathbb{R}$.

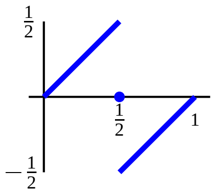
Beweis: Wir schreiben $a^x = e^{x \cdot \log a}$. Die lineare Funktion $x \mapsto x \cdot \log a$ ist stetig und nach dem gerade bewiesenen Lemma auch die Exponentialfunktion. Somit ist auch f als Hintereinanderausführung dieser beiden Funktionen stetig. ∎

Man beachte, daß wir hier nicht den Logarithmus als Funktion, sondern nur den Wert $\log a$ als Konstante verwendet haben. Trotzdem wollen wir natürlich wissen, ob die Logarithmusfunktion stetig ist. Diese Frage soll aber erst einmal zurückgestellt werden, denn die nun folgenden allgemeinen Aussagen über Stetigkeit werden uns Ergebnisse liefern, mit denen sich insbesondere auch dies einfach entscheiden läßt.

Wir untersuchen dazu den Wertebereich einer Funktion. Betrachten wir etwa die Funktion $f(x) = x$ auf dem offenen Intervall $(0, 1)$, so ist deren Bild einfach wieder dasselbe Intervall. Dessen Infimum ist die Null, Supremum die Eins, und beides liegt nicht im Intervall. Hätten wir die Funktion stattdessen auf dem abgeschlossenen Intervall $[0, 1]$ betrachtet, wären sowohl Infimum als auch Supremum Funktionswerte gewesen.

Ein abgeschlossenes Intervall allein genügt jedoch noch nicht, um dies sicherzustellen: Die Funktion

$$f\colon \begin{cases} [0, 1] \to \mathbb{R} \\ x \mapsto \begin{cases} x & \text{falls } x < \frac{1}{2} \\ 0 & \text{falls } x = \frac{1}{2} \\ x-1 & \text{falls } x > \frac{1}{2} \end{cases} \end{cases}$$



hat in ihrem Wertebereich alle Zahlen zwischen $-\frac{1}{2}$ und $\frac{1}{2}$, nicht aber diese beiden Zahlen selbst. Sie ist allerdings bei $x = \frac{1}{2}$ nicht stetig.

Wenn wir sowohl Stetigkeit fordern als auch von einem abgeschlossenen Intervall ausgehen, kann so etwas nicht passieren:

Lemma: $f\colon D \to \mathbb{R}$ sei eine stetige Funktion und das abgeschlossene Intervall $[a, b]$ liege ganz in der offenen Menge D. Dann gilt:

a) f ist beschränkt auf $[a, b]$

b) Ist $m = \inf\{f(x) \mid x \in [a, b]\}$ die größte untere Schranke für die Funktionswerte auf $[a, b]$ und $M = \sup\{f(x) \mid x \in [a, b]\}$ die kleinste obere Schranke, so gibt es Zahlen $x_m, x_M \in [a, b]$, für die $f(x_m) = m$ und $f(x_M) = M$ ist.

Beweis: a) Angenommen, der Betrag von $f(x)$ wäre nicht beschränkt auf $[a, b]$. Dann gäbe es insbesondere zu jedem $n \in \mathbb{N}$ ein $x_n \in [a, b]$, so daß $|f(x_n)| > n$ wäre. Da alle Glieder der Folge $(x_n)_{n\in\mathbb{N}}$ im Intervall $[a, b]$ liegen, wäre die Folge beschränkt; nach dem Satz von BOLZANO-WEIERSTRASS gäbe es daher eine konvergente Teilfolge $(y_n)_{n\in\mathbb{N}}$ mit $y_n = x_{\nu_n}$, wobei ν_n eine streng monoton ansteigende Folge natürlicher Zahlen ist. Der Grenzwert y der Folge $(y_n)_{n\in\mathbb{N}}$ müßte die Ungleichung $a \le y \le b$ erfüllen, da diese von allen Folgengliedern y_n erfüllt wird.

Nun betrachten wir die Folge der Zahlen $f(y_n)/\nu_n$. Die Folge der $1/\nu_n$ ist eine Nullfolge, die der $f(y_n)$ konvergiert wegen der Stetigkeit von f gegen $f(y)$; nach den Rechenregeln für Grenzwerte wäre daher auch die Produktfolge eine Nullfolge. Andererseits hat aber $f(y_n) = f(x_{\nu_n})$ einen Betrag größer ν_n, so daß $|f(y_n)/\nu_n| > 1$ ist für alle n, so daß die Folge unmöglich gegen Null konvergieren kann. Dieser Widerspruch zeigt die Beschränktheit von f auf $[a, b]$.

b) Auch hier argumentieren wir indirekt: Angenommen, es gäbe kein $x_m \in [a, b]$, so daß $f(x_m) = m$ wäre. Da m als Infimum insbesondere eine untere Schranke ist, wäre dann $f(x) > m$ für alle $x \in [a, b]$, also wäre die Funktion

$$g(x) = \frac{1}{f(x) - m}$$

stetig und positiv in allen Punkten aus $[a, b]$. Nach *a)* gäbe es daher eine obere Schranke N für diese Funktion. Die Ungleichung $g(x) \le N$ ist aber gleichbedeutend mit

$$f(x) - m \ge \frac{1}{N} \quad \text{oder} \quad f(x) \ge m + \frac{1}{N}$$

für alle $x \in [a, b]$. Dies widerspricht der Definition von m als größter unterer Schranke. Dieser Widerspruch zeigt, daß die Annahme, es gäbe kein $x_m \in [a, b]$ mit $f(x_m) = m$, falsch war.

Für die kleinste obere Schranke M können wir entweder genauso argumentieren, oder aber wir nutzen aus, daß $-M$ die größte untere Schranke von $-f$ ist. Wie gerade gezeigt, gibt es daher ein $x_M \in [a, b]$, so daß $-f(x_M) = -M$, also $f(x_M) = M$ ist. ■

Als nächstes wollen wir zeigen, daß unter den Voraussetzungen dieses Lemmas auch alle Werte *zwischen* m und M angenommen werden. Dazu beginnen wir mit einem Spezialfall, den zumindest einige wohl schon von Kurvendiskussionen aus der Schule kennen:

Lemma: $f: D \to \mathbb{R}$ sei eine stetige Funktion und das abgeschlossene Intervall $[a, b]$ liege ganz in der offenen Menge D. Außerdem sei $f(a)f(b) < 0$, d.h. $f(a)$ und $f(b)$ haben verschiedene Vorzeichen. Dann gibt es ein $x \in (a, b)$, so daß $f(x) = 0$ ist.

Beweis: Wir konstruieren einen solchen Punkt x durch eine Intervallschachtelung aus Intervallen $[a_n, b_n]$ mit der Eigenschaft, daß $f(a_n)f(b_n) \le 0$ für alle $n \in \mathbb{N}$. Als erstes solches Intervall können wir einfach $[a_1, b_1] = [a, b]$ nehmen.

Wenn wir das Intervall $[a_n, b_n]$ konstruiert haben, betrachten wir dessen Mittelpunkt $c_n = \frac{1}{2}(a_n+b_n)$. Wäre $f(a_n)f(c_n) > 0$ und $f(b_n)f(c_n) > 0$, so wären entweder alle drei Zahlen a_n, b_n, c_n positiv oder alle drei negativ, so daß auch $f(a_n)f(b_n) > 0$ wäre, im Widerspruch zur Voraussetzung. Also ist mindestens eine der beiden Zahlen kleiner oder gleich Null; falls $f(a_n)f(c_n) \le 0$ ist, nehmen wir die linke Hälfte $[a_n, c_n]$ des Intervalls als neues Intervall $[a_{n+1}, b_{n+1}]$, andernfalls die rechte Hälfte $[c_n, b_n]$.

Es ist klar, daß dadurch eine Intervallschachtelung definiert wird, denn jedes Intervall ist als linke oder rechte Hälfte seines Vorgängers insbesondere in diesem enthalten, und da jedes Intervall nur noch halb so lang ist wie sein Vorgänger, bilden die Intervallängen eine Nullfolge. Also definiert diese Konstruktion eine reelle Zahl $x \in [a, b]$. Wegen der

Stetigkeit von f ist

$$f(x) = f\left(\lim_{n\to\infty} a_n\right) = \lim_{n\to\infty} f(a_n) \quad \text{und}$$

$$f(x) = f\left(\lim_{n\to\infty} b_n\right) = \lim_{n\to\infty} f(b_n)\,, \text{ also}$$

$$f(x)^2 = \lim_{n\to\infty} f(a_n) \cdot \lim_{n\to\infty} f(b_n) = \lim_{n\to\infty} \big(f(a_n)f(b_n)\big)\,.$$

Für alle $n \in \mathbb{N}$ ist $f(a_n)f(b_n) \le 0$, also ist auch der Grenzwert $f(x)^2$ kleiner oder gleich null. Als Quadrat einer reellen Zahl kann er aber nicht negativ sein, also ist $f(x) = f(x)^2 = 0$, und wir haben eine Nullstelle gefunden. ■

Die Verallgemeinerung dieses Satzes auf andere Werte als null ist nun nur noch eine reine Formalität; wir erhalten den

Zwischenwertsatz: $f: D \to \mathbb{R}$ sei eine stetige Funktion und das abgeschlossene Intervall $[a, b]$ liege ganz in der offenen Menge D. Weiter seien $m = \inf\{f(x) \mid x \in [a, b]\}$ die größte untere Schranke für die Funktionswerte auf $[a, b]$ und $M = \sup\{f(x) \mid x \in [a, b]\}$ die kleinste obere Schranke. Dann gibt es zu jeder reellen Zahl c mit $m \le c \le M$ ein $x \in [a, b]$, so daß $f(x) = c$ ist.

Beweis: Für $c = m$ und für $c = M$ haben wir das bereits bewiesen; wir können also zwei Werte $x_m, x_M \in [a, b]$ finden, so daß $f(x_m) = m$ und $f(x_M) = M$ ist. Wir wollen uns überlegen, daß es für $m < c < M$ sogar zwischen x_m und x_M ein x gibt mit $f(x) = c$. Dazu wenden wir das vorige Lemma an auf die Funktion $g(x) = f(x) - c$ auf diesem Intervall $[a^*, b^*]$, d.h. für $a^* = \min(x_m, x_M)$ und $b^* = \max(x_m, x_M)$. Die Voraussetzungen sind erfüllt, denn

$$g(a)g(b) = g(x_m)g(x_M) = \big(f(x_m) - c\big)\big(f(x_M) - c\big) = (m - c)(M - c)$$

ist negativ, da $m < c$ und $M > c$ ist. Somit gibt es ein x im Intervall $[a^*, b^*] \subseteq [a, b]$ mit $g(x) = f(x) - c = 0$, also $f(x) = c$, wie behauptet. ■

Als nächsten Schritt in Richtung Umkehrfunktionen stetiger Funktionen wollen wir injektive Funktionen genauer untersuchen. Hier gilt:

Lemma: $f\colon D \to \mathbb{R}$ sei eine stetige Funktion und das abgeschlossene Intervall $[a, b]$ liege ganz in der offenen Menge D. Die Funktion f ist genau dann injektiv auf $[a, b]$, wenn sie dort streng monoton ist.

Beweis: Falls f auf $[a, b]$ streng monoton ist, können zwei verschiedene Werte $x, y \in [a, b]$ nicht auf denselben Funktionswert abgebildet werden, denn wegen der strengen Monotonie muß eine der beiden Ungleichungen $f(x) < f(y)$ oder $f(x) > f(y)$ gelten.

Ist umgekehrt f injektiv auf $[a, b]$, so müssen wir uns zunächst überlegen, *welche* Monotonie wir beweisen wollen. Dazu betrachten wir die Funktionswerte an den Intervallenden; falls $f(a) < f(b)$, ist f auf $[a, b]$ sicherlich nicht monoton fallend; wir müssen also zeigen, daß f dann auf ganz $[a, b]$ monoton wächst.

Wir überlegen uns zunächst, daß $f(a) < f(x) < f(b)$ ist für alle x aus dem offenen Intervall (a, b). Wäre nämlich $f(x) < f(a)$, so müßte f nach dem Zwischenwertsatz im Intervall $[x, b]$ den Wert $f(a)$ annehmen, im Widerspruch zur Injektivität auf $[a, b]$. Entsprechend müßte f im Intervall $[a, x]$ den Wert $f(b)$ annehmen, falls $f(x) > f(b)$ wäre.

Gäbe es nun im Intervall (a, b) zwei Punkte $x < y$ mit $f(x) > f(y)$, so gäbe es wegen der Ungleichung $f(a) < f(y) < f(x)$ nach dem Zwischenwertsatz ein $z \in [a, x]$ mit $f(z) = f(y)$, wieder im Widerspruch zur Injektivität. Somit ist f monoton wachsend auf $[a, b]$.

Falls $f(a) > f(b)$ ist, ist $-f(a) < -f(b)$, wir können also das gerade bewiesene auf die Funktion $-f$ anwenden. Da diese somit streng monoton wächst, muß f selbst streng monoton fallen. ∎

Satz: $f\colon D \to \mathbb{R}$ sei eine stetige Funktion und das abgeschlossene Intervall $[a, b]$ liege ganz in der offenen Menge D. Falls f auf $[a, b]$ streng monoton wächst, gilt

a) f bildet das Intervall $[a, b]$ ab auf das Intervall $[m, M]$ mit $m = f(a)$ und $M = f(b)$.

b) f hat eine streng monoton wachsende Umkehrfunktion, d.h. eine Funktion $g\colon [m, M] \to [a, b]$, für die $f \circ g$ die Identität auf $[m, M]$ ist und $g \circ f$ die Identität auf $[a, b]$.

c) g ist stetig auf (m, M).

Beweis: a) Wegen der vorausgesetzten Monotonie ist natürlich $f(a)$ das Minimum und $f(b)$ das Maximum der Funktionswerte auf dem Intervall $[a, b]$. Nach dem Zwischenwertsatz wird auch jede Zahl zwischen diesen beiden Werten angenommen.

b) Als streng monoton wachsende Funktion ist f insbesondere injektiv; daher gibt es zu jedem $y \in [m, M]$ genau ein $x \in [a, b]$, so daß $f(x) = y$ ist. Natürlich setzen wir $g(y) = x$, also ist $f \circ g(y) = f\bigl(g(y)\bigr) = f(x) = y$. Entsprechend ist $g \circ f(x) = g\bigl(f(x)\bigr)$ das einzige Element des Intervalls $[a, b]$, das auf $f(x)$ abgebildet wird, also x. Auch die strenge Monotonie von g ist klar: Wäre nämlich für $y_1 < y_2$ nicht $g(y_1) < g(y_2)$, so wäre $g(y_2) \le g(y_1)$ und damit wegen der Monotonie von f auch $y_2 = f\bigl(g(y_2)\bigr) \le f\bigl(g(y_1)\bigr) = y_1$.

c) Auch die Stetigkeit von g folgt leicht aus der Monotonie von f: Sind $y = f(x_0) \in (m, M)$ und $\varepsilon > 0$ vorgegeben, so können wir ε natürlich problemlos verkleinern; wir nehmen an, daß sowohl $y - \varepsilon$ als auch $y + \varepsilon$ in (m, M) liegen. Dann gibt es Elemente x_0, x_1 und x_2 aus $[a, b]$, so daß $f(x_1) = y - \varepsilon$, $f(x) = y$ und $f(x_2) = y + \varepsilon$. Wir definieren δ als das Minimum der beiden Zahlen $x - x_1$ und $x_2 - x_1$. Ist $|x - x_0| < \delta$, so ist insbesondere $x_1 < x < x_2$, also $y - \varepsilon < f(x) < y + \varepsilon$, also $|f(x) - y| < \varepsilon$. ∎

§6: Summen und Reihen

Wenn wir per Computer einen Funktionswert näherungsweise berechnen lassen, können je nach Funktion im Hintergrund sehr verschiedene Formeln ausgewertet werden. In vielen Fällen aber werden Polynome und ähnliche Summen ausgewertet. Im nächsten Kapitel werden wir sehen, wie man die meisten interessanten Funktionen als Grenzwerte einer Folge von Polynomen darstellen kann; hier soll es zunächst nur um einige erste Grundlagen gehen.

Reihen sind spezielle Folgen, deren Glieder nicht direkt definiert werden, sondern über die Differenz zu ihrem Vorgänger. Ist also $x_1 = a_1$ das erste Folgenglied und $x_n = x_{n-1} + a_n$, so ist

$$x_n = \sum_{k=1}^{n} a_k$$

die Summe der Zahlen von a_1 bis a_n. Falls diese Folge gegen einen Grenzwert x konvergiert, schreiben wir

$$x = \sum_{k=1}^{\infty} a_k \,.$$

Die Zahlen a_k und damit auch x_n können dabei wahlweise reelle oder komplexe Zahlen sein.

Als Beispiel einer Reihe betrachten wir wieder ein Geldanlage mit einem festen Zinssatz von p %, gehen aber jetzt davon aus, daß jedes Jahr am ersten Januar eine feste Summe a einbezahlt wird. Wir interessieren uns für den Kontostand am ersten Januar n Jahre nach Kontoeröffnung.

Bei der Kontoeröffnung wurde erstmalig der jährliche Anlagebetrag a einbezahlt: diese Summe stand n Jahre lang auf dem Konto, wurde als n mal mit $q = 1 + p/100$ multipliziert und wuchs damit auf aq^n.

Am ersten Januar des Folgejahres wurde wieder die gleiche Summe a einbezahlt; da sie nur $n-1$ Jahre auf dem Konto stand, wird sie nur mit q^{n-1} multipliziert. Der zwölf Monate später einbezahlte Betrag wir entsprechend nur mit q^{n-2} multipliziert, und so weiter bis zum ersten Januar n Jahre später, der gerade erst einbezahlt wurde und damit natürlich noch keine Zinsen abwirft.

Der Betrag, der nach n Jahren zur Verfügung steht, ist also

$$x_n = aq^n + aq^{n-1} + \cdots + aq + a = \sum_{k=0}^{n} aq^k \,.$$

Diesen Ausdruck hätten wir gerne in einer einfacheren Form. Dazu verhilft ein einfacher Trick: Wir multiplizieren ihn mit q:

$$qx_n = aq^{n+1} + aq^n + \cdots + aq^2 + aq = \sum_{k=1}^{n+1} aq^k \,.$$

Vergleichen wir dies mit der Darstellung von x_n, so kommen fast alle Terme in beiden Summen vor; lediglich am Anfang und am Ende gibt es Diskrepanzen. Damit können wir mit einer deutlichen Vereinfachung rechnen, wenn wir die beiden Summen voneinander subtrahieren:

$$qx_n - x_n = aq^{n+1} - a \quad \text{also ist} \quad x_n = a\frac{q^{n+1}-1}{q-1} \,,$$

wobei letztere Formel natürlich nur für $q \neq 1$ gilt. (Die Formel für $q = 1$ kann sich hoffentlich jeder Leser selbst überlegen.) Auch ohne diese Formel ist klar, daß die Folge x_n im bei einer Geldanlage angestrebten Fall $q > 1$ unbeschränkt wächst.

Was passiert, wenn $q < 1$ ist? Unter dem Gesichtspunkt einer Geldanlage mag dieser Fall unrealistisch erscheinen, wenn wir aber bedenken, daß gerade bei einer langfristigen Geldanlage auch die Inflation berücksichtigt werden muß und nicht zu allen Zeiten jede Anlage einen über der Inflationsrate liegenden Zinssatz erzielt, kann auch er zumindest nicht ausgeschlossen werden. Die Annahme einer über lange Zeit konstanten Inflationsrate ist zwar natürlich noch unrealistischer als die einer über lange Zeit konstanten Zinssatzes, aber bei der Anwendung mathematischer Methoden tastet man sich meist langsam vor von extrem idealisierten (sprich: völlig unrealistischen) Modellen zu immer realitätsnäheren und im Idealfall schließlich auch praktisch nützlichen.

Gehen wir also aus von einer abstrakt mathematischen Situation: Wir haben zwei reelle oder komplexe Zahlen a und $q \neq 1$, und wir interessieren uns für das Verhalten von

$$x_n = \sum_{k=0}^{n} aq^k = a\frac{q^{n+1}-1}{q-1} \,.$$

Aus §4 wissen wir, daß die Folge $(q^n)_{n\in\mathbb{N}}$ für $|q| > 1$ betragsmäßig unbeschränkt wächst, für $|q| = 1$ aber $q \neq 1$ unbestimmt divergiert und für $|q| < 1$ gegen null konvergiert. Damit ist nach den Rechenregeln für Grenzwerte klar, daß in diesem Fall die Folge der

$$x_n = \sum_{k=1}^{n} aq^k = a\frac{q^{n+1}-1}{q-1} = a\frac{1-q^{n+1}}{1-q}$$

gegen $a/(1-q)$ konvergiert. Für $|q| < 1$ ist somit

$$\sum_{k=0}^{\infty} aq^k = \frac{a}{1-q} \,.$$

Damit haben wir zumindest eine unendliche Reihe ausgerechnet; wie sich bald zeigen wird, ist dieses zunächst eher künstlich und unbedeutend erscheinende spezielle Resultat ein Ausgangspunkt für deutlich allgemeinere, interessantere und nützlichere Ergebnisse.

Wenn eine Summe aus unendlich vielen Termen einen endlichen Wert haben soll, müssen die einzelnen Summanden natürlich klein werden; insbesondere gilt

Lemma: Falls die Summe mit Summanden $(a_k)_{k\in\mathbb{N}}$ konvergiert, muß $(a_k)_{k\in\mathbb{N}}$ eine Nullfolge sein.

Beweis: Wenn die Folge der Teilsummen $x_n = \sum_{k=1}^{n} a_k$ konvergiert, ist sie insbesondere eine CAUCHY-Folge, es gibt also zu jedem $\varepsilon > 0$ ein $n_0 \in \mathbb{N}$, so daß $|x_n - x_m| < \varepsilon$ für alle $m, n \geq n_0$. Dies gilt insbesondere für $m = n + 1$; für $n \geq n_0$ ist daher $|x_{n+1} - x_n| = |a_{n+1}| < \varepsilon$. Somit ist $(a_k)_{k\in\mathbb{N}}$ eine Nullfolge. ∎

Umgekehrt betrachten wir die Summe, deren Summanden die Glieder unserer Standard-Nullfolge $1/n$ sind, wir setzen also $a_k = 1/k$ und

$$x_n = \sum_{k=1}^{n} \frac{1}{k}\,.$$

Um einen ersten Eindruck über die Konvergenz dieser Reihe zu bekommen, berechnen wir einige Glieder (näherungsweise):

$$\begin{aligned}
x_{10} &\approx 2{,}928968253968254\\
x_{100} &\approx 5{,}187377517639621\\
x_{1000} &\approx 7{,}485470860550343\\
x_{10000} &\approx 9{,}787606036044345\\
x_{100000} &\approx 12{,}09014612986328\\
x_{1000000} &\approx 14{,}39272672286478\\
x_{10000000} &\approx 16{,}69531136585671\\
x_{100000000} &\approx 18{,}9978964138477
\end{aligned}$$

Die Teilsummen steigen also nur recht langsam an, allerdings ist die Differenz zwischen $x_{10^{n+1}}$ und x_{10^n} bei den hier betrachteten Werten stets ungefähr gleich; wenn man nachrechnet kommt man auf etwa 2,302585.

Das spricht nicht dafür, daß die Summe gegen einen endlichen Grenzwert konvergiert, und in der Tat divergiert diese sogenannte *harmonische Reihe.* Das können wir am einfachsten dadurch sehen, daß wir die Differenzen zwischen den Teilsummen zu aufeinanderfolgenden Zweierpotenzen abschätzen: Da die Folge der Zahlen $1/k$ monoton fällt, ist

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

Damit ist

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

und das wächst unbegrenzt mit wachsendem n. Die harmonische Reihe divergiert also (bestimmt) gegen unendlich.

Sie schrammt allerdings nur haarscharf an der Konvergenz vorbei: Im Kapitel über Integralrechnung werden wir sehen, daß die Summe der Zahlen $1/n^r$ für jedes *reelle* $r > 1$ konvergiert. Für $r = 2$ etwa kann man (mit Methoden jenseits der Analysis I wie beispielsweise der Theorie der FOURIER-Reihen) zeigen, daß die Reihe gegen $\pi^2/6$ konvergiert; entsprechend ergeben sich auch für andere gerade natürliche Zahlen r Summen der Form π^r mal rationale Zahl. Die Grenzwerte für ungerade r sind erheblich schlechter verstanden.

Wenn wir bei Summen und Reihen sind, sollten wir auch endlich eine Summenformel beweisen, die zwar nichts mit unendlichen Summen zu tun hat und eigentlich auch nichts mit Analysis, die aber doch auch in der Analysis immer wieder gebraucht wird. Es geht darum, die n-te Potenz einer Summe $(a + b)$ durch Terme der Form $a^k b^\ell$ auszudrücken.

Der Ansatz ist einfach: Wir schreiben

$$(a+b)^n = \underbrace{(a+b)(a+b)\cdots(a+b)}_{n \text{ Faktoren}}$$

und multiplizieren dies aus nach dem Distributivgesetz. Danach erhalten wir die Summe aller Produkte aus n Faktoren, wobei der i-te jeweils einer der beiden Summanden aus der i-te Klammer ist. Insgesamt erhalten wir also 2^n Summanden; wegen der Kommutativität der Multiplikation haben diese jedoch allesamt einen der $n + 1$ Werte $a^k b^{n-k}$ mit einem k

zwischen 0 und n. Wir müssen zählen, wie oft jeder dieser Terme vorkommt.

Dazu überlegen wir uns, auf wie viele Arten wir aus n Faktoren k auswählen können (aus denen wir dann jeweils den Summanden a nehmen). Wenn wir einfach der Reihe nach Faktoren hernehmen, haben wir für den ersten noch die volle Auswahl von n Faktoren, beim zweiten haben wir nur noch $n-1$, nämlich alle, außer dem bereits ausgewählten, für den dritten $n-2$, und so weiter. Insgesamt haben wir also

$$n(n-1)\dots(n-k+1)$$

Möglichkeiten. Für die Berechnung von $(a+b)^n$ ist es allerdings ohne Bedeutung, ob wir den i-ten Faktor als ersten oder als letzten berücksichtigt haben bei der Auswahl von k Faktoren, aus denen wir das a nehmen wollen; für jede Wahl von k-Faktoren müssen wir also noch dividieren durch die Anzahl der möglichen Anordnungen einer Menge von k Elementen. Wie wir in den Übungen gesehen haben, gibt es dafür $k!$ Möglichkeiten; der Term $a^k b^{n-k}$ tritt daher

$$\binom{n}{k} \underset{\text{def}}{=} \frac{n(n-1)\cdots(n-k+1)}{k!} = \frac{n!}{k!\,(n-k)!}$$

Mal auf. Für $k=0$ und $k=n$ kommt hier auch $0!$ vor; dafür verwenden wir die übliche Konvention, daß ein leeres Produkt gleich eins sein soll, d.h. $0! = 1$. Somit ist

$$\binom{n}{0} = \frac{n!}{0!\,n!} = 1 \quad \text{und} \quad \binom{n}{n} = \frac{n!}{n!\,0!} = 1\,,$$

wie es auch sein soll: Eine Menge von n Elementen hat genau eine Teilmenge mit null Elementen, nämlich die leere Menge, und genau eine Teilmenge mit n Elementen, nämlich sich selbst.

Das Symbol $\binom{n}{k}$ bezeichnen wir als den *Binomialkoeffizienten* $\binom{n}{k}$; gesprochen n *über* k. (Im Englischen sagt man n *choose* k; hier steckt der Begriff der Auswahl also schon in der Sprechweise.) Damit folgt

Binomischer Lehrsatz: Für zwei Elemente a, b eines Körpers k ist

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k} \quad \text{mit} \quad \binom{n}{k} = \frac{n!}{k!\,(n-k)!}$$

■

Die kompakte Schreibweise $\dfrac{n!}{k!\,(n-k)!}$ ist natürlich nicht zum wirklichen Rechnen gedacht; da ist der längere Ausdruck

$$\binom{n}{k} = \frac{n(n-1)\cdots(n-k+1)}{k!}$$

erheblich günstiger.

Der Name *binomischer Lehrsatz* kommt daher, daß man eine Summe aus zwei Termen als *Binom* bezeichnet; er hat nichts mit irgendwelchen Mathematikern zu tun.

Die obige Herleitung dieses Satzes war wahrscheinlich sehr gut verständlich für die Leser, die aus der Schule mit elementarer Kombinatorik und Wahrscheinlichkeitstheorie vertraut sind; für den Rest sieht das Ganze vielleicht noch etwas dubios aus. Deshalb möchte ich den Satz noch einmal ohne Kombinatorik beweisen, und zwar durch vollständige Induktion:

Für $n=1$ ist $\binom{1}{0} = \binom{1}{1} = 1$, der Satz wird also zur trivialen Aussage

$$a+b = a+b\,.$$

Wenn wir den Satz für ein festes n bewiesen haben, können wir $(a+b)^{n+1}$ berechnen als

$$\begin{aligned}
(a+b)^{n+1} = (a+b)^n \cdot (a+b) &= \left(\sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}\right)(a+b) \\
&= \sum_{k=0}^{n} \binom{n}{k} a^{k+1} b^{n-k} + \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k+1} \\
&= \sum_{k=1}^{n+1} \binom{n}{k-1} a^k b^{n+1-k} + \sum_{k=0}^{n} \binom{n}{k} a^k b^{n+1-k} \\
&= \underbrace{b^{n+1}}_{k=0} + \sum_{k=1}^{n} C_k a^k b^{n+1-k} + \underbrace{a^{n+1}}_{k=n+1}
\end{aligned}$$

mit $C_k = \binom{n}{k-1} + \binom{n}{k}$, denn wie wir bereits gesehen haben, ist

$$\binom{n}{0} = \binom{n}{n} = 1\,.$$

Für den Induktionsschluß müssen wir C_k noch weiter umformen:

$$C_k = \binom{n}{k-1} + \binom{n}{k} = \frac{n!}{(k-1)!\,(n+1-k)!} + \frac{n!}{k!\,(n-k)!}$$
$$= \frac{n!\,k + n!\,(n+1-k)}{k!\,(n+1-k)!} = \frac{(n+1)!}{k!\,(n+1-k)!} = \binom{n+1}{k}.$$

Damit gilt der binomische Lehrsatz auch für $n+1$, also für alle $n \in \mathbb{N}$. ■

Die beiden im Beweis verwendeten Regeln

$$\binom{n}{0} = \binom{n}{n} = 1 \quad \text{und} \quad \binom{n}{k-1} + \binom{n}{k} = \binom{n+1}{k}$$

lassen sich auch dazu verwenden, für ein festes, nicht zu großes n die sämtlichen Binomialkoeffizienten $\binom{n}{k}$ einfach zu berechnen: Im Dreieck

$$\begin{array}{ccccccccccc}
 & & & & \binom{1}{0} & & \binom{1}{1} & & & & \\
 & & & \binom{2}{0} & & \binom{2}{1} & & \binom{2}{2} & & & \\
 & & \binom{3}{0} & & \binom{3}{1} & & \binom{3}{2} & & \binom{3}{3} & & \\
 & \binom{4}{0} & & \binom{4}{1} & & \binom{4}{2} & & \binom{4}{3} & & \binom{4}{4} & \\
\binom{5}{0} & & \binom{5}{1} & & \binom{5}{2} & & \binom{5}{3} & & \binom{5}{4} & & \binom{5}{5} \\
 & & & & & \cdots & & & & &
\end{array}$$

stehen außen lauter Einsen, und innen ist jeder Eintrag die Summe der beiden darüberstehenden. Damit läßt sich das Dreieck leicht rekursiv berechnen:

$$\begin{array}{ccccccccccc}
 & & & & 1 & & 1 & & & & \\
 & & & 1 & & 2 & & 1 & & & \\
 & & 1 & & 3 & & 3 & & 1 & & \\
 & 1 & & 4 & & 6 & & 4 & & 1 & \\
1 & & 5 & & 10 & & 10 & & 5 & & 1 \\
 & & & & & \cdots & & & & &
\end{array}$$

Die offensichtliche Symmetrie dieses sogenannten PASCALschen Dreiecks in Bezug auf seine Mittelachse können wir leicht allgemein beweisen:

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!} = \frac{n!}{(n-k)!\,k!} = \binom{n}{n-k}.$$

BLAISE PASCAL (1623–1662) ging nie zur Schule, da ihm sein Vater alles selbst beibringen wollte. Mathematik war erst ab dem Alter 15 vorgesehen, jedoch war er durch Andeutungen seines Vaters so neugierig geworden, daß er die Geometrie im Alter von 12 Jahren selbst erfand. Zwei Jahre später nahm er an mathematischen Treffen in Paris teil; mit 16 präsentierte er seine erste eigene Arbeit. Um seinem Vater, einem Steuereinnehmer, zu helfen, baute er eine der ersten mechanischen Rechenmaschinen. Im späteren Leben befaßte er sich hauptsächlich mit Philosophie und Theologie, kam aber immer wieder zur Mathematik zurück, wo er unter anderem die Wahrscheinlichkeitstheorie begründete.

Nach diesem Einschub kehren wir wieder zurück zu unendlichen Summen und betrachten als nächstes Beispiel die Summe

$$S = \sum_{k=1}^{\infty} (-1)^{k+1} = 1 - 1 + 1 - 1 + 1 - \cdots.$$

Durch Zusammenfassen aufeinanderfolgender Terme können wir sie leicht ausrechnen: Fassen wir jeweils die Terme mit $k = 2\ell - 1$ und $k = 2\ell$ zusammen, erhalten wir die Formel

$$S = \sum_{\ell=1}^{\infty} \left((-1)^{2\ell} + (-1)^{2\ell+1}\right) = \sum_{\ell=1}^{\infty} (1-1) = (1-1) + (1-1) + \cdots = 0.$$

Wir können freilich auch die Terme mit $k = 2\ell$ und $k = 2\ell + 1$ zusammenfassen und erhalten dann

$$S = 1 + \sum_{\ell=1}^{\infty} \left((-1)^{2\ell+1} + (-1)^{2\ell+2}\right) = 1 + (-1+1) + (-1+1) + \cdots = 1.$$

Eine dritte Möglichkeit wäre, daß wir nur über die ungeraden k summieren und den Term $(-1)^{k+1}$ mit $(-1)^{2k+1}$ kombinieren. Damit erreichen wir allerdings nur diejenigen geraden Zahlen, die das Doppelte einer ungeraden Zahl sind. Was noch fehlt, sind die geraden Zahlen, die das Doppelte einer geraden Zahl sind; das sind genau die durch vier teilbaren Zahlen. Somit ist

$$\begin{aligned} S &= \sum_{\substack{k=1\\ k \text{ ungerade}}}^{\infty} \big((-1)^{k+1} + (-1)^{2k+1}\big) + \sum_{\ell=1}^{\infty}(-1)^{4\ell+1} \\ &= \sum_{\substack{k=1\\ k \text{ ungerade}}}^{\infty} (1-1) + \sum_{\ell=1}^{\infty}(-1) = 0 - \sum_{\ell=1}^{\infty} 1\,, \end{aligned}$$

was gegen $-\infty$ divergiert.

Von diesen drei Ergebnissen kann offensichtlich höchstens eines richtig sein; welches?

Gehen wir zurück zur Definition: Eine unendliche Summe konvergiert, wenn die Summe ihrer Teilsummen konvergiert, und ihr Wert ist dann der Grenzwert dieser Folge. In unserem Fall haben wir die Teilsummen

$$s_n = \sum_{k=1}^{n}(-1)^{k+1} = \begin{cases} 1 & \text{falls } n \text{ ungerade} \\ -1 & \text{falls } n \text{ gerade} \end{cases} .$$

Somit ist keines der drei gerade „bewiesenen“ Ergebnisse richtig; die Folge der s_n und damit die Summe ist unbestimmt divergent.

Der Fehler in den obigen Rechnungen lag darin, daß wir zwar in endlichen Summen beliebig umordnen und klammern können; dazu müssen wir nur das Kommutativ- und das Assoziativgesetz hinreichend oft anwenden. Daß wir diese Gesetze beliebig oft anwenden können, folgt aus dem Prinzip der vollständigen Induktion, aber damit können wir nur beweisen, daß wir jede *endliche* Anzahl von Umordnungen vornehmen können, nicht aber unendlich viele. In unendlichen Summen dürfen wir also nicht beliebig umordnen; wenn wir es trotzdem tun, erhalten wir mit etwas Geschick die verschiedensten Reihen, die nicht nur gegen verschiedene Grenzwerte konvergieren, sondern teilweise auch bestimmt divergent sind.

Das Problem liegt offensichtlich hauptsächlich darin, daß sich Terme gegenseitig wegheben können. Das funktioniert offensichtlich nur dann, wenn es Summanden mit verschiedenen Vorzeichen gibt; wir sollten also erwarten, daß wir weniger Probleme haben bei Reihen, deren Summanden allesamt dasselbe Vorzeichen haben. Das ist insbesondere dann der Fall, wenn wir die Summanden einer beliebigen Reihe durch ihre Beträge ersetzen. Deshalb definieren wir

Definition: Die unendliche Reihe $\sum_{k=1}^{\infty} a_k$ heißt *absolut konvergent*, wenn die Reihe $\sum_{k=1}^{\infty} |a_k|$ konvergiert.

Die beiden Summen $\sum_{k=1}^{\infty} a_k$ und $\sum_{k=1}^{\infty} |a_k|$ werden, selbst wenn beide existieren, oft nicht das geringste miteinander zu tun haben: Betrachten wir etwa für eine reelle Zahl $-1 < q < 0$ die geometrische Reihe

$$\sum_{k=1}^{\infty} q^k = \frac{1}{1-q}\,.$$

Die Summe der Beträge ist hier

$$\sum_{k=1}^{\infty} |q|^k = \frac{1}{1-|q|} = \frac{1}{1+q}\,.$$

Für $q = -\frac{1}{2}$ etwa ist die erste Summe gleich $\frac{2}{3}$, die zweite aber zwei.

Trotzdem folgt aus der absoluten Konvergenz einer Reihe deren Konvergenz. Um dies einzusehen, betrachten wir das CAUCHYsche Konvergenzkriterium für die Folge der Teilsummen

$$s_n = \sum_{k=1}^{\infty} a_k$$

einer unendlichen Reihe. Nach CAUCHY konvergiert diese Folge genau dann, wenn es zu jedem $\varepsilon > 0$ ein $n_1 \in \mathbb{N}$ gibt, so daß $|s_m - s_n| < \varepsilon$ ist für alle $n, m \geq n_1$. Hier ist für $m \geq n$

$$|s_m - s_n| = \left|\sum_{k=1}^{m} a_k - \sum_{k=1}^{n} a_k\right| = \left|\sum_{k=n+1}^{m} a_k\right| .$$

Setzen wir $n_0 = n_1 + 1$, so können wir die Bedingung auch so formulieren, daß es ein $n_0 \in \mathbb{N}$ gibt so daß

$$\left|\sum_{k=n}^{m} a_k\right| < \varepsilon \quad \text{für alle } m \geq n \geq n_0\,.$$

Somit gilt:

Cauchysches Konvergenzkriterium für Reihen: Die unendliche Reihe $\sum_{k=1}^{\infty} a_k$ mit reellen oder komplexen Summanden a_k konvergiert genau dann, wenn es zu jedem $\varepsilon > 0$ ein $n_0 \in \mathbb{N}$ gibt, so daß

$$\left|\sum_{k=n}^{m} a_k\right| < \varepsilon \quad \text{für alle } m \geq n \geq n_0\,.$$

Daraus folgt nun leicht

Satz: Ist eine unendliche Reihe $\sum_{k=1}^{\infty} a_k$ mit reellen oder komplexen Summanden a_k absolut konvergent, so ist sie auch konvergent.

Beweis: Da die Summe der Beträge konvergiert, gibt es nach CAUCHY zu jedem $\varepsilon > 0$ ein $n_0 \in \mathbb{N}$, so daß

$$\left|\sum_{k=n}^{m} |a_k|\right| = \sum_{k=n}^{m} |a_k| < \varepsilon \quad \text{für alle } m \geq n \geq n_0\,.$$

Nach der Dreiecksungleichung ist dann aber auch

$$\left|\sum_{k=n}^{m} a_k\right| \leq \sum_{k=n}^{m} |a_k| < \varepsilon \quad \text{für alle } m \geq n \geq n_0\,,$$

so daß die gegebene Reihe nach CAUCHY konvergiert. ∎

Um zu sehen, ob auch die Umkehrung dieses Satzes gilt, betrachten wir die sogenannte *alternierende harmonische Reihe*

$$\sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \cdots\,.$$

Sie ist natürlich nicht absolut konvergent, denn wenn wir alle Summanden durch ihre Beträge ersetzen, erhalten wir die divergente harmonische Reihe.

Um einen Eindruck von möglicher Konvergenz oder Divergenz der alternierenden harmonischen Reihe zu bekommen, können wir wieder einige der Teilsummen $s_n = \sum_{k=1}^{n} \frac{(-1)^k}{k}$ berechnen (lassen):

$$\begin{aligned} s_{10} &\approx 0{,}6456349206 \\ s_{100} &\approx 0{,}6881721793 \\ s_{1000} &\approx 0{,}6926474306 \\ s_{10000} &\approx 0{,}6930971831 \end{aligned}$$

Diese Werte legen den Verdacht nahe, daß die Folge gegen einen Wert nahe 0,7 konvergieren sollte. Zumindest die Tatsache, daß sie konvergiert, können wir auch leicht beweisen mit Hilfe eines Kriteriums von LEIBNIZ:

Definition: Eine Reihe der Form $\sum_{k=1}^{\infty} (-1)^{k+1} a_k$ mit reellen Summanden heißt *alternierend*, wenn entweder alle $a_k \geq 0$ oder alle $a_k \leq 0$ sind.

Die Summanden sind also abwechselnd größer oder gleich Null und kleiner oder gleich Null. Für solche Reihen haben wir das

Konvergenzkriterium von Leibniz: Eine alternierende Reihe

$$\sum_{k=1}^{\infty} (-1)^{k+1} a_k$$

konvergiert, falls die Folge $\big(|a_k|\big)_{k\in\mathbb{N}}$ eine monoton fallende Nullfolge ist.

Beweis: Wenn wir alle Summanden mit -1 multiplizieren, ändert sich nichts an der Konvergenz: Der Grenzwert wird einfach ebenfalls mit mit -1 multipliziert. Daher können wir uns beim Beweis auf den Fall beschränken, daß alle $a_k \geq 0$ sind. Bei den Teilsummen

$$s_n = \sum_{k=1}^{n} (-1)^{k+1} a_k$$

ist dann für gerades n stets $s_n \leq s_{n+1}$, denn

$$s_{n+1} = s_n + (-1)^{n+2} a_{n+1} = s_n + a_{n+1} \,.$$

Wir betrachten die Folge der Intervalle $\big([s_{2n},\, s_{2n+1}]\big)_{n\in\mathbb{N}}$ und wollen uns überlegen, daß sie eine Intervallschachtelung definieren.

Dazu müssen wir als erstes zeigen, daß jedes Intervall in seinem Vorgänger liegt, daß also $s_{2n} \leq s_{2n+2}$ und $s_{2n+3} \geq s_{2n+1}$ ist. Da $(a_n)_{n\in\mathbb{N}}$ eine monoton fallende Nullfolge ist, folgt das sofort durch Berechnung der Differenzen:

$$s_{2n+2} - s_{2n} = \big((-1)^{2n+2} a_{2n+1} + (-1)^{2n+3} a_{2n+2}\big) = a_{2n+1} - a_{2n+2} > 0$$

und

$$s_{2n+3} - s_{2n+1} = \big((-1)^{2n+3} a_{2n+2} + (-1)^{2n+4} a_{2n+3}\big) = a_{2n+3} - a_{2n+2} < 0 \,.$$

Schließlich muß noch gezeigt werden, daß die Intervallängen eine Nullfolge bilden; das ist aber klar, denn wie wir zu Beginn des Beweises gesehen haben, hat das n-te Intervall die Länge a_{2n+1}. ∎

Insbesondere konvergiert also die alternierende harmonische Reihe; damit ist klar, daß es auch konvergente Reihen gibt, die nicht absolut konvergent sind. Im nächsten Kapitel werden wir übrigens sehen, daß sie gegen den natürlichen Logarithmus von zwei konvergiert.

BARON GOTTFRIED WILHELM VON LEIBNIZ (1646–1716) gilt als der letzte Universalgelehrte, der das gesamte Wissen seiner Zeit überblickte. In der Mathematik ist er vor allem berühmt durch die Entwicklung der Infinitesimalrechnung (bezüglich derer es einen langen Prioritätsstreit mit NEWTON gab); Bezeichnungen wie $\frac{dy}{dx}$ und $\int f(x)\,dx$ gehen auf ihn zurück. Durch seine Begründung der symbolischen Logik legte er auch einen wesentlichen Grundstein der späteren Informatik. Weitere Arbeiten befassen sich mit den Naturwissenschaften und der Technik, der Philosophie, Theologie und der Geschichte.

Wir werden Reihen im weiteren Verlauf der Vorlesung hauptsächlich brauchen, um uns unbekannte Funktionswerte wie $\log 2, e^3, \sin 2$ oder auch Konstanten wie π näherungsweise zu berechnen. Die Konvergenz solcher Reihen können wir natürlich nicht dadurch beweisen, daß wie die Differenzen zum (uns unbekannten) Grenzwert betrachten; stattdessen brauchen wir als erstes Kriterien, die uns die Konvergenz einer Reihe garantieren; danach erst können wir uns in einem zweiten Schritt Gedanken über die Berechnung des Grenzwerts machen.

Ein solches Kriterium, das von LEIBNIZ, haben gerade kennengelernt; die meisten anderen hängen in irgendeiner Weise zusammen mit geometrischen Reihen, mit denen wir eine gegebene Reihe vergleichen. Die Vergleiche beruhen auf dem folgenden

Majorantenkriterium: Zur unendlichen Reihe $\sum\limits_{k=1}^{\infty} a_k$ mit reellen oder komplexen Summanden a_k gebe es eine konvergente Reihe $\sum\limits_{k=1}^{\infty} b_k$ mit reellen Summanden $b_k \geq 0$ derart, daß $|a_k| \leq b_k$ für alle k. Dann konvergiert $\sum\limits_{k=1}^{\infty} a_k$ absolut.

Zum *Beweis* können wir genauso vorgehen wie beim Beweis, daß jede absolut konvergente Reihe konvergiert: Wegen der Konvergenz von $\sum\limits_{k=1}^{\infty} b_k$ gibt es nach CAUCHY zu jedem $\varepsilon > 0$ ein $n_0 \in \mathbb{N}$, so daß

$$\left|\sum_{k=n}^{m} b_k\right| = \sum_{k=n}^{m} b_k < \varepsilon \quad \text{für alle } m \geq n \geq n_0 \,.$$

Nach der Dreiecksungleichung ist dann auch

$$\left|\sum_{k=n}^{m} a_k\right| \leq \sum_{k=n}^{m} |a_k| \leq \sum_{k=n}^{m} b_k < \varepsilon \quad \text{für alle } m \geq n \geq n_0 \,,$$

so daß die gegebene Reihe nach CAUCHY konvergiert. ∎

Die Reihe $\sum_{k=1}^{\infty} b_k$ bezeichnet man in diesem Zusammenhang als eine *konvergente Majorante.*

Als erste Anwendung wollen wir uns überlegen, daß die Reihe $\sum_{k=1}^{n} \frac{1}{k^2}$ konvergiert: Für $k > 1$ ist $\frac{1}{k^2} < \frac{1}{k(k-1)} = \frac{1}{k-1} - \frac{1}{k}$ und

$$\sum_{k=2}^{n} \frac{1}{k(k-1)} = \sum_{k=2}^{n} \left(\frac{1}{k-1} - \frac{1}{k} \right) = \sum_{k=1}^{n-1} \frac{1}{k} - \sum_{k=2}^{n} \frac{1}{k} = 1 - \frac{1}{n}$$

konvergiert gegen eins. Somit ist die Summe $1 + \sum_{k=2}^{\infty} \frac{1}{k(k-1)}$ eine konvergente Majorante, die betrachtete Summe konvergiert also absolut und ist damit insbesondere konvergent.

Als weitere Anwendung erhalten wir sofort, daß für jede reelle Zahl $s \geq 2$ die Reihe

$$\zeta(s) = \sum_{k=1}^{\infty} \frac{1}{k^s}$$

konvergiert, denn $1/k^s \leq 1/k^2$ für alle $k \in \mathbb{N}$. Die so definierte Funktion heißt RIEMANNsche ζ-Funktion. (ζ ist der griechische Buchstabe *zeta.*)

GEORG FRIEDRICH BERNHARD RIEMANN (1826-1866) war Sohn eines lutherischen Pastors und schrieb sich 1846 auf Anraten seines Vaters an der Universität Göttingen für das Studium der Theologie ein. Schon bald wechselte an die Philosophische Fakultät, um dort unter anderem bei GAUSS Mathematikvorlesungen zu hören. Nach Promotion 1851 und Habilitation 1854 erhielt er dort 1857 einen Lehrstuhl. Trotz seines frühen Todes initiierte er grundlegende auch noch heute fundamentale Entwicklungen in der Geometrie, der Zahlentheorie und über abelsche Funktionen. Seine Vermutung über die Nullstellen der (heute als RIEMANNsch bezeichneten) ζ-Funktion ist die berühmteste offene Vermutung der heutigen Mathematik.

Tatsächlich konvergiert die Reihe sogar für alle $s > 1$, aber das werden wir erst im Kapitel über Intergralrechnung beweisen können. Mehr noch:

In der *Funktionentheorie*, der Analysis mit komplexen Zahlen, zeigt man, daß die so definierte Funktion fortgesetzt werden kann zu einer Funktion mit beliebigen komplexen Argumenten $s \neq 1$. Eine berühmte Vermutung von RIEMANN besagt, daß alle ihre nichtreellen Nullstellen den Imaginärteil $\frac{1}{2}$ haben. Dies hätte wichtige Konsequenzen beispielsweise für Sätze über die Verteilung der Primzahlen.

Nicht zuletzt deshalb wählte die CLAY Foundation dieses Problem im Jahre 2000 als eines ihrer sieben Millenium-Probleme, für deren Lösung sie jeweils eine Million Dollar ausgesetzt hat. Bislang ist erst eines dieser Probleme, die POINCARÉ-Vermutung, gelöst (von GRISHA PERELMAN aus St. Petersburg, der den Preis ablehnte); um die restlichen sechs, darunter die RIEMANN-Vermutung, können sich die Leser noch bemühen. Für Einzelheiten siehe `http://www.claymath.org/millennium/`.

Wie bereits erwähnt, werden wir das Majorantenkriterium hauptsächlich verwenden, um vorgegebene Reihen mit geometrischen Reihen zu vergleichen; schließlich sind das die einzigen, bei denen wir einen vollständigen Überblick sowohl über ihr Konvergenzverhalten als auch ihren Grenzwert haben. Als erstes solches Vergleichskriterium betrachten wir das

Wurzelkriterium: $\sum_{k=1}^{\infty} a_k$ sei eine Reihe mit reellen oder komplexen Summanden. Falls es eine reelle Zahl $q < 1$ gibt und ein $k_0 \in \mathbb{N}$, so daß $\sqrt[k]{|a_k|} \leq q$ ist für alle $k \geq k_0$, ist die Reihe absolut konvergent.

Beweis: Für $k \geq k_0$ ist $|a_k| \leq q^k$; somit ist für die Reihe

$$\sum_{k=1}^{\infty} b_k \quad \text{mit} \quad b_k = \begin{cases} |a_k| & \text{falls } k < k_0 \\ q^k & \text{falls } k \geq k_0 \end{cases}$$

stets $|a_k| \leq b_k$. Diese Reihe ist konvergent mit Grenzwert

$$\sum_{k=1}^{\infty} b_k = \sum_{k=1}^{k_0-1} |a_k| + \sum_{k=k_0}^{\infty} q^k = \sum_{k=1}^{k_0-1} |a_k| + q^{k_0} \sum_{k=0}^{\infty} q^k = \sum_{k=1}^{k_0-1} |a_k| + \frac{q^{k_0}}{1-q} .$$

Somit haben wir eine konvergente Majorante gefunden; die gegebene Reihe konvergiert daher nach dem Majorantenkriterium absolut. ■

In der Literatur wird das Wurzelkriterium meist etwas spezieller, dafür aber kompakter formuliert:

Wurzelkriterium, 2. Fassung: $\sum_{k=1}^{\infty} a_k$ sei eine Reihe mit reellen oder komplexen Summanden. Falls $\lim_{k\to\infty} \sqrt[k]{|a_k|}$ existiert und echt kleiner als eins ist, konvergiert die Reihe absolut.

Beweis: Sei $\lim_{k\to\infty} \sqrt[k]{|a_k|} = q_1 < 1$. Dann ist $\varepsilon = \frac{1}{2}(1-q_1) > 0$, und nach Definition eines Grenzwerts gibt es ein $k_0 \in \mathbb{N}$, so daß

$$\left|\sqrt[k]{|a_k|} - q_1\right| < \varepsilon = \frac{1-q_1}{2} \quad \text{für alle } k \geq k_0 \,.$$

Für diese $k \geq k_0$ ist dann insbesondere

$$\sqrt[k]{|a_k|} < q_1 + \varepsilon = q_1 + \frac{1-q_1}{2} = \frac{q_1+1}{2} \underset{\text{def}}{=} q \,.$$

Als Mittelwert von $q_1 < 1$ und der Eins ist auch $q < 1$, also können wir das Wurzelkriterium in seiner ersten Form anwenden mit dieser Zahl q und k_0. ■

Eine zweite Möglichkeit zum Vergleich mit einer geometrischen Reihe besteht darin, daß wir Quotienten zweier aufeinanderfolgender Summanden betrachten. Da dabei eine Division durch Null auftreten könnte, formulieren wir es zunächst quotientenfrei:

Quotientenkriterium: $\sum_{k=1}^{\infty} a_k$ sei eine Reihe mit reellen oder komplexen Summanden. Falls es eine reelle Zahl $q < 1$ gibt und ein $k_0 \in \mathbb{N}$, so daß $|a_{k+1}| \leq q\,|a_k|$ für alle $k \geq k_0$, konvergiert die Reihe absolut.

Beweis: Wir setzen $a = a_{k_0}$. Durch vollständige Induktion folgt sofort, daß $|a_k| \leq q^{k-k_0} a$ ist für alle $k \geq k_0$. Damit können wir wieder im wesentlichen genauso verfahren wie beim Beweis des Wurzelkriteriums: Für die Reihe

$$\sum_{k=1}^{\infty} b_k \quad \text{mit} \quad b_k = \begin{cases} |a_k| & \text{falls } k < k_0 \\ aq^k & \text{falls } k \geq k_0 \end{cases}$$

ist stets $|a_k| \leq b_k$, und diese Reihe ist konvergiert gegen den Grenzwert

$$\sum_{k=1}^{\infty} b_k = \sum_{k=1}^{k_0-1} |a_k| + \sum_{k=k_0}^{\infty} aq^k = \sum_{k=1}^{k_0-1} |a_k| + aq^{k_0} \sum_{k=0}^{\infty} q^k = \sum_{k=1}^{k_0-1} |a_k| + \frac{aq^{k_0}}{1-q} \,.$$

Somit haben wir eine konvergente Majorante gefunden; die gegebene Reihe konvergiert daher nach dem Majorantenkriterium absolut. ■

Auch dieses Kriterium können wir etwas spezieller mit einem Grenzwert formulieren:

Quotientenkriterium, 2. Fassung: $\sum_{k=1}^{\infty} a_k$ sei eine Reihe mit reellen oder komplexen Summanden. Falls $\lim_{k\to\infty} \frac{|a_{k+1}|}{|a_k|}$ existiert und echt kleiner als eins ist, konvergiert die Reihe absolut.

Beweis: Wir können diese zweite Fassung im wesentlichen auf die gleiche Weise aus der ersten Fassung folgern wie beim Wurzelkriterium: Sei $\lim_{k\to\infty} \frac{|a_{k+1}|}{|a_k|} = q_1 < 1$. Dann ist $\varepsilon = \frac{1}{2}(1-q_1) > 0$, und nach Definition eines Grenzwerts gibt es ein $k_0 \in \mathbb{N}$, so daß

$$\left|\frac{|a_{k_1}|}{|a_k|} - q_1\right| < \varepsilon = \frac{1-q_1}{2} \quad \text{für alle } k \geq k_0 \,.$$

Für diese $k \geq k_0$ ist dann insbesondere

$$\frac{|a_{k_1}|}{|a_k|} < q_1 + \varepsilon = q_1 + \frac{1-q_1}{2} = \frac{q_1+1}{2} \underset{\text{def}}{=} q < 1 \,.$$

Damit können wir das Quotientenkriterium in seiner obigen Form anwenden. ■

Man beachte, daß die jeweils zweite Fassung dieser Kriterien schwächer ist als die erste, denn diese auch anwendbar, wenn der in der zweiten Fassung betrachtete Grenzwert nicht existiert.

Für Anwendungen dieser beiden Kriterien sei auf die Übungen verwiesen und vor allem auf die Behandlung der Potenzreihen im nächsten Kapitel.