



Friedrich-Alexander-Universität
Naturwissenschaftliche Fakultät



Vorlesungsskript

Mathematik für Ingenieure II

Sommersemester 2010

Sommersemester 2012

Sommersemester 2014

Sommersemester 2015

Sommersemester 2018

Sommersemester 2020

Sommersemester 2022

Prof. Dr. Serge Kräutle

am

Lehrstuhl für Angewandte Mathematik (Modellierung und Numerik)

AM1

— Version vom 16. März 2022 —

Friedrich–Alexander–Universität Erlangen–Nürnberg

Inhaltsverzeichnis

1	Folgen	1
1.1	Einführung	1
1.2	Monotonie, Beschränktheit, Supremum	2
1.3	Konvergenz von Folgen	4
2	Reihen	14
2.1	Konvergenzkriterien für Reihen	15
2.2	Potenzreihen: Konvergenzradius	25
2.3	Einige Eigenschaften der Exponentialfunktion und von Sinus und Kosinus	29
3	Funktionen in einem Argument	36
3.1	Funktionengrenzwerte	36
3.2	Stetigkeit	39
3.3	Eigenschaften stetiger Funktionen	44
3.4	Funktionsfolgen, gleichmäßige Konvergenz, und nochmal Potenzreihen	47
3.5	Differenzierbarkeit	52
3.6	Eigenschaften differenzierbarer Funktionen: Folgen differenzierbarer Funktionen, Extremstellensuche, Mittelwert- satz, Regel von l'Hospital, Taylor-Formel, Newton-Verfahren	59
3.6.1	Differenzierbarkeit von Funktionsfolgen, insbesondere Potenz- reihen	59
3.6.2	Extremstellen und Mittelwertsätze	60
3.6.3	Die Regel von l'Hospital	64
3.6.4	Die Formel von Taylor	67
3.6.5	Das Newton-Verfahren (und das Bisektionsverfahren)	77
3.7	Integration	82
3.7.1	Stammfunktionen	82
3.7.2	Das Riemann-Integral	84
3.7.3	Uneigentliche Riemann-Integrale	90
3.7.4	Partialbruchzerlegung	92
4	Grundlagen der Analysis im \mathbb{R}^n	95
4.1	Folgen im \mathbb{R}^n	96
4.2	Stetigkeit	98
4.3	Differenzierbarkeit	101
4.3.1	Richtungsableitung, partielle Ableitung, Jacobi-Matrix	101
4.3.2	Totale Ableitung	105
4.3.3	Weitere Differentialoperatoren und Anwendungen	110
4.3.4	Die Formel von Taylor im \mathbb{R}^n	113

1 Folgen

1.1 Einführung

Eine Folge ist eine 'Aufzählung' von unendlich vielen Zahlen a_1, a_2, a_3, \dots
Z.B. durch

$$a_n := n^2 \quad \forall n \in \mathbb{N} \quad (1.1)$$

ist eine Folge $a_1 = 1, a_2 = 4, a_3 = 9, \dots$ erklärt; die Folge der Quadratzahlen. Eine Folge muss natürlich nicht aus natürlichen Zahlen $a_n \in \mathbb{N}$ bestehen; auch reelle Zahlen ($a_n \in \mathbb{R}$) oder komplexe Zahlen ($a_n \in \mathbb{C}$) sind möglich; z.B. die komplexe Folge

$$a_n := \left(\frac{1}{2}(1+i)\right)^n \quad \forall n \in \mathbb{N}. \quad (1.2)$$

Das mathematische Symbol für eine Folge, bestehend aus den Folgengliedern a_n ist

$$(a_n)_{n \in \mathbb{N}}$$

oder kurz: (a_n) .

Mathematisch betrachtet wird jedem Index n eine Zahl a_n zugeordnet, d.h. wir können den Begriff der Folge *zurückführen* auf den Begriff 'Abbildung':

Eine Folge ist eine Abbildung von der Menge der natürlichen Zahlen \mathbb{N} in eine Menge M von Zahlen; üblicherweise $M = \mathbb{Z}$ oder $M = \mathbb{R}$ oder $M = \mathbb{C}$; jedes $n \in \mathbb{N}$ wird abgebildet auf $a_n \in M$. Da $\mathbb{N} \subset \mathbb{Z} \subset \mathbb{R} \subset \mathbb{C}$, reicht es im Grunde genommen, $M = \mathbb{C}$ zu betrachten.¹

Als Indexmenge muss man nicht zwingend \mathbb{N} nehmen; auch \mathbb{N}_0 ist möglich (oder z.B. auch $\{2, 3, 4, 5, \dots\}$); so ist z.B. durch

$$a_n := (n+1)^2 \quad \forall n \in \mathbb{N}_0$$

ebenfalls die Folge der Quadratzahlen beschrieben. Man schreibt dann $(a_n)_{n \in \mathbb{N}_0}$ bzw. $(a_n)_{n \geq 2}$ ²

Neben der 'expliziten' Angabe der Rechenvorschrift zur Berechnung von a_n (wie in den Beispielen oben) kann man Folgen auch *rekursiv* erklären. Ein Beispiel ist, für vorgegebenes $z \in \mathbb{R}^+$

$$a_1 := 1, \quad a_{n+1} := \frac{1}{2} \left(a_n + \frac{z}{a_n} \right).$$

Diese Rekursionsvorschrift beschreibt das aus dem 1. Semester bekannte Heron-Verfahren;³ wir hatten uns anschaulich überlegt, dass, wenn diese Folge 'konvergiert',

¹Später werden wir auch noch andere Mengen M , wie z.B. Vektorräume \mathbb{R}^m , betrachten, also Folgen von *Vektoren*.

²Im Folgenden werden wir bei mathematischen Sätzen und Definitionen der Einfachheit halber \mathbb{N} als Indexmenge annehmen; es lässt sich alles unmittelbar auch auf Folgen mit anderen Indexmengen wie \mathbb{N}_0 oder $\mathbb{N} \setminus \{1\}$ übertragen.

³Das Heron-Verfahren muss nicht bei $a_1 = 1$ anfangen werden; man sollte für a_1 eine Zahl nehmen, die in der Nähe von \sqrt{z} liegt.

sie gegen \sqrt{z} konvergiert, also ein a_n , für n hinreichend groß, eine Näherung an \sqrt{z} darstellt.

Ein weiteres Beispiel für eine rekursiv definierte Folge ist die sog. *Fibonacci-Folge*⁴ $(a_n)_{n \in \mathbb{N}_0}$

$$a_0 := 0, \quad a_1 := 1, \quad a_{n+1} := a_n + a_{n-1} \quad \forall n \in \mathbb{N}$$

bei der jedes Folgenglied mittels *zweier* vorangegangener Folgenglieder erklärt wird; der Anfang der Folge lautet also 0,1,1,2,3,5,8,13,21,34,...

1.2 Monotonie, Beschränktheit, Supremum

Definition (Monotonie). Eine Folge (a_n) reeller(!) Zahlen heißt *monoton wachsend*, falls

$$\forall n \in \mathbb{N} : a_{n+1} \geq a_n$$

und *monoton fallend*, falls

$$\forall n \in \mathbb{N} : a_{n+1} \leq a_n.$$

Bemerkung: Monotonie beweist man oft, indem man zeigt, dass $a_{n+1} - a_n$ größer gleich bzw. kleiner gleich null ist, oder, bei positiven Folgen, indem man zeigt, dass $\frac{a_{n+1}}{a_n}$ größer gleich oder kleiner gleich eins ist (s.Ü.).

Definition (Beschränktheit). Eine Folge (a_n) von reellen oder komplexen Zahlen heißt *beschränkt*, wenn es eine Zahl $s \in \mathbb{R}$, genannt *Schranke*, gibt, so dass für alle $n \in \mathbb{N}$ gilt: $|a_n| \leq s$. In Formeln (Quantorschreibweise):

$$\exists s \in \mathbb{R} \forall n \in \mathbb{N} : |a_n| \leq s$$

Andernfalls heißt die Folge *unbeschränkt*.

Für Folgen von reellen(!) Zahlen sagt man, die Folge ist *nach oben beschränkt*, falls es ein $s \in \mathbb{R}$, genannt *obere Schranke*, gibt, so dass für alle $n \in \mathbb{N}$ gilt $a_n \leq s$. Sie heißt *nach unten beschränkt*, wenn es ein s , genannt *untere Schranke*, gibt, mit $a_n \geq s$ für alle Indizes n .⁵

Beispiele: Die Folge $(a_n)_{n \in \mathbb{N}}$, $a_n := \frac{1}{n}$, ist beschränkt mit Schranke $s = 1$; eine untere Schranke ist z.B. null, aber auch jede negative Zahl; eine obere Schranke ist eins, aber auch jede größere Zahl. Schranken sind offensichtlich nicht eindeutig bestimmt. Die Folge ist außerdem monoton fallend.

Die Folge (1.1) ist unbeschränkt; sie ist jedoch nach unten beschränkt, z.B. durch eins. Die komplexe Folge $a_n = (1+i)^{-n}$ ist beschränkt mit Schranke $s = \frac{1}{\sqrt{2}}$, denn

⁴benannt nach Leonardo Fibonacci, 1180-1241, der damit das Wachstum einer Kaninchenpopulation beschrieb oder zu beschreiben glaubte

⁵Für *komplexe* Folgen haben die Begriffe 'nach oben beschränkt' und 'nach unten beschränkt' offensichtlich keinen Sinn, dementsprechend gibt es da auch kein Supremum, Infimum.

$$|a_n| = (\sqrt{2})^{-n} \leq \frac{1}{\sqrt{2}} \text{ für alle } n \in \mathbb{N}.$$

Definition (Supremum, Infimum). Die kleinste obere Schranke einer reellen(!) Folge wird *Supremum* genannt, die größte untere Schranke heißt *Infimum*.⁶ Die Bezeichnungen sind 'inf' bzw. 'sup':

$$\begin{aligned} \sup_{n \in \mathbb{N}} a_n &:= \min\{s \in \mathbb{R} \mid s \text{ ist obere Schranke von } (a_n)\}, \\ \inf_{n \in \mathbb{N}} a_n &:= \max\{s \in \mathbb{R} \mid s \text{ ist untere Schranke von } (a_n)\}, \end{aligned}$$

Falls es keine obere (bzw.: keine untere) Schranke gibt (die Folge also nach oben bzw. nach unten *unbeschränkt* ist), so legt man fest, dass das Supremum $+\infty$ (bzw.: das Infimum $-\infty$) ist. Es ist immer

$$\inf_{n \in \mathbb{N}} a_n \in [-\infty, \infty), \quad \sup_{n \in \mathbb{N}} a_n \in (-\infty, \infty]$$

So wie für Folgen kann man ganz analog auch für *Mengen* von Zahlen festlegen: Eine Menge von reellen oder komplexen Zahlen $\emptyset \neq M \subseteq \mathbb{C}$ heißt beschränkt, wenn es eine Schranke $s \in \mathbb{R}$ gibt, so dass $|a| \leq s$ für alle $m \in M$ gilt. Eine Menge reeller(!) Zahlen $\emptyset \neq M \subseteq \mathbb{R}$ heißt nach oben (bzw: nach unten) beschränkt, wenn es ein $s \in \mathbb{R}$ gibt so dass $a \leq s$ (bzw.: $a \geq s$ gilt für alle $a \in M$); die kleinste obere und die größte untere Schranke heißen *Supremum* bzw. *Infimum* der Menge; man schreibt $\inf(M)$, $\sup(M)$.

Offensichtlich passen diese Begriffsbildungen für Folgen und für Mengen zusammen, in dem Sinne dass

$$\sup_{n \in \mathbb{N}} a_n = \sup\{a_n \mid n \in \mathbb{N}\};$$

analog für 'inf'.

Ist das Supremum einer Menge oder Folge von reellen Zahlen das gleiche wie das Maximum? Nicht unbedingt! Das *Maximum* einer Menge/Folge ist die größte darin enthaltene Zahl (falls eine solche existiert):

$$m = \max(M) \Leftrightarrow (m \in M \wedge \forall x \in M : m \geq x).$$

Eine Menge $(a, b) := \{x \in \mathbb{R} \mid a < x < b\}$ hat kein Maximum! Die Menge $[a, b] := \{x \in \mathbb{R} \mid a \leq x \leq b\}$ hat b als Maximum. *Beide* Mengen jedoch haben als Supremum (kleinste obere Schranke) b . Also: Falls es ein Maximum gibt, so ist es gleich dem Supremum; falls es kein Maximum gibt, so kann es trotzdem ein Supremum geben. (Analog: Minimum \leftrightarrow Infimum). Beispiele:

$$\begin{aligned} \max [a, b] &= b, & \sup [a, b] &= b \\ \max (a, b) &\text{ ex. nicht} & \sup (a, b) &= b \\ \max [a, \infty) &\text{ ex. nicht} & \sup (a, \infty) &= +\infty \end{aligned}$$

⁶Strenggenommen müsste man beweisen, dass es, sofern es überhaupt obere Schranken gibt, es eine kleinste obere Schranke gibt; analog für untere Schranken.

Das Supremum und die Vollständigkeit von \mathbb{R} . Wir hatten im 1. Semester einen Unterschied zwischen der Menge der rationalen Zahlen \mathbb{Q} (Menge aller Brüche) und der Menge \mathbb{R} festgehalten: Während die Menge \mathbb{Q} in gewisser Weise 'Lücken' hat, 'unvollständig' ist, nämlich derart, dass gewisse Zahlen wie $\sqrt{2}$ (auch $\sqrt[3]{2}$, π) nicht in \mathbb{Q} liegen, so hat \mathbb{R} solche Lücken nicht; man sagt: \mathbb{R} ist *vollständig*. Man kann diesen Begriff der Vollständigkeit mit Hilfe des Supremums präzise fassen:

Jede nichtleere Menge reeller Zahlen $\emptyset \neq M \subseteq \mathbb{R}$, die nach oben beschränkt ist, hat ein Supremum in $s \in \mathbb{R}$. ('Vollständigkeitsaxiom')

Anders in \mathbb{Q} :

Nicht jede Menge rationaler Zahlen $M \subseteq \mathbb{Q}$, die nach oben beschränkt ist, hat ein Supremum $s \in \mathbb{Q}$ (das Supremum s kann nämlich in $\mathbb{R} \setminus \mathbb{Q}$ liegen).

Beispiel (s. auch 1. Sem.): $\{x \in \mathbb{R} \mid x^2 < 2\} \subseteq \mathbb{R}$ hat das Supremum $\sqrt{2} \in \mathbb{R}$. Die Menge $\{x \in \mathbb{Q} \mid x^2 < 2\} \subseteq \mathbb{Q}$ hat das Supremum $\sqrt{2} \in \mathbb{R}$, genauer, in $\mathbb{R} \setminus \mathbb{Q}$; sie hat also kein Supremum in \mathbb{Q} .

Eine farbenfrohe Anwendung zum Thema 'Beschränktheit von Folgen'. Vielleicht haben Sie schonmal etwas von der sog. *Mandelbrot-Menge*,⁷ umgangssprachlich auch *Apfelmännchen* genannt, gehört. Bei dieser Menge handelt es sich um eine extrem komplizierte und atemberaubend schöne Struktur, deren Bildungsgesetz aber sehr einfach ist:

Für eine feste Zahl $c \in \mathbb{C}$ betrachte die rekursiv definierte Folge $z_0 := 0$, $z_{n+1} := z_n^2 + c$. Falls diese Folge beschränkt ist, so färbe den Punkt c der komplexen Ebene schwarz ein, andernfalls weiß. Um buntere Bilder zu bekommen, kann man statt weiß hier eine Farbe in Abhängigkeit davon wählen, 'wie schnell' die Werte der Zahlenfolge anwachsen, beispielsweise kann man die Farbe als Funktion von n wählen, wobei z_n das erste Folgenglied ist mit $|z_n| > 10^6$. Siehe z.B. <http://de.wikipedia.org/wiki/Mandelbrot-Menge>. Wer ein bisschen programmieren kann, kann sich solche Grafiken leicht selbst erstellen. Das Kriterium der Beschränktheit prüft man numerisch dann nur approximativ nach; falls z.B. die ersten 1000 Folgenglieder alle einen Betrag kleiner als 10^6 haben, geht man von Beschränktheit aus. Die Mandelbrot-Menge wurde erst 1978 'entdeckt', d.h. erstmals von einem Computer mit Drucker dargestellt. Eine Beschreibung mit vielen Grafiken findet man in *The beauty of fractals*, H.O. Peitgen & P. Richter, 1986. Siehe auch www.youtube.com, z.B. Suchbegriffe 'Mandelbrot Zoom'.

1.3 Konvergenz von Folgen

Für das oben bereits erwähnte Heron-Verfahren hatten wir uns im 1. Semester anschaulich überlegt, dass die so konstruierte Folge, sofern sie 'konvergiert', gegen \sqrt{z} 'konvergiert'. In diesem Semester wollen wir den Begriff der Konvergenz mathematisch präzise fassen:

⁷Benoit Mandelbrot, geb. 1924

Definition (Konvergenz, Grenzwert). Sei (a_n) eine Folge in \mathbb{C} . Eine Zahl $a \in \mathbb{C}$ heißt *Grenzwert* oder *Limes* von (a_n) , falls

$$\boxed{\forall \epsilon > 0 \exists n_0 \in \mathbb{N} \forall n \geq n_0 : |a_n - a| \leq \epsilon.} \quad \text{(Epsilon-Kriterium der Konvergenz)} \quad (1.3)$$

Falls (a_n) gegen a konvergiert, so schreibt man

$$\lim_{n \rightarrow \infty} a_n = a$$

oder auch kurz

$$\lim_n a_n = a \quad \text{oder} \quad a_n \xrightarrow{(n \rightarrow \infty)} a.$$

Falls die Folge einen Grenzwert $a \in \mathbb{C}$ hat, heißt sie *konvergent*, andernfalls *divergent*. Anschaulich/in Worten: Für jeden noch so kleinen Kreis (im Fall *reeller* Folgen ersetze man 'Kreis' durch 'Intervall') gilt: Es gibt einen Folgenindex n_0 , so dass ab diesem Index alle Folgenglieder in dem Kreis liegen. Beachte: Das n_0 darf natürlich von ϵ abhängen: $n_0 = n_0(\epsilon)$

Bsp.: Die Folge $a_n = \frac{1}{n}$ konvergiert gegen $a=0$, d.h. es ist $\lim_{n \rightarrow \infty} \frac{1}{n} = 0$.

Beweis: Es ist zu zeigen:

$$\forall \epsilon > 0 \exists n_0 \in \mathbb{N} \forall n \geq n_0 : \left| \frac{1}{n} - 0 \right| \leq \epsilon.$$

Sei dazu $\epsilon > 0$ beliebig vorgegeben. Wir müssen ein n_0 (in Abhängigkeit von ϵ) angeben, so dass für alle $n \geq n_0$ gilt: $\frac{1}{n} \leq \epsilon$. Da die Folge $\frac{1}{n}$ monoton fallend ist, reicht es dazu, dass $\frac{1}{n_0} \leq \epsilon$ ist. Die dazu hinreichende Bedingung finden wir durch Multiplikation mit n_0 und Division durch ϵ : $n_0 \geq \frac{1}{\epsilon}$.

Nachdem wir diese Überlegungen durchgeführt haben, können wir den Beweis in der folgenden Form aufschreiben:

Sei $\epsilon > 0$ beliebig vorgegeben. Sei n_0 eine natürliche Zahl größer oder gleich $\frac{1}{\epsilon}$, z.B. $n_0 := \lceil \frac{1}{\epsilon} \rceil$.⁸ Dann ist für alle $n \geq n_0$

$$|a_n - a| = \left| \frac{1}{n} - 0 \right| = \frac{1}{n} \leq \frac{1}{n_0} \leq \epsilon$$

Die Folge konvergiert also gegen null. □

Man kann leicht zeigen: Jede Folge hat höchstens einen Grenzwert.

Der Beweis dazu ist ein *Beweis durch Widerspruch* ('Indirekter Beweis'): Man nimmt an, dass es zwei verschiedene Grenzwerte a und \tilde{a} gibt, $a \neq \tilde{a}$, und mit setzen von $\epsilon := \frac{1}{3}|\tilde{a} - a|$ müssten 'fast alle' Folgenglieder sowohl im ϵ -Kreis um a als auch im ϵ -Kreis um \tilde{a} liegen, was nicht möglich ist, da die Kreisscheiben disjunkt sind. □

⁸ $\lceil x \rceil$ bezeichnet die kleinste ganze Zahl, die größer gleich x ist.

Def. (uneigentliche Grenzwerte, 'bestimmte Divergenz'). Falls für eine reelle(!) Folge (a_n) gilt

$$\forall \epsilon > 0 \exists n_0 \in \mathbb{N} \forall n \geq n_0 : a_n > \epsilon,$$

so sagen wir, die Folge (a_n) divergiert(!) gegen $+\infty$ und schreiben $\lim_n a_n = +\infty$.

Falls für eine reelle Folge (a_n) gilt

$$\forall \epsilon > 0 \exists n_0 \in \mathbb{N} \forall n \geq n_0 : a_n < -\epsilon,$$

so sagen wir, die Folge (a_n) divergiert gegen $-\infty$ und schreiben $\lim_n a_n = -\infty$.

Neben dem obigen Epsilon-Kriterium der Konvergenz gibt es noch ein weiteres, das sogenannte Cauchy-Kriterium, das zu diesem äquivalent ist:

Satz (Cauchy-Kriterium der Konvergenz).⁹ Eine Folge (a_n) von komplexen (bzw. reellen) Zahlen ist genau dann konvergent gegen ein $a \in \mathbb{C}$ (bzw. $a \in \mathbb{R}$), wenn gilt:

$$\boxed{\forall \epsilon > 0 \exists n_0 \in \mathbb{N} \forall m > n > n_0 : |a_m - a_n| \leq \epsilon}$$

Anschaulich: Der Abstand von Folgengliedern a_n und a_m wird, für n, m hinreichend groß, beliebig klein.

Vorteile/Anwendungsbereiche des Cauchy-Kriteriums: Das Interessante an dieser Charakterisierung der Konvergenz ist, dass in diesem Kriterium der Grenzwert a selbst gar nicht vorkommt! Das hat den Vorteil, dass man den Grenzwert nicht a priori erraten muss, bevor man Konvergenz zeigen kann (so wie wir im Beispiel oben als Grenzwert null geraten hatten). Auch wenn man Divergenz einer Folge mit dem Epsilon-Kriterium der Konvergenz (1.3) zeigen will, so steht man vor der Schwierigkeit, dass man für *alle* Zahlen a zeigen müsste, dass a nicht das Epsilon-Kriterium erfüllt. Verwendung des Cauchy-Kriteriums ist hier oft einfacher:

Bsp. Zeige: Die Folge $a_n = (-1)^n$ ist divergent.

Beweis: Wir verwenden das Cauchy-Kriterium der Konvergenz; genauer, wir zeigen, dass die Negation des Cauchy-Kriteriums erfüllt ist. Die *Negation des Cauchy-Kriteriums* lautet ('Umdrehen' aller Quantoren, s. 1. Sem.):

$$\exists \epsilon > 0 \forall n_0 \in \mathbb{N} \exists m > n > n_0 : |a_m - a_n| > \epsilon \quad (1.4)$$

Wir wollen zeigen, dass (1.4) gilt. Wir wählen dazu $\epsilon := 1$ ¹⁰. Sei $n_0 \in \mathbb{N}$ beliebig. Wähle $n := n_0 + 1$ und $m := n_0 + 2$. Es ist dann

$$\begin{aligned} |a_m - a_n| &= |(-1)^m - (-1)^n| = |(-1)^{n_0+2} - (-1)^{n_0+1}| = |(-1)^{n_0+1} (-1 - 1)| \\ &= |(-1)^{n_0+1}| \cdot |-2| = 2 > \epsilon \end{aligned}$$

⁹Augustin Louis Cauchy, 1789-1857, einer der wichtigsten Väter der modernen Analysis

¹⁰wir können für ϵ hier jede Zahl zwischen 0 und 2 nehmen

□

Zum Beweis des Cauchy-Kriteriums.

Wir wollen hier nur die eine Richtung der Äquivalenz zeigen, dass aus dem Erfülltsein des Epsilon-Kriterium der Konvergenz das Erfülltsein des Cauchy-Kriteriums folgt:

Sei (a_n) also eine Folge in \mathbb{C} (oder in \mathbb{R}), die das Epsilon-Kriterium der Konvergenz erfüllt, und sei $a \in \mathbb{C}$ der Grenzwert. Für jede beliebig vorgegebene Zahl $\epsilon > 0$ gibt es also ein $n_0 \in \mathbb{N}$, so dass $|a - a_n| \leq \epsilon$ für alle $n > n_0$. Nun ein kleiner Trick: Da dies für alle positiven Zahlen ϵ gilt, gilt diese Eigenschaft, wenn $\epsilon > 0$, auch für $\frac{\epsilon}{2}$! Also: Sei $\epsilon > 0$ beliebig. Dann gibt es ein $n_0 \in \mathbb{N}$, so dass für alle $n > n_0$ gilt: $|a - a_n| \leq \frac{\epsilon}{2}$. Wir wollen zeigen, dass (a_n) die Cauchy-Eigenschaft erfüllt. Seien dazu $m > n > n_0$ beliebig. Es gilt unter Verwendung der Dreiecksungleichung:

$$|a_m - a_n| \leq |a_m - a| + |a - a_n| \leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

Damit ist die Aussage bewiesen. □

Eine Anmerkung zur umgekehrten Richtung: Um zu zeigen, dass aus dem Cauchy-Kriterium das Epsilon-Kriterium folgt, verwendet man die Vollständigkeit von \mathbb{R} . (Das bedeutet insbesondere: In \mathbb{Q} ist eine entsprechende Aussage nicht wahr: Nicht jede Folge in \mathbb{Q} , die das Cauchy-Kriterium erfüllt, hat einen Grenzwert in \mathbb{Q} : Der Grenzwert kann in $\mathbb{R} \setminus \mathbb{Q}$ liegen. Bsp.: Das ist z.B. für die Heron-Folge so; deren Folgenglieder sind, sofern $a_1 \in \mathbb{Q}$, alle in \mathbb{Q} ; der Grenzwert \sqrt{z} ist i.a. nicht in \mathbb{Q} .)

Die Tatsache, dass jede Folge in \mathbb{R} , die das Cauchy-Kriterium erfüllt, einen Grenzwert in \mathbb{R} hat, aber nicht jede solche Folge in \mathbb{Q} einen Grenzwert in \mathbb{Q} hat, ist, neben der Supremum-Charakterisierung der Vollständigkeit auf S. 4 eine weitere Möglichkeit, die Vollständigkeit von \mathbb{R} zu charakterisieren; aus jeder der beiden Charakterisierungen kann man die jeweils andere Charakterisierung herleiten.

Zusammenhänge zwischen Beschränktheit, Monotonie und Konvergenz von Folgen liefern die beiden folgenden Sätze:

Satz. Jede konvergente Folge ist beschränkt.

Beweis: s.Ü.

Die Umkehrung gilt natürlich nicht (Gegenbeispiel: $a_n := (-1)^n$ ist beschränkt, aber nicht konvergent). Aber immerhin gilt:

Satz (Konvergenz monotoner beschränkter Folgen) Jede monoton wachsende, nach oben beschränkte reelle Folge ist (in \mathbb{R}) konvergent, und zwar gegen ihr Supremum:

$$\lim_{n \rightarrow \infty} a_n = \sup_{n \in \mathbb{N}} a_n$$

Jede monoton fallende, nach unten beschränkte reelle Folge ist konvergent, und zwar

gegen ihr Infimum:

$$\lim_{n \rightarrow \infty} a_n = \inf_{n \in \mathbb{N}} a_n$$

Beweis: Sei die reelle Folge (a_n) monoton wachsend und beschränkt. Dann hat sie ein Supremum $a \in \mathbb{R}$. Sei $\epsilon > 0$ beliebig vorgegeben. Da a kleinste obere Schranke von (a_n) ist, gibt es ein Folgenglied a_{n_0} , das zwischen $a - \epsilon$ und a liegt (denn andernfalls wäre $a - \epsilon$ eine obere Schranke, also a nicht die kleinste Schranke). Da (a_n) monoton wachsend ist, gilt $a_n \geq a_{n_0} \forall n \geq n_0$, also $|a_n - a| = a - a_n \leq a - a_{n_0} \leq \epsilon$. \square

Bsp. $a_n := \frac{1}{1 - \frac{1}{\sqrt{n}}}$, $n \geq 2$. Der Nenner ist monoton wachsend, also (a_n) monoton fallend; außerdem ist der Nenner kleiner als eins, also a_n größer als eins, d.h. durch 1 nach unten beschränkt. Nach dem Satz ist also die Folge konvergent. Wenn wir uns klar gemacht haben, dass 1 sogar das Infimum der Folge ist, folgt mit dem Satz, dass $\lim_n a_n = 1$.

Bsp.: (ehem. Klausuraufgabe)

$$a_0 := 0, \quad a_{n+1} := \frac{1}{2} a_n + 6$$

Wir zeigen per Vollständiger Induktion, (a) dass (a_n) beschränkt mit Schranke=12 und (b) monoton wachsend ist:

(a) Es ist $a_0 = 0 \leq 12$. Es sei $a_n \leq 12$ für ein $n \in \mathbb{N}_0$. Dann ist $a_{n+1} = \frac{1}{2} a_n + 6 \leq \frac{1}{2} \cdot 12 + 6 = 12$. Per Induktion folgt, dass die Folge nach oben beschränkt ist.

(b) Wir zeigen per Induktion, dass $a_{n+1} - a_n \geq 0$ ist für alle $n \in \mathbb{N}$:

Es ist $a_1 - a_0 = (\frac{1}{2} \cdot 0 + 6) - 0 = 6 > 0$. Unter der Annahme, dass $a_{n+1} - a_n > 0$ ist für ein $n \in \mathbb{N}$, folgt: $a_{n+2} - a_{n+1} = (\frac{1}{2} a_{n+1} + 6) - (\frac{1}{2} a_n + 6) = \frac{1}{2} \underbrace{(a_{n+1} - a_n)}_{>0} > 0$.

Per Induktion folgt die behauptete Monotonie.

Aus Monotonie und Beschränktheit folgt, dass die Folge konvergent ist. Können wir auch den Grenzwert berechnen? Leider ist nicht so ohne weiteres klar, was das Supremum der Folge ist. Stattdessen gehen wir wie folgt vor: Es sei a der Grenzwert. Auf beiden Seiten der Rekursionsgleichung $a_{n+1} = \frac{1}{2} a_n + 6$ gehen wir zum Grenzwert $n \rightarrow \infty$ über. Die linke Seite konvergiert gegen a . Die rechte Seite konvergiert gegen $\frac{1}{2} a + 6$. Es gilt also

$$a = \frac{1}{2} a + 6.$$

Daraus folgt sofort: $a = 12$

Definition (Teilfolge). Sei $(a_n)_{n \in \mathbb{N}}$ eine Folge, und sei n_1, n_2, n_3, \dots eine Folge von natürlichen Zahlen ('Indizes') mit $n_1 < n_2 < n_3 < \dots$. Dann heißt die Folge $(a_{n_k})_{k \in \mathbb{N}}$ *Teilfolge* der Folge $(a_n)_{n \in \mathbb{N}}$.

Beispiel. Die Folge $a_n := (-1)^n \cdot n$, also $-1, +2, -3, +4, \dots$ hat als Teilfolge z.B. die Folge $2, 4, 6, \dots$, aber auch die Teilfolge $-1, -5, -9, -13, \dots$ oder auch $2, 4, 8, 16, 32, \dots$, aber

4, 2, 8, 6, 12, 10, ... ist *keine* Teilfolge, da die Reihung der Folgenglieder verändert wurde.

Definition (Häufungspunkt). Sei (a_n) eine Folge in \mathbb{C} (oder \mathbb{R}). Ein $a \in \mathbb{C} \cup \{\infty\}$ (bzw. $a \in \mathbb{R} \cup \{-\infty, +\infty\}$) heißt *Häufungspunkt (HP)* von (a_n) , wenn die Folge eine Teilfolge hat, die gegen a konvergiert (bzw. im Fall $a = (\pm)\infty$: bestimmt divergiert). Im Falle der bestimmten Divergenz einer Teilfolge bezeichnet man einen HP $(\pm)\infty$ auch als *uneigentlich*.

Beispiele. Die Folge $a_n = (-1)^n$ hat zwei HP $+1$ und -1 .

Die Folge $a_n = (-1)^n \cdot (1 - \frac{1}{n})$ hat ebenfalls zwei HP ± 1 .

Die komplexe Folge $a_n = i^n$ hat vier HP $\pm i, \pm 1$.

Die Folge $a_n = \frac{(-1)^n}{n}$ hat genau einen HP 0 .

Die Folge $a_n := (1 - (-1)^n)^n$, also die Folge $2, 0, 8, 0, 32, 0, 128, \dots$ hat zwei Häufungspunkte, davon einer uneigentlich: 0 und $+\infty$.

Bemerkung. Äquivalente Charakterisierung von Häufungspunkten:

a ist genau dann HP von (a_n) , wenn in jedem ϵ -Kreis um a unendlich viele Folgenglieder liegen.

(Nämlich: Die Glieder der Teilfolge aus der Definition des HP)

Bemerkung: Dafür dass a ein HP von (a_n) ist, ist es sogar hinreichend, wenn in jedem ϵ -Kreis um a nur mindestens *ein* Folgenglied a_n mit $a_n \neq a$ liegt! Grund: Wähle immer kleinere Kreise um a ; in jedem liegt wieder ein Folgenglied; diese Folgenglieder bilden eine Teilfolge, die gegen a konvergiert (Skizze).

Definition (Limes Superior, Limes Inferior) Für eine reelle(!) Folge, die mindestens einen HP hat, heißt der größte (ggf. auch uneigentliche) HP *Limes Superior* "lim sup"; der kleinste (ggf. uneigentliche) HP heißt *Limes Inferior* "lim inf".

Beispiele:

Die Folge $a_n = (-1)^n$ hat $\limsup a_n = 1$ und $\liminf a_n = -1$.

Die Folge $a_n := (1 - (-1)^n)^n$ (s.o.) hat $\limsup a_n = +\infty$ und $\liminf a_n = 0$.

Satz (Zusammenhang Grenzwert/Häufungspunkt). Der Grenzwert einer (konvergenten) Folge ist immer auch HP (ist trivial; als Teilfolge einfach die Folge selbst nehmen). Die Umkehrung ist falsch: Ein HP braucht nicht Grenzwert zu sein (Gegenbeispiel: $a_n = (-1)^n$ hat zwei HP, die beide nicht Grenzwert sind).

Eine komplexe (reelle) Folge ist genau dann konvergent, wenn sie genau einen Häufungspunkt im \mathbb{C} (bzw. in \mathbb{R}) (d.h. $\neq \pm\infty$) hat.

Zur Begründung: Sei (a_n) eine Folge mit zwei HP $a, \tilde{a} \in \mathbb{C}$; $a \neq \tilde{a}$. Dann ist $|a - \tilde{a}| > 0$. Wir setzen $\epsilon := \frac{1}{3}|a - \tilde{a}| > 0$. Dann liegen unendlich viele Folgenglieder im ϵ -Kreis um a , und es liegen unendlich viele Folgenglieder im ϵ -Kreis um \tilde{a} . Man kann dann keinen Kreis mit Radius ϵ derart finden, dass

nur endlich viele Folgenglieder außerhalb des Kreises liegen. Konvergenz aber bedeutet insbesondere, dass nur endlich viele Folgenglieder a_1, \dots, a_{n_0-1} außerhalb eines ϵ -Kreises um den Grenzwert liegen.

Satz (Satz von Bolzano-Weierstraß,¹¹ Zusammenhang Beschränktheit/Häufungspunkte). Jede beschränkte komplexe (oder reelle) Folge hat (mindestens) einen Häufungspunkt.

Zur Begründung: Die Beweisidee ist sehr anschaulich. Sei (a_n) beschränkt. Dann kann man ein Quadrat finden (in \mathbb{C} , bzw. Intervall in \mathbb{R}) in dem alle Folgenglieder liegen. Dieses Quadrat kann man in 4 gleich große Teilquadrate teilen (2 Teilintervalle im reellen Fall); in mindestens einem der Teilquadrate müssen unendlich viele Folgenglieder liegen. Ein solches Teilquadrat, in dem unendlich viele Folgenglieder liegen, zerteilt man wieder in 4 gleich große Teilquadrate, in mindestens einem davon müssen unendlich viele Folgenglieder liegen. So bekommt man eine Folge Q_k von ineinanderliegenden Quadraten, deren Seitenlänge gegen null geht, und in jedem der Quadrate liegen unendlich viele Folgenglieder. Man wählt aus jedem dieser Quadrate Q_k ein darinliegendes Folgenglied a_{n_k} aus und erhält so eine Teilfolge $(a_{n_k})_{k \in \mathbb{N}}$ von $(a_n)_{n \in \mathbb{N}}$. Man sieht leicht ein, dass diese Cauchy-Folge ist. Somit hat sie einen Grenzwert, und dieser ist nach Definition ein HP von (a_n) . \square

Sowohl das Epsilon- als auch das Cauchy-Kriterium der Konvergenz stoßen schnell an ihre Grenzen, wenn es um die praktische Überprüfung von komplizierten Folgen geht. (beispielsweise $a_n := \frac{1+n^2}{3-4n+5n^2}$) Da sich komplizierte Folgen oft als Summen oder Produkte oder Quotienten von einfacheren Folgen auffassen lassen, sind die gleich folgenden Rechenregeln für Grenzwerte sehr nützlich. Auch schon weiter oben hatten wir einfach behauptet, dass wenn a_n gegen ein a konvergiert, dann $\frac{1}{2} a_n + 6$ gegen $\frac{1}{2} a + 6$ geht. Wieso ist das so?

Satz (wichtige Rechenregeln für konvergente Folgen). Seien (a_n) und (b_n) konvergente Folgen mit $\lim_n a_n = a$, $\lim_n b_n = b$. Dann sind auch die Folgen $(a_n + b_n)$ und $a_n b_n$ konvergent, und zwar gilt $\lim_n (a_n + b_n) = a + b$, $\lim_n a_n b_n = ab$. Sind ferner alle $b_n \neq 0$ und ist $b \neq 0$, so ist auch $\left(\frac{a_n}{b_n}\right)$ konvergent mit $\lim_n \frac{a_n}{b_n} = \frac{a}{b}$.
Im Falle der Existenz von $\lim_n a_n$ und $\lim_n b_n$ kann man also schreiben

$$\lim_n (a_n + b_n) = \lim_n a_n + \lim_n b_n, \quad \lim_n (a_n b_n) = \lim_n a_n \cdot \lim_n b_n, \quad \lim_n \frac{a_n}{b_n} = \frac{\lim_n a_n}{\lim_n b_n}$$

Beweis: Wir wollen hier die Aussage exemplarisch (nur) für ”+” zeigen:
Seien (a_n) und (b_n) konvergente Folgen mit $a = \lim_n a_n$, $b = \lim_n b_n$. Sei $\epsilon > 0$ beliebig

¹¹Karl Weierstraß, 1815-1897, einer der wichtigsten Mathematiker des 19. Jahrhunderts, soweit es um die präzise Formulierung der Grundlagen der Analysis geht.

vorgegeben. Dann gibt es ein n_0 und ein \tilde{n}_0 in \mathbb{N} , so dass ($\frac{\epsilon}{2}$ -Trick!)

$$|a - a_n| \leq \frac{\epsilon}{2} \quad \forall n \geq n_0 \quad \text{und} \quad |b - b_n| \leq \frac{\epsilon}{2} \quad \forall n \geq \tilde{n}_0.$$

Es folgt für alle $n \geq N_0 := \max\{n_0, \tilde{n}_0\}$:

$$|(a+b) - (a_n+b_n)| = |(a-a_n) + (b-b_n)| \leq |a-a_n| + |b-b_n| \leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

□

Ergänzung zum Satz: Die Rechenregeln gelten zum Teil auch dann, wenn eine der beiden Folgen oder beide bestimmt divergent ist/sind. So gilt also: Ist $\lim_n a_n = +\infty$ und $\lim_n b_n = +\infty$, so ist $(a_n+b_n)_n$ bestimmt divergent und es gilt $\lim_n a_n+b_n = \infty$. Kurz (symbolisch) kann man dies schreiben als " $\infty+c=\infty$ ". In einem solchen Sinne gelten auch die entsprechenden Regel " $\infty+\infty=\infty$ ", " $\infty \cdot (\pm\infty) = \pm\infty$ ", " $\infty \cdot c = \text{sgn}(c) \cdot \infty$ " für $c \neq 0$, " $\frac{c}{\infty} = 0$ ". Aber: Über Grenzwerte vom Typ " $\infty-\infty$ ", " $\frac{\infty}{\infty}$ " kann man keine allgemeinen Aussagen treffen; erst durch Umformungen kann man diese Fälle oft auf die Rechenregeln zurückführen:

Anwendungsbeispiel für den Fall " $\frac{\infty}{\infty}$ ", d.h. genauer: $\lim_n \frac{a_n}{b_n}$, wobei $\lim_n a_n = \infty$, $\lim_n b_n = \infty$:

$$\begin{aligned} \lim_n \frac{1+2n^2}{3-4n+5n^2} &= \lim_n \frac{n^2(\frac{1}{n^2}+2)}{n^2(\frac{3}{n^2}-\frac{4}{n}+5)} = \lim_n \frac{\frac{1}{n^2}+2}{\frac{3}{n^2}-\frac{4}{n}+5} \stackrel{(*)}{=} \frac{\lim_n(\frac{1}{n^2}+2)}{\lim_n(\frac{3}{n^2}-\frac{4}{n}+5)} \\ &\stackrel{(*)}{=} \frac{\lim_n \frac{1}{n^2} + \lim_n 2}{\lim_n \frac{3}{n^2} - \lim_n \frac{4}{n} + \lim_n 5} = \frac{0+2}{0+0+5} = \frac{2}{5}, \end{aligned}$$

dabei wurde an den Stellen (*) die Grenzwert-Rechenregeln verwendet.

Anwendungsbeispiel für den Fall " $\infty-\infty$ ", d.h. genauer: $\lim_n a_n-b_n$, wobei $\lim_n a_n = \infty$, $\lim_n b_n = \infty$:

$$\begin{aligned} \lim_n \sqrt{n+2} - \sqrt{n+1} &= \lim_n \frac{(\sqrt{n+2}-\sqrt{n+1})(\sqrt{n+2}+\sqrt{n+1})}{\sqrt{n+2}+\sqrt{n+1}} \\ &= \lim_n \frac{(n+2)-(n+1)}{\sqrt{n+2}+\sqrt{n+1}} = \lim_n \frac{1}{\sqrt{n+2}+\sqrt{n+1}} = 0 \end{aligned}$$

wobei im letzten Schritt die Rechenregel für den Fall " $\frac{c}{\infty}$ " verwendet wurde. Die Vorgehensweise beruht darauf, dass " $\infty+\infty$ " weit weniger 'schlimm' ist als " $\infty-\infty$ ", denn es gibt die Rechenregel " $\infty+\infty=\infty$ ".

Eine weitere Anwendung/Folgerung der Rechenregeln: Aus obigem Satz folgt: Wenn (a_n) eine konvergente Folge ist und $\alpha \in \mathbb{C}$ beliebige Zahl, so ist auch (αa_n) konvergent, und $\lim_n \alpha a_n = \alpha \lim_n a_n$. (Ist Spezialfall vom Satz mit konstanter Folge $(b)_n$.)

Es gilt also auch eine *Linearität des Grenzwertprozesses*: Sind (a_n) und (b_n) konvergente Folgen, so ist auch die Folge $(\alpha a_n + \beta b_n)$ konvergent, und es gilt

$$\lim_n (\alpha a_n + \beta b_n) = \lim_n \alpha a_n + \lim_n \beta b_n = \alpha \lim_n a_n + \beta \lim_n b_n$$

Satz (Rechenregel für konvergente Folgen). Sei (a_n) eine Folge, die gegen null konvergiert (eine solche Folge nennt man auch *Nullfolge*). Sei (b_n) eine beschränkte Folge, dann ist auch $(a_n b_n)$ konvergent, und es ist $\lim_n a_n b_n = 0$.

Beweis: Sei $b \in \mathbb{R}$ eine Schranke von (b_n) . O.B.d.A. sei $b > 0$. Sei $\epsilon > 0$ beliebig. Da (a_n) gegen null konvergiert, gibt es ein n_0 , so dass für alle $n \geq n_0$ gilt: $|a_n - 0| = |a_n| \leq \frac{\epsilon}{b}$. (Beachte: Begründung für das $\frac{\epsilon}{b}$ wie beim Epsilon-Halbe-Trick). Es folgt

$$|a_n b_n - 0| = |a_n| \cdot |b_n| \leq \frac{\epsilon}{b} \cdot b = \epsilon$$

für alle $n \geq n_0$. Es folgt die Behauptung □.

Anwendungsbeispiel: $a_n = \frac{\sin(\sqrt{n})}{n} = \frac{1}{n} \cdot \sin(\sqrt{n})$ konvergiert gegen null, denn $\frac{1}{n}$ konvergiert gegen null und $\sin(\sqrt{n})$ ist (durch 1) beschränkt.

Satz ("Einschachtelung durch konvergente Folgen", "Vergleichskriterium"). Seien (a_n) und (c_n) konvergente reelle(!) Folgen mit dem gleichen Grenzwert $\lim_n a_n = b = \lim_n c_n$. Sei (b_n) eine Folge mit der Eigenschaft

$$a_n \leq b_n \leq c_n \quad \forall n \in \mathbb{N}.$$

Dann ist auch (b_n) konvergent mit $\lim_n b_n = b$.

Ist (a_n) eine Folge mit $\lim_n a_n = +\infty$ und $b_n \geq a_n$ für alle $n \in \mathbb{N}$, so ist auch $\lim_n b_n = +\infty$. Analog: Ist (a_n) Folge mit $\lim_n a_n = -\infty$ und $b_n \leq a_n$ für alle $n \in \mathbb{N}$, so ist auch $\lim_n b_n = -\infty$.

Die Aussagen leuchtet anschaulich ein; auf einen Beweis verzichten wir.

Beispiel: Untersuche das Konvergenzverhalten von $a_n := \frac{n + \sin(\cos(n^2))}{n+1}$. Wir schachteln diese recht komplizierte Folge durch zwei einfachere Folgen ein: Es ist

$$\frac{n-1}{n+1} \leq a_n \leq \frac{n+1}{n+1} = 1$$

(Einschachtelung), und $\lim_n \frac{n-1}{n+1} = \lim_n \frac{n(1-\frac{1}{n})}{n(1+\frac{1}{n})} = \lim_n \frac{1-\frac{1}{n}}{1+\frac{1}{n}} = 1$, also nach obigem Satz ist (a_n) konvergent und es ist $\lim_n a_n = 1$.

Weiteres Beispiel: In der Übung wird mit dem Einschachtelungskriterium gezeigt: Für $q \in \mathbb{R}$ mit $|q| < 1$ ist $\lim_n q^n = 0$.

Noch einige elementare/wichtige Grenzwerte:

$$(a) \quad \lim_n n^\alpha = \begin{cases} \infty & , \alpha > 0 \\ 1 & , \alpha = 0 \\ 0 & , \alpha < 0 \end{cases} \quad (b) \quad \lim_n \alpha^n = \begin{cases} \infty & , \alpha > 1 \\ 1 & , \alpha = 1 \\ 0 & , 0 \leq \alpha < 1 \\ \text{s.Ü.} & , \alpha < 0 \end{cases}$$

$$(c) \quad \lim_n \sqrt[n]{n} = 1, \quad (d) \quad \lim_n \sqrt[n]{n!} = \infty.$$

Beachte: Die Aussage (a) folgt im Fall $\alpha \in \mathbb{Z}$ direkt aus den Rechenregeln.

Aussage (d) wird in der Übung gezeigt.

Beweis der Aussage (c) (beachte: Leider ist keine unserer Rechenregeln auf $\sqrt[n]{n}$ anwendbar!):

Wir betrachten die Hilfsfolge $x_n := \sqrt[n]{n} - 1 \geq 0$ und wollen zeigen, dass $x_n \xrightarrow{(n \rightarrow \infty)} 0$.

Es ist offensichtlich $(1+x_n)^n = n$.

Ferner, unter Verwendung der Binomischen Formel (1. Sem.) unter Berücksichtigung der Nichtnegativität der Summanden:

$$n = (1+x_n)^n = \sum_{k=0}^n \binom{n}{k} 1^{n-k} x_n^k \geq \binom{n}{0} 1^{n-0} x_n^0 + \binom{n}{2} 1^{n-2} x_n^2 = 1 + \frac{n(n-1)}{2} x_n^2$$

Es folgt $n \geq 1 + \frac{n(n-1)}{2} x_n^2$, somit $\frac{2}{n} \geq x_n^2 \geq 0$. Mit dem Einschachtelungskriterium folgt, dass $\lim_n x_n^2 = 0$. Mit Rechenregel also $0 = (\lim_n x_n)^2$, also $0 = \lim_n x_n$. \square

Allgemeine Schlussbemerkung: Bei der Prüfung einer Folge auf Konvergenz sollte man nach Möglichkeit versuchen, mit Hilfe der Rechenregeln zu einem Schluss zu kommen, d.h. mit ihrer Hilfe die Berechnung von Grenzwerten auf einfachere/bekanntere Grenzwerte zurückführen, und nur im 'Notfall', wenn keine Rechenregel greift, zum "Epsilon" greifen. Die Bedeutung der Epsilon-Kriterien liegt hauptsächlich darin, dass wir mit ihrer Hilfe die obigen Rechenregeln herleiten konnten. Außerdem werden wir die Epsilon-Kriterien auch im folgenden Kapitel über *Reihen* verwenden müssen, um praktikablere Konvergenzkriterien für Reihen herzuleiten.

2 Reihen

Sei $(a_n)_{n \in \mathbb{N}}$ eine Folge. Dann nennt man $s_n := \sum_{k=1}^n a_k$ ¹² die n -te Partialsumme der Folge. Die Partialsummen bilden offensichtlich ebenfalls eine Folge, (s_n) , also die Folge $s_1 = a_1, s_2 = a_1 + a_2, s_3 = a_1 + a_2 + a_3, \dots$

Die Folge von Partialsummen wird *Reihe* genannt.

Eine Reihe ist also nicht anderes als eine Folge mit einer speziellen Struktur. Aus jeder Folge $(a_n)_{n \in \mathbb{N}}$ kann man also eine Reihe $(s_n)_{n \in \mathbb{N}}$ machen, indem man s_n als die n -te Partialsumme setzt, also die Summe der ersten n Folgenglieder.

Bsp. (Geometrische Reihe). Wir betrachten, für ein vorgegebenes $q \in \mathbb{C}$, die Folge $a_n := q^n, n \in \mathbb{N}_0$; also die Folge $q^0, q^1, q^2, q^3, \dots$. Die zugehörige Reihe (Folge der Partialsummen) ist dann die (aus dem ersten Semester bekannte) sog. *Geometrische Reihe*

$s_n = \sum_{k=0}^n q^k$. Im ersten Semester haben wir berechnet:

$$s_n = \sum_{k=0}^n q^k = \frac{1 - q^{n+1}}{1 - q}, \quad (2.1)$$

sofern $q \neq 1$ (Summenformel der geometrischen Reihe).

Da es sich bei Reihen letztendlich um Folgen handelt, liegt es nahe, auch Reihen auf Konvergenz zu untersuchen. Zunächst zur Schreibweise: Ist eine Reihe $s_n = \sum_{k=1}^n a_k$ konvergent, so bezeichnet man den Grenzwert mit

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

Leider trennt man, was die Schreibweise angeht, nicht immer sauber zwischen der Reihe und dem Grenzwert der Reihe; die eigentlich korrekte Schreibweise für eine Reihe, $(s_n)_{n \in \mathbb{N}}$, also $\left(\sum_{k=1}^n a_k \right)_{n \in \mathbb{N}}$, findet selten Verwendung; stattdessen bezeichnet man die Reihe selbst i.a. auch mit $\sum_{k=1}^{\infty} a_k$. Aus dem Zusammenhang geht hervor, ob die Reihe oder ihr Grenzwert gemeint ist.¹³

¹²oder auch $s_n := \sum_{k=0}^n a_k$, falls $(a_n)_{n \in \mathbb{N}_0}$

¹³Kommt dieser Ausdruck in irgendwelchen *Gleichungen* vor, so kann man davon ausgehen, dass der Grenzwert gemeint ist; kommt dagegen in einem Text oder Aufgabenstellung " ... die Reihe $\sum_{k=1}^{\infty} a_k$..." vor, so ist offensichtlich die Reihe gemeint.

Bsp. (Grenzwert der Geometrischen Reihe). Nach den Rechenregeln für Folgen ist unter Verwendung von (2.1)

$$\sum_{k=0}^{\infty} q^k = \lim_n \sum_{k=0}^n q^k = \lim_n \frac{1 - q^{n+1}}{1 - q} = \frac{\lim_n 1 - q^{n+1}}{1 - q} = \frac{1 - \lim_n q^{n+1}}{1 - q},$$

sofern $\lim_n q^n$ existiert. Im Fall $|q| < 1$ existiert dieser Grenzwert; es ist $\lim_n q^n = 0$, also:

$$\sum_{k=0}^{\infty} q^k = (\text{s.o.}) = \frac{1 - \lim_n q^n}{1 - q} = \frac{1}{1 - q}, \quad |q| < 1.$$

Man kann zeigen, dass für alle $q \in \mathbb{C}$ mit $|q| \geq 1$ die geometrische Reihe divergent ist.

Also z.B. ist $\sum_{k=0}^{\infty} \left(\frac{1}{2}\right)^k = \frac{1}{1 - \frac{1}{2}} = 2$, und $\sum_{k=0}^{\infty} (-3)^k$ existiert nicht.

Historische Anmerkung/Motivation. Vor Beginn des 19. Jahrhunderts hat man generationenlang diskutiert, ob denn die folgende "Summation von unendlich vielen Zahlen" $\sum_{k=0}^{\infty} (-1)^k = +1 - 1 + 1 - 1 + 1 - 1 \dots$ gleich null oder gleich eins ist. Argumente lauteten: Bei Klammerung $(1 - 1) + (1 - 1) + (1 - 1) + \dots = 0 + 0 + 0 + \dots$ kommt null heraus. Bei Klammerung $1 + (-1 + 1) + (-1 + 1) + \dots = 1 + 0 + 0 + \dots$ kommt eins heraus. Noch schlimmer: Bald bemerkte man: Wenn man die Summanden *umordnet* zu $+1 + 1 - 1 + 1 + 1 - 1 + 1 + 1 - 1 + 1 + 1 - 1 + \dots$, so bekommt man $(2 - 1) + (2 - 1) + (2 - 1) + \dots = 1 + 1 + \dots = \infty$! Wir können dazu folgende präzise Aussage machen: Wir kennen keine Summation von unendlich vielen Zahlen. Was wir kennen, sind Reihen. Der Grenzwert einer Reihe ist als Grenzwert der n -ten Partialsummen festgelegt, d.h. wir legen damit eine Reihenfolge fest, in der die Summanden addiert werden; umordnen der Reihenglieder ist i.a. nicht erlaubt. Die Reihe $\sum_{k=0}^{\infty} (-1)^k$ ist divergent, (denn) sie hat zwei verschiedenen Häufungspunkte 1 und 0. Das sieht man daran, dass für ungerades n die n -te Partialsumme $s_n = (1 - 1) + (1 - 1) + \dots + (1 - 1) = 0 + 0 \dots + 0 = 0$ ist, und für gerades n die n -te Partialsumme $s_n = (1 - 1) + (1 - 1) + \dots + (1 - 1) + 1 = 0 + 0 \dots + 0 + 1 = 1$ ist; es gibt also jeweils unendlich viele Folgenglieder s_n die in jedem beliebigen ϵ -Kreis um 1 bzw. um 0 liegen.

Man soll an diesem Beispiel sehen: Es ist ziemlich sinnlos, unendlich viele Zahlen 'gleichzeitig' addieren zu wollen; das Ergebnis kann je nach Klammerung oder Umsortierung der Summanden ziemlich beliebig werden. Man kann immer nur *endlich* viele Zahlen addieren. Eine sinnvolle mathematische Vorgehensweise ist, nur n viele Summanden zu addieren, und zwar bei vorher festgelegter Reihenfolge der Glieder, und dann zu prüfen, ob es einen Grenzwert für $n \rightarrow \infty$ gibt. Dies motiviert, warum man Reihen als Folgen der Partialsummen definiert hat.

2.1 Konvergenzkriterien für Reihen

Da Reihen nichts anderes als Folgen von Partialsummen sind, kann man natürlich versuchen, alles, was für Folgen gilt, auch für Reihen zu formulieren. So bekommen

wir einige ad-hoc-Konvergenzkriterien für Reihen:

Das **Epsilon-Kriterium** der Konvergenz **für Reihen**:

Eine Reihe $\sum_{k=1}^{\infty} a_k$ ist genau dann konvergent gegen eine Zahl $a \in \mathbb{C}$, wenn

$$\forall \epsilon > 0 \exists n_0 \in \mathbb{N} \forall n > n_0 : \left| a - \sum_{k=1}^n a_k \right| \leq \epsilon$$

Dieses Kriterium ist in der Praxis i.a. wenig nützlich (denn kaum jemals kann man das passende a erraten).

Das **Cauchy-Kriterium** der Konvergenz **für Reihen** lautet offensichtlich (indem wir das Cauchy-Kriterium für Folgen auf die Folge der Partialsummen $(s_n)_{n \in \mathbb{N}}$ anwenden):

$$\forall \epsilon > 0 \exists n_0 \in \mathbb{N} \forall m > n > n_0 : \left| \sum_{k=1}^m a_k - \sum_{k=1}^n a_k \right| \leq \epsilon,$$

was man kürzer schreiben kann als

$$\forall \epsilon > 0 \exists n_0 \in \mathbb{N} \forall m > n > n_0 : \left| \sum_{k=n+1}^m a_k \right| \leq \epsilon,$$

Auch die **Rechenregel für die Summe (und Linearkombination)**¹⁴ von konvergenten Folgen lassen sich auf Reihen anwenden: Sind $\sum_{k=1}^n a_k$ und $\sum_{k=1}^n b_k$ konvergente Reihen, so ist auch die Reihe über die Folgenglieder $\alpha a_k + \beta b_k$ konvergent, und es gilt für den Grenzwert:

$$\sum_{k=1}^{\infty} (\alpha a_k + \beta b_k) = \alpha \sum_{k=1}^{\infty} a_k + \beta \sum_{k=1}^{\infty} b_k.$$

Nun noch ein **Divergenz-Kriterium** für Reihen:

Satz. Sei $\sum_{k=0}^{\infty} a_k$ eine konvergente Reihe. Dann ist die Folge (a_n) eine Nullfolge.

Dies ist ein Divergenzkriterium für Reihen, denn: Ist (a_n) nicht gegen null konvergent, so muss die Reihe $\sum_{k=1}^{\infty} a_k$ divergent sein.

Beweis des Satzes: Sei die Reihe konvergent. Dann erfüllt sie das Cauchy-Kriterium der Konvergenz. Im Cauchy-Kriterium wählen wir speziell $m := n+1$ und bekommen:

$$\forall \epsilon > 0 \exists n_0 \in \mathbb{N} \forall n > n_0 : |a_{n+1}| < \epsilon$$

¹⁴Um das *Produkt* von zwei Reihen kümmern wir uns später, siehe: Cauchy-Produkt

Mit $\tilde{n}_0 := n_0 + 1$ wird also das Epsilon-Kriterium der Konvergenz gegen null erfüllt. \square

Wichtig: Die Umkehrung des obigen Kriteriums gilt nicht:

Aus $\lim_n a_n = 0$ folgt i.a. *nicht*, dass $\sum_{k=1}^{\infty} a_k$ konvergent ist.

Dazu gibt es das folgende berühmte

Gegenbeispiel (Harmonische Reihe). Obwohl $\lim_n \frac{1}{n} = 0$ ist, ist die Reihe $\sum_{k=1}^{\infty} \frac{1}{k}$ divergent.

Beweis: Wir zeigen, dass das Cauchy-Kriterium nicht erfüllt wird, dass also die Negation des Cauchy-Kriteriums für Reihen erfüllt wird:

Setze $\epsilon := \frac{1}{4}$. Sei $n_0 \in \mathbb{N}$ beliebig vorgegeben. Sei n eine Zweierpotenz, die größer als n_0 ist: $n = 2^{\tilde{n}}$. Sei $m = 2^{\tilde{n}+1}$. Dann ist

$$s_m - s_n = \sum_{k=n+1}^m \frac{1}{k} = \frac{1}{n+1} + \dots + \frac{1}{m} \geq (m-n) \cdot \frac{1}{m} = (2 \cdot 2^{\tilde{n}} - 2^{\tilde{n}}) \cdot \frac{1}{2 \cdot 2^{\tilde{n}}} = \frac{1}{2} > \epsilon$$

\square

Bemerkung zur Idee obigen Beweises: Im Beweis wird die Idee verwendet, jeweils 'Pakete' zu packen aus den Summanden $\frac{1}{k}$, wobei k von einer Zweierpotenz bis zur nächsten Zweierpotenz läuft; jedes solches Paket ist $\geq \frac{1}{2}$. Anschaulich kann man dementsprechend argumentieren, $\sum_{k=1}^{\infty} \frac{1}{k}$ wächst über alle Schranken (ist bestimmt divergent gegen ∞), da

$$\sum_{k=1}^{\infty} \frac{1}{k} = 1 + \frac{1}{2} + \underbrace{\left(\frac{1}{3} + \frac{1}{4}\right)}_{\geq 2 \cdot \frac{1}{4} = \frac{1}{2}} + \underbrace{\left(\frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8}\right)}_{\geq 4 \cdot \frac{1}{8} = \frac{1}{2}} + \underbrace{\left(\frac{1}{9} + \dots + \frac{1}{16}\right)}_{\geq 8 \cdot \frac{1}{16} = \frac{1}{2}} + \dots$$

Bemerkung: Die Harmonische Reihe divergiert gegen unendlich, jedoch extrem langsam. So ist z.B. $s_{1.000.000} \approx 14,39272672$.

Bemerkung: Neben obiger Rechnung kann man, im Vorgriff auf spätere Kapitel bzw. im Rückgriff auf Schulwissen, die bestimmte Divergenz der Harmonischen Reihe auch anders zeigen: Es ist offensichtlich (Skizze) $\sum_{k=1}^n \frac{1}{k} \geq \int_1^{n+1} \frac{1}{x} dx = \ln x \Big|_1^{n+1} = \ln(n+1) \xrightarrow{(n \rightarrow \infty)} \infty$

Bevor wir zu weiteren Konvergenz-/Divergenzkriterien kommen, hier eine Definition eines 'verschärften' Konvergenzbegriffes für Reihen:

Def. (Absolute Konvergenz). Eine Reihe $\sum_{k=1}^{\infty} a_k$ heißt *absolut konvergent*, wenn die Reihe über die Beträge, also $\sum_{k=1}^{\infty} |a_k|$, konvergent ist.

Man kann leicht einsehen, dass aus absoluter Konvergenz die Konvergenz folgt:

Sei die Reihe $\sum_{k=1}^{\infty} a_k$ absolut konvergent, also $\sum_{k=1}^{\infty} |a_k|$ ist konvergent. Sei $\epsilon > 0$ beliebig. Nach dem Cauchy-Kriterium für die Reihe $\sum_{k=1}^{\infty} |a_k|$ gibt es ein n_0 , so dass für alle $m > n > n_0$ gilt $||a_{n+1}| + \dots + |a_m|| \leq \epsilon$. Mit der Dreiecksungleichung folgt $|a_{n+1} + \dots + a_m| \leq |a_{n+1}| + \dots + |a_m| \leq \epsilon$, d.h. die Reihe $\sum_{k=1}^{\infty} a_k$ erfüllt das Cauchy-Kriterium der Konvergenz. \square

Nun weitere Kriterien (Vergleichskriterien für Reihen):

Satz (Minoranten-Kriterium). Sei $\sum_{k=1}^{\infty} a_k$ eine reelle Reihe und $a_k \geq b_k \geq 0$, und sei die reelle Reihe $\sum_{k=1}^{\infty} b_k$ divergent. Dann ist auch $\sum_{k=1}^{\infty} a_k$ divergent. (Man nennt die Reihe $\sum_{k=1}^{\infty} b_k$ eine *divergente Minorante* der Reihe $\sum_{k=1}^{\infty} a_k$.)

Satz (Majoranten-Kriterium). Sei $\sum_{k=1}^{\infty} a_k$ eine (komplexe oder reelle) Reihe und $|a_k| \leq b_k$, und sei $\sum_{k=1}^{\infty} b_k$ konvergent (da die $b_k \geq 0$ sind: sogar absolut konvergent). Dann ist auch die Reihe $\sum_{k=1}^{\infty} a_k$ absolut konvergent. (Man nennt die Reihe $\sum_{k=1}^{\infty} b_k$ eine *konvergente Majorante* der Reihe $\sum_{k=1}^{\infty} a_k$.)

Zum Beweis (wir beschränken uns auf das Majoranten-Kriterium): Da die Reihe $\sum_{k=1}^{\infty} b_k$ konvergent ist, gilt nach dem Cauchy-Kriterium für jedes $\epsilon > 0$: Es gibt ein n_0 , so dass für alle $m > n > n_0$ gilt: $b_{n+1} + \dots + b_m \leq \epsilon$. Da $|a_k| \leq b_k \forall k$ gilt, folgt $|a_{n+1}| + \dots + |a_m| \leq b_{n+1} + \dots + b_m \leq \epsilon$, also ist die Reihe $\sum_{k=1}^{\infty} |a_k|$ konvergent. \square

Die bisher genannten Konvergenzkriterien sind meist recht umständlich anzuwenden; die Epsilon-Kriterien sowieso, und die obigen Vergleichskriterien erfordern Geschick bei der Konstruktion einer geeigneten Minorante/Majorante. Um den praktischen Nachweis von Konvergenz oder Divergenz einer Reihe einfacher führen zu können, leiten wir daher weitere Kriterien her, die in der Praxis leicht anwendbar sind: Das Wurzel- sowie das Quotientenkriterium. Als Hilfsmittel zu deren Herleitung verwenden wir das Majorantenkriterium sowie unser Wissen über die Konvergenz der Geometrischen Reihe; wir benutzen nämlich die geometrische Reihe als (konvergente) Majorante:

Satz (Wurzelkriterium, 1. Version).¹⁵ (a) Die Reihe $\sum_{k=0}^{\infty} a_k$ ist absolut konvergent, falls es eine Zahl $q \in (0, 1)$ gibt und ein $n_0 \in \mathbb{N}$ so dass

$$\sqrt[n]{|a_n|} \leq q \quad \forall n \geq n_0$$

gilt. (b) Falls

$$\sqrt[n]{|a_n|} \geq 1 \quad \forall n \geq n_0,$$

so ist die Reihe divergent.

Zum Beweis: Die Bedingung schreiben wir in der Form $|a_n| \leq q^n$. (a) Für $q \in (0, 1)$ ist $\sum_{k=0}^{\infty} q^n$ konvergente Majorante der gegebenen Reihe. (b) Für $q = 1$ ist wegen $\sqrt[n]{|a_n|} \geq 1$ auch $|a_n| \geq 1$, somit kann (a_n) keine Nullfolge sein, und somit muss die Reihe nach dem Divergenzkriterium divergent sein. \square

Satz (Wurzelkriterium, 2. Version). (a) Die Reihe $\sum_{k=0}^{\infty} a_k$ ist absolut konvergent, falls

$$\limsup \sqrt[n]{|a_n|} < 1$$

gilt. (b) Falls

$$\limsup \sqrt[n]{|a_n|} > 1$$

so ist die Reihe divergent. Falls $\lim_n \sqrt[n]{|a_n|}$ existiert, so kann man in den obigen Bedingungen "lim sup" durch "lim" ersetzen.

Beweisidee: Die (a)-Teilaussage der 2. Version lässt sich auf die 1. Version unmittelbar zurückführen. Für die (b)-Teilaussage überlegt man sich, dass die Existenz eines Häufungspunktes größer als eins unendlich viele Folgenglieder größer als eins zur Folge hat, was der für die Konvergenz der Reihe notwendige Bedingung $\lim_n a_n = 0$ widerspricht. \square

Beachte:

- Falls die Folge $\sqrt[n]{|a_n|}$ konvergent ist oder bestimmt divergent gegen unendlich, so ist $\limsup \sqrt[n]{|a_n|} = \lim_n \sqrt[n]{|a_n|}$, d.h. im Kriterium lässt sich dann "lim sup" durch "lim" ersetzen. Man kann also knapp auch schreiben

$$\begin{aligned} \lim_{n \rightarrow \infty} \sqrt[n]{|a_n|} > 1 &\Rightarrow \sum_{k=1}^{\infty} a_k \text{ divergent} \\ \lim_{n \rightarrow \infty} \sqrt[n]{|a_n|} < 1 &\Rightarrow \sum_{k=1}^{\infty} a_k \text{ (abs.) konvergent} \end{aligned}$$

Die Formulierung mit "lim sup" ist allerdings etwas allgemeiner. Die Formulierung mit "lim" reicht jedoch oft aus.

¹⁵Es geht zurück auf A. Cauchy und J. Hadamard (1865-1963).

- Zur 1. Formulierung; eine 'beliebte' Fehlerquelle ist:
Es reicht *nicht* aus, wenn $\sqrt[n]{|a_n|} < 1$ ist, um auf Konvergenz der Reihe zu schließen. Aus dieser Eigenschaft folgt nämlich i.a. nicht, dass es ein q gibt mit $\sqrt[n]{|a_n|} \leq q < 1 \forall n$.
Gegenbeispiel: $a_n := \left(1 - \frac{1}{n}\right)^n \Rightarrow \sqrt[n]{|a_n|} = 1 - \frac{1}{n}$ ist echt kleiner als 1 für alle n , aber trotzdem gibt es kein q , welches sich zwischen 1 und die Terme $1 - \frac{1}{n}$ schieben lässt, da $1 - \frac{1}{n}$ der Zahl 1 beliebig nahe kommt. Und in der Tat: Man kann zeigen (späteres Kapitel), dass $a_n := \left(1 - \frac{1}{n}\right)^n \xrightarrow{(n \rightarrow \infty)} \frac{1}{e} \neq 0$, woraus mit dem Divergenzkriterium folgt, dass die Reihe divergent ist.
- zur 2. Formulierung: Der Fall $\limsup \sqrt[n]{|a_n|} = 1$ ist (mit diesem Kriterium) unentscheidbar. In diesem Fall ist, falls alle $\sqrt[n]{|a_n|} \geq 1$ sind (man sich also von oben der eins nähert), der Divergenz-Fall des 1. Kriteriums anwendbar; falls es jedoch (unendlich viele) n gibt mit $\sqrt[n]{|a_n|} < 1$ (man sich von unten der eins nähert), ist mit beiden Versionen des Wurzelkriteriums keine Entscheidung möglich.

Anwendungsbeispiel: Untersuche das Konvergenzverhalten der Reihe $\sum_{k=0}^{\infty} \frac{1}{k!}$.

Es ist $\lim_n \sqrt[n]{|a_n|} = \lim_n \frac{1}{\sqrt[n]{n!}} = 0$, denn $\lim_{n \rightarrow \infty} \sqrt[n]{n!} = \infty$. Die Reihe ist also konvergent.

Auf diesem Konvergenzresultat aufbauend definieren wir:

Def. (Euler'sche Zahl) Die Zahl

$$e := \sum_{k=0}^{\infty} \frac{1}{k!}$$

heißt *Euler'sche Zahl*.¹⁶

Bemerkung: Die Euler'sche Zahl ist irrational. Ihr Zahlenwert ist näherungsweise $e \approx 2.7182818284590452355\dots$. Sie ist, neben π , eine der wichtigsten Konstanten der Mathematik. Wir werden Sie bei der Definition der Exponentialfunktion wiedertreffen.

Satz (Quotientenkriterium).¹⁷ (a) Die Reihe $\sum_{k=0}^{\infty} a_k$ ist absolut konvergent, falls es eine Zahl $q \in (0, 1)$ gibt und ein $n_0 \in \mathbb{N}$ so dass $a_n \neq 0 \forall n \geq n_0$ und

$$\left| \frac{a_{n+1}}{a_n} \right| \leq q \quad \forall n \geq n_0$$

¹⁶Leonhard Euler, 1707-1783.

Die Reihe konvergiert recht schnell, da $\frac{1}{n!}$ schnell klein wird; man kann also mit dem Taschenrechner/Computer unter Verwendung von recht wenigen Reihengliedern eine gute Näherung für e berechnen.

¹⁷Das Kriterium geht zurück auf J.-B. d'Alembert (1717-1783).

gilt. Äquivalent zu dieser Bedingung ist die Bedingung

$$\limsup \left| \frac{a_{n+1}}{a_n} \right| < 1.$$

(b) Falls

$$\left| \frac{a_{n+1}}{a_n} \right| \geq 1 \quad \forall n \geq n_0,$$

so ist die Reihe divergent.

Im Fall $\limsup \left| \frac{a_{n+1}}{a_n} \right| > 1$ ist, anders als beim Wurzelkriterium, keine sichere Aussage möglich. Was jedoch reicht ist: Aus $\lim_{n \rightarrow \infty} \left| \frac{a_{n+1}}{a_n} \right| > 1$ folgt Divergenz der Reihe, oder, noch etwas schwächer, aus $\liminf \left| \frac{a_{n+1}}{a_n} \right| > 1$ folgt Divergenz der Reihe.

Jedenfalls kann man auch hier im Fall der Existenz eines Grenzwertes (auch uneigentlich) kurz formulieren:

$$\begin{aligned} \lim_{n \rightarrow \infty} \left| \frac{a_{n+1}}{a_n} \right| > 1 &\Rightarrow \sum_{k=1}^{\infty} a_k \text{ divergent} \\ \lim_{n \rightarrow \infty} \left| \frac{a_{n+1}}{a_n} \right| < 1 &\Rightarrow \sum_{k=1}^{\infty} a_k \text{ (abs.) konvergent} \end{aligned} \quad (2.2)$$

Die obige Formulierung aus dem Quotientenkriterium ist jedoch etwas allgemeiner als (2.2).

Anschaulich: Die Quotientenbedingung quantifiziert, ob die Reihenglieder 'hinreichend schnell', nämlich so schnell wie eine geometrische Reihe, abklingt. Dies liefert die Beweisidee für die erste Aussage des Satzes: Nämlich zu zeigen, dass die geometrische Reihe eine konvergente Majorante ist. Die Divergenzaussage zeigt man so: Aus der angegebenen Bedingung folgt $|a_{n+1}| \geq |a_n|$, d.h. die Folge $(|a_n|)$ ist ab $a \geq n_0$ monoton wachsend. Damit ist $\lim_n |a_n| \neq 0$, woraus man $\lim_n a_n \neq 0$ folgern kann. Daraus folgt die Divergenz der Reihe. \square

Anwendungsbeispiel: Untersuche die Reihe $\sum_{k=1}^{\infty} \frac{(2k)!}{(k!)^2}$ auf Konvergenz.

Es ist

$$\left| \frac{a_{n+1}}{a_n} \right| = \frac{(2n+2)! \cdot n!^2}{(n+1)!^2 \cdot (2n)!} = \frac{(2n+1) \cdot (2n+2)}{(n+1)^2} = \frac{n^2 \overbrace{\left(2 + \frac{1}{n}\right)}^{\rightarrow 2} \overbrace{\left(2 + \frac{2}{n}\right)}^{\rightarrow 2}}{n^2 \underbrace{\left(1 + \frac{1}{n}\right)^2}_{\rightarrow 1}} \xrightarrow{(n \rightarrow \infty)} 4$$

Da dieser Term gegen $4 > 1$ konvergiert, ist die Reihe, nach dem Zusatz (2.2) zum Quotientenkriterium, divergent; oder, ohne Zusatz direkt mit dem Quotientenkriterium argumentiert, für alle n ab einem hinreichend großen n_0 sind die Terme $\left| \frac{a_{n+1}}{a_n} \right|$, da sie gegen 4 gehen, größer als eins (z.B. $\epsilon = 3$ oder kleiner wählen). Nach dem

Quotientenkriterium ist die Reihe also divergent.

Und nun noch ein letztes Konvergenzkriterium. Dieses liefert, anders als Wurzel- und Quotientenkriterium, nur die Konvergenz, nicht die absolute Konvergenz. Es betrifft alternierende Reihen, also Reihen, deren Glieder abwechselnd positiv und negativ sind:

Satz (Leibniz-Kriterium) Seien alle $a_k \geq 0$ (Nichtnegativität)¹⁸ mit $a_k \geq a_{k+1}$ (Monotonie) und $\lim_k a_k = 0$ (Nullfolge). Dann ist die alternierende Reihe $\sum_{k=0}^{\infty} (-1)^k a_k$ konvergent.

Anwendungsbeispiel (Alternierende Harmonische Reihe): Die Reihe $\sum_{k=1}^{\infty} (-1)^k \frac{1}{k}$ ist konvergent.

Grund: Die Folge $a_k = \frac{1}{k}$ ist nichtnegativ, monoton fallend, und Nullfolge, also liefert das Leibniz-Kriterium die Konvergenz der Reihe.¹⁹ Die Alternierende Harmonische Reihe ist also ein Beispiel für eine Reihe, die konvergent, aber nicht absolut konvergent ist: Sie kann nicht absolut konvergent sein, da wir uns von der Divergenz von $\sum_{k=1}^{\infty} \left| \frac{(-1)^k}{k} \right| = \sum_{k=1}^{\infty} \frac{1}{k}$ bereits überzeugt haben.

Beweis des Satzes: Wir verwenden das Cauchy-Kriterium für Reihen, um die Konvergenz zu zeigen. Wir müssen also zeigen, dass $s_m - s_n$ (für $m > n$) 'klein' wird für 'große' m, n . Dazu betrachten wir $s_m - s_n$ in zwei verschiedene Klammerungen:²⁰ (Wir beschränken uns hier auf den Fall, dass $m - n$ eine *gerade* Zahl ist; der Fall, dass $m - n$ ungerade ist, kann ähnlich behandelt werden.)

Erste Klammerung:

$$\begin{aligned} (-1)^{n+1} (s_m - s_n) &= (-1)^{n+1} \sum_{k=n+1}^m (-1)^k a_k = a_{n+1} - a_{n+2} + a_{n+3} - \dots - a_m \\ &= \underbrace{(a_{n+1} - a_{n+2})}_{\geq 0 \text{ wg. Monot.}} + \underbrace{(a_{n+3} - a_{n+4})}_{\geq 0 \text{ wg. Monot.}} + \dots + \underbrace{(a_{m-1} - a_m)}_{\geq 0 \text{ wg. Monot.}} \end{aligned}$$

Zweite Klammerung:

$$(-1)^{n+1} (s_m - s_n) = a_{n+1} - \underbrace{(a_{n+2} - a_{n+3})}_{\geq 0 \text{ wg. Monot.}} - \underbrace{(a_{n+4} - a_{n+5})}_{\geq 0 \text{ wg. Monot.}} - \dots - \underbrace{a_m}_{\geq 0 \text{ wg. Pos.}}$$

Aus der ersten Klammerung schließen wir, dass $(-1)^{n+1} (s_m - s_n) \geq 0$. Aus der zweiten Klammerung schließen wir, dass $(-1)^{n+1} (s_m - s_n) \leq a_{n+1} \leq a_n$. Insgesamt ist also

¹⁸Die Nichtnegativität kann man hier auch weglassen, da sie sich automatisch aus den beiden anderen Voraussetzungen ergibt.

¹⁹Man kann, unter Zuhilfenahme von Methoden aus der Differential- und Integralrechnung ('Satz von Taylor', siehe auch (3.18)) zeigen, dass der Grenzwert $\sum_{k=1}^{\infty} (-1)^k \frac{1}{k} = -\ln 2$ ist.

²⁰In *endlichen* Summen darf man natürlich beliebig klammern.

$$|s_m - s_n| \leq a_n.$$

Für beliebig vorgegebenes $\epsilon > 0$ können wir, da (a_n) Nullfolge ist, ein n_0 finden, so dass $a_{n_0} \leq \epsilon$. Mit obiger Abschätzung ist dann auch $|s_m - s_n| \leq a_n \leq a_{n_0} \leq \epsilon \forall m > n > n_0$, wobei die Monotonie nochmal verwendet wurde. \square

Bemerkung über die Konvergenzkriterien für Reihen: Die am einfachsten anzuwendenden Kriterien, um Konvergenz oder Divergenz von Reihen zu zeigen, sind das Divergenz-, das Wurzel- und das Quotientenkriterium; bei alternierenden Reihen auch das Leibniz-Kriterium. Wenn diese Kriterien keine Entscheidung liefern, sollte man es mit Majoranten-/Minoranten- oder Cauchy-Kriterium versuchen. Auch wenn man erkennt, dass die gegebene Reihe sich nur wenig von einer Reihe, deren Konvergenzverhalten man kennt (z.B. die Bauart $\sum_k (1/k)^\alpha$), unterscheidet, bietet es sich an zu probieren, diese Reihe als konvergente Majorante bzw. divergente Minorante zu benutzen.

Anders als bei der Konvergenzuntersuchung von *Folgen* liefert die Konvergenzuntersuchung von Reihen mit Hilfe der in diesem Kapitel dargestellten Methoden nicht den konkreten Grenzwert, sondern nur *die Existenz* eines Grenzwertes. Die Ermittlung eines konkreten Zahlenwertes für den Grenzwert ist i.a. für Reihen schwieriger als für Folgen. Um für einige Reihen die Grenzwerte konkret zu bestimmen, hilft manchmal der folgende Satz, der sich mit der Multiplikation von Reihen beschäftigt. Wir wollen, für gegebene absolut konvergente Reihen $\sum_{k=0}^{\infty} a_k$ und $\sum_{k=0}^{\infty} b_k$ eine Formel für das Produkt der Grenzwerte $(\sum_{k=0}^{\infty} a_k) \cdot (\sum_{k=0}^{\infty} b_k)$ herleiten, und zwar soll das Ergebnis die Form einer *Reihe* haben. Um zu einer Vermutung zu kommen, schauen wir uns an, was wir bei *endlichen* Summen $\sum_{k=0}^n a_k$ und $\sum_{k=0}^n b_k$ bekommen. In diesem Fall müssen wir jeden Summanden des einen Terms mit jedem Summanden des anderen Terms multiplizieren und die Produkte $a_j b_k$ aufaddieren:

$$\left(\sum_{k=0}^n a_k \right) \left(\sum_{k=0}^n b_k \right) = \sum_{k=0}^n \sum_{j=0}^n a_j b_k.$$

Wenn wir hier einfach n durch ∞ ersetzen, dann hätten wir auf der rechten Seite eine Reihe, deren Glieder wiederum Reihen sind; das ist unangenehm (unpraktisch). Auf der rechten Seite müssen wir über alle Terme $a_j b_k$ summieren, wobei $j, k = 0, \dots, \infty$. Stellen wir uns die Terme $a_j b_k$ als Einträge einer (unendlich großen) Matrix (=Tabelle) vor, so können wir die Verschachtelung von zwei unendlichen Summationen vermeiden, indem wir die Summationsreihenfolge ändern: Wir addieren, anstatt entlang von Zeilen oder Spalten, entlang von *Diagonalen*, und summieren anschließend über alle Diagonalen. Das ergibt:

Satz (Cauchy-Produkt). Seien $\sum_{k=0}^{\infty} a_k$ und $\sum_{k=0}^{\infty} b_k$ absolut konvergente Reihen. Dann ist die Reihe $\sum_{n=0}^{\infty} \sum_{k=0}^n a_k b_{n-k}$ ebenfalls absolut konvergent, und es gilt die folgende Gleichheit der Grenzwerte:

$$\left(\sum_{k=0}^{\infty} a_k \right) \cdot \left(\sum_{k=0}^{\infty} b_k \right) = \sum_{n=0}^{\infty} \left(\sum_{k=0}^n a_k b_{n-k} \right) \quad (2.3)$$

Auf einen strengen Beweis wollen wir verzichten.

Anwendungsbeispiel: Berechne das Cauchy-Produkt der geometrischen Reihe $\sum_{k=0}^{\infty} q^k$ für $|q| < 1$ mit sich selbst: Wir wenden den Satz vom Cauchy-Produkt an mit $a_k = q^k$ und $b_k = q^k$. Die innere Summation auf der rechten Seite von (2.3) lautet also

$$\sum_{k=0}^n a_k b_{n-k} = \sum_{k=0}^n q^k q^{n-k} = \sum_{k=0}^n q^n = (n+1) q^n$$

Es folgt also mit (2.3):

$$\left(\sum_{k=0}^{\infty} q^k \right) \cdot \left(\sum_{k=0}^{\infty} q^k \right) = \sum_{n=0}^{\infty} (n+1) q^n,$$

wobei die Reihe auf der rechten Seite absolut konvergent ist. Wir haben somit für eine Reihe, deren Grenzwert wir bisher nicht kannten, den Grenzwert berechnet: Es ist

$$\sum_{n=0}^{\infty} (n+1) q^n = \left(\sum_{k=0}^{\infty} q^k \right) \cdot \left(\sum_{k=0}^{\infty} q^k \right) = \frac{1}{1-q} \cdot \frac{1}{1-q} = \frac{1}{(1-q)^2}.$$

Daraus kann man auch noch mit einer der üblichen Rechenregeln den Grenzwert

$$\sum_{n=0}^{\infty} n q^n = \sum_{n=0}^{\infty} (n+1) q^n - \sum_{n=0}^{\infty} 1 \cdot q^n = \frac{1}{(1-q)^2} - \frac{1}{1-q}$$

bekommen. Auf diese Art und Weise hilft der Satz über das Cauchy-Produkt, aus bekannten Grenzwerten gewisser Reihen Grenzwerte für gewisse andere Reihen zu berechnen. Indem man von den Reihen $\sum_{k=0}^{\infty} k q^k$ und $\sum_{k=0}^{\infty} q^k$ das Cauchy-Produkt bildet, kann man auf ähnliche Weise den Grenzwert von $\sum_{k=0}^{\infty} k^2 q^k$ berechnen, etc. Eine weitere Anwendung für das Cauchy-Produkt ist der Beweis der Formel (2.6).

Zum Schluss noch der sog. *Umordnungssatz für Reihen*. Dieser macht die Bedeutung der *absoluten* Konvergenz von Reihen klar. Insbesondere der zweite Teil des Satzes ist erstaunlich!

Satz (Umordnungssatz für Reihen). Für eine absolut konvergente Reihe ändert sich der Grenzwert nicht, wenn man die Reihenglieder permutiert (vertauscht) oder wenn man mehrere Reihengliedern mittels Klammerung zusammenfasst. Bei einer Reihe, die nicht absolut konvergent ist, kann man durch Umordnung der Reihenglieder jeden beliebigen Grenzwert oder auch Divergenz der Reihe erreichen.

Bemerkung. Vergleiche die Betrachtung der Reihe $\sum_{k=0}^{\infty} (-1)^k$ vom Beginn dieses Kapitels, wo wir durch Umordnung der Reihenglieder oder Klammersetzungen verschiedenes Konvergenzverhalten produzieren konnten (Umordnung der Glieder oder Klammerung macht aus einer Reihe eine *andere* Reihe). Die dort betrachtete Reihe war divergent. Der Umordnungssatz sagt, dass die gleichen Effekte auftreten bei Reihen, die zwar konvergent, aber nicht absolut konvergent sind. Nur für absolut konvergente Reihen dagegen ist die Reihenfolge der Summation oder das Zusammenfassen von mehreren Reihengliedern ('Klammerung') beliebig erlaubt, sagt der Umordnungssatz.

2.2 Potenzreihen: Konvergenzradius

Def. (Potenzreihe). Eine Potenzreihe ist eine Reihe, die von einem Parameter $x \in \mathbb{R}$ oder $x \in \mathbb{C}$ abhängt, und zwar in der folgenden Form: Sei $x_0 \in \mathbb{C}$ 'fest' vorgegeben und $x \in \mathbb{C}$ 'variabel'. Sei (a_k) eine Folge in \mathbb{C} . Dann nennt man die von x abhängende Reihe

$$\sum_{k=0}^{\infty} a_k (x - x_0)^k$$

eine *Potenzreihe*. Für diejenigen $x \in \mathbb{C}$, für die die Reihe konvergiert, kann man

$$\sum_{k=0}^{\infty} a_k (x - x_0)^k =: P(x)$$

setzen. Dies definiert eine Funktion $P : D \rightarrow \mathbb{C}$, wobei $D \subseteq \mathbb{C}$ die Menge aller x ist, für die die Potenzreihe konvergiert.

Bsp. Für $x_0 := 0$ und $a_k := \frac{1}{k!}$ bekommen wir die Potenzreihe $P(x) = \sum_{k=0}^{\infty} \frac{x^k}{k!}$, $x \in D$. Wichtigste Frage wird für uns sein, wie groß der Konvergenzbereich $D \subseteq \mathbb{C}$ ist.

Anmerkungen.

- Die Menge D kann ganz \mathbb{C} umfassen, sie kann eine echte Teilmenge von \mathbb{C} sein, und sie kann sogar nur aus $D = \{x_0\}$ bestehen.
- Man kann " \mathbb{C} " durch " \mathbb{R} " ersetzen, d.h. wenn die a_k reelle Zahlen sind und $x_0 \in \mathbb{R}$, kann man (muss man aber nicht) die Potenzreihe auch als Abbildung $P : D \rightarrow \mathbb{R}$, wobei $D \subseteq \mathbb{R}$, ist, betrachten.

- Die n -te Partialsumme s_n einer Potenzreihe ist ein *Polynom* vom Grad $\leq n$
- Die große Bedeutung, die Potenzreihen in der Analysis haben, rührt im wesentlichen daher:
Viele wichtige Funktionen, wie z.B. \sin , \cos , \exp , \ln , lassen sich als Potenzreihen (d.h.: als Grenzwert/Grenzfunktion einer Potenzreihe) schreiben. Wenn wir also allgemeine Sätze finden, die etwas über Potenzreihen aussagen (Stetigkeit, Differenzierbarkeit,...), so liefern diese Sätze Aussagen über viele wichtige Funktionen. Außerdem, im Vorgriff auf Zukünftiges: Wir werden uns (später) davon überzeugen, dass man Potenzreihen 'gliedweise' ableiten und integrieren darf. Wenn sich also eine Funktion als Potenzreihe schreiben lässt, dann können wir durch gliedweises Ableiten die Ableitung oder Stammfunktion dieser Funktion leicht berechnen; gliedweises Ableiten/Integrieren ist leicht, da man nur Terme der Form $a_k (x-x_0)^k$ zu betrachten hat!

Unsere erste Aufgabe wird sein, herauszufinden, für welche $x \in \mathbb{C}$ (oder $x \in \mathbb{R}$) eine Potenzreihe konvergiert. Es stellt sich heraus, dass der Konvergenzbereich D immer ein *Kreis* um den Punkt x_0 ist (mit Radius $r \in [0, \infty]$); im reellen Fall also ein Intervall mit Randpunkten $x_0 - r$, $x_0 + r$:

Satz (Konvergenzbereich einer Potenzreihe) und Definition (Konvergenzradius einer Potenzreihe).

Sei $\sum_{k=0}^{\infty} a_k (x-x_0)^k$ eine Potenzreihe. Dann gibt es eine Zahl (Radius) $R \in [0, \infty]$, so dass die Potenzreihe

- (a) für alle $x \in K_R := \{x \in \mathbb{C} \mid |x-x_0| < R\}$ die Potenzreihe konvergent (sogar: absolut konvergent) ist, und
- (b) für alle $x \in \mathbb{C}$ mit $|x-x_0| > R$ divergent ist.

Die Zahl R heißt *Konvergenzradius* der Potenzreihe.

Zur Ausdrucksweise: Mit der Schreibweise " $R = \infty$ " wollen wir ausdrücken, dass die Potenzreihe für alle $x \in \mathbb{C}$ konvergiert.

Für den Konvergenzradius R gilt:

$$R = \begin{cases} \frac{1}{\limsup \sqrt[n]{|a_n|}} & , \text{ falls } \limsup \sqrt[n]{|a_n|} \in (0, \infty) \\ \infty & , \text{ falls } \limsup \sqrt[n]{|a_n|} = 0 \\ 0 & , \text{ falls } \limsup \sqrt[n]{|a_n|} = \infty \end{cases}$$

Das können wir kurz (und nicht ganz mathematisch präzise) zusammenfassen zu:

$$\boxed{R = \frac{1}{\limsup \sqrt[n]{|a_n|}}}$$

sofern wir zur Vereinfachung der Schreibweise die Vereinbarung " $\frac{1}{0} = \infty$ " und " $\frac{1}{\infty} = 0$ " zulassen.

Anmerkung: Über das Verhalten *auf dem Rande* der Kreisscheibe/des Intervalls ($|x-x_0|=R$) kann man keine allgemeinen Aussagen treffen; dort ist sowohl Konvergenz als auch Divergenz möglich.

Beweis des Satzes. Wir brauchen nichts weiter zu tun, als das Wurzelkriterium auf die Potenzreihe anzuwenden: Laut Wurzelkriterium ist die Reihe konvergent, falls $\limsup \sqrt[k]{|a_k(x-x_0)^k|} < 1$, und divergent, wenn dieser Ausdruck > 1 ist. Dies ist gleichbedeutend damit, dass $|x-x_0| \cdot \limsup \sqrt[k]{|a_k|}$ kleiner bzw. größer eins ist. Division durch $\limsup \sqrt[k]{|a_k|}$ liefert die Behauptung. \square

Anwendungsbeispiele. Berechne den Konvergenzradius der Potenzreihen

$$(1.) \sum_{k=0}^{\infty} \frac{1}{k^2} x^k \quad (2.) \sum_{k=0}^{\infty} \frac{1}{k!} x^k \quad (3.) \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k)!} x^{2k} \quad (4.) \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k+1)!} x^{2k+1}$$

zu 1.: $\sqrt[k]{\frac{1}{k^2}} = \frac{1}{(\sqrt[k]{k})^2}$

Dies konvergiert gegen 1, da $\lim_k \sqrt[k]{k} = 1$. Also ist auch der Limes Superior gleich eins. Kehrwertbildung: $R = \frac{1}{1} = 1$

zu 2.: $\sqrt[k]{\frac{1}{k!}} = \frac{1}{\sqrt[k]{k!}}$

Dies konvergiert gegen 0, da $\lim_k \sqrt[k]{k!} = \infty$. Also ist auch der Limes Superior gleich null. Es folgt $R = \infty$, d.h. die Potenzreihe konvergiert für alle $x \in \mathbb{C}$.

zu 3.: Auch diese Reihe können wir als Potenzreihe auffassen, bei der allerdings die ungeraden Potenzen 'fehlen', d.h. den Koeffizienten $=0$ haben:

$$|a_k| = \begin{cases} 0 & , k \text{ ungerade} \\ \frac{1}{k!} & , k \text{ gerade} \end{cases}$$

Wir suchen die Häufungspunkte der Folge $\sqrt[k]{|a_k|}$. Die Teilfolge der Folgenglieder mit ungeraden Indizes konvergiert offenbar gegen null. Die Teilfolge der Folgenglieder mit geraden Indizes konvergiert ebenfalls gegen 0. Es ist also $\limsup \sqrt[n]{|a_n|} = 0$, also $R = \infty$.

zu 4: Analog zu 3. haben wir hier

$$|a_k| = \begin{cases} \frac{1}{k!} & , k \text{ ungerade} \\ 0 & , k \text{ gerade} \end{cases} ,$$

und wir erhalten auch hier $R = \infty$.

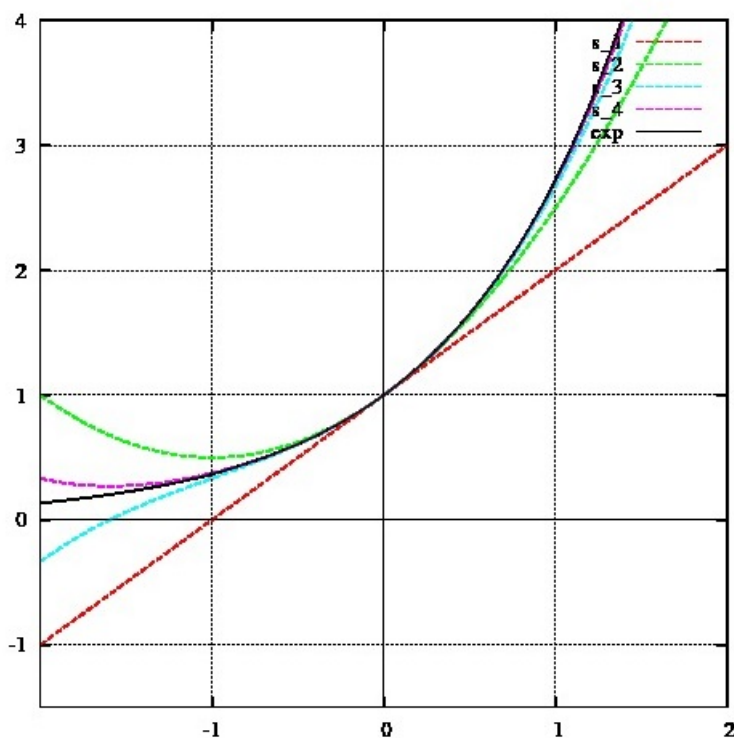


Abbildung 1: Die (reelle) Exponentialreihe.

Fazit: Die Potenzreihen 2.-4. definieren also Grenzfunktionen *auf ganz* \mathbb{C} , $P : \mathbb{C} \rightarrow \mathbb{C}$; wir benutzen sie, um **die Exponential-, die Kosinus- und die Sinus-Funktion**

$$\exp, \cos, \sin : \mathbb{C} \rightarrow \mathbb{C}$$

zu **definieren**:

$$\boxed{\exp(x) := \sum_{k=0}^{\infty} \frac{1}{k!} x^k}, \quad \boxed{\cos(x) := \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k)!} x^{2k}}, \quad \boxed{\sin(x) := \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k+1)!} x^{2k+1}} \quad (2.4)$$

Für *reelle* x bekommt man, da die Koeffizienten a_k reell sind, offensichtlich reelle Funktionswerte, d.h. durch Einschränkung der obigen Funktionen auf $\mathbb{R} \subset \mathbb{C}$ bekommt man die Funktionen

$$\exp, \cos, \sin : \mathbb{R} \rightarrow \mathbb{R}. \quad (2.5)$$

Man spricht daher auch, je nachdem, welchen Definitionsbereich man meint, z.B. von der *komplexen* oder der *reellen* Exponential-/Kosinus-/Sinus-Funktion.

Dass obige Funktionen tatsächlich die uns aus der Schule bekannten (dort, zumindest für \sin, \cos , vermutlich über geometrische Sachverhalte (rechtwinklige Dreiecke) definierte) Funktionen sind, ist zunächst mal völlig unklar. Dass dem so ist, kann man zunächst mal plausibel machen, indem man sich die zugehörige Partialsummen $s_n(x)$ visualisiert (s. Abb. 1-2). Eine genauere Untersuchung, dass die in (2.4)-(2.5) definierten Funktionen die aus der Schule bekannten Funktionen sind (bzw. deren Eigenschaften) haben, erfolgt in Kap. 2.3; in der Tat kann man aus der Sinus- und

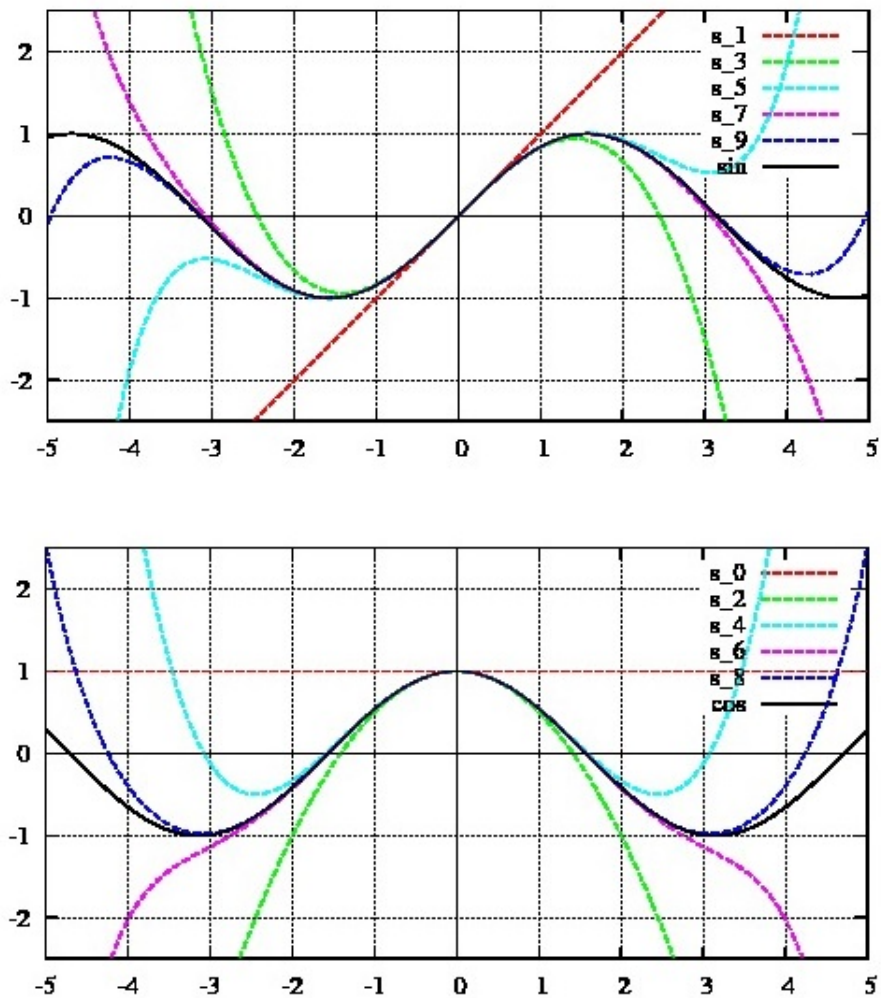


Abbildung 2: Die Sinusreihe (oben) und die Kosinusreihe (unten).

Kosinus-Definition als Reihe alle relevanten Eigenschaften von Sinus und Kosinus relativ leicht herleiten, wie wir im folgenden Kapitel sehen werden.

2.3 Einige Eigenschaften der Exponentialfunktion und von Sinus und Kosinus

Man kann an obiger Definition (2.4), zusammen mit der Definition der Euler'schen Zahl e , sofort ablesen:

$$\exp(0) = 1, \quad \exp(1) = e, \quad \cos(0) = 1, \quad \sin(0) = 0,$$

$$\cos(-x) = \cos(x), \quad \sin(-x) = -\sin(x),$$

Eine Funktion mit der Eigenschaft $f(-x) = f(x) \forall x \in \mathbb{R}$ heißt *gerade Funktion*, eine Funktion mit $f(-x) = -f(x) \forall x \in \mathbb{R}$ heißt *ungerade Funktion*.

Ganz zentral ist das **Additionstheorem der Exponentialfunktion** (auch genannt: **Funktionalgleichung der Exponentialfunktion**):

$$\boxed{\exp(x+y) = \exp(x) \cdot \exp(y) \quad \forall x, y \in \mathbb{C}} \quad (2.6)$$

Beweis. Wir verwenden den Satz über das Cauchy-Produkt, um die rechte Seite $(\sum_{k=0}^{\infty} \frac{x^k}{k!}) \cdot (\sum_{k=0}^{\infty} \frac{y^k}{k!})$ auszurechnen. Wir haben $a_k = \frac{x^k}{k!}$ und $b_k = \frac{y^k}{k!}$, also

$$\sum_{k=0}^n a_k b_{n-k} = \sum_{k=0}^n \frac{1}{k!(n-k)!} x^k y^{n-k} = \frac{1}{n!} \sum_{k=0}^n \underbrace{\frac{n!}{k!(n-k)!}}_{=\binom{n}{k}} x^k y^{n-k} = \frac{1}{n!} (x+y)^n$$

wobei im letzten Schritt die *Binomische Formel* aus dem 1. Semester verwendet wurde. Nach dem Satz über das Cauchy-Produkt folgt somit die Gleichheit $(\sum_{k=0}^{\infty} \frac{x^k}{k!}) \cdot (\sum_{k=0}^{\infty} \frac{y^k}{k!}) = \sum_{n=0}^{\infty} \frac{(x+y)^n}{n!}$, also (2.6). \square

Ganz analog, ebenfalls mit Hilfe des Cauchy-Produkts kann man zeigen, dass

$$\boxed{\sin^2 x + \cos^2 x = 1} \quad (2.7)$$

ist für alle $x \in \mathbb{C}$ (dabei steht $\sin^\alpha x$ für $((\sin(x))^\alpha)$; diese Rechnung erfordert allerdings einige technische Tricks, wie z.B. Indextransformationen. Viel einfacher ist es (2.7) auf die Euler'sche Formel (2.10) zurückzuführen (dazu einfach (2.10) auf x und auf $-x$ anwenden, und beide Gleichungen miteinander multiplizieren).

Aus (2.7) folgen nun sofort für reelle x die Schranken $\sin^2 x, \cos^2 x \in [0, 1]$, also

$$\sin(x), \cos(x) \in [-1, 1] \quad \forall x \in \mathbb{R}.$$

exp(x) als Potenz von e geschrieben. Eine direkte Folgerung aus der Funktionalgleichung ist $\exp(x+x) = \exp(x) \cdot \exp(x)$, also $\exp(2x) = \exp(x)^2$. Durch Iterieren (eigentlich: Vollständige Induktion) folgt, dass

$$\exp(nx) = \exp(x)^n \quad \forall n \in \mathbb{N}_0, x \in \mathbb{C} \quad (2.8)$$

ist. Für $n \in \mathbb{N}_0$ folgt mit $x := 1$: $\exp(n) = \exp(1)^n = e^n$, d.h. wir haben die *Funktionsauswertung* $\exp(n)$ als *Potenz* e^n geschrieben, sofern $n \in \mathbb{N}_0$. Gilt dies auch für andere Exponenten außerhalb von \mathbb{N} ? Dazu:

Es ist $\underbrace{\exp(x + (-x))}_{=\exp(0)=1} = \exp(x) \cdot \exp(-x)$, woraus insbesondere folgt, dass $\exp(x) \neq 0$

ist für alle $x \in \mathbb{C}$. Wir können also dividieren:

$$\exp(-x) = \frac{1}{\exp(x)},$$

somit $\exp(-n) = \frac{1}{\exp(n)} = \frac{1}{e^n} = e^{-n} \quad \forall n \in \mathbb{N}$. Insgesamt also:

$$\exp(n) = e^n \quad \forall n \in \underline{\underline{\mathbb{Z}}}$$

Gilt diese Formel auch für nicht-ganzzahlige (reelle) n ? Dazu rechnen wir zunächst für $q \in \mathbb{N}$

$$e = \exp(1) = \exp\left(\underbrace{q}_{\in \mathbb{N}} \cdot \frac{1}{q}\right) \stackrel{(2.8)}{=} \exp\left(\frac{1}{q}\right)^q \quad |\sqrt[q]{}$$

$$\Rightarrow \exp\left(\frac{1}{q}\right) = \sqrt[q]{e} =: e^{\frac{1}{q}}.$$

Nun nehmen wir als Exponenten einen Bruch $\frac{p}{q}$ und erhalten

$$\exp\left(\frac{p}{q}\right) = \exp\left(\underbrace{p}_{\in \mathbb{N}} \cdot \underbrace{\frac{1}{q}}_{=:x}\right) = \exp\left(\frac{1}{q}\right)^p \stackrel{(2.8)}{=} \left(e^{\frac{1}{q}}\right)^p = e^{\frac{p}{q}}$$

Wir haben also

$$\exp(x) = e^x \quad \forall x \in \underline{\underline{\mathbb{Q}}}$$

gezeigt, wobei $e := \exp(1)$ definiert war (Euler'sche Zahl).

Für *irrationale* $x \in \mathbb{R} \setminus \mathbb{Q}$ definieren(!) wir einfach die Potenz zur Basis e als

$$e^x := \exp(x). \tag{2.9}$$

(Zur Definition der Potenz bei anderen Basen kommen wir noch.) Auch für komplexe, nicht-reelle Zahlen definieren wir die Potenz zur Basis e als

$$e^z := \exp(z) \quad \forall z \in \mathbb{C};$$

obige Funktionalgleichung (2.6) für \exp können wir somit als Potenzgesetz

$$e^{x+y} = e^x \cdot e^y \quad \forall x, y \in \mathbb{C}$$

formulieren.

Es gibt einen Zusammenhang zwischen der komplexen Exponentialfunktion und Sinus und Kosinus. Um den herzuleiten, berechnen wir $\exp(ix)$, wobei $x \in \mathbb{R}$:

$$e^{ix} = \exp(ix) = \sum_{k=0}^{\infty} \frac{i^k x^k}{k!}$$

Wir betrachten nun in dieser Summe jeweils vier aufeinanderfolgende Folgenglieder:

$$\Rightarrow e^{ix} = \sum_{\substack{k=0, \dots, \infty \\ k \text{ durch } 4 \text{ teilbar}}} \left(\frac{x^k}{k!} + i \frac{x^{k+1}}{(k+1)!} - \frac{x^{k+2}}{(k+2)!} - i \frac{x^{k+3}}{(k+3)!} \right)$$

Wir fassen die Terme mit geradem Exponenten sowie die Terme mit ungeraden Exponenten jeweils zusammen:²¹

$$\Rightarrow e^{ix} = \sum_{\substack{k=0, \dots, \infty \\ k \text{ durch } 4 \text{ teilbar}}} \frac{x^k}{k!} - \frac{x^{k+2}}{(k+2)!} + i \sum_{\substack{k=0, \dots, \infty \\ k \text{ durch } 4 \text{ teilbar}}} \frac{x^{k+1}}{(k+1)!} - \frac{x^{k+3}}{(k+3)!}$$

Die erste der beiden Summationen kann man schreiben als $\sum_{n=0}^{\infty} (-1)^n \frac{x^{2n}}{(2n)!}$, also als $\cos(x)$.

Die hintere Summation kann man schreiben als $\sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{(2n+1)!}$, also als $\sin(x)$. Wir erhalten somit die sog. *Euler'sche Formel*

$$\boxed{e^{ix} = \cos(x) + i \sin(x)} \quad \forall x \in \mathbb{R}. \quad (2.10)$$

Für den Betrag folgt:

$$|e^{ix}| = |\cos x + i \sin x| = \sqrt{\cos^2 x + \sin^2 x} = 1 \quad \forall x \in \mathbb{R}.$$

d.h. der Punkt e^{ix} , für $x \in \mathbb{R}$, befindet sich immer auf dem Einheitskreis (Skizze!).

Wir wollen nun noch e^z für beliebige $z \in \mathbb{C}$ berechnen. Dazu zerlegen wir z in Real- und Imaginärteil $z = a + bi$, $a, b \in \mathbb{R}$, und nutzen das Potenzgesetz und (2.10):

$$\boxed{e^{a+bi} = e^a \cdot e^{bi} = e^a (\cos b + i \sin b)} \quad (2.11)$$

Diese Formeln erinnern uns (hoffentlich) an die **Polardarstellung von komplexen Zahlen** aus dem 1. Semester: Wir hatten eine beliebige Komplexe Zahl z mittels ihres Betrags $r = |z|$ und ihres Arguments (=Winkels zur positiven reellen Achse) $\varphi = \arg(z)$ dargestellt als

$$z = |z| (\cos(\arg z) + i \sin(\arg z)) \quad \text{bzw.} \quad z = r \cdot (\cos \varphi + i \sin \varphi). \quad (2.12)$$

(2.12) können wir kürzer schreiben unter Verwendung von Formel (2.10) als

$$\boxed{z = |z| e^{i \arg z}} \quad \text{bzw.} \quad \boxed{z = r e^{i\varphi}} \quad \forall z \in \mathbb{C}.$$

Beispiele.

Die Zahl $z = 1 + i$ hat Argument $\varphi = \frac{\pi}{4}$ und Betrag $r = \sqrt{2}$. Also ist

$$1 + i = \sqrt{2} (\cos \frac{\pi}{4} + i \sin \frac{\pi}{4}) = \sqrt{2} e^{i\frac{\pi}{4}}.$$

Die Zahl -1 hat Argument $\varphi = \pi$ und Betrag $r = 1$. Es ist also

$$e^{\pi i} = (-1) \cdot (\cos \pi + i \sin \pi) = -1.$$

²¹Wir dürfen nach dem Umordnungssatz klammern und umordnen, da die Exponentialreihe *absolut* konvergent ist.

(Weitere) Anwendungen der Euler'schen Formel (2.10) zur Untersuchung von Sinus und Kosinus.

Man definiert nun π als die kleinste echt positive Nullstelle von $\sin : \mathbb{R} \rightarrow \mathbb{R}$ ²² Da $\sin(\pi) = 0$, folgt mit obigem Additionstheorem, das $\cos(\pi)$ plus oder minus eins sein muss; mit Hilfe der Def. (2.4) kann man sich überlegen, dass nur ein negatives Ergebnis möglich ist, also $\cos(\pi) = -1$.

Mit der Euler'schen Formel können wir Additionstheoreme für Sinus und Kosinus herleiten:

$$\boxed{\sin(x+y) = \sin x \cos y + \cos x \sin y}, \quad \boxed{\cos(x+y) = \cos x \cos y - \sin x \sin y}, \quad x, y \in \mathbb{R} \quad (2.13)$$

Beweis: Es ist mit der Euler'schen Formel einerseits

$$\exp(i(x+y)) = \cos(x+y) + i \sin(x+y),$$

andererseits mit der Funktionalgleichung

$$\begin{aligned} \exp(i(x+y)) &= \exp(ix) \cdot \exp(iy) = (\cos x + i \sin x) \cdot (\cos y + i \sin y) \\ &= \cos x \cos y - \sin x \sin y + i (\sin x \cos y + \cos x \sin y) \end{aligned}$$

Durch Vergleich von Real- und Imaginärteilen folgt die Behauptung (2.13). \square

Spezialfälle von (2.13) für $x=y$ sind

$$\sin 2x = 2 \sin x \cos x, \quad \cos 2x = \cos^2 x - \sin^2 x, \quad x \in \mathbb{R}. \quad (2.14)$$

Es folgt insbesondere $\sin(2\pi) = 2 \sin \pi \cos \pi = 2 \cdot 0 \cdot 1 = 0$, $\cos(2\pi) = \cos^2 \pi - \sin^2 \pi = 1^2 - 0^2 = 1$,

und somit auch

$$\sin(x+2\pi) = \sin x \cos 2\pi + \cos x \sin 2\pi = 1 \cdot \sin x + 0 \cdot \cos x = \sin x,$$

$$\cos(x+2\pi) = \cos x \cos 2\pi - \sin x \sin 2\pi = 1 \cdot \cos x - 0 \cdot \sin x = \cos x,$$

somit haben wir die **Periodizität von Sinus und Kosinus** hergeleitet.

z.Ü.:

Zeigen Sie:

$$\sin(x+\pi) = -\sin x,$$

$$\cos(x+\pi) = -\cos x,$$

²²Strenggenommen muss man erst zeigen, dass es eine solche Nullstelle gibt. Dazu fehlt uns aber noch der Begriff der Stetigkeit. Die grundlegende Idee ist: Wenn man weiß, dass \sin stetig ist, braucht man nur noch ein x ausfindig zu machen, für das $\sin(x)$ positiv ist (z.B. $x=1$), sowie ein y , für das $\sin(y)$ negativ ist (dazu die Def. (2.4) verwenden), woraus man mit der Stetigkeit folgern kann (unter Verwendung des sog. Zwischenwertsatzes, siehe Kap. 3.3), dass es ein z gibt mit $\sin(z) = 0$.

$$\cos\left(\frac{\pi}{4} + x\right) = \sin\left(\frac{\pi}{4} - x\right),$$

$$\sin\frac{\pi}{4} = \frac{1}{2}\sqrt{2},$$

$$\cos\frac{\pi}{4} = \frac{1}{2}\sqrt{2}$$

Leiten Sie eine Formel für $\sin 3x$ und für $\cos 3x$ her (dazu (2.13) mehrfach anwenden). Mittels dieser Formeln berechnen Sie $\sin\frac{\pi}{3}$, $\cos\frac{\pi}{3}$.

Hier noch einige tabellierte Werte von Sinus und Kosinus (Herleitung z.T. siehe Übung):

x	0	$\frac{\pi}{6}$	$\frac{\pi}{4}$	$\frac{\pi}{3}$	$\frac{\pi}{2}$	$\frac{2\pi}{3}$	$\frac{3}{4}\pi$	$\frac{5}{6}\pi$	π
$\cos x$	1	$\frac{1}{2}\sqrt{3}$	$\frac{1}{2}\sqrt{2}$	$\frac{1}{2}$	0	$-\frac{1}{2}$	$-\frac{1}{2}\sqrt{2}$	$-\frac{1}{2}\sqrt{3}$	-1
$\sin x$	0	$\frac{1}{2}$	$\frac{1}{2}\sqrt{2}$	$\frac{1}{2}\sqrt{3}$	1	$\frac{1}{2}\sqrt{3}$	$\frac{1}{2}\sqrt{2}$	$\frac{1}{2}$	0

x	π	$\frac{7}{6}\pi$	$\frac{5}{4}\pi$	$\frac{4}{3}\pi$	$\frac{3}{2}\pi$	$\frac{5}{3}\pi$	$\frac{7}{4}\pi$	$\frac{11}{6}\pi$	2π
$\cos x$	-1	$-\frac{1}{2}\sqrt{3}$	$-\frac{1}{2}\sqrt{2}$	$-\frac{1}{2}$	0	$\frac{1}{2}$	$\frac{1}{2}\sqrt{2}$	$\frac{1}{2}\sqrt{3}$	1
$\sin x$	0	$-\frac{1}{2}$	$-\frac{1}{2}\sqrt{2}$	$-\frac{1}{2}\sqrt{3}$	-1	$-\frac{1}{2}\sqrt{3}$	$-\frac{1}{2}\sqrt{2}$	$-\frac{1}{2}$	0

Monotonie der Exponentialfunktion und Definition des Logarithmus.

Abschließend noch eine kurze Argumentation, die uns die *Monotonie* von $\exp : \mathbb{R} \rightarrow \mathbb{R}$ liefert:

Def. (Monotonie). Eine Abbildung $f : D \rightarrow \mathbb{R}$ mit $D \subseteq \mathbb{R}$ heißt streng monoton wachsend (bzw.: fallend), falls für alle $x, y \in D$ mit $x < y$ gilt: $f(x) < f(y)$ (bzw.: $f(x) > f(y)$). Falls nur $f(x) \leq f(y)$ bzw. $f(x) \geq f(y)$ zutrifft, spricht man von *schwacher Monotonie*.

Beachte: Eine Abbildung, die streng monoton ist (egal ob wachsend oder fallend) ist offensichtlich immer injektiv.

Für die Exponentialfunktion, angewandt auf *positives* $x \in \mathbb{R}^+$ ist laut Reihenentwicklung offensichtlich $\exp(x) > 1$, denn nach dem ersten Reihenglied, das = 1 ist, folgen nur positive Summanden. Da $\exp(-x) = \frac{1}{\exp(x)}$ ist, ist demnach auch für negative reelle Argumente $\exp(x) > 0$. Es folgt mit der Funktionalgleichung der Exponentialfunktion, das für alle $\epsilon > 0$ gilt

$$\exp(x+\epsilon) = \exp(x) \cdot \underbrace{\exp(\epsilon)}_{>1} > \exp(x),$$

d.h. $\exp : \mathbb{R} \rightarrow \mathbb{R}$ ist streng monoton wachsend. Wenn wir außerdem noch das Bild $\exp(\mathbb{R})$ von \exp bestimmen, ist $\exp : \mathbb{R} \rightarrow \exp(\mathbb{R})$ bijektiv und wir können eine Umkehrfunktion $\exp(\mathbb{R}) \rightarrow \mathbb{R}$ definieren. Wie sieht das Bild, $\exp(\mathbb{R})$, aus? Aufgrund der Eigenschaft $\exp(n) = e^n$ und $e > 1$ übersteigt $\exp(x)$ für $x \rightarrow \infty$ alle Schranken. Wegen

$\exp(-x) = \frac{1}{\exp(x)}$ nähert sich $\exp(x)$ für $x \rightarrow -\infty$ ('von oben') dem Wert null. Daraus kann man folgern, dass $\exp(\mathbb{R}) = (0, \infty) = \mathbb{R}^+$ ist.²³ Es ist also

$$\exp : \mathbb{R} \rightarrow \mathbb{R}^+$$

bijektiv,²⁴

d.h. diese Funktion besitzt eine Umkehrfunktion. Diese Umkehrfunktion heißt (*natürlicher*) *Logarithmus* (*logarithmus naturalis*), \ln :

$$\ln : \mathbb{R}^+ \rightarrow \mathbb{R}.$$

Der natürliche Logarithmus ist definiert als die Umkehrfunktion der reellen Exponentialfunktion.

Es gilt also

$$\begin{aligned} \exp(\ln x) &= x \quad \forall x > 0 \\ \ln(\exp x) &= x \quad \forall x \in \mathbb{R} \end{aligned}$$

Daraus, dass \exp streng monoton wachsend ist, kann man folgern, dass auch \ln streng monoton wachsend ist (s.Ü.).

Durch Anwendung von 'ln' auf die Potenzgesetze $\exp(a+b) = \exp(a) \cdot \exp(b)$ und $\exp(-a) = \frac{1}{\exp(a)}$ folgen mit $a = \ln x$ und $b = \ln y$ die Rechenregeln für den Logarithmus

$$\boxed{\ln x + \ln y = \ln xy}, \quad \boxed{\ln \frac{1}{x} = -\ln x}$$

da $\exp(0) = 1$, ist $\ln(1) = 0$, d.h. $x = 1$ ist die einzige Nullstelle des \ln .

Allgemeine Definition der Potenz.

Wir definieren die allgemeine Potenz

$$a^b := \exp(b \ln a), \quad a > 0, b \in \mathbb{R}. \quad (2.15)$$

Wegen $\ln(e) = 1$ ist diese Definition konsistent mit (2.9). Wegen (2.8) ist diese Definition der Potenz auch mit der klassischen, für ganzzahlige Exponenten b , $a^b = \underbrace{a \cdot a \cdot \dots \cdot a}_{b \text{ mal}}$, konsistent, denn

$$\exp(b \ln a) \stackrel{(2.8)}{=} \exp(\ln a)^b = a^b \text{ für } b \in \mathbb{N}.$$

Auch die aus der Schule bekannten Potenzgesetze $a^{b+c} = a^b \cdot a^c$ und $a^{-b} = \frac{1}{a^b}$ und $a^{bc} = (a^b)^c$ lassen sich aus den Potenzgesetzen für Potenzen zur Basis e herleiten unter Verwendung von (2.15). Und Logarithmieren von (2.15) liefert die Logarithmus-Rechenregel

$$\ln(a^b) = b \ln a.$$

²³Hier braucht man eigentlich (wie schon beim Nachweis, dass die Sinus-Funktion eine Nullstelle hat), die Stetigkeit der Exponentialfunktion und den Zwischenwertsatz.

²⁴Bemerkung: Die komplexe Exponentialfunktion $\exp : \mathbb{C} \rightarrow \mathbb{C}$ ist *nicht* injektiv. Durch Einschränkung auf einen gewissen Teilbereich von \mathbb{C} kann man jedoch noch eine bijektive Exponentialfunktion erhalten, so dass man, als deren Umkehrfunktion, auch eine komplexe Logarithmusfunktion definieren kann, die aber i.a. eher selten Verwendung findet.

3 Funktionen in einem Argument

3.1 Funktionengrenzwerte

In diesem Kapitel 3.1 wollen wir folgendes untersuchen: Gegeben sei eine Funktion $f : D_f \rightarrow \mathbb{R}$, mit $\emptyset \neq D_f \subseteq \mathbb{R}$. Wenn man sich mit dem Argument $x \in D_f$ einer festen Stelle x_* nähert, wie verhalten sich dann die zugehörigen Funktionswerte $f(x)$?

Vorab wollen wir klären, welche Werte dabei für x_* in Frage kommen. Muss $x_* \in D_f$ gelten? Nicht zwingend; z.B. für $D_f := (1, 5)$ ergibt es durchaus Sinn, $x_* := 5$ zu setzen und zu fragen, wie sich $f(x)$ verhält, wenn sich x dem Wert 5 annähert. Umgekehrt ist nicht notwendigerweise jeder Punkt $x_* \in D_f$ ein sinnvoller Punkt zur Untersuchung des Funktionengrenzwertes. Z.B. wenn wir etwas exotischere Definitionsbereiche wie z.B. $D_f = (1, 2) \cup \{3\}$ zulassen. Dann ergibt es wenig Sinn, sich mit $x \in D_f$ dem Wert $x_* := 3$ nähern zu wollen.

Also: Für x_* kommen solche Punkte in Frage, denen man sich 'näher' kann durch Punkte $x \in D_f$. Einen solchen Punkt x_* wollen wir *Häufungspunkt* der Menge D_f nennen:

Def. (Häufungspunkt einer Menge). Sei $M \subset \mathbb{R}$. Ein Punkt $x_* \in \mathbb{R}$ heißt *Häufungspunkt (HP)* von M , wenn es eine Folge (x_n) in $M \setminus \{x_*\}$ gibt mit $\lim_{n \rightarrow \infty} x_n = x_*$; anders ausgedrückt, wenn es für jedes $\epsilon > 0$ ein $x \in M \setminus \{x_*\}$ gibt mit $|x - x_*| < \epsilon$. Man sagt $+\infty$ bzw. $-\infty$ ist ein *uneigentlicher Häufungspunkt* von M , falls es eine Folge in M gibt, die bestimmt divergiert gegen $+\infty$ bzw. $-\infty$.

Dies lässt sich genau so auch für Mengen $M \subseteq \mathbb{C}$ definieren; lediglich die uneigentlichen Häufungspunkte $+\infty$ und $-\infty$ fallen dann zusammen zu ' ∞ '.

Nun die zentrale Definition dieses Kapitels:

Def. (Funktionengrenzwert). Sei $f : D_f \rightarrow \mathbb{R}$, $D_f \subset \mathbb{R}$. Sei x_* ein HP oder ein uneigentlicher HP von D_f . Falls für alle Folgen (x_n) , die in D_f liegen, und für die $\lim_{n \rightarrow \infty} x_n = x_*$ gilt, der Grenzwert

$$\lim_{n \rightarrow \infty} f(x_n)$$

existiert, ggf. auch nur als uneigentlicher Grenzwert, in $\mathbb{R} \cup \{+\infty, -\infty\}$, und für alle diese Folgen ein und derselbe ist, so bezeichnen wir diesen (ggf. uneigentlichen) Grenzwert als

$$\lim_{x \rightarrow x_*} f(x);$$

Sprich: Der Grenzwert von $f(x)$ für x gegen x_* .

Kurz:

$$\lim_{x \rightarrow x_*} f(x) := \lim_{n \rightarrow \infty} f(x_n), \quad \text{wobei } x_n \xrightarrow{(n \rightarrow \infty)} x_*.$$

Wir haben durch diese Definition den Begriff des Funktionengrenzwerts zurückgeführt auf Grenzwerte von Folgen.

Beispiel.

Für $f(x) := \frac{1}{x}$, $D_f = \mathbb{R} \setminus \{0\}$ ist ganz \mathbb{R} die Menge der HP und $\pm\infty$ sind uneigentliche HP. Wir können also fragen, ob $\lim_{x \rightarrow 0} f(x)$, $\lim_{x \rightarrow +\infty} f(x)$, $\lim_{x \rightarrow -\infty} f(x)$ existieren (ggf. als uneigentliche Grenzwerte). Es ist $\lim_{x \rightarrow +\infty} f(x) = 0$ (denn für alle Folgen (x_n) mit $\lim_{n \rightarrow \infty} x_n = +\infty$ ist $\lim_{n \rightarrow \infty} \frac{1}{x_n} = 0$), es ist $\lim_{x \rightarrow -\infty} f(x) = 0$, und $\lim_{x \rightarrow 0} f(x)$ existiert nicht (dazu einmal eine Folge (x_n) aus positiven Zahlen und einmal eine Folge (x_n) aus negativen Zahlen betrachten).

Beispiel. Es soll das Verhalten von $f(x) = \sqrt{x + \sqrt{x}} - \sqrt{x}$; $D_f = [0, \infty)$, für $x \rightarrow \infty$ untersucht werden. Sei also (x_n) eine beliebige Folge mit $\lim_{n \rightarrow \infty} x_n = +\infty$. Wir sollen

$$\underbrace{\lim_{x \rightarrow +\infty} f(x)}_{\text{Funktionengrenzwert}} = \underbrace{\lim_{n \rightarrow \infty} f(x_n)}_{\text{Folngengrenzwert}} = \lim_{n \rightarrow \infty} \sqrt{x_n + \sqrt{x_n}} - \sqrt{x_n} \text{ berechnen. Wir erweitern so,}$$

dass die dritte Binomische Formel angewendet werden kann:

$$f(x_n) = \frac{(\sqrt{x_n + \sqrt{x_n}} - \sqrt{x_n})(\sqrt{x_n + \sqrt{x_n}} + \sqrt{x_n})}{\sqrt{x_n + \sqrt{x_n}} + \sqrt{x_n}} = \frac{(x_n + \sqrt{x_n}) - x_n}{\sqrt{x_n + \sqrt{x_n}} + \sqrt{x_n}} = \frac{\sqrt{x_n}}{\sqrt{x_n + \sqrt{x_n}} + \sqrt{x_n}}$$

Nun hat man (im Nenner) die Situation " $\infty + \infty$ ", was 'besser' ist als " $\infty - \infty$ "; es besteht aber noch das Problem, dass sowohl Zähler als auch Nenner gegen ∞ gehen. Daher noch ein Ausklammern und Kürzen einer geeigneten x_n -Potenz:

$$f(x_n) = \frac{1}{\sqrt{1 + \frac{1}{\sqrt{x_n}}} + 1}$$

Unter Verwendung der Rechenregeln für Folgen geht für $x_n \rightarrow +\infty$ der Nenner gegen 2; der Zähler ist konstant eins, also

$$\lim_{x \rightarrow \infty} f(x) = \frac{1}{2}.$$

Da Funktionengrenzwerte mittels Grenzwerten von *Folgen* definiert sind, übertragen sich auch die Rechenregeln von Folgen auf Funktionengrenzwerte:

Satz (Vererbung von Funktionengrenzwerten). Seien $f, g : D \rightarrow \mathbb{R}$ Funktionen, $D \subset \mathbb{R}$, sei $x_* \in \mathbb{R} \cup \{\pm\infty\}$, ein ggf. uneigentlicher HP von D , und es existieren die Funktionengrenzwerte $\lim_{x \rightarrow x_*} f(x)$, $\lim_{x \rightarrow x_*} g(x)$. Dann existieren auch die Funktionengrenzwerte $\lim_{x \rightarrow x_*} f(x) + g(x)$, $\lim_{x \rightarrow x_*} f(x) \cdot g(x)$, und es gilt

$$\lim_{x \rightarrow x_*} (f(x) + g(x)) = \lim_{x \rightarrow x_*} f(x) + \lim_{x \rightarrow x_*} g(x), \quad \lim_{x \rightarrow x_*} (f(x) \cdot g(x)) = \lim_{x \rightarrow x_*} f(x) \cdot \lim_{x \rightarrow x_*} g(x)$$

Eine entsprechende Aussage gilt auch wieder für $\frac{f(x)}{g(x)}$, sofern der Nenner nicht gegen null geht.

Falls $\lim_{x \rightarrow x_*} f(x)$ und/oder $\lim_{x \rightarrow x_*} g(x)$ nur uneigentlich existieren, so existieren die obigen Funktionengrenzwerte in einigen Fällen ebenfalls, und zwar im Sinne von " $\infty + \infty = +\infty$ ", " $\infty + c = \infty$ ", " $(-\infty) + (-\infty) = -\infty$ ", " $-\infty + c = -\infty$ ", " $\infty - \infty$ " ist nicht definiert(!); " $(+\infty) \cdot (+\infty) = +\infty$ ", " $(-\infty) \cdot (-\infty) = +\infty$ ", " $(+\infty) \cdot (-\infty) = -\infty$ ", und " $(\pm\infty) \cdot c$ " ist, $\pm\infty$, je nach Vorzeichen von $c \neq 0$. Es ist " $\frac{c}{\pm\infty} = 0$ ", und " $\frac{c}{0}$ " ist plus oder minus unendlich, sofern ob Zähler und Nenner gleiches bzw. unterschiedliches Vorzeichen haben für x hinreichend nahe bei x_* .

Für die Fälle " $\infty - \infty$ " und " $(\pm\infty) \cdot 0$ ", " $\frac{\pm\infty}{\pm\infty}$ ", " $\frac{0}{0}$ " gibt es keine allgemeine Rechenregeln. Beachten Sie, dass wir speziell für die drei letzten Situationen in Kap. 3.6.3 noch eine spezielle Rechentechnik (Regel von l'Hospital) kennenlernen werden.

Dass wir die Rechenregeln für Folgentrenzwerte auf Funktionengrenzwerte übertragen haben, erleichtert es uns, die Berechnung von Funktionengrenzwerten durchzuführen/aufzuschreiben. Dazu zurück zum letzten Beispiel:

Beispiel. Es soll $\lim_{x \rightarrow +\infty} f(x)$ für $f(x) = \sqrt{x + \sqrt{x}} - \sqrt{x}$ berechnet werden. Anstatt alles mit Folgen zu formulieren, schreiben wir direkt

$$\lim_{x \rightarrow +\infty} f(x) = \dots (\text{analog zu oben, mit 3. Bin. Formel}) \dots = \lim_{x \rightarrow +\infty} \frac{1}{\sqrt{1 + \frac{1}{\sqrt{x}}} + 1} \stackrel{(*)}{=} \frac{1}{2}$$

wobei wir an der Stelle (*) nun die Rechenregeln für *Funktionengrenzwerte* benutzt haben.

Einseitige Grenzwerte. Wenn man lediglich solche Folgen (x_n) in D_f mit $\lim_{n \rightarrow \infty} x_n = x_*$ zulässt, die $x_n > x_*$ erfüllen ('von rechts' bzw. 'von oben' gegen x_* konvergieren, und die Grenzwerte all solcher Folgen $(f(x_n))$ existieren und gleich sind, so spricht man vom *rechtsseitigen (Funktionen-)Grenzwert*, bezeichnet mit

$$\lim_{x_n \searrow x_*} f(x),$$

sowie analog, wenn man nur Folgen mit $x_n < x_*$ betrachtet, vom *linksseitigen Grenzwert*

$$\lim_{x_n \nearrow x_*} f(x).^{25}$$

Es gilt:

Wenn rechtsseitiger und linksseitiger Funktionengrenzwert an einer Stelle x_* existieren und gleich sind, dann existiert auch der Funktionengrenzwert an dieser Stelle und ist

²⁵Eine andere gebräuchliche Schreibweise ist $\lim_{x_n \rightarrow x_* + 0} f(x)$ für den rechtsseitigen und $\lim_{x_n \rightarrow x_* - 0} f(x)$ für den linksseitigen Grenzwert.

gleich den beiden einseitigen Funktionengrenzwerten.

Diese Aussage ist nicht so völlig trivial, wie sie auf den ersten Blick erscheinen mag. Sie besagt nämlich, dass man zur Berechnung des Funktionengrenzwertes all solche Folgen (x_n) , deren Folgenglieder auf beide Seiten von x_* verteilt sind, getrost ignorieren kann. Dies macht Grenzwertberechnungen für abschnittsweise definierte Funktionen deutlich einfacher!

Sind rechts- und linksseitiger Funktionengrenzwert ungleich, so existiert der Funktionengrenzwert an dieser Stelle nicht.

3.2 Stetigkeit

In diesem Kapitel sei $D_f \subseteq \mathbb{R}$ der Einfachheit halber stets ein Intervall.

Beim Berechnen von Funktionengrenzwerten macht man folgende Beobachtung: Falls (1.) die Funktion $f : D_f \rightarrow \mathbb{R}$ "glatt" ist (was auch immer das heißen mag) und (2.) $x_* \in D_f$, dann bekommt man i.a. als Funktionengrenzwert $\lim_{x \rightarrow x_*} f(x)$ gerade $f(x_*)$ heraus. Diese Eigenschaft einer Funktion bezeichnet man als *Stetigkeit*:

Def. (Stetigkeit). Sei $x_* \in D_f$. Falls für alle Folgen (x_n) in D_f mit $\lim_{n \rightarrow \infty} x_n = x_*$ gilt, dass

$$\lim_{n \rightarrow \infty} f(x_n) = f(x_*) \quad (3.1)$$

gilt, dann ist f stetig an der Stelle x_* .

f heißt *stetig (auf D_f)*, falls f an allen Stellen $x_* \in D_f$ stetig ist. Die Menge aller stetigen Funktionen von D nach \mathbb{R} bezeichnet man auch als $C^0(D)$ oder $\mathcal{C}^0(D)$.²⁶

Mit Worten ausgedrückt: Für alle Folgen (x_n) mit $x_n \xrightarrow{(n \rightarrow \infty)} x_*$ müssen also die Folgen der zugehörigen Bilder $f(x_n)$ konvergieren, und zwar gegen ein und den selben Wert, nämlich $f(x_*)$: $f(x_n) \xrightarrow{(n \rightarrow \infty)} f(x_*)$

Man beachte, dass auf der linken Seite von (3.1) ein Funktionengrenzwert $\lim_{x \rightarrow x_*} f(x)$ steht. Man kann also die obige Definition der Stetigkeit wie folgt ausdrücken:

Satz (Stetigkeit ausgedrückt mit Funktionengrenzwert). Sei $f : D_f \rightarrow \mathbb{R}$, $D_f \subset \mathbb{R}$ und $x_* \in D_f$.
 f ist genau dann stetig an der Stelle x_* , wenn

$$\lim_{x \rightarrow x_*} f(x) = f(x_*)$$

gilt.

²⁶" \mathcal{C} " für engl. continuous=stetig

Indem man die beiden Grenzwertaussagen aus der Definition der Stetigkeit ineinander einsetzt, kann man für stetige Funktionen f auch die Gleichung

$$\boxed{\lim_{n \rightarrow \infty} f(x_n) = f\left(\lim_{n \rightarrow \infty} x_n\right)}$$

erhalten (die für alle Folgen (x_n) , für die der Grenzwert auf der rechten Seite existiert und in D_f liegt, gelten muss). Man kann die **Stetigkeit** also deuten als **Vertauschbarkeit** von **Funktionsanwendung** und **Grenzwertbildung**; als 'Herausziehen' oder 'Hereinziehen' von 'lim'.

Beispiel (Stetigkeit mit Folgenkriterium).

Sei a eine beliebige fest vorgegebene reelle Zahl und

$$f(x) := \begin{cases} \cos \frac{1}{x} & , x \neq 0 \\ a & , x = 0 \end{cases} \quad (3.2)$$

Dann ist f an der Stelle $x_* = 0$ unstetig. Begründung: Für die Folge $x_n = \frac{1}{2\pi n}$ ist $f(x_n) = \cos(2\pi n) = 1$; für die Folge $x_n = \frac{1}{(2n+1)\pi}$ ist $f(x_n) = \cos((2n+1)\pi) = -1$. Somit haben nicht alle Folgen $(f(x_n))$ mit $\lim_{n \rightarrow \infty} x_n = x_*$ den gleichen Grenzwert. \square

Es gibt noch eine weitere Charakterisierung der Stetigkeit, die zur obigen äquivalent ist. Da wir die obige Charakterisierung über Folgen als Definition verwendet haben, müssen wir die nun folgende Charakterisierung als *Satz* formulieren:

Satz (Epsilon-Delta-Charakterisierung der Stetigkeit). Eine Funktion f ist genau dann an einer Stelle $x_* \in D_f$ stetig, wenn

$$\forall \epsilon > 0 \exists \delta > 0 \forall x \in D_f : |x - x_*| \leq \delta \Rightarrow |f(x) - f(x_*)| \leq \epsilon$$

Veranschaulichung: Mittels Skizze.

Grob gesprochen ist die Idee hinter der Epsilon-Delta-Charakterisierung der Stetigkeit: Stetigkeit bedeutet, dass wenn man das Argument 'wenig' (um höchstens δ) verändert, dann darf sich auch der Funktionswert nur 'wenig' ändern (um höchstens ϵ). Hat die Funktion bei $x = x_*$ einen 'Sprung', so ist offenbar diese Bedingung verletzt; die Funktion ist unstetig.

Zu beachten ist, dass die ϵ -Schranke für die Funktionswerte vorgegeben wird, und dass in Abhängigkeit von diesem ϵ die Schranke δ der Argumente existieren muss, also $\delta = \delta(\epsilon)$ darf von ϵ abhängen (und natürlich von den gegebenen 'Daten', also von x_* und der Funktion f , aber nicht von x).

Beispiel (Stetigkeit mit ϵ - δ -Kriterium). Wir wollen mit dem ϵ - δ -Kriterium zeigen, dass die Funktion $f(x) = x^2$ an der Stelle $x_* \in \mathbb{R}$ stetig ist.

Vorüberlegung:

Wir geben also eine Schranke $\epsilon > 0$ beliebig vor und wollen zeigen, dass für x hinreichend

nahe bei x_* , nämlich $|x - x_*| \leq \delta = \delta(\epsilon)$, gilt: $|f(x) - f(x_*)| \leq \epsilon$. Frage: Gibt es so ein $\delta = \delta(\epsilon)$, wie sieht es aus? Dazu schreiben wir um:

$$|f(x) - f(x_*)| = |x^2 - x_*^2| = \underbrace{|x - x_*|}_{\leq \delta} \cdot |x + x_*| \stackrel{!}{\leq} \epsilon. \text{ Wir müssen dazu sicherstellen, dass}$$

$|x + x_*| \leq \frac{\epsilon}{\delta}$ ist; denn wenn wir dies hinbekommen, ist die mit dem "!" gekennzeichnete Stelle offensichtlich erfüllt. Dies legt nahe, $\delta > 0$ als $\delta := \frac{\epsilon}{|x + x_*|}$ zu wählen; jedoch darf δ nur von ϵ und x_* , jedoch nicht von x abhängen. Wenn wir unser δ so wählen, dass sicher $\delta < 1$ gilt, so können wir den hinteren Faktor mittels $|x + x_*| \leq |x| + |x_*| \leq |x_*| + |x_*| + \delta \leq 2|x_*| + 1$ abschätzen, also $\delta \cdot (2|x_*| + 1) \stackrel{!}{\leq} \epsilon$. Diese Bedingung wird erfüllt, wenn δ kleiner oder gleich $\frac{\epsilon}{2|x_*| + 1}$ gewählt wird.

Wir schreiben nun den Beweis 'ins Reine':

Sei $\epsilon > 0$ beliebig vorgegeben. Setze $\delta := \min\{1, \frac{\epsilon}{2|x_*| + 1}\}$. Dann ist für alle $x \in \mathbb{R}$ mit $|x - x_*| \leq \delta$:

$$\begin{aligned} |f(x) - f(x_*)| &= |x^2 - x_*^2| = |x - x_*| \cdot |x + x_*| \leq \delta \cdot (|x_*| + |x|) \leq \delta \cdot (2|x_*| + \delta) \\ &\leq \frac{\epsilon}{2|x_*| + 1} \cdot (2|x_*| + 1) = \epsilon \end{aligned} \quad \square$$

Wie schon beim Thema 'Konvergenz von Folgen' sieht man auch bei der Stetigkeit, dass es bei komplizierteren Funktionen sehr schwierig werden kann, mit der ϵ -Technik Beweise zu führen. Einfacher ist es, die Folgen-Charakterisierung der Stetigkeit zu verwenden,²⁷ denn für Folgen haben wir diverse Rechenregeln. Hier für obiges Beispiel: Sei (x_n) eine Folge mit $\lim_{n \rightarrow \infty} x_n = x_*$. Dann ist $\lim_{n \rightarrow \infty} f(x_n) = \lim_{n \rightarrow \infty} x_n^2 \stackrel{(*)}{=} (\lim_{n \rightarrow \infty} x_n)^2 = x_*^2$, d.h. $f(x) = x^2$ ist an jeder Stelle $x_* \in \mathbb{R}$ stetig. Dabei wurde an der Stelle (*) eine Rechenregel für Folgen angewendet. \square

Da komplizierte Funktionen häufig als Verkettungen/Verknüpfungen von einfacheren Funktionen, deren Stetigkeit man kennt, aufgebaut sind, kann man Stetigkeit häufig besonders einfach mit folgendem Satz begründen:

Satz (Vererbung von Stetigkeit bei Verknüpfung und Verkettung von Funktionen). Seien $f, g : D \rightarrow \mathbb{R}$ Funktionen, $D \subset \mathbb{R}$, $x_* \in D$, und seien f und g an der Stelle x_* stetig. Dann sind auch die Funktionen $f + g$ und $f \cdot g$ an der Stelle x_* stetig. Falls außerdem $g(x_*) \neq 0$, so ist auch $\frac{f}{g}$ stetig an der Stelle x_* .

Seien $f : D_f \rightarrow \mathbb{R}$ und $g : D_g \rightarrow \mathbb{R}$ Funktionen, $x_* \in D_g$ und $g(x_*) \in D_f$, und sei g stetig an der Stelle x_* , und sei f stetig an der Stelle $g(x_*)$. Dann ist die Verkettung $f \circ g$ an der Stelle x_* stetig.

Zum Beweis: Die obigen Aussagen folgen leicht aus den entsprechenden Rechenregeln für Grenzwerte von Folgen und der Charakterisierung der Stetigkeit mittels Folgen.

²⁷Um Sätze über stetige Funktionen zu beweisen, ist die ϵ - δ -Charakterisierung der Stetigkeit hingegen oft sehr nützlich.

Beweis. Folgt leicht aus den Rechenregeln für Funktionengrenzwerte in Kap. 3.1, so z.B. $\lim_{x \rightarrow x_*} (f+g)(x) = \lim_{x \rightarrow x_*} f(x) + g(x) \stackrel{(\text{RR.})}{=} \lim_{x \rightarrow x_*} f(x) + \lim_{x \rightarrow x_*} g(x) \stackrel{\text{f.g stet.}}{=} f(x_*) + g(x_*) = (f+g)(x_*) \Rightarrow f+g$ stetig. \square

Anwendung. Die identische Funktion $f : \mathbb{R} \rightarrow \mathbb{R}, f(x) = x$, ist offensichtlich stetig (wähle dazu $\delta := \epsilon$ im ϵ - δ -Kriterium). Auch die konstante Funktion $f : \mathbb{R} \rightarrow \mathbb{R}, f(x) = c$, ist offensichtlich stetig (wähle z.B. $\delta := 1$ oder auch wieder $\delta := \epsilon$). Der Satz über die Vererbung der Stetigkeit besagt dann, dass auch jede Polynomfunktion

$f(x) = \sum_{k=0}^n a_k x^k$ stetig ist. Auch jede gebrochen-rationale Funktion $f(x) = \frac{\sum_{k=0}^m a_k x^k}{\sum_{k=0}^n b_k x^k}$ ist an allen Stellen, an denen der Nenner ungleich null ist, stetig.

Zur Stetigkeit von sin, cos, exp, ln. Interessant ist natürlich in diesem Zusammenhang, ob auch die anderen 'elementaren' Funktionen, die wir kennen, wie sin, cos, exp, ln stetig sind, da diese Funktionen als Bausteine häufig in anderen Funktionen vorkommen. Wir wollen hier exemplarisch lediglich zeigen, dass die Exponentialfunktion stetig ist.

Die Stetigkeit von Sinus und Kosinus wollen wir hier nicht beweisen, denn in Kap. 3.4 werden wir eine viel einfachere Argumentation kennenlernen, die die Stetigkeit sämtlicher Potenzreihen, somit also auch von sin, cos, exp, liefert.

Satz. Die Exponentialfunktion $\exp : \mathbb{R} \rightarrow \mathbb{R}$ ist stetig.

Beweis. Wir wollen zeigen, dass $e^{x_*+h} - e^{x_*}$ gegen null geht für $h \rightarrow 0$ (denn dann geht e^{x_*+h} gegen e^{x_*} , d.h. exp stetig an der Stelle x_*).²⁸

Wir beschränken uns zunächst einmal auf positive h und bekommen

$$e^{x_*+h} - e^{x_*} = e^{x_*} \cdot e^h - e^{x_*} = \underbrace{e^{x_*}}_{\text{fest}} \underbrace{(e^h - 1)}_{\rightarrow 0?}$$

Wegen $h > 0$ und der Monotonie von exp ist²⁹

$$0 \leq e^h - 1 = \left(1 + \frac{h^1}{1!} + \frac{h^2}{2!} + \frac{h^3}{3!} + \dots\right) - 1 = h \cdot \left(1 + \frac{h}{2!} + \frac{h^2}{3!} + \dots\right) \leq h \cdot \left(1 + \frac{h}{1!} + \frac{h^2}{2!} + \dots\right) = h \cdot e^h$$

Wir brauchen nur h nahe bei 0, also nehmen wir $h \leq 1$ an, dann ist wegen der Monotonie von exp der zweite Faktor durch e^1 beschränkt, also haben wir $0 \leq e^h - 1 \leq h \cdot e$. Für $h \rightarrow 0$ geht also $e^h - 1$ gegen null ('Einschachtelung').

²⁸Genau genommen betrachten wir statt einer Folge (x_n) mit $x_n \rightarrow x_*$ eine Folge $h_n := x_n - x_* \rightarrow 0$.

²⁹Beachte: In so einer Summation von unendlich vielen Summanden ist es i.a. *nicht* erlaubt zu argumentieren: "Jeder Summand geht gegen null, also auch die Summe"; so kann man nur bei endlich vielen Summanden argumentieren.

Nun noch für negative h : Statt $|e^{x_*+h} - e^{x_*}|$ mit $h < 0$ betrachten wir, was gleichwertig ist, $|e^{x_*-h} - e^{x_*}|$ mit $h > 0$. Wir bekommen

$$|e^{x_*-h} - e^{x_*}| = e^{x_*} - e^{x_*-h} = e^{x_*} \cdot (1 - e^{-h}) = e^{x_*} \cdot \frac{e^h - 1}{e^h}$$

Unter Verwendung der bereits hergeleiteten Abschätzung von $e^h - 1 \leq h \cdot e^h$ haben wir

$$|e^{x_*-h} - e^{x_*}| \leq e^{x_*} h,$$

was für $h \rightarrow 0$ gegen null geht. □

Man kann zeigen: Ist eine Funktion $f : D_f \rightarrow \mathbb{R}$, wobei $D_f \subset \mathbb{R}$ ein Intervall(!) ist, stetig, und besitzt f eine Umkehrfunktion $f^{-1} : \text{Bild}(f) \rightarrow D_f$, so ist auch die Umkehrfunktion stetig.³⁰

Insbesondere ist also auch $\ln : \mathbb{R}^+ \rightarrow \mathbb{R}$ stetig.

Stetigkeit und einseitige Funktionengrenzwerte, abschnittsweise definierte Funktionen.

Sei x_* im Inneren des Intervalls D_f .

Wir haben definiert: f ist an der Stelle x_* stetig, wenn der Funktionengrenzwert $\lim_{x \rightarrow x_*} f(x)$ mit dem Funktionswert $f(x_*)$ übereinstimmt. Im Kapitel 3.1 hatten wir festgehalten, dass man zur Berechnung des Funktionengrenzwertes den rechts- und linksseitigen Grenzwert ausrechnen kann (was sich besonders bei abschnittsweise definierten Funktionen anbietet). Insgesamt ergibt sich:

f ist an der Stelle x_* , die im Inneren von D_f liege, genau dann stetig, wenn rechtseitiger und linksseitiger Funktionengrenzwert und auch der Funktionswert an der Stelle übereinstimmen.

Dies ist ein nützliches Kriterium, um Stetigkeit von abschnittsweise definierten Funktionen an der Übergangsstelle zu untersuchen.

Stetige Fortsetzung von Funktionen. Für Stellen x_* , die zwar Häufungspunkte von D_f sind, aber nicht zu D_f selbst gehören, z.B. Nullstellen des Nenners bei gebrochenrationalen Funktionen $f(x) = \frac{x^2-4}{x-2}$ kann man die Frage stellen, ob es ein $a \in \mathbb{R}$ gibt, so dass durch Festsetzung $f(x_*) := a$ die Funktion zu einer an der Stelle x_* stetigen Funktion gemacht werden kann. Man bezeichnet dies als *stetige Fortsetzung*. Das Beispiel (3.2) war nichts anderes als die Frage nach stetiger Fortsetzung der Funktion $f : \mathbb{R} \setminus \{0\} \rightarrow \mathbb{R}$, $f(x) = \cos \frac{1}{x}$.

³⁰Wenn man die Voraussetzung, dass D_f ein Intervall ist, fallenlässt, wird die Aussage überraschenderweise falsch. Gegenbeispiel: $f : [0, 1] \cup (2, 3] \rightarrow \mathbb{R}$ mit $f(x) = x$ für $x \in [0, 1]$ und $f(x) = x - 1$ für $x \in (2, 3]$ ist auf D_f in der Tat stetig. Die Umkehrfunktion, die man sich am besten grafisch durch Spiegeln an der Winkelhalbierenden verschafft, ist auf $[0, 2]$ definiert, jedoch an der Stelle $x=1$ nicht stetig, denn $f^{-1}(x) = x$ für $x \in [0, 1]$ und $f^{-1}(x) = x + 1$ für $x \in (1, 2]$.

3.3 Eigenschaften stetiger Funktionen

Als Vorbereitung für den eigentlichen Inhalt dieses Kapitles benötigen wir einige *topologische Grundbegriffe*:

Def. u. Satz (Rand, Abschluss und Inneres einer Menge; Abgeschlossenheit, Offenheit, Kompaktheit von Mengen). Sei $M \subseteq \mathbb{R}$. Der *Abschluss* von M ist die Menge $\overline{M} := M \cup \{x \in \mathbb{R} \mid x \text{ ist HP}\}$. Man nennt M *abgeschlossen*, falls $M = \overline{M}$. Der *Rand* ∂M von M ist die Menge $\partial M := \{x \in \mathbb{R} \mid \forall \epsilon > 0 \exists x_1 \in M, x_2 \in \mathbb{R} \setminus M : |x - x_1| \leq \epsilon, |x - x_2| \leq \epsilon\}$. Man kann zeigen: $\partial M = \overline{M} \cap \overline{\mathbb{R} \setminus M}$ und $\overline{M} = M \cup \partial M$. Man nennt die Menge $\overset{\circ}{M} := M \setminus \partial M$ das *Innere* von M , und M heißt *offen*, wenn $M = \overset{\circ}{M}$. Man kann zeigen:

M ist genau dann offen (bzw. abgeschlossen), wenn $\mathbb{R} \setminus M$ abgeschlossen (bzw. offen) ist. Und:

M ist genau dann offen, wenn $\forall x \in M \exists \epsilon > 0 : K_\epsilon(x) \subset M$. (s.Ü.)

Eine Menge, die abgeschlossen und beschränkt ist, heißt *kompakt*.

Dies lässt sich genau so auch für Mengen $M \subseteq \mathbb{C}$ definieren; lediglich die uneigentlichen Häufungspunkte $+\infty$ und $-\infty$ fallen dann zusammen zu ' ∞ '.

Beispiele:

Das Intervall $(0, 1)$ ist offen, das Intervall $[0, 1]$ ist abgeschlossen; für beide gilt: Der Rand ist $\{0, 1\}$, der Abschluss ist $[0, 1]$, das Innere ist $(0, 1)$, die Menge der HP ist $[0, 1]$.

Das Intervall $(0, \infty)$ ist offen; das Intervall $[0, \infty)$ ist abgeschlossen; beide haben ∞ als uneigentlichen HP, beide haben als Rand $\{0\}$.

Die Menge $M := \{\frac{1}{n} \mid n \in \mathbb{N}\}$ hat als Menge der Häufungspunkte nur $\{0\}$, somit ist $\overline{M} = \{0\} \cup M$, $\overset{\circ}{M} = \emptyset$, $\partial M = \overline{M}$; M ist weder offen noch abgeschlossen.

z.Ü: sind \emptyset und \mathbb{R} offen und/oder abgeschlossen?

Bemerkung. Man kann den Begriff der Offenheit von Mengen benutzen, um die Stetigkeit von Funktionen zu charakterisieren:

Satz: Eine Funktion $f : \mathbb{R} \rightarrow \mathbb{R}$ ist genau dann stetig, wenn für jede *offene* Menge $M \subseteq \mathbb{R}$ die zugehörige Urbildmenge $f^{-1}(M)$ *offen* ist.

Dies ist, neben der Folgen- und der Epsilon-Delta-Charakterisierung der Stetigkeit eine dritte Möglichkeit, Stetigkeit zu erklären. Diese Art der Definition wird im mathematischen Gebiet der *Topologie* bevorzugt, da sie sich auf sehr abstrakte Situationen, nämlich auf Funktionen zwischen sog. 'topologischen Räumen' (anstelle von \mathbb{R}) verallgemeinern lässt. Diesen Ansatz werden wir in dieser Vorlesung jedoch nicht weiter verfolgen.

Zur Übung: Die Funktion $f : \mathbb{R} \rightarrow \mathbb{R}$ mit $f(x) = 1$ für $x \geq 0$ und $f(x) = 0$ für $x < 0$ ist offensichtlich unstetig. Geben Sie eine offene Menge $M \subseteq \mathbb{R}$ an, so dass $f^{-1}(M)$ nicht offen ist.

Das Thema dieses Kapitels. In diesem Kapitel wollen wir zwei wesentliche Eigenschaften stetiger Funktionen thematisieren:

1. Stetige Funktionen auf kompakten Mengen (Definitionsbereichen) nehmen Maximum und Minimum an.
2. Stetige Funktionen, die sowohl einen positiven als auch einen negativen Funktionswert haben, und deren Definitionsbereich ein Intervall ist, haben auch eine Nullstelle ("Nullstellensatz").³¹

Satz (Stetige Funktionen auf kompakten Mengen). Sei $D \subset \mathbb{R}$ kompakt und $f : D \rightarrow \mathbb{R}$ stetig.

Dann ist f beschränkt, d.h. es gibt ein $s \in \mathbb{R}$ so dass $|f(x)| \leq s$ für alle $x \in D$.³²

Ferner werden Infimum und Supremum 'angenommen', d.h. es gibt $x_* \in D$ mit $f(x_*) = \inf_{x \in D} f(x)$ und $x_{**} \in D$ mit $f(x_{**}) = \sup_{x \in D} f(x)$, d.h. die Menge der Funktionswerte

hat ein Minimum $\min_{x \in D} f(x) = \inf_{x \in D} f(x)$ und ein Maximum $\max_{x \in D} f(x) = \sup_{x \in D} f(x)$.

Kurz: **Stetige Funktionen nehmen auf kompakten Mengen Maximum/Minimum an.**

Beweis. Wir zeigen zuerst die Beschränktheit. Angenommen, f ist nicht nach oben beschränkt. Dann gibt es eine Folge (x_n) in D mit $\lim_n f(x_n) = \infty$. Da D beschränkt ist, gibt es nach dem Satz von Bolzano-Weierstraß eine Teilfolge $(x_{n_k})_{k \in \mathbb{N}}$, die konvergent ist; nennen wir den Grenzwert $x_* \in \mathbb{R}$. Da (x_{n_k}) in D liegt, muss ihr Limes zumindest in \overline{D} liegen; wegen der vorausgesetzten Abgeschlossenheit von D also $x_* \in \overline{D} = D$. Das bedeutet, es gibt einen Funktionswert $f(x_*) \in \mathbb{R}$. Aber: Wir hatten $\lim_k f(x_{n_k}) = \infty \neq f(x_*)$, was ein Widerspruch zur vorausgesetzten Stetigkeit von f ist. Also ist die Annahme falsch, somit f nach oben beschränkt und besitzt somit ein Supremum s .

Nun zur Existenz des Maximums. Da das Supremum definitionsgemäß die kleinste obere Schranke der Menge der Funktionswerte $\{f(x) \mid x \in D\}$ ist, gibt es in dieser Menge Elemente, die beliebig nahe an s heranrücken, d.h. es gibt eine Folge $f(x_1), f(x_2), \dots$ die gegen s konvergiert. Wie schon oben beim Beweis der Beschränktheit von f zeigen wir auch hier wieder, dass die Folge (x_n) dann eine Teilfolge haben muss, die gegen einen Wert $x_* \in D$ konvergiert. Wegen der Stetigkeit von f an der Stelle x_* muss dann $f(x_*) = \lim_k f(x_{n_k})$ sein, also $= s$: Damit ist das Supremum s Element der Menge der Funktionswerte $\{f(x) \mid x \in D\}$ und ist somit ihr Maximum.

Analog zeigt man, dass f nach unten beschränkt ist und dass das Infimum gleich dem

³¹Diesen Sachverhalt haben wir bereits verwendet, nämlich bei der Definition von π als Nullstelle von \sin .

³²Dann gibt es also insbesondere $\inf_{x \in D} f(x)$ und $\sup_{x \in D} f(x)$; eine größte untere und eine kleinste obere Schranke der Funktionswerte.

Minimum ist.³³

□

Beispiele. $f(x) := \frac{\exp(x)}{x}$ nimmt auf $[1, 5]$ Maximum und Minimum an, denn f ist auf $[1, 5]$ stetig und $[1, 5]$ ist kompakt.

$f(x) := \frac{1}{x}$ nimmt auf $[1, 2]$ Minimum und Maximum an. Dagegen die Mengen $(0, 1]$ sowie $[1, \infty)$ sind nicht kompakt, d.h. der obige Satz ist nicht anwendbar, und in der Tat hat f auf diesen Mengen kein Maximum bzw. kein Minimum. Man mache sich klar, dass auch auf die *Stetigkeit* in obigem Satz nicht verzichtet werden kann (abschnittsweise definierte Funktion betrachten, die kein Min./Max. hat)!

Nun zum Nullstellensatz:

Satz (Nullstellensatz von Bolzano). Sei $f : [a, b] \rightarrow \mathbb{R}$ stetig; $a < b$, und es sei $f(a) \cdot f(b) < 0$ (damit drückt man in knapper Form aus, dass einer der beiden Werte positiv und der andere negativ ist). Dann besitzt f in (a, b) (mindestens) eine Nullstelle, d.h. ein x_* mit $f(x_*) = 0$.

Beweis. O.B.d.A. sei $f(a) < 0$ und $f(b) > 0$ (der umgekehrte Fall geht analog). Sei $M := \{x \in [a, b] \mid f(x) < 0\}$. Die Menge M ist nicht leer, denn $a \in M$. Die Menge M ist beschränkt, denn $M \subseteq [a, b]$. Sie besitzt daher ein Supremum $x_* \in [a, b]$. Wir zeigen nun, dass x_* eine Nullstelle von f ist:

Angenommen, $f(x_*) > 0$. Dann können wir wegen der Stetigkeit von f , durch die Wahl $\epsilon := \frac{1}{2}f(x_*)$ ein $\delta > 0$ finden, so dass alle $x \in [x_* - \delta, x_* + \delta]$ einen Funktionswert haben, der um höchstens ϵ von $f(x_*)$ abweicht, somit ebenfalls echt positiv ist (Skizze!). Da aber die Menge M 'bis an x_* ' heranreicht (genauer: x_* ist als Supremum ein HP von M), ist $M \cap [x_* - \delta, x_* + \delta] \neq \emptyset$, für x aus M ist allerdings $f(x) < 0$ nach Definition von M ; Widerspruch. Analog geht's mit der Annahme, dass $f(x_*)$ strikt negativ ist (man setzt dann $\epsilon := \frac{1}{2}|f(x_*)|$). Es bleibt also nur: $f(x_*) = 0$ □.

Anwendungsbeispiel. Für $f(x) := \exp(x) + x$ ist $f(-1) = \frac{1}{e} - 1 < 0$ und $f(0) = 1 > 0$. Da f stetig ist, hat f im Intervall $(-1, 0)$ (mindestens) eine Nullstelle. (Da f außerdem streng monoton wachsend ist (denn 'exp' ist monoton wachsend und $x \mapsto x$ ist monoton wachsend), hat f höchstens eine Nullstelle, somit sogar genau eine Nullstelle.)

Eine elementare Folgerung des Nullstellensatz ist der sog. *Zwischenwertsatz*:

Satz (Zwischenwertsatz von Bolzano). Sei $f : [a, b] \rightarrow \mathbb{R}$ stetig und $f(a) \neq f(b)$. Dann wird jeder Wert ('Zwischenwert') ξ , der zwischen $f(a)$ und $f(b)$ liegt, als Funktionswert (mindestens) einmal angenommen, d.h. es gibt ein $x_* \in (a, b)$ mit $f(x_*) = \xi$.

Der Beweis ist elementar: Sei ξ zwischen $f(a)$ und $f(b)$ liegend vorgegeben. Wir be-

³³Dieser Beweis ist eine typische Argumentationsweise aus der Analysis, genannt *Kompaktheitschluss* (so in etwa: man konstruiert gewisse beschränkte Folgen, folgert, dass es HP, d.h. konvergente Teilfolge, gibt; für den Grenzwert dieser Teilfolge leitet man gewisse Eigenschaften her).

trachten die Funktion $h(x) := f(x) - \xi$ (d.h. Graph von f um ξ nach unten verschoben). h ist dann ebenfalls auf $[a, b]$ stetig, und da ξ zwischen $f(a)$ und $f(b)$ liegt, ist einer der beiden Werte $h(a), h(b)$ positiv und der andere negativ. Es kann also auf h der Nullstellensatz angewendet werden. Der liefert ein $x_* \in (a, b)$ mit $h(x_*) = 0$, also $f(x_*) = \xi$. \square

3.4 Funktionenfolgen, gleichmäßige Konvergenz, und nochmal Potenzreihen

Folgen von Funktionen(!) zu betrachten ist für uns nicht völlig neu: Bereits im Zusammenhang mit *Potenzreihen* $\sum_{k=0}^{\infty} a_k (x-x_0)^k$ hatten wir die Partialsummen $s_n(x) := \sum_{k=0}^n a_k (x-x_0)^k$ betrachtet; die Grenzfunktion f war "punktweise", d.h. für jede Stelle x erklärt als

$$f(x) := \lim_{n \rightarrow \infty} s_n(x),$$

für solche x , für die der Grenzwert existiert. Wir hatten bereits gelernt, dass es immer einen Konvergenzradius $R \in [0, \infty]$ gibt, so dass für alle $x \in \mathbb{R}$ oder $x \in \mathbb{C}$ mit $|x-x_0| < R$ Konvergenz (sogar absolute Konvergenz) gilt und für alle x mit $|x-x_0| > R$ Divergenz. Wir können den oben zugrunde gelegten Konvergenzbegriff für Funktionenfolgen in eine Definition packen:

Definition (punktweise Konvergenz einer Folge von Funktionen). Sei $D \subseteq \mathbb{C}$ oder $D \subseteq \mathbb{R}$, und seien $f_n : D \rightarrow \mathbb{C}$ (od. \mathbb{R}) eine Folge von Funktionen.³⁴ Wir sagen, die Folge (f_n) konvergiert punktweise auf D gegen eine Grenzfunktion $f : D \rightarrow \mathbb{R}$, wenn für jedes $x \in D$ (also 'punktweise') gilt

$$\lim_{n \rightarrow \infty} f_n(x) = f(x).$$

In Quantoren ausgedrückt:

$$\boxed{\forall x \in D \forall \epsilon > 0 \exists n_0 \in \mathbb{N} \forall n \geq n_0 : |f_n(x) - f(x)| \leq \epsilon}$$

Wichtig: Hier darf – beachte die Reihenfolge der Quantoren – das n_0 sowohl von ϵ als auch von x abhängen: $n_0 = n_0(\epsilon, x)$.

Neben der *punktweisen* Annäherung einer Funktionenfolge an eine Grenzfunktion kann man stattdessen auch die Annäherung mittels des maximalen Abstandes $\max_{x \in D} |f_n(x) - f(x)|$ 'messen', d.h. man legt fest, dass die Funktionenfolge (f_n) nur dann gegen f konvergiert, wenn der maximale Abstand gegen null geht, also f_n in einem

³⁴ D muss nicht der 'gesamte' (der maximale) Definitionsbereich der f_n sein; man kann hier auch eine Teilmenge einsetzen.

ϵ -Schlauch um f liegt (Skizze!). Da jedoch $\max_{x \in D} |f_n(x) - f(x)|$ "seltener" existiert als $\sup_{x \in D} |f_n(x) - f(x)|$, ist es sinnvoll diesen letzteren Ausdruck als Abstandsmaß zu benutzen.³⁵

Definition (gleichmäßige Konvergenz einer Folge von Funktionen). Seien D , f_n , f wie oben. Wir sagen, dass die Funktionenfolge *gleichmäßig gegen f konvergiert*, wenn

$$\lim_{n \rightarrow \infty} \left(\sup_{x \in D} |f_n(x) - f(x)| \right) = 0. \quad (3.3)$$

gilt. In Quantoren ausgedrückt:

$$\boxed{\forall \epsilon > 0 \exists n_0 \in \mathbb{N} \forall n \geq n_0 \underbrace{\forall x \in D : |f_n(x) - f(x)| \leq \epsilon}_{\cong \sup_{x \in D} |f_n(x) - f(x)| \leq \epsilon}} \quad (3.4)$$

Wichtig: Hier darf n_0 nur von ϵ abhängen: $n_0 = n_0(\epsilon)$. Man sagt: n_0 wird 'gleichmäßig' bzgl. $x \in D$ gewählt.

Offensichtlich gilt: Aus der gleichmäßigen Konvergenz folgt immer die punktweise Konvergenz, denn man kann das $n_0 = n_0(\epsilon)$ aus der Definition der glm. Konvergenz einfach für alle $x \in D$ als $n_0 = n_0(\epsilon, x)$ in der Definition der punktweisen Konvergenz verwenden. Die Umkehrung ist jedoch nicht richtig; in diesem Sinne ist die Eigenschaft "gleichmäßige Konvergenz" 'stärker' als die Eigenschaft "punktweise Konvergenz". Dass die Umkehrung nicht gilt, kann man leicht anhand des folgenden Gegenbeispiels sehen: Sei (Skizze!)

$$f_n(x) := \begin{cases} 0 & , x \leq 0 \\ 1 & , x \geq \frac{1}{n} \\ nx & , 0 < x < \frac{1}{n} \end{cases}$$

Für jedes feste x ist offensichtlich

$$\lim_{n \rightarrow \infty} f_n(x) = \begin{cases} 0 & , x \leq 0 \\ 1 & , x > 0 \end{cases} =: f(x),$$

d.h. die Funktionenfolge (f_n) konvergiert punktweise gegen f auf $D = \mathbb{R}$. Die Funktionenfolge (f_n) konvergiert jedoch nicht gleichmäßig gegen f , denn der Abstand zwischen f_n und f , gemessen im Supremum, ist, für jedes $n \in \mathbb{N}$, immer 1, geht also nicht gegen null.

Anmerkung zur Veranschaulichung der x -(un-)abhängigen Wahl von n_0 :

Ist $\epsilon > 0$ fest vorgegeben, so muss man in diesem Beispiel das n_0 , für welches $|f_n(x) - f(x)| \leq \epsilon$ ist, in der Tat abhängig von x wählen: Je näher x gegen null rückt

³⁵Wir wissen: Wenn D kompakt ist und sowohl alle f_n als auch f auf D stetig sind, dann existiert obiges Maximum und ist gleich dem Supremum, d.h. in (3.3) darf man dann "sup" durch "max" ersetzen.

(für positive x), desto größer muss man n_0 wählen; es ist nicht möglich, n_0 gleichmäßig bzgl. x zu wählen.

Zur Übung: Überlege, ob die Folge $f_n(x) = \sin \frac{x}{n}$ punktweise und/oder gleichmäßig konvergiert, (a) auf $D = \mathbb{R}$, (b) auf $D = [0, 1]$.

Wozu braucht man gleichmäßige Konvergenz?

Dann, wenn man eine Folge von stetigen Funktionen gegeben hat, und die Frage stellt, ob die Grenzfunktion dann ebenfalls stetig ist.

Im allgemeinen ist so eine Schlussfolgerung bei punktweiser Konvergenz unzulässig: Siehe obiges Beispiel! Die f_n sind stetig, aber die (punktweise) Grenzfunktion f ist unstetig!

Aber: Nehmen wir als stärkere Voraussetzung die gleichmäßige Konvergenz, so bekommen wir ein Resultat:

Satz (Stetigkeit der Grenzfunktion). Seien D , (f_n) , f wie oben, seien die f_n stetig, und sei (f_n) auf D gegen f gleichmäßig(!) konvergent. Dann ist auch die Grenzfunktion f stetig.

Beweis. Wir wollen die Stetigkeit von f an einer Stelle $x_* \in D$ zeigen.

Der Beweis verwendet den sog. $\frac{\epsilon}{3}$ -Trick und die ϵ - δ -Charakterisierung der Stetigkeit:

Sei $\epsilon > 0$ vorgegeben. [Wir müssen ein $\delta > 0$ finden, so dass für alle $|x - x_*| \leq \delta$ $|f(x) - f(x_*)| \leq \epsilon$ gilt.]

Wegen der gleichmäßigen Konvergenz von (f_n) gegen f gibt es ein $n_0 = n_0(\epsilon) \in \mathbb{N}$, so dass für alle $n \geq n_0$ und alle $x \in D$ gilt:³⁶

$$|f_n(x) - f(x)| \leq \frac{\epsilon}{3}. \quad (3.5)$$

Wir wählen so ein $n \geq n_0$, z.B. $n := n_0$.

Da f_n stetig ist, gibt es ein $\delta = \delta(\epsilon, n_0(\epsilon)) = \delta(\epsilon) > 0$, so dass für alle $x \in D$ mit $|x - x_*| \leq \delta$ gilt $|f_n(x) - f_n(x_*)| \leq \frac{\epsilon}{3}$.

Wir schätzen nun mit der Dreiecksungleichung ab, verwenden dabei (3.5) an zwei verschiedenen Stellen x und x_* : Es folgt

$$|f(x) - f(x_*)| \leq |f(x) - f_n(x)| + |f_n(x) - f_n(x_*)| + |f_n(x_*) - f(x_*)| \leq \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon$$

für alle $x \in D$ mit $|x - x_*| \leq \delta$. □

Nun zurück zu den Potenzreihen. Wir wissen, dass Potenzreihen im Inneren ihres Konvergenzradius *punktweise* konvergent sind. Sind sie dort auch gleichmäßig

³⁶Beachte: Hätten wir nur *punktweise* Konvergenz, so müssten wir ggf. für jedes x ein anderes n_0 nehmen; es wäre nicht sicher, dass wir ein gemeinsames n finden, so dass (3.5) für alle betrachteten x gilt.

konvergent? Wenn dem so wäre, wüssten wir mit obigem Satz, dass die Grenzfunktionen stetig sind, dass also z.B. $\sin, \cos, \exp : \mathbb{R} \rightarrow \mathbb{R}$ (und auch als $\mathbb{C} \rightarrow \mathbb{C}$) stetig sind!

Satz (Gleichmäßige Konvergenz von Potenzreihen). Sei $\sum_{k=0}^{\infty} a_k (x - x_0)^k = \lim_n s_n(x)$ eine Potenzreihe mit Partialsummen $s_n(x) = \sum_{k=0}^n a_k (x - x_0)^k$. Sei der Konvergenzradius strikt positiv, also $R \in (0, \infty]$ (d.h. $R = \infty$ ist zugelassen). Sei $0 < r < R$. Dann ist die Folge der Partialsummen (s_n) auf der Kreisscheibe $K_r := \{x \in \mathbb{C} \mid |x - x_0| \leq r\}$ gleichmäßig konvergent gegen die Grenzfunktion.

Beweis. Man überlegt sich zunächst, dass um die gleichmäßige Konvergenz einer x -abhängigen Reihe zu zeigen, es reicht, eine konvergente Majorante zu finden, die unabhängig von x ist.

Wir wählen dann irgendein x mit $|x - x_0| = r$. Da dann $|x - x_0| < R$ ist, wissen wir, dass die Potenzreihe für dieses x absolut konvergent ist (nach dem Satz über den Konvergenzbereich von Potenzreihen, S. 26), dass also die Reihe $\sum_{k=0}^{\infty} a_k r^k$ absolut konvergent, ist, also $\sum_{k=0}^{\infty} |a_k| r^k$ konvergent ist. Da für alle $x \in K_r$ $|x - x_0| \leq r$ ist, gilt für diese x dass $|a_k (x - x_0)^k| \leq |a_k| r^k$ ist, d.h. die Reihe $\sum_{k=0}^{\infty} |a_k| r^k$ ist eine konvergente Majorante für unsere Potenzreihe an jeder Stelle $x \in K_r$. Daraus folgt die gleichmäßige Konvergenz der Potenzreihe. \square

Folgerung: Potenzreihen mit Konvergenzradius $R \in (0, \infty]$ sind auf dem Konvergenzintervall $I := (x_0 - R, x_0 + R)$ stetig.

Beweis. Um zu zeigen, dass eine Potenzreihe an einer Stelle $x \in I$ stetig ist, wähle r aus obigem Satz hinreichend groß, nämlich $|x - x_0| < r < R$. Mit dem Satz folgt die Stetigkeit an dieser Stelle.

Also:

$$\boxed{\sin, \cos, \exp : \mathbb{C} \rightarrow \mathbb{C}, \quad \sin, \cos, \exp : \mathbb{R} \rightarrow \mathbb{R} \text{ sind stetig.}}$$

Arkusfunktionen. $\sin : \mathbb{R} \rightarrow \mathbb{R}$ ist zwar nicht injektiv, jedoch durch Einschränkung ('Restriktion') des Sinus auf ein geeignetes Intervall können wir den Sinus injektiv 'machen' und somit eine Umkehrfunktion finden:

Auf dem Intervall $I := [-\frac{\pi}{2}, \frac{\pi}{2}]$ ist der Sinus streng monoton wachsend, denn: Für $z \in I$ und positives $h \in (0, \pi)$ ist

$$\begin{aligned} \sin(z+h) - \sin(z-h) &= (\sin z \cos h + \cos z \sin h) - (\sin z \cos h + \cos z \sin(-h)) \\ &= 2 \underbrace{\cos z}_{>0} \underbrace{\sin h}_{>0} > 0. \end{aligned}$$

Für vorgegebenes $x, y \in I$, $y > x$, folgt mit Setzung $z := \frac{x+y}{2}$ und $h := \frac{y-x}{2}$, dass $z \in I$ und $h \in (0, \frac{\pi}{2}]$; somit folgt mit obiger Rechnung die behauptete strenge Monotonie.

Aus der strengen Monotonie folgt die Injektivität, somit die Bijektivität von

$$\sin : D \longrightarrow \sin(D), \quad D := \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$$

Es bleibt noch die Bildmenge $\sin(D)$ zu identifizieren: Es ist $\sin(-\frac{\pi}{2}) = -1$ und $\sin \frac{\pi}{2} = +1$; wegen der Monotonie ist somit $\sin x \in [-1, 1]$ für alle $x \in [-\frac{\pi}{2}, \frac{\pi}{2}]$, d.h. $\sin(D) \subseteq [-1, 1]$. Das Argument, dass die Bildmenge sogar *gleich* der Menge $[-1, 1]$ ist, liefert der Zwischenwertsatz, angewandt auf das Intervall D : Da \sin stetig ist, gibt es für jedes $\xi \in [-1, 1]$ ein Urbild in D . Es ist also

$$\sin : \left[-\frac{\pi}{2}, \frac{\pi}{2}\right] \longrightarrow [-1, 1]$$

bijektiv. Somit gibt es eine Umkehrfunktion, genannt \arcsin (sprich: Arkussinus):

$$\arcsin : [-1, 1] \longrightarrow \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$$

Diese ist ebenfalls stetig und ebenfalls streng monoton wachsend.

Analog kann man auch für \cos sowie für $\tan := \frac{\sin}{\cos}$ Intervalle strenger Monotonie identifizieren und zugehörige Umkehrfunktionen \arccos , \arctan finden (siehe Übung).

3.5 Differenzierbarkeit

Die Stetigkeit besagt, grob gesprochen, dass sich Funktionswerte, wenn man das Argument 'wenig' variiert, nur 'wenig' verändern. Beim Differenzieren (oder Ableiten) geht es darum, genauer zu quantifizieren, wie stark sich Funktionswerte ändern, wenn man das Argument variiert. Die folgende Motivation ist sicherlich bereits aus der Schule bekannt:

Bei einer 'affin linearen' Funktion³⁷ $f(x) := ax + b$ können wir uns leicht darauf verständigen, dass die 'Steigung' der Funktion gleich a ist, dh. variiert man das Argument um einen Wert Δx , so variieren die Funktionswerte um den Wert $a \cdot \Delta x$. Wie lässt sich der Begriff der Steigung auf andere Funktionen verallgemeinern? Zunächst einmal haben beliebige Funktionen offenbar die Eigenschaft, dass ihr 'Steigungsverhalten' von Punkt zu Punkt unterschiedlich sein kann, d.h. die Steigung ist eine *lokale* Eigenschaft in dem Sinne, dass wir an einer vorgegebenen *Stelle* x_* die zugehörige Steigung bestimmen wollen. Die grundlegende Idee ist nun, dass man als Steigung der Funktion f an der Stelle x_* die Steigung der durch den Punkt $(x_*, f(x_*))$ gehenden *Tangente* an den Graphen von f festlegt (Skizze!). Die Tangente ist eine lineare Funktion, d.h. deren Steigung ist klar definiert. Es bleibt nur noch das Problem, wie man die Tangente an den Graphen einer Funktion bestimmt. Dazu die Vorgehensweise: Man startet mit *Sekanten*, also mit Geraden, die den Graphen nicht nur an der Stelle x_* , sondern auch noch an (mindestens) einer weiteren Stelle x schneiden. Dann lässt man x gegen x_* gehen und beobachtet dabei, wie sich die Steigung der Sekante verändert. Falls für $x \rightarrow x_*$ die Steigung der Sekanten einem Grenzwert entgegen gehen, so definiert man diesen Grenzwert als die Steigung, mathematisch, die *Ableitung*, von f an der Stelle x_* :

Zur Vereinfachung der Darstellung wollen wir uns auf den Fall konzentrieren, dass die Funktion auf einem offenen Intervall $D_f = (a, b)$, $-\infty \leq a < b \leq \infty$, definiert sind und wir Stellen $x_* \in D_f$ betrachten. (Man könnte auch etwas allgemeiner vorgehen und beliebige Mengen $D_f \subset \mathbb{R}$ zulassen, und x_* Punkte zulassen, die in D_f liegen und die gleichzeitig HP von D_f sind.)

Definition (Ableitung). Sei $D_f = (a, b) \subseteq \mathbb{R}$ und $f : D_f \rightarrow \mathbb{R}$. Sei $x_* \in D_f$. Wir sagen, f ist an der Stelle x_* differenzierbar (kurz: diff'bar), wenn der (Funktionen-)Grenzwert

$$\lim_{\substack{x \rightarrow x_* \\ x \in D_f \setminus \{x_*\}}} \frac{f(x) - f(x_*)}{x - x_*} =: f'(x_*)$$

existiert (als reelle Zahl). Der Grenzwert wird als *Ableitung von f an der Stelle x_** bezeichnet, kurz: $f'(x_*)$. Indem man die Differenz $h := x - x_*$ einführt (Substitution)

³⁷Beachte: In der Analysis wird eine solche Funktion manchmal als 'linear' bezeichnet; das ist etwas kritisch, denn im Sinne der Linearen Algebra handelt es sich nicht um eine lineare Funktion.

und $x = x_* + h$ eliminiert, bekommt man die alternative Darstellung

$$f'(x_*) = \lim_{\substack{h \rightarrow 0 \\ h \neq 0}} \frac{f(x_* + h) - f(x_*)}{h}$$

(Man kann dann auch den '*' weglassen und 'x' statt 'x*' schreiben).³⁸

Die Funktion f heißt *differenzierbar*, wenn f an allen Stellen $x \in D_f$ differenzierbar ist. Falls f differenzierbar ist, so ist $f'(x)$ für alle $x \in D_f$ definiert, d.h. durch $f' : x \mapsto f'(x)$ ist wieder eine Funktion $f' : D_f \rightarrow \mathbb{R}$ definiert, die *Ableitung(sfunktion) von f* .

Der Ausdruck $\frac{f(x) - f(x_*)}{x - x_*}$ wird als *Differenzenquotient* bezeichnet, in den Ingenieurwissenschaften oft (etwas ungenau) als $\frac{\Delta f}{\Delta x}$ bezeichnet; sein Grenzwert wird als *Differenzialquotient* $\frac{df}{dx}$ bezeichnet. Neben der Schreibweise $f'(x)$ ist auch noch $\frac{d}{dx}f(x)$ gebräuchlich.

Natürlich kann man die obigen Funktionengrenzwerte auch wieder mittels Folgen ausdrücken:

f ist an der Stelle x_* differenzierbar, wenn für alle Folgen (x_n) in $D \setminus \{x_*\}$ mit $\lim_{n \rightarrow \infty} x_n = x_*$ der Grenzwert $\lim_{n \rightarrow \infty} \frac{f(x_n) - f(x_*)}{x_n - x_*}$ existiert und für alle diese Folgen (x_n) der selbe ist. Oder: Wenn für alle Nullfolgen (h_n) mit $h_n \neq 0$ der Grenzwert $\lim_{n \rightarrow \infty} \frac{f(x_* + h_n) - f(x_*)}{h_n}$ existiert und immer der gleiche ist.

Beispiel. Berechne die Ableitung der Funktion $f(x) := x^2$ an der Stelle $x \in \mathbb{R}$.

$$\lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h} = \lim_{h \rightarrow 0} \frac{(x+h)^2 - x^2}{h} = \lim_{h \rightarrow 0} \frac{2hx + h^2}{h} = \lim_{h \rightarrow 0} 2x + h = 2x$$

Also: Die Funktion $f(x) = x^2$ ist differenzierbar, und die Ableitung lautet $f'(x) = 2x$. Diese Rechnung funktioniert ganz analog für $f(x) = x^n$ mit $n \in \mathbb{N}_0$; man muss dabei den Term $(x+h)^n$ mit der Binomischen Formel (1. Sem.) ausrechnen. Man erhält:

$$\boxed{f'(x) = n x^{n-1} \quad \forall n \in \mathbb{N}.} \quad (3.6)$$

Beispiel. Untersuche, ob die Funktion $f(x) = \sqrt{|x|}$ an der Stelle $x=0$ differenzierbar ist.

$$\frac{f(0+h) - f(0)}{h} = \frac{\sqrt{|h|} - \sqrt{0}}{h} = \operatorname{sgn}(h) \frac{1}{\sqrt{|h|}}$$

Für $h \rightarrow 0$ geht $\sqrt{|h|}$ gegen null; der Grenzwert des Differenzenquotienten für $h \rightarrow 0$ existiert also nicht, d.h. f ist an der Stelle $x=0$ nicht differenzierbar.

Historisches zur Analysis im allgemeinen und dem Ableitungsbegriff im besonderen.

In dieser Form formal präzise gestellt und gelöst wurde das Problem des Differenzierens erstmalig, und zwar unabhängig voneinander, von Isaac Newton und Gottfried Wilhelm Leibniz gegen Ende des 17. Jahrhunderts. Beide haben erbittert gestritten um den Ruhm der Erstentdeckung. Bekannt

³⁸Meist ist diese Darstellung "mit h " für praktische Berechnungszwecke besser geeignet als die erste.

war die Problemstellung bereits in der Antike als 'Tangentenproblem'. Newton ging das Problem 'physikalisch' mit Ableitung einer Orts-Zeit-Kurve als Momentangeschwindigkeit an, Leibniz dagegen losgelöst von Geometrie und Physik als rein mathematisches Problem. Man beachte, dass zur Zeit Newtons und Leibniz' weder der Begriff der Funktion noch der Begriff des Grenzwerts formal präzise gefasst war. Die Präzision, mit der wir hier die Analysis kennenlernen, wurde erst im (frühen) 19. Jahrhundert möglich. Die Notwendigkeit einer solchen Präzisierung lag lange auf der Hand, da vorher ständig eher diffus von 'unendlich kleinen' oder 'unendlich großen' Größen die Rede war, was oft wegen der innewohnenden Ungenauigkeit der Ausdrucksweise zu Widersprüchen führte. So wurde beispielsweise vor dem 19. Jh. nicht strikt zwischen punktweiser und gleichmäßiger Konvergenz unterschieden.

Es ist wieder offensichtlich, dass das Berechnen der Ableitung mittels Differenzenquotienten für komplizierte Funktionen sehr aufwändig ist. Fürs einfache Differenzieren brauchen wir also wieder *Rechenregeln*; diese dürften aus der Schule bekannt sein:

Satz (Rechenregeln fürs Differenzieren). Seien $f, g : (a, b) \rightarrow \mathbb{R}$, $x \in D$, und f und g differenzierbar an der Stelle x . Dann ist

- (i) $f+g$ an der Stelle x diff'bar, mit

$$\boxed{(f+g)'(x) = f'(x) + g'(x)}$$

(Summenregel),

- (ii) $f \cdot g$ an der Stelle x diff'bar, mit

$$\boxed{(f \cdot g)'(x) = f'(x)g(x) + f(x)g'(x)}$$

(Produktregel, auch: "Regel des wandernden Strichs"),

- (iii) $\frac{f}{g}$ ist an der Stelle x diff'bar, mit

$$\boxed{\left(\frac{f}{g}\right)'(x) = \frac{f'(x)g(x) - f(x)g'(x)}{g(x)^2}},$$

falls $g(x) \neq 0$.

Beweis. Der Beweis von (i) ist trivial; bei (ii) und (iii) muss man geeignete Terme einschieben; welche dies sind, kann man sich überlegen aus dem, was herauskommen soll:

- (i)

$$\begin{aligned} \lim_{h \rightarrow 0} \frac{[f(x+h) + g(x+h)] - [f(x) + g(x)]}{h} &= \lim_{h \rightarrow 0} \left(\frac{f(x+h) - f(x)}{h} + \frac{g(x+h) - g(x)}{h} \right) \\ &= \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h} + \lim_{h \rightarrow 0} \frac{g(x+h) - g(x)}{h} = f'(x) + g'(x) \end{aligned}$$

d.h. $(f+g)'(x)$ existiert und ist gleich $f'(x) + g'(x)$.

(ii)

$$\begin{aligned} & \lim_{h \rightarrow 0} \frac{f(x+h)g(x+h) - f(x)g(x)}{h} \\ &= \lim_{h \rightarrow 0} \frac{f(x+h)g(x+h) - f(x+h)g(x) + f(x+h)g(x) - f(x)g(x)}{h} \\ &= \lim_{h \rightarrow 0} \left(f(x+h) \frac{g(x+h) - g(x)}{h} + g(x) \frac{f(x+h) - f(x)}{h} \right) \\ &= \lim_{h \rightarrow 0} f(x+h) \lim_{h \rightarrow 0} \frac{g(x+h) - g(x)}{h} + g(x) \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h} = f(x)g'(x) + g(x)f'(x) \end{aligned}$$

wobei im letzten Schritt die Stetigkeit von f ausgenutzt wurde; jede differenzierbare Funktion ist in der Tat stetig, s.u..

(iii) Für (iii) beweist man zunächst, dass für $u(x) := \frac{1}{g(x)}$ gilt $u'(x) = -\frac{g'(x)}{g(x)^2}$:

$$\begin{aligned} \lim_{h \rightarrow 0} \frac{u(x+h) - u(x)}{h} &= \lim_{h \rightarrow 0} \frac{\frac{1}{g(x+h)} - \frac{1}{g(x)}}{h} \\ &= \lim_{h \rightarrow 0} \frac{1}{g(x+h)g(x)} \cdot \frac{g(x) - g(x+h)}{h} = \frac{1}{g(x)^2} \cdot (-g'(x)) \end{aligned}$$

Zusammen mit der Produktregel folgt nun die Behauptung. \square

Anwendungsbeispiel.

Für $n \in \mathbb{N}$ ist $\frac{d}{dx}x^{-n} = \frac{d}{dx}\frac{1}{x^n} = -\frac{nx^{n-1}}{x^{2n}} = -nx^{-n-1}$. Zusammen mit dem vorherigen Ergebnis über Potenzen mit *positivem* Exponenten folgt also:

$$\frac{d}{dx}x^n = nx^{n-1} \quad \forall n \in \underline{\mathbb{Z}}$$

Folgerung aus den Rechenregeln. Jedes skalares Vielfache einer differenzierbaren Funktion ist offensichtlich wieder differenzierbar, $(\alpha f)' = \alpha f'$; die Summe zweier differenzierbarer Funktionen ist ebenfalls differenzierbar, $(f+g)' = f'+g'$; die Menge aller auf einer festen Menge $I \subset \mathbb{R}$ definierten reellwertigen Funktionen erfüllt somit die Unterraumaxiome (siehe 1. Semester); sie bildet somit einen **Vektorraum**, nämlich einen Unterraum des Vektorraums $\text{Abb}(I, \mathbb{R})$ aller Funktionen von I nach \mathbb{R} .

Satz (Kettenregel). Seien $D_f, D_g \subseteq \mathbb{R}$ offene Mengen, $g : D_g \rightarrow \mathbb{R}$, $f : D_f \rightarrow \mathbb{R}$ mit $g(D_g) \subseteq D_f$. Sei $x \in D_g$ und $y := g(x) \in D_f$. Ferner seien g an der Stelle $x \in D_g$ und f an der Stelle $y := g(x)$ differenzierbar. Dann ist die Verkettung $f \circ g$ an der Stelle x differenzierbar, und es gilt

$$(f \circ g)'(x) = f'(y) \cdot g'(x) = f'(g(x)) \cdot g'(x).$$

Kurz, für überall diff'bare Funktionen: $(f \circ g)' = (f' \circ g) \cdot g'$. Oder in 'Differenzial-Schreibweise': $\frac{df}{dx} = \frac{df}{dg} \cdot \frac{dg}{dx}$

Um weitere Funktionen, wie z.B. $f(x) = e^{-x^2}$ oder $f(x) = \sin(5x + 3(\frac{x+1}{x^2-1})^7)$ mit Hilfe der Rechenregeln abzuleiten, brauchen wir zunächst die Ableitung von den elementaren Funktionen exp, sin, cos:

Satz. Die Funktionen exp, sin, cos : $\mathbb{R} \rightarrow \mathbb{R}$ sind differenzierbar, und es ist

$$\boxed{\exp' x = \exp x}, \quad \boxed{\sin' x = \cos x}, \quad \boxed{\cos' x = -\sin x}. \quad (3.7)$$

Beweis.

- (i) Wir wollen $\lim_{h \rightarrow 0} \frac{\exp(x+h) - \exp(x)}{h} = \exp(x)$ zeigen, also, gleichbedeutend, $\lim_{h \rightarrow 0} \frac{\exp(x+h) - \exp(x)}{h} - \exp(x) = 0$. Dazu die Untersuchung:

$$\begin{aligned} \left| \frac{\exp(x+h) - \exp(x)}{h} - \exp(x) \right| &= e^x \cdot \left| \frac{e^h - 1 - h}{h} \right| \leq e^x \cdot \frac{1}{|h|} \left(\frac{|h|^2}{2!} + \frac{|h|^3}{3!} + \frac{|h|^4}{4!} + \dots \right) \\ &= e^x |h| \left(\frac{1}{2!} + \frac{|h|}{3!} + \frac{|h|^2}{4!} + \dots \right) \leq e^x |h| \left(\frac{1}{0!} + \frac{|h|}{1!} + \frac{|h|^2}{2!} + \dots \right) = e^x \underbrace{|h|}_{\rightarrow 0} \underbrace{e^{|h|}}_{\rightarrow e^0=1} \xrightarrow{(h \rightarrow 0)} 0 \end{aligned}$$

wobei am Ende die Stetigkeit von exp verwendet wurde.

- (ii) Es ist

$$\frac{\sin(x+h) - \sin(x)}{h} = \frac{\sin x \cos h + \cos x \sin h - \sin x}{h} = \sin x \cdot \underbrace{\frac{\cos h - 1}{h}}_{\rightarrow 0?} + \cos x \cdot \underbrace{\frac{\sin h}{h}}_{\rightarrow 1?}$$

Um zu zeigen, dass dies gegen $\cos x$ geht für $h \rightarrow 0$, müssen wir nur zeigen, dass

$$\boxed{\lim_{h \rightarrow 0} \frac{\sin h}{h} = 1}, \quad \lim_{h \rightarrow 0} \frac{\cos h - 1}{h} = 0. \quad (3.8)$$

Um die erste Aussage von (3.8) zu zeigen, betrachten wir, analog zu (i),

$$\begin{aligned} \left| \frac{\sin h}{h} - 1 \right| &= \frac{1}{|h|} \cdot \left| \left(h - \frac{h^3}{3!} + \frac{h^5}{5!} - \frac{h^7}{7!} + \dots \right) - h \right| \leq |h|^2 \cdot \left(\frac{1}{3!} + \frac{h^2}{5!} + \frac{h^4}{7!} + \dots \right) \\ &\leq |h|^2 \cdot \left(1 + \frac{h^2}{2!} + \frac{h^4}{4!} + \frac{h^6}{6!} + \dots \right) \leq |h|^2 \cdot e^{|h|} \xrightarrow{(h \rightarrow 0)} 0. \end{aligned}$$

Die zweite Aussage von (3.8) zeigt man ganz analog.

- (iii) z.Ü: Zeigen Sie (iii). Sie dürfen (3.8) verwenden. □

In Kap. 3.6.1 werden wir eine einfachere Methode kennenlernen, Potenzreihen wie sin, cos, exp abzuleiten!

Beispiele (Anwendung der Kettenregel und von (3.7)):

Es ist

$$\begin{aligned}\frac{d}{dx}e^{5x} &= 5e^{5x}, \\ \frac{d}{dx}e^{-x^2} &= e^{-x^2} \cdot (-2x), \\ \frac{d}{dx}\sin\left(5x + 3\left(\frac{x+1}{x^2-1}\right)^7\right) &= \cos\left(5x + 3\left(\frac{x+1}{x^2-1}\right)^7\right) \cdot \left(5 + 3 \cdot 7 \cdot \left(\frac{x+1}{x^2-1}\right)^6 \cdot \frac{1 \cdot (x^2-1) - (x+1) \cdot 2x}{(x^2-1)^2}\right)\end{aligned}$$

Höhere Ableitungen. Da für differenzierbares $f : D \rightarrow \mathbb{R}$ die Ableitung f' wieder eine Funktion von D nach \mathbb{R} darstellt, kann man untersuchen, ob f' wiederum auf D differenzierbar ist. Wenn ja, so bezeichnet man die Ableitung $f'' := (f')'$ von f' als *zweite Ableitung von f* . So weiter erhält man, falls sie existieren, $f', f'', f''', f^{(4)}, f^{(5)}, \dots$

z.Ü.: Berechnen Sie $(f \cdot g \cdot h)', (f \cdot g)''', (f \circ g)''$.

Satz (Differenzieren der Umkehrfunktion). Seien $A, B \subseteq \mathbb{R}$ offene Mengen und $f : A \rightarrow B$ bijektiv und differenzierbar an der Stelle $x \in A$. Dann ist die Umkehrfunktion $f^{-1} : B \rightarrow A$ differenzierbar an der Stelle $y = f(x) \in B$, sofern $f'(x) \neq 0$, und es gilt:

$$(f^{-1})'(y) = \frac{1}{f'(x)} = \frac{1}{f'(f^{-1}(y))}. \quad (3.9)$$

Zum Beweis: Wir wollen hier nicht die Differenzierbarkeit von f^{-1} zeigen, sondern uns nur davon überzeugen, dass die Formel (3.9) gilt. Dazu differenzieren wir die Gleichung $f \circ f^{-1} = \text{Id}$, also $f(f^{-1}(y)) = y$, nach der Kettenregel. Wir erhalten

$$f'(f^{-1}(y)) \cdot (f^{-1})'(y) = 1$$

Per Division folgt (3.9).³⁹

Bemerkungen:

– Die Formel (3.9) braucht man sich nicht unbedingt zu merken; es ist *einfacher*, wenn man sich merkt, wie man diese Formel *herleitet*.

– Indem man die Gleichung $f(f^{-1}(y)) = y$ ein zweites, drittes, ... mal ableitet, kann man Formeln für die zweite, dritte, ... Ableitung von f^{-1} herleiten (s.Ü.)

Anwendungen der Regel (3.9): Bestimme die Ableitungen von \sqrt{x} , \arcsin , \ln , a^x , x^x . Hier exemplarisch für den Logarithmus durchgeführt:

” \ln ” ist die Umkehrfunktion von $f := \exp$. Wir wissen, dass f' ebenfalls \exp ist. Wir erhalten:

$$\ln'(y) = \frac{1}{\exp'(\ln(y))} = \frac{1}{\exp(\ln(y))} = \frac{1}{y}$$

Ergebnis:

$$\boxed{\ln'(x) = \frac{1}{x}} \quad \forall x > 0$$

z.Ü.: Berechnen Sie für $x > 0$, $\alpha \in \mathbb{R}$ die Ableitung von $f(x) = x^\alpha$, $f(x) = x^x$, $f(x) = x^{x^x}$.

³⁹In (m.E. nicht empfehlenswerter) Ingenieur-Kurzschreibweise (Leibniz-Notation) kurz: $\frac{dx}{dy} = \frac{1}{\frac{dy}{dx}}$.

Hinweis: Verwenden Sie $a^b = e^{b \ln a}$. Beachten Sie auch, dass a^{b^c} immer im Sinne von $a^{(b^c)}$ zu verstehen ist. (Grund: Die andere Klammerung $(a^b)^c$ ließe sich einfacher als a^{bc} schreiben.)

Satz (Beziehung zwischen Differenzierbarkeit und Stetigkeit).

Sei $D \subseteq \mathbb{R}$ offen und Sei $f : D \rightarrow \mathbb{R}$ differenzierbar an der Stelle $x_* \in D$. Dann ist f stetig an der Stelle x_* .

Kurz: Jede diffbare Funktion ist stetig.

Die Diff'barkeit ist in diesem Sinne eine 'stärkere' Eigenschaft als die Stetigkeit.

Beweis. Diff'barkeit an der Stelle x_* bedeutet, dass für jede Folge (x_n) mit $\lim_n x_n = x_*$ gilt $\lim_n \frac{f(x_n) - f(x_*)}{x_n - x_*} = f'(x_*)$, also dass $\frac{f(x_n) - f(x_*)}{x_n - x_*} - f'(x_*) \xrightarrow{(n \rightarrow \infty)} 0$. Multiplikation mit der Folge $(x_n - x_*)$ liefert, da diese Folge ebenfalls gegen null konvergiert, nach den Rechenregeln für konvergente Folgen, dass die Produktfolge ebenfalls gegen null geht, also $f(x_n) - f(x_*) - f'(x_*) \cdot (x_n - x_*) \xrightarrow{(n \rightarrow \infty)} 0$. Addition der Nullfolge $(f'(x_*) \cdot (x_n - x_*))$ liefert nach den Rechenregeln, dass $f(x_n) - f(x_*)$ gegen null geht, also $\lim_n f(x_n) = f(x_*)$.

Ist die Ableitung einer differenzierbaren Funktion immer stetig? Bei den "meisten" diff'baren Funktionen ist die Ableitung in der Tat stetig. Jedoch kann man auch Gegenbeispiele finden.

Ein Gegenbeispiel ist die Funktion

$$f(x) := \begin{cases} x^2 \sin\left(\frac{1}{x}\right) & , x \neq 0 \\ 0 & , x = 0 \end{cases}$$

Für diese Funktion existiert $f'(0)$, denn

$$f'(0) = \lim_{h \rightarrow 0} \frac{f(h) - f(0)}{h} = \lim_{h \rightarrow 0} \frac{h^2 \sin \frac{1}{h} - 0}{h} = \lim_{h \rightarrow 0} \underbrace{h \cdot \sin \frac{1}{h}}_{\text{beschr.}} = 0,$$

aber es ist, mit

$$f'(x) = 2x \sin \frac{1}{x} - \cos \frac{1}{x} \quad \forall x \neq 0,$$

$\lim_{x \rightarrow 0} f'(x) \neq 0$ (der Limes existiert nicht), also $\lim_{x \rightarrow 0} f'(x) \neq f'(0)$, f' ist somit unstetig.

Bezeichnung. Die Menge aller auf einem Intervall $I = (a, b)$, $-\infty \leq a < b \leq \infty$, definierten Funktionen $f : I \rightarrow \mathbb{R}$, die auf I differenzierbar sind, und deren Ableitung auf I stetig ist, wird mit $C^1(I)$ oder $\mathcal{C}^1(I)$ bezeichnet.

Die Menge aller Funktionen $f : I \rightarrow \mathbb{R}$, die auf I n -mal differenzierbar sind, und deren n -te Ableitung stetig ist, wird mit $C^n(I)$ oder $\mathcal{C}^n(I)$ bezeichnet. Die Menge der Funktion $f : I \rightarrow \mathbb{R}$, für die *alle* Ableitungen $f^{(n)}$, $n \in \mathbb{N}$, existieren, wird mit $C^\infty(I)$ bezeichnet.

Zum Zusammenhang von Ableitung und Monotonieverhalten einer Funktion (quasi: zur Anschaulichen Bedeutung der Ableitung) siehe S.63.

3.6 Eigenschaften differenzierbarer Funktionen: Folgen differenzierbarer Funktionen, Extremstellensuche, Mittelwertsatz, Regel von l'Hospital, Taylor-Formel, Newton-Verfahren

3.6.1 Differenzierbarkeit von Funktionenfolgen, insbesondere Potenzreihen

Wir hatten in Kap. 3.4 bewiesen, dass bei einer Folge von stetigen Funktionen, die *gleichmäßig* konvergent ist, die Grenzfunktion ebenfalls stetig ist. Der Beweis verwendete die $\frac{\epsilon}{3}$ -Technik. Es gibt einen entsprechenden Satz auch für die Differenzierbarkeit (hier ohne Beweis):

Satz (Differenzierbarkeit einer Grenzfunktion). Seien $f_n : (a, b) \rightarrow \mathbb{R}$ differenzierbare Funktionen; die Folge (f_n) sei punktweise konvergent; die Grenzfunktion heie f . Sei die Folge der Ableitungen (f'_n) gleichmig konvergent auf (a, b) . Dann ist auch die Grenzfunktion f differenzierbar, und fr ihre Ableitung gilt

$$f'(x) = \lim_{n \rightarrow \infty} f'_n(x) \quad \forall x \in (a, b).$$

Wir wollen prfen, ob dieser Satz fr Potenzreihen $\sum_{k=0}^{\infty} a_k (x-x_0)^k$ allgemein anwendbar ist. Wir setzen also $f_n(x) := s_n(x) = \sum_{k=0}^n a_k (x-x_0)^k$ und setzen voraus, dass die Folge $f_n(x)$ auf einem Intervall $(x_0 - R, x_0 + R)$, mit Konvergenzradius $R > 0$, konvergent ist. Wir mssen, um obigen Satz anwenden zu knnen, prfen, ob die Folge der Ableitungen, also

$$s'_n(x) = \sum_{k=1}^n k a_k (x-x_0)^{k-1} = \sum_{k=0}^{n-1} (k+1) a_{k+1} (x-x_0)^k$$

gleichmig konvergent ist. Wir sehen, dass neben $s_n(x)$ auch $s'_n(x)$ eine Potenzreihe ist, somit nach Satz (s.S. 50) auf jedem kompakten Teilintervall des Konvergenzbereichs $(x_0 - \tilde{R}, x_0 + \tilde{R})$, wobei $\tilde{R} \in [0, \infty]$ der Konvergenzradius von $s'_n(x)$ ist, gleichmig konvergent ist. Wir mssen nur noch \tilde{R} berechnen (wir hoffen, dass $\tilde{R} = R$ ist):

$$\tilde{R} = \frac{1}{\limsup \sqrt[n+1]{(n+1)|a_n|}} = \frac{1}{\underbrace{\lim_n \sqrt[n]{n} \cdot \limsup \sqrt[n+1]{|a_n|}}_{=1}} = R$$

Somit ist die Folge der Ableitungen also auf dem ganzen offenen Konvergenzintervall der Reihe konvergent; auf jedem kompakten Teilintervall gleichmig konvergent, somit ist nach obigem Satz die Grenzfunktion differenzierbar (und zwar, indem wir das kompakte Teilintervall fr vorgegebenes x hinreichend gro whlen, auf ganz $(x_0 - R, x_0 + R)$, und

man bekommt die Ableitung der Grenzfunktion, indem man den Grenzwert von $s'_n(x)$ nimmt:

$$\frac{d}{dx} \underbrace{\left(\sum_{k=0}^{\infty} a_k (x-x_0)^k \right)}_{=f(x)} \stackrel{\text{Satz}}{=} \lim_{n \rightarrow \infty} \underbrace{\frac{d}{dx} \sum_{k=0}^n a_k (x-x_0)^k}_{=s'_n(x)} \stackrel{(*)}{=} \lim_{n \rightarrow \infty} \sum_{k=0}^n \frac{d}{dx} a_k (x-x_0)^k = \sum_{k=1}^{\infty} a_k k (x-x_0)^{k-1}$$

(an der Stelle (*) haben wir die Rechenregel benutzt, dass wir bei endlichen(!) Summen gliedweise differenzieren dürfen)

Kurz: Potenzreihen darf man 'gliedweise' ableiten:

$$\boxed{\frac{d}{dx} \sum_{k=0}^{\infty} a_k (x-x_0)^k = \sum_{k=0}^{\infty} \frac{d}{dx} a_k (x-x_0)^k}$$

Anwendungsbeispiel: Die Funktionen \exp , \sin , \cos sind differenzierbar, und es gilt

$$\exp'(x) = \frac{d}{dx} \left(\sum_{k=0}^{\infty} \frac{x^k}{k!} \right) = \sum_{k=0}^{\infty} \frac{d}{dx} \frac{x^k}{k!} = \sum_{k=1}^{\infty} \frac{k x^{k-1}}{k!} \stackrel{(*)}{=} \sum_{k=0}^{\infty} \frac{x^k}{k!} = \exp(x),$$

wobei wir an der Stelle (*) eine Indextransformation vorgenommen haben.

Analog: Es ist

$$\begin{aligned} \sin'(x) &= \frac{d}{dx} \left(\sum_{k=0}^{\infty} (-1)^k \frac{x^{2k+1}}{(2k+1)!} \right) = \sum_{k=0}^{\infty} \frac{d}{dx} (-1)^k \frac{x^{2k+1}}{(2k+1)!} \\ &= \sum_{k=0}^{\infty} (-1)^k \frac{(2k+1) x^{2k}}{(2k+1)!} = \sum_{k=0}^{\infty} (-1)^k \frac{x^{2k}}{(2k)!} = \cos(x) \end{aligned}$$

analog: $\cos'(x) = -\sin(x)$.

Diese kurzen Rechnungen ersetzen die etwas mühsameren Rechnungen auf S. 56.

3.6.2 Extremstellen und Mittelwertsätze

Def. (relative und globale Extremstellen, Extrema). Sei $f : D \rightarrow \mathbb{R}$, $D \subseteq \mathbb{R}$.

Ein $x_* \in D$ heißt *globale Maximalstelle* (*globale Minimalstelle*), falls $f(x_*) \geq f(x)$ [bzw.: $f(x_*) \leq f(x)$] für alle $x \in D$. Der Wert $f(x_*)$ heißt dann das (*globale*) *Maximum* (bzw.: das (*globale*) *Minimum*).

Ein $x_* \in D$ heißt *relative Maximalstelle* (*relative Minimalstelle*), falls $f(x_*) \geq f(x)$ [bzw.: $f(x_*) \leq f(x)$] für alle $x \in D \cap (x_* - \epsilon, x_* + \epsilon)$ für ein $\epsilon > 0$. Der Wert $f(x_*)$ heißt dann *relatives Maximum* (bzw.: *relatives Minimum*).

Ein Intervall $U_\epsilon(x_*) := (x_* - \epsilon, x_* + \epsilon)$ heißt auch (*offene*) ϵ -*Umgebung* von x_* .

Statt *relatives* Extremum ist auch der Begriff *lokales* Extremum gebräuchlich.

Wie berechnet man Extremstellen? Ein notwendiges Kriterium für eine (relative) Extremstelle ist (für differenzierbare Funktionen), wie bereits aus der Schule bekannt sein dürfte, $f'(x_*)=0$; allerdings

- natürlich nur wenn f differenzierbar ist und
- nur für Punkte *im Inneren* von D ; nicht für Randpunkte⁴⁰:

Satz (notwendiges Kriterium für Extremstellen). Sei $D \subset \mathbb{R}$ und sei $x_* \in \overset{\circ}{D}$ eine relative Extremstelle. Dann ist

$$\boxed{f'(x_*)=0.} \quad (3.10)$$

Beweis. Sei $x_* \in \overset{\circ}{D}$ relative Maximalstelle (für Minimalstellen geht's analog), also $f(x) - f(x_*) \leq 0$ für alle x aus einer ϵ -Umgebung von x_* . Division durch $x - x_*$ liefert

$$\frac{f(x) - f(x_*)}{x - x_*} \begin{cases} \leq 0 & , \text{für } x > x_* \\ \geq 0 & , \text{für } x < x_* \end{cases} \quad (*)$$

Nach Voraussetzung existiert $f'(x_*) = \lim_n \frac{f(x_n) - f(x_*)}{x_n - x_*}$ für alle Folgen (x_n) mit $x_n \xrightarrow{(n \rightarrow \infty)} x_*$. Nehmen wir nun eine Folge (x_n) aus der ϵ -Umgebung, die von rechts gegen x_* geht, so ist wegen (*) $\lim_n \frac{f(x_n) - f(x_*)}{x_n - x_*} \leq 0$. Neben wir eine Folge (x_n) aus der ϵ -Umgebung, die von links gegen x_* geht, so ist wegen (*) $\lim_n \frac{f(x_n) - f(x_*)}{x_n - x_*} \geq 0$. Es muss also $f'(x_*) = \lim_n \frac{f(x_n) - f(x_*)}{x_n - x_*}$ sowohl ≤ 0 als auch ≥ 0 sein, also ist $f'(x_*)=0$. \square

Eine Stelle x , die (3.10) erfüllt, wird auch als *kritische Stelle* der Funktion f bezeichnet. Wie man zuverlässig entscheidet, ob es sich bei einer kritischen Stelle um eine *relative* Maximal- oder Minimalstelle handelt, werden wir später noch diskutieren (Kriterium, das f'' verwendet).

Beachte auch, dass (3.10) nicht *hinreichend* ist für das Vorliegen einer relativen Extremstelle; Gegenbeispiel $f(x) := x^3$ mit $f'(0) = 0$, obwohl $x_* = 0$ keine relative Extremstelle ist.

Anwendungsbeispiel. Berechne das (globale) Minimum und Maximum von $f(x) := e^x - x$, $f : [-1, 1] \rightarrow \mathbb{R}$.

Vorbemerkung: Wir wissen, dass es Minimum und Maximum geben muss, denn f ist stetig und $[-1, 1]$ ist kompakt.

Wir bestimmen zunächst alle 'potenziellen' relativen Extremstellen; also alle kritischen Stellen, das sind alle Stellen, an denen das notwendige Kriterium (3.10) erfüllt ist:

$$f'(x) = e^x - 1, \text{ also } f'(x) = 0 \Leftrightarrow e^x = 1 \Leftrightarrow x = \ln(1) = 0.$$

Nun Vorsicht: Das Kriterium ist nur notwendig für Extremstellen, die *im Inneren* des

⁴⁰Klar, denn betrachte $f(x) := x$ auf $[0, 1]$: Extrema existieren (am Rande), obwohl $f'(x)$ immer ungleich null ist.

Gebietes liegen; es sind also neben $x = 0$ auch $x = -1$ und $x = +1$ möglicherweise relative Extremstellen.

Wir wissen nun: Sowohl (globales) Minimum als auch Maximum wird irgendwo an den (endlich vielen!) Stellen $0, -1, +1$ angenommen. Um die globalen Extrema zu finden, müssen wir nur f an diesen drei Stellen ausrechnen und untereinander vergleichen:

$$f(-1) = \frac{1}{e} + 1, \quad f(0) = 1, \quad f(1) = e - 1.$$

Es folgt:

$$\max_{x \in [-1, 1]} f(x) = e - 1, \quad \min_{x \in [-1, 1]} f(x) = 1$$

Falls man auf *offenen* oder *unbeschränkten* Intervallen nach globalen Extrema sucht, so muss man die *Funktionengrenzwerte* von f an diesen Rändern berechnen und mit den Funktionswerten an den kritischen Stellen vergleichen, um zu entscheiden, ob es ein globales Maximum/Minimum gibt und wie groß es ist; also hat man z.B. auf $D := (a, b)$ nur einen kritischen Punkt $x \in (a, b)$ mit $f(x) = 3$ und $\lim_{x \rightarrow a} f(x) = 5$ und $\lim_{x \rightarrow b} f(x) = 4$, so ist das globale Minimum gleich 3, und ein globales Maximum existiert nicht.

Der folgende Satz von Rolle ist ein Hilfssatz; seine Bedeutung liegt darin, dass man aus ihm sofort den sog. Mittelwertsatz folgern kann.

Satz (Satz von Rolle). Sei $f : [a, b] \rightarrow \mathbb{R}$ stetig und auf (a, b) differenzierbar. Sei $f(a) = f(b)$. Dann gibt es ein $\xi \in (a, b)$ mit $f'(\xi) = 0$.

Beweis. Der Satz von Rolle folgt sofort aus dem obigen Kriterium $f'(x_*) = 0$ für Extremstellen:

Zunächst mal ist der Trivialfall möglich, dass $f = \text{const}$ ist. In dem Fall ist $f'(x) = 0$ für alle $x \in (a, b)$ und die Behauptung ist offensichtlich erfüllt.

Nun zum Fall, dass f nicht konstant ist. In dem Fall wird, wegen der Stetigkeit, auf $[a, b]$ ein Maximum und ein Minimum angenommen, und das Maximum ist echt größer als das Minimum. Da $f(a) = f(b)$ können nicht sowohl Maximum als auch Minimum am Rand angenommen werden; es gibt also eine Extremstelle im Inneren des Intervalls. Nach dem obigen notwendigen Kriterium muss dort f' den Wert null annehmen. \square

Nun der eigentliche Satz, weswegen wir den Satz von Rolle brauchen:

Satz (Mittelwertsatz). Sei $f : [a, b] \rightarrow \mathbb{R}$ stetig und auf (a, b) differenzierbar. Dann gibt es ein $\xi \in (a, b)$ mit

$$f'(\xi) = \frac{f(b) - f(a)}{b - a}.$$

Anschaulich: Es gibt eine Stelle ξ , an der die Steigung der Kurve gleich der "mittleren Steigung" der Kurve (d.h. gleich der Steigung der direkten Verbindungslinie der beiden Kurvenendpunkte) ist.

Beweis. Der Mittelwertsatz kann direkt aus dem Satz von Rolle gefolgert werden. Wir müssen dazu nur unsere Funktion f "verbiegen" (scheren), so dass die Funktionswerte an den Stellen a und b gleich groß werden; wir ziehen dazu von f eine geeignete lineare Funktion ab: Die Funktion

$$F(x) := f(x) - \frac{f(b) - f(a)}{b - a} \cdot (x - a)$$

erfüllt in der Tat $F(a) = f(a) = F(b)$; also gibt es nach dem Satz von Rolle ein $\xi \in (a, b)$ mit $0 = F'(\xi) = f'(\xi) - \frac{f(b) - f(a)}{b - a}$. Es folgt die Behauptung. \square

Anwendung des / Folgerung aus dem Mittelwertsatz (Ableitung und Monotonieverhalten von Funktionen):

Ist $f : [a, b] \rightarrow \mathbb{R}$ differenzierbar auf (a, b) mit $f' \geq 0$ (bzw. > 0 bzw. ≤ 0 bzw. < 0) auf (a, b) , so ist f auf $[a, b]$ schwach monoton wachsend (bzw. streng monoton wachsend bzw. schwach monoton fallend bzw. streng monoton fallend.).

Ist umgekehrt $f : [a, b] \rightarrow \mathbb{R}$ schwach monoton wachsend (fallend) auf $[a, b]$ und ist f differenzierbar mit stetiger Ableitung, so ist $f' \geq 0$ ($f' \leq 0$) auf (a, b) .⁴¹

Begründung: Die erste Teilaussage folgt, indem man $x_1, x_2 \in [a, b]$, $x_1 < x_2$ vorgibt, und auf das Intervall $[x_1, x_2]$ den Mittelwertsatz anwendet. Die zweite Teilaussage folgt sofort aus der ersten, indem man einen Widerspruchsbeweis führt und die Stetigkeit von f' ausnutzt, die dafür sorgt, dass wenn f' an einer Stelle ein bestimmtes Vorzeichen hat, auch in einer (kleinen) Umgebung dieser Stelle dieses Vorzeichen gilt.

Der folgende Satz ist wieder eher als Hilfssatz zu sehen. Wir werden ihn verwenden, um die Regel von l'Hospital sowie den Satz von Taylor herzuleiten.

Satz (Verallgemeinerter Mittelwertsatz). Seien $f, g : [a, b] \rightarrow \mathbb{R}$ stetig und auf (a, b) differenzierbar, und es sei $g'(x) \neq 0$ für alle $x \in (a, b)$. Dann gibt es ein $\xi \in (a, b)$ mit⁴²

$$\frac{f'(\xi)}{g'(\xi)} = \frac{f(b) - f(a)}{g(b) - g(a)}.$$

Beweis: Wir wenden den Satz von Rolle auf die Funktion

$$F(x) := [f(b) - f(a)]g(x) - [g(b) - g(a)]f(x)$$

an. Es ist $F(a) = f(b)g(a) - f(a)g(b)$, $F(b) = f(b)g(a) - f(a)g(b)$. Es gibt also ein $\xi \in (a, b)$ mit $0 = F'(\xi) = [f(b) - f(a)]g'(\xi) - [g(b) - g(a)]f'(\xi)$. Umstellen liefert die Behauptung. \square

⁴¹Bachte, dass man aus der *strengen* Monotonie nicht etwa ein *striktes* Vorzeichen von f' folgern kann, siehe Bsp. $f(x) = x^3$.

⁴²Technische Randbemerkung: Wir müssen $g(b) - g(a) \neq 0$ nicht extra fordern; dies folgt aus $g'(\xi) \neq 0 \forall \xi \in (a, b)$ mit dem Satz von Rolle.

3.6.3 Die Regel von l'Hospital

Die Regel von l'Hospital⁴³ liefert Vorgehensweisen zur Berechnung von Grenzwerten $\lim_{x \rightarrow x_*} \frac{f(x)}{g(x)}$ der Form " $\frac{0}{0}$ " (z.B. $\lim_{x \rightarrow 0} \frac{\sin x}{x}$) oder auch " $\frac{\infty}{\infty}$ " (z.B. $\lim_{x \rightarrow +\infty} \frac{e^x}{x}$), für die unsere bisherigen Grenzwert-Rechenregeln (S. 38) i.a. kein Ergebnis liefern:

Satz (Regel von l'Hospital). Sei $-\infty \leq a < b \leq +\infty$, sei $x_* \in [a, b]$, und seien die Funktionen $f, g : (a, b) \rightarrow \mathbb{R}$ differenzierbar auf $(a, b) \setminus \{x_*\}$ mit $g'(x) \neq 0 \forall x \in (a, b) \setminus \{x_*\}$. Es gelte ferner, dass

$$\lim_{x \rightarrow x_*} f(x) = 0, \quad \lim_{x \rightarrow x_*} g(x) = 0, \quad (3.11)$$

und der Grenzwert

$$\lim_{x \rightarrow x_*} \frac{f'(x)}{g'(x)} \in \mathbb{R} \cup \{\pm\infty\}$$

existiere, ggf. auch als uneigentlicher Grenzwert.

Dann existiert auch der Grenzwert $\lim_{x \rightarrow x_*} \frac{f(x)}{g(x)}$ (ggf. uneigentlich), und es ist

$$\lim_{x \rightarrow x_*} \frac{f(x)}{g(x)} = \lim_{x \rightarrow x_*} \frac{f'(x)}{g'(x)}$$

Die Aussage des Satzes bleibt gültig, wenn man die Voraussetzung (3.11) ersetzt durch

$$\lim_{x \rightarrow x_*} f(x) = \pm\infty, \quad \lim_{x \rightarrow x_*} g(x) = \pm\infty.$$

Beispiele:

(a) $\lim_{x \rightarrow 0} \frac{\sin x}{x}$ (siehe (3.8)):

Da sowohl $\lim_{x \rightarrow 0} \sin x = 0$ (Stetigkeit des Sinus) als auch $\lim_{x \rightarrow 0} x = 0$, ist die Regel von l'Hospital anwendbar, und es gilt

$$\lim_{x \rightarrow 0} \frac{\sin x}{x} = \lim_{x \rightarrow 0} \frac{\cos x}{1} = 1$$

wobei im letzten Schritt die Stetigkeit des Kosinus verwendet wurde.

Beachte: Das war viel einfacher als die Rechnung auf S. 56 zur Berechnung dieses Grenzwerts!

(b) $\lim_{x \rightarrow +\infty} \frac{e^x}{x}$:

Es gilt sowohl $\lim_{x \rightarrow +\infty} e^x = \infty$ als auch $\lim_{x \rightarrow +\infty} x = \infty$, also ist die Regel von l'Hospital anwendbar, und es gilt

$$\lim_{x \rightarrow +\infty} \frac{e^x}{x} = \lim_{x \rightarrow +\infty} \frac{e^x}{1} = +\infty$$

⁴³Guillaume François Antoine Marquis de l'Hospital, 1661-1704, Autor eines der frühesten Bücher über Analysis

Man kann die Regel u.U. auch mehrfach hintereinander anwenden; z.B. ist

$$\lim_{x \rightarrow +\infty} \frac{e^x}{x^n} = \lim_{x \rightarrow +\infty} \frac{e^x}{nx^{n-1}} = \dots = \lim_{x \rightarrow +\infty} \frac{e^x}{n!} = +\infty \quad \forall n \in \mathbb{N},$$

kurz: "Die Exponentialfunktion wächst für $n \rightarrow \infty$ stärker als jede Potenz (und somit auch: als jedes Polynom)".

Analog kann man mit der Regel von l'Hospital zeigen: **"Die Logarithmusfunktion wächst für $x \rightarrow \infty$ langsamer als jede Potenz".**

- (c) Beachte: Auch für Grenzwerte der Form " $0 \cdot (\pm\infty)$ " kann die Regel von l'Hospital nützlich sein, z.B. kann man $\lim_{x \rightarrow 0} x \ln x = \lim_{x \rightarrow 0} \frac{\ln x}{\frac{1}{x}}$ schreiben, was vom Typ " $\frac{\infty}{\infty}$ " ist; man bekommt

$$\lim_{x \rightarrow 0} x \ln x = \lim_{x \rightarrow 0} \frac{\ln x}{\frac{1}{x}} = \lim_{x \rightarrow 0} \frac{\frac{1}{x}}{-\frac{1}{x^2}} = \lim_{x \rightarrow 0} (-x) = 0$$

Kurz: Das x dominiert das $\ln x$, zwingt das Produkt gegen 0, den Grenzwert von x . Auch hier wieder gilt das entsprechende ebenfalls für Potenzen von x , wie man durch wiederholten anwenden von l'Hospital leicht zeigen kann:

$$\boxed{\lim_{x \rightarrow 0} x^n \ln x = 0 \quad \forall n \in \mathbb{N}}$$

- (d) Zur Übung (dies sind Klausuraufgaben von 2009):

$$\lim_{x \rightarrow \infty} \frac{e^{\frac{2}{x}} - 1}{\frac{1}{x+3}}, \quad \lim_{x \rightarrow 0} \frac{\sin(e^{x^2} - 1)}{e^{\cos x} - e}$$

Hinweis: Neben l'Hospital auch (3.8) verwenden!⁴⁴
Weitere Beispiele in den Übungen.

Hinweis: Man darf nicht vergessen, vor der Anwendung von l'Hospital *die Voraussetzungen zu überprüfen* (d.h. ob wirklich der Fall " $\frac{0}{0}$ " oder " $\frac{\pm\infty}{\pm\infty}$ " vorliegt); andernfalls kann das Ergebnis falsch werden! Auch sollte man, bevor man die Regel von l'Hospital (ein- oder mehrfach) anwendet, immer zuvor prüfen, ob man Terme (Faktoren), deren Verhalten man kennt, herausziehen kann; z.B. für

$$\lim_{x \rightarrow 0} \frac{\sin(x^2) (e^x - 1 - x)}{x^3 \sin x}$$

wäre die unmittelbare Anwendung der Regel von l'Hospital etwas aufwändig, da beim Ableiten von Zähler und Nenner längere Terme entstehen. Wenn man sieht, dass man

⁴⁴Dieser Hinweis wurde in der Klausur nicht gegeben

den Term $\lim_{x \rightarrow 0} \frac{\sin(x^2)}{x^2} = \lim_{y \rightarrow 0} \frac{\sin y}{y} = 1$ herausziehen kann, muss man l'Hospital nur auf den Rest anwenden:

$$\begin{aligned} \lim_{x \rightarrow 0} \frac{\sin(x^2) (e^x - 1 - x)}{x^3 \sin x} &= \lim_{x \rightarrow 0} \frac{\sin(x^2)}{x^2} \cdot \lim_{x \rightarrow 0} \frac{e^x - 1 - x}{x \sin x} \stackrel{(*)}{=} 1 \cdot \lim_{x \rightarrow 0} \frac{e^x - 1}{\sin x + x \cos x} \\ &\stackrel{(*)}{=} \lim_{x \rightarrow 0} \frac{e^x}{2 \cos x - x \sin x} = \frac{e^0}{2 \cos 0 - 0} = \frac{1}{2} \end{aligned}$$

Dabei wurde an den Stellen (*) l'Hospital angewendet.⁴⁵

Noch kurz zum **Beweis** (Beweisskizze) der Regel von l'Hospital: Wir führen Sie auf den verallgemeinerten Mittelwertsatz zurück:

Wir betrachten zunächst den Fall " $x_* \in \mathbb{R}$ ": Wir können nach Voraussetzung f und g stetig in x_* durch $f(x_*) := 0$, $g(x_*) := 0$ fortsetzen. Wir erhalten nach dem verallgemeinerten Mittelwertsatz für jedes x ein $\xi = \xi(x)$ zwischen x und x_* , so dass

$$\lim_{x \rightarrow x_*} \frac{f(x)}{g(x)} = \lim_{x \rightarrow x_*} \frac{f(x) - \overbrace{f(x_*)}^{=0}}{g(x) - \underbrace{g(x_*)}_{=0}} = \lim_{x \rightarrow x_*} \frac{f'(\xi)}{g'(\xi)} = \lim_{\xi \rightarrow x_*} \frac{f'(\xi)}{g'(\xi)}$$

Nun lassen wir x gegen x_* gehen. Dann muss offensichtlich auch ξ gegen x_* gehen, was den letzten Schritt in obiger Rechnung begründet, und es folgt die Behauptung.

Den Fall $x_* = +\infty$ kann man ebenfalls auf den verallgemeinerten Mittelwertsatz zurückführen, indem man $\tilde{f}(x) := f(\frac{1}{x})$, $\tilde{g}(x) := g(\frac{1}{x})$ betrachtet. Den Fall " $\frac{\pm\infty}{\pm\infty}$ " mit $x_* \in \mathbb{R}$ kann man unter Betrachtung von $\tilde{f}(x) := \frac{1}{f(x)}$, $\tilde{g}(x) := \frac{1}{g(x)}$ auf den verallgemeinerten Mittelwertsatz zurückführen.

Noch eine wichtige Anwendung, die zu folgendem wichtigen Folgentwert führt:

$$\boxed{\lim_{n \rightarrow \infty} \left(1 + \frac{x}{n}\right)^n = e^x} \quad \forall x \in \mathbb{R} \quad (3.12)$$

Insbesondere:

$$\boxed{e = \lim_{n \rightarrow \infty} \left(1 + \frac{1}{n}\right)^n}$$

⁴⁵Ohne dieses Herausziehen am Anfang hätte man hier viermal(!) hintereinander l'Hospital anwenden müssen, mit unangenehmen Ableitungen!

Herleitung unter Verwendung der Regel von l'Hospital an der Stelle (*):

$$\begin{aligned}
 \lim_{n \rightarrow \infty} \left(1 + \frac{x}{n}\right)^n &= \lim_{n \rightarrow \infty} \exp\left(n \cdot \ln\left(1 + \frac{x}{n}\right)\right) = \exp\left(\lim_{n \rightarrow \infty} \underbrace{n}_{\rightarrow \infty} \cdot \underbrace{\ln\left(1 + \frac{x}{n}\right)}_{\rightarrow 0}\right) \\
 &= \exp\left(\lim_{n \rightarrow \infty} \frac{\ln\left(1 + \frac{x}{n}\right)}{\frac{1}{n}}\right) \stackrel{(*)}{=} \exp\left(\lim_{n \rightarrow \infty} \frac{\frac{1}{1 + \frac{x}{n}} \cdot \left(-\frac{x}{n^2}\right)}{-\frac{1}{n^2}}\right) \\
 &= \exp\left(\lim_{n \rightarrow \infty} \frac{x}{1 + \frac{x}{n}}\right) = e^x
 \end{aligned}$$

Formel (3.12) spielt eine Rolle in der Zinseszinsrechnung: Ein Kapital K , das zum Ende des Jahres zu einem Zinssatz x verzinst wird, steigt auf den Wert $K \cdot (1+x)$. Zahlt man dagegen *halbjährlich* den Zinssatz $\frac{x}{2}$, so bekommt man (Zinseszins!) $K \cdot (1 + \frac{x}{2})^2$, und allgemein, bei n -maliger Verzinsung zum Zinssatz von jeweils $\frac{x}{n}$ auf $K \cdot (1 + \frac{x}{n})^n$. Im Grenzwert $n \rightarrow \infty$ erhält man $K \cdot e^x$.

Bemerkung: Die Formel (3.12) ist ungeeignet zur praktischen näherungsweise Berechnung von e oder e^x , denn die Folge konvergiert nur sehr langsam. Die praktisch-numerische Berechnung von e oder e^x mittels der Exponentialreihe ist viel effizienter.

3.6.4 Die Formel von Taylor

Im Folgenden geht es darum, Funktionen (die einige male differenzierbar sind), in einer (ggf. kleinen) Umgebung eines festen Punktes x_* durch eine einfachere Funktion, nämlich durch ein Polynom, zu *approximieren*. Eine solche Approximation hat vielfältige Anwendungsmöglichkeiten; z.B. möchte man ggf. überschlagen, wie viel ist $\ln(1.0002)$ oder $\sin(\pi + \frac{1}{1000})$, oder wir können an Operationen wie das Integrieren (Kap. 3.7) denken, die für komplizierte Funktionen nur schwer oder gar nicht exakt durchführbar ist, für Polynome jedoch leicht ist. Eine weitere Anwendung (s.S. 76 sowie die Übung) ist die näherungsweise Bestimmung von Umkehrfunktionen in Fällen, in denen wir nicht in der Lage sind, die Umkehrfunktion exakt zu bestimmen, wie beispielsweise $f(x) = x \ln x$. Wir werden am Ende dieses Kapitels die Taylor-Entwicklung benutzen, um ein hinreichendes Kriterium für Maximal- und Minimalstellen (in Abhängigkeit von f', f'', \dots) aufzustellen. Außerdem werden wir in Kap. 3.6.5 die Taylor-Entwicklung benutzen, um zu analysieren, warum das Newton-Verfahren so schnell konvergiert. Und es gibt unzählige weitere Anwendungen. Die Taylor-Entwicklung ist aus den Ingenieurwissenschaften nicht wegzudenken, egal um welche Fachrichtung es sich handelt.

Eine erste Idee kann es sein, eine gegebene, einmal differenzierbare Funktion f in einer Umgebung von x_* durch eine *lineare Approximation*, nämlich durch ihre Tangente im

Punkte x_* zu approximieren, also

$$f(x) \approx \underbrace{f(x_*) + f'(x_*)(x-x_*)}_{=:T_1(x), \text{Tangentengleichung}} .$$

Wir haben hier ausgenutzt, dass die Tangente (nach Definition der Ableitung von f) die gleiche Steigung hat wie f an der Stelle x_* : Die Bedingungen, die T_1 definieren, sind also: T_1 ist ein Polynom vom Grad (kleiner gleich) eins, es ist $T(x_*) = f(x_*)$ und $T'(x_*) = f'(x_*)$.

Vermutlich wird die obige Approximation von f durch T_1 nicht besonders gut sein, und man wird vermutlich eine bessere Approximation bekommen, wenn man f durch ein quadratisches Polynom T_2 approximiert. Das Polynom soll an der Stelle $x = x_*$ nicht nur die gleiche Steigung sondern auch das gleiche 'Krümmungsverhalten' haben wie f ;⁴⁶ das soll heißen, wir fordern, dass T_2 ein Polynom zweiten Grades sei mit $T(x_*) = f(x_*)$, $T'(x_*) = f'(x_*)$ und $T''(x_*) = f''(x_*)$. Ein Ansatz $T_2(x) = \alpha_0 + \alpha_1(x-x_*) + \alpha_2(x-x_*)^2$ (wir haben also 3 Bedingungen für die 3 Koeffizienten des Polynoms) führt nach zweimaligem Ableiten zu $T_2'(x_*) = \alpha_1 + 2\alpha_2(x-x_*)$, also zu $\alpha_0 = f(x_*)$, $\alpha_1 = f'(x_*)$, $\alpha_2 = \frac{1}{2}f''(x_*)$, also

$$T_2(x) = f(x_*) + f'(x_*)(x-x_*) + \frac{f''(x_*)}{2}(x-x_*)^2,$$

für differenzierbares f und $|x-x_*|$ 'klein' haben wir also die Näherung

$$f(x) \approx T_2(x) = f(x_*) + f'(x_*)(x-x_*) + \frac{f''(x_*)}{2}(x-x_*)^2.$$

Für n -mal differenzierbares f bekommt man so, indem man die $n+1$ Bedingungen $T_n^{(k)}(x_*) \stackrel{!}{=} f^{(k)}(x_*)$, $k = 0, \dots, n$, stellt an ein Polynom $T_n(x) = \sum_{k=0}^n \alpha_k (x-x_*)^k$ die Näherung

$$\begin{aligned} T_n(x) &= f(x_*) + f'(x_*)(x-x_*) + \frac{f''(x_*)}{2!}(x-x_*)^2 + \frac{f'''(x_*)}{3!}(x-x_*)^3 + \dots + \frac{f^{(n)}(x_*)}{n!}(x-x_*)^n \\ &= \boxed{\sum_{k=0}^n \frac{f^{(k)}(x_*)}{k!}(x-x_*)^k}, \quad n \in \mathbb{N}_0. \end{aligned} \tag{3.13}$$

Dieses Polynom heißt n -tes *Taylor-Polynom* von f für den Entwicklungspunkt x_* .⁴⁷ Indem wir die Substitution $h := x - x_*$ vornehmen, bekommen wir die äquivalente

⁴⁶Ein völlig anderer Ansatz wäre, zu fordern, dass $T_2(x_i) = f(x_i)$ für drei verschiedene Punkte x_i (allgemeiner: dass $T_n(x_i) = f(x_i)$ für $n+1$ verschiedene Punkte x_i gilt); dies ist ebenfalls möglich und wird durchaus praktiziert, jedoch verfolgen wir diese Idee hier nicht weiter (Stichwort 'Polynomapproximation/-interpolation' in der Literatur).

⁴⁷Konvention: die 'nullte Ableitung' $f^{(0)}$ sei nichts anderes als die Funktion f selbst.

Darstellung⁴⁸

$$\begin{aligned} f(x_*+h) &\approx f(x_*) + f'(x_*)h + \frac{f''(x_*)}{2}h^2 + \dots + \frac{f^{(n)}}{n!}h^n \\ &= \sum_{k=0}^n \frac{f^{(k)}(x_*)}{k!}h^k = \tilde{T}_n(h) \quad \text{für } |h| \text{ 'klein'}. \end{aligned} \quad (3.14)$$

Bsp. Für $f(x) = \ln x$ und den Entwicklungspunkt $x_* := 1$ haben wir $f'(x) = \frac{1}{x}$, $f''(x) = -\frac{1}{x^2}$, also $f(1) = 0$, $f'(1) = 1$, $f''(1) = -1$, somit

$$\tilde{T}_2(h) = 0 + h - \frac{1}{2}h^2,$$

$$\ln x \approx 0 + (x-1) - \frac{1}{2}(x-1)^2 \quad \text{für } |x-1| \text{ 'klein'}$$

bzw

$$\ln(1+h) \approx h - \frac{1}{2}h^2 \quad \text{für } |h| \text{ 'klein'}$$

Eine Näherung von $\ln(1.002)$ ist also, \tilde{T}_1 verwendend, 0.002 , und \tilde{T}_2 verwendend, $0.002 - 0.5 \cdot 0.002^2 = 0.001998$. Und in der Tat; der exakte Wert ist laut Taschenrechner: $\ln(1.002) = 0.001998002663\dots$; der Approximationsfehler $\tilde{T}_2(0.002) - f(1.002) = T_2(1.002) - f(1.002)$ ist also mit $T_2(1.002) - f(1.002) \approx 2.663 \cdot 10^{-9}$ ziemlich klein.

In der Praxis hat man natürlich normalerweise den exakten Wert (hier: Aus dem Taschenrechner) nicht zur Verfügung (denn wenn man den exakten Wert "hat" braucht man überhaupt keine Approximation mehr durch Polynome).

Eine nützliche Sache wäre also: Kann man für den (*Approximations-*)Fehler $f(x) - T_n(x)$ eine Formel aufstellen, die einem etwas über die Größe dieses Fehlers sagt, ohne dass man den exakten Wert $f(x)$ kennt? Darauf liefert uns obiges Konstruktionsprinzip von T_n leider keine Antwort.

Man bekommt dagegen eine solche *Darstellung des (Approximations-)Fehlers* auf folgende, nicht ganz naheliegende Weise:

Wir definieren, für (fest aber beliebig) vorgegebenes x_* und h , die Hilfsfunktionen $z \mapsto F(z)$, $g \mapsto g(z)$, durch

$$F(z) := \sum_{k=0}^n \frac{f^{(k)}(z)}{k!} (x_* - z + h)^k, \quad g(z) := -(x_* - z + h)^{n+1},$$

⁴⁸Bemerkung zum Polynom \tilde{T}_n aus (3.13): Es ist $\tilde{T}_n(h) = T_n(x)$, wobei $h = x - x_*$, also $\tilde{T}_n(x - x_*) = T_n(x)$ bzw. $\tilde{T}_n(h) = T_n(x_* + h)$. Formal sind die Polynomfunktionen $x \mapsto T_n(x)$ und $h \mapsto \tilde{T}_n(h)$ nicht die selben Abbildungen (was man z.B. daran sehen kann, dass \tilde{T}_n bei 0 eine Nullstelle hat und T_n bei x_*), daher haben wir die Bezeichnungen T_n , \tilde{T}_n unterschieden.

und wenden auf F und g den *verallgemeinerten Mittelwertsatz* auf dem Intervall $[a, b]$ mit den Grenzen $a := x_*$, $b := x_* + h$ an. (Hier wird $h > 0$ angenommen; im Fall $h < 0$ sollte man die Rollen von a und b vertauschen.) Wir bekommen

$$F'(z) = \sum_{k=0}^n \frac{f^{(k+1)}(z)}{k!} (x_* - z + h)^k - \underbrace{\sum_{k=1}^n \frac{f^{(k)}(z)}{k!} k (x_* - z + h)^{k-1}}_{= \sum_{k=0}^{n-1} \frac{f^{(k+1)}(z)}{k!} (x_* - z + h)^k} = \frac{f^{(n+1)}(z)}{n!} (x_* - z + h)^n,$$

$$g'(z) = (n+1) (x_* - z + h)^n,$$

$$F(x_*) = \sum_{k=0}^n \frac{f^{(k)}(x_*)}{k!} h^k = \tilde{T}_n(h),$$

$$g(x_*) = -h^{n+1}$$

$$F(x_* + h) = f(x_* + h),$$

$$g(x_* + h) = 0$$

und⁴⁹ somit die Existenz eines $\xi \in (a, b)$ mit $\frac{F(b)-F(a)}{g(b)-g(a)} = \frac{F'(\xi)}{g'(\xi)}$, also

$$\frac{f(x_* + h) - \tilde{T}_n(h)}{0 - (-h^{n+1})} = \frac{f^{(n+1)}(\xi) \cdot (x_* - \xi + h)^n}{n! (n+1) \cdot (x_* - \xi + h)^n} = \frac{f^{(n+1)}(\xi)}{(n+1)!}.$$

Der Fehler an der Stelle $x = x_* + h$ ist also

$$\tilde{R}_{n,x_*}(h) := f(x_* + h) - \tilde{T}_n(h) = \frac{f^{(n+1)}(\xi)}{(n+1)!} \cdot h^{n+1}.$$

Dabei ist ξ eine Stelle zwischen x_* und $x = x_* + h$, also $\xi \in (x_*, x)$ für $x > x_*$ und $\xi \in (x, x_*)$ für $x < x_*$. Beide Fälle kann man zusammenfassen, indem man $\xi = x_* + \theta \cdot (x - x_*)$ einsetzt, wobei $\theta \in (0, 1)$.

Man nennt den Fehlerterm $\tilde{R}(x_*, h)$ auch *Restglied*; die obige Restgliedformel geht auf Lagrange⁵⁰ zurück (es gibt noch andere Darstellungen des Restgliedes). Da man das ξ bzw. θ nicht genau kennt, kann man das Restglied immer noch nicht exakt angeben, aber man kann die Darstellung des Restliedes immerhin benutzen, um eine Schranke für den Fehler ('Fehlerabschätzung') zu bekommen.

⁴⁹Zum Problem "0⁰" bei der Auswertung von $F(x_* + h)$: Aus Gründen der Stetigkeit ist eine Funktion $u(x) := x^k$ für festes $k=0$, die an allen Stellen $x \neq 0$ den Wert 1 hat, auch an der Stelle $x=0$ mit dem Funktionswert 1 zu interpretieren. (Anders die Funktion $u(x) := k^x$ für festes $k=0$: Diese hat für alle $x \neq 0$ den Wert 0, somit ist auch ihr Funktionswert an der Stelle $x=0$ als null zu interpretieren.)

⁵⁰Joseph-Louis Lagrange, 1736-1813.

Bsp. (Restgliedbetrachtung/-abschätzung): Oben hatten wir für $f(x) := \ln x$ und $x_* := 1$ das Taylor-Polynom zweiten Grades $\tilde{T}_2(h) = h - \frac{1}{2}h^2$ hergeleitet, also $\ln(1+h) \approx h - \frac{1}{2}h^2$. Ferner ist $f'''(x) = \frac{2}{x^3}$. Für den Fehler (das Restglied) gilt also

$$\tilde{R}_{2,x_*}(h) = \frac{f'''(\xi)}{3!} h^3 = \frac{1}{6} \cdot \frac{2}{\xi^3} h^3,$$

somit

$$|\tilde{R}_{2,x_*}(h)| \leq \frac{1}{3} \left(\frac{|h|}{|\xi|} \right)^3.$$

Suchen wir eine Schranke für den Fehler z.B. an der **festen Stelle** $x = \frac{1}{3}$, d.h. $h = -\frac{2}{3}$, so wissen wir, dass $\xi \in (x, x_*) = (\frac{1}{3}, 1)$. Aus Monotoniegründen ist $\xi^3 = |\xi|^3 \in [\frac{1}{27}, 1]$, somit $|\xi|^3 \geq \frac{1}{27}$ verwendend erhalten wir

$$|\tilde{R}_{2,x_*}(-\frac{2}{3})| \leq \frac{1}{3} \cdot 27 \cdot \left(\frac{2}{3}\right)^3 = \frac{8}{3}$$

(Dass wir hier mit so großen Fehlern rechnen müssen, hängt damit zusammen, dass wir bei $x = \frac{1}{3}$ nicht weit von der Singularität des Logarithmus bei $x=0$ entfernt sind.) Für $x = \frac{3}{2}$, also $h = \frac{1}{2}$, ist $\xi \in (1, \frac{3}{2})$, also $|\xi| \geq 1$, somit

$$|\tilde{R}_{2,x_*}(\frac{1}{3})| \leq \frac{1}{3} \cdot 1 \cdot (1/2)^3 = \frac{1}{24}$$

Etwas allgemeiner, für **variables** x bzw. h :

Für $x > 1$, also $h > 0$, ist $\xi \in (1, 1+h)$, somit $|\xi| \geq 1$, somit

$$|\tilde{R}_{2,x_*}(h)| \leq \frac{1}{3} \cdot h^3,$$

und für $x < 1$, also $h < 0$, ist $\xi \in (1-|h|, 1)$, somit $|\xi| \geq 1-|h|$, somit

$$|\tilde{R}_{2,x_*}(h)| \leq \frac{1}{3} \cdot \frac{|h|^3}{(1-|h|)^3},$$

Für das $h = 2 \cdot 10^{-3}$ aus obigem Beispiel bekommen wir also $|\tilde{R}_{2,x_*}(h)| \leq \frac{1}{3} \cdot \frac{8 \cdot 10^{-9}}{0.998^3}$ (Der Fehler kann also kaum größer sein als $\frac{8}{3} \cdot 10^{-9}$, was eine sehr gute Abschätzung für den Fehler ist, den wir auf S. 69 mittels Taschenrechner ermittelt hatten, nämlich $2.663 \cdot 10^{-9}$.)

Wir fassen unsere Erkenntnisse zusammen im Satz von Taylor:

Satz (Satz von Taylor, Satz über die Taylor-Entwicklung, Formel von Taylor).

Sei $f : (a, b) \rightarrow \mathbb{R}$ $(n+1)$ -mal differenzierbar und $x_*, x \in (a, b)$. Dann erfüllt das n -te **Taylor-Polynom**

$$\tilde{T}_n(h) := \sum_{k=0}^n \frac{f^{(k)}(x_*)}{k!} h^k$$

die Gleichung

$$f(x_* + h) = \tilde{T}_n(h) + \tilde{R}_{n,x_*}(h),$$

wobei das **Restglied** (=Fehlerterm=Approximationsfehler zwischen Taylor-Polynom und f) die Form

$$\tilde{R}_{n,x_*}(h) = \frac{f^{(n+1)}(\xi)}{(n+1)!} h^{n+1} = \frac{f^{(n+1)}(\xi)}{(n+1)!} (x - x_*)^{n+1} \quad (3.15)$$

hat, wobei ξ zwischen x_* und $x_* + h$ liegt, kurz:

$$f(x_* + h) = \underbrace{\sum_{k=0}^n \frac{f^{(k)}(x_*)}{k!} h^k}_{=\tilde{T}_n(h)} + \underbrace{\frac{f^{(n+1)}(\xi)}{(n+1)!} h^{n+1}}_{=\tilde{R}_{n,x_*}(h)}.$$

Alle obigen Terme kann man mittels einer Substitution $h = x - x_*$ statt in x_* , h auch in x_* , x ausdrücken:

$$f(x) = \underbrace{\sum_{k=0}^n \frac{f^{(k)}(x_*)}{k!} (x - x_*)^k}_{=T_n(x)} + \underbrace{\frac{f^{(n+1)}(\xi)}{(n+1)!} (x - x_*)^{n+1}}_{=:R_{n,x_*}(x)},$$

wobei ξ zwischen x und x_* liegt.

Bemerkung: Für $n=0$ ist die Formel von Taylor nichts anderes als der Mittelwertsatz.

z.Ü.: Leiten sie die sog. *Siebziger-Regel* her, d.h. eine Formel aus der Finanzmathematik, mit der man näherungsweise berechnen kann, nach wie vielen Jahren sich ein Kapital, bei gegebener Verzinsung, verdoppelt hat. Hinweis: Nehmen Sie kontinuierliche Verzinsung an, also $K(t) = K_0 \cdot e^{\tilde{p}t}$, wobei der kontinuierliche Zinssatz \tilde{p} so definiert, ist, dass nach $t = 1$ Jahr ein Kapital $K(1) = K_0 \cdot (1 + \frac{p}{100})$ erreicht ist; p ist also der jährliche Zinssatz (in Prozent). Drücken Sie $K(t)$ mittels p aus. Stellen Sie eine Gleichung auf, nach wie vielen Jahren t sich das Kapital verdoppelt hat (als Funktion von p). Ersetzen/Approximieren Sie die dabei auftretende ln-Funktion durch ihr Taylor-Polynom erster Ordnung; p "klein" voraussetzend. Warum wird die Regel Siebziger-Regel genannt?

Zusammenhang von Taylor-Entwicklung und Potenzreihen: Die Taylor-Reihe

Das Restglied geht offenbar, wenn $x \rightarrow x_*$ (d.h. $h \rightarrow 0$), gegen null, sofern $f^{(n+1)}$ stetig ist. Dabei ist $n \in \mathbb{N}$ fest. Nun kann man andererseits auch untersuchen, was mit dem Restglied passiert, wenn man x festhält und n gegen unendlich gehen lässt, d.h. man erhöht den Grad des Taylor-Polynoms immer weiter. Dies setzt natürlich voraus, dass f beliebig oft differenzierbar ist zumindest an der Stelle x_* , damit man die T_n für alle

$n \in \mathbb{N}_0$ bilden kann. Die Folge der Taylor-Polynome T_0, T_1, T_2, \dots , also die Reihe mit den Partialsummen T_0, T_1, T_2, \dots , also die Reihe

$$\sum_{k=0}^{\infty} \frac{f^{(k)}(x_*)}{k!} h^k = \sum_{k=0}^{\infty} \frac{f^{(k)}(x_*)}{k!} (x-x_*)^k,$$

bezeichnet man als *Taylor-Reihe*.

Die Taylor-Reihe ist offensichtlich eine Potenzreihe; die n -te Partialsumme ist das n -te Taylor-Polynom. Wir hatten Taylor-Polynome hoher Ordnung betrachtet in der Hoffnung/Vermutung, dass bei höherer Ordnung n die Approximation von T_n an f immer besser wird. Eine berechtigte Vermutung ist also, dass

$$T_n(x) \xrightarrow{(n \rightarrow \infty)} f(x) \quad (x \text{ fest}).$$

Gleichbedeutend ist, dass die Differenz, also das Restglied

$$R_{n,x_*}(x) \xrightarrow{(n \rightarrow \infty)} 0,$$

denn $R_{n,x_*}(x)$ ist gerade die Differenz zwischen der n -ten Partialsumme der Reihe und ihrem Grenzwert (sofern dieser existiert). Wir wissen: Die Taylor-Reihe hat, da sie Potenzreihe ist, einen Konvergenzradius $R \in [0, \infty]$. Eine Vermutung könnte nun sein, dass die Taylor-Reihe innerhalb ihres Konvergenzbereichs immer gegen f konvergiert; dies wäre wünschenswert, denn das hieße, dass man, durch Erhöhung von n grundsätzlich eine immer bessere Approximation T_n an f bekommt. Jedoch: Es gibt Gegenbeispiele! Die Funktion

$$f(x) := \begin{cases} \exp(-\frac{1}{x^2}) & , x \neq 0 \\ 0 & , x = 0 \end{cases} \quad (3.16)$$

ist (auch wenn das auf den ersten Blick nicht so offensichtlich ist) unendlich oft differenzierbar, aber *alle* Ableitungen an der Stelle $x_* = 0$ haben den Wert null, somit ist jedes T_n gleich der Nullfunktion, somit konvergiert die Taylor-Reihe gegen die Nullfunktion und somit, für $x \neq 0$, nicht gegen f !

Für vorgegebenes x_*, x sind also, bei einer unendlich oft differenzierbaren Funktion f , folgende drei Fälle möglich:

1. Die Taylor-Reihe an der Stelle x konvergiert gegen $f(x)$.
Gleichbedeutend: $R_{n,x_*}(h) \xrightarrow{(n \rightarrow \infty)} 0$
2. Die Taylor-Reihe an der Stelle x konvergiert, aber nicht gegen $f(x)$.
Gleichbedeutend: Das Restglied konvergiert für $n \rightarrow \infty$, jedoch nicht gegen null.
3. Die Taylor-Reihe an der Stelle x divergiert.
Gleichbedeutend: Das Restglied divergiert.

Aus der Sicht desjenigen, der eine Approximation an f sucht, ist der Fall 1 also der wünschenswerte Fall. Um zu prüfen, ob die Taylor-Reihe an einer Stelle x gegen f konvergiert, untersucht man also das Restglied (3.15).

Eine Funktion $f : \mathbb{R} \rightarrow \mathbb{R}$, die unendlich oft differenzierbar ist, und bei der für beliebigen Entwicklungspunkt x_* ein Intervall um x_* existiert, so dass auf diesem Intervall die Taylor-Reihe gegen die Funktion konvergiert, heißt *analytisch*. Die Funktion (3.16) ist also nicht analytisch. Man kann die Differenzierbarkeit sowie den Begriff 'analytisch' auch für Funktionen $f : \mathbb{C} \rightarrow \mathbb{C}$ erklären und Reihenentwicklungen analytischer komplexer Funktionen untersuchen. Das zugehörige Teilgebiet der Mathematik ist die sog. *Funktionentheorie* oder *Komplexe Analysis*. Funktionen $f : \mathbb{C} \rightarrow \mathbb{C}$ haben erstaunliche Eigenschaften, so folgt z.B. aus der einmaligen Differenzierbarkeit die n -malige Differenzierbarkeit für beliebiges $n \in \mathbb{N}$, was für Funktionen $f : \mathbb{R} \rightarrow \mathbb{R}$ nicht wahr ist.

Beispiel für das Aufstellen einer Taylor-Reihe und das Überprüfen ihres Konvergenzverhaltens. Für $f(x) := \ln x$ ist $f'(x) = x^{-1}$, $f''(x) = -x^{-2}$, $f'''(x) = 2x^{-3}, \dots$; allgemein, wie man leicht (strenggenommen per Vollständiger Induktion) zeigen kann: $f^{(k)}(x) = (-1)^{k+1}(k-1)! \frac{1}{x^k} \forall k \in \mathbb{N}$. Es folgt für den Entwicklungspunkt $x_* = 1$: $f^{(k)}(1) = (-1)^{k+1}(k-1)!$ sowie $f^{(0)}(1) = 0$. Die Taylor-Reihe von $f = \ln$ lautet also

$$\sum_{k=1}^{\infty} \frac{(-1)^{k+1}(k-1)!}{k!} (x-1)^k = \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} (x-1)^k$$

Der Konvergenzradius ist

$$R = \frac{1}{\limsup \sqrt[k]{|a_k|}} = \frac{1}{\limsup \sqrt[k]{\frac{1}{k}}} = \limsup \sqrt[k]{k} = 1,$$

d.h. die Taylor-Reihe ist konvergent für $h \in (-1, 1)$ bzw. für $x \in (0, 2)$, und divergent für $x > 2$ und für $x < 0$.

Wir können hier sogar Aussagen treffen über das Verhalten *am Rand* des Konvergenzintervalls: Bei $x = 2$ ist die Taylor-Reihe nichts anderes als die alternierende harmonische Reihe, somit konvergent. Bei $x = 0$ ist die Taylor-Reihe bis aufs Vorzeichen die nicht-alternierende harmonische Reihe, somit divergent. Die Taylor-Reihe konvergiert also genau dann wenn $x \in (0, 2]$.

Diese Rechnung (Bestimmung des Konvergenzintervalls) beantwortet aber nur die Frage, ob für gegebenes x einerseits Fall 1 oder 2 oder andererseits Fall 3 vorliegt; sie beantwortet nicht, ob Fall 1 vorliegt. Um zwischen Fall 1 einerseits und den Fällen 2 und 3 andererseits zu unterscheiden, also um zu überprüfen, ob die Taylor-Reihe (für $|h| < 1$; nur solche h sind nach obiger Rechnung sinnvoll) gegen f konvergiert, muss man prüfen, ob das Restglied für $n \rightarrow \infty$ gegen null geht:

Restgliedabschätzung:

Das Restglied lautet betragsmäßig:

$$|\tilde{R}_{n,x_*}(h)| = \frac{|f^{(n+1)}(\xi_n)|}{(n+1)!} |h|^{n+1} = \frac{1}{(n+1)} \frac{|h|^{n+1}}{|\xi_n|^{n+1}} = \frac{1}{n+1} \cdot \left(\frac{|h|}{|\xi_n|} \right)^{n+1}$$

Für $x > 1$, also $h > 0$, ist $\xi_n \in (1, x)$, somit $|\xi_n| \geq 1$, somit

$$|\tilde{R}_{n,x^*}(h)| \leq \underbrace{\frac{1}{n+1}}_{\rightarrow 0} \cdot \underbrace{h^{n+1}}_{\text{beschr.}} \xrightarrow{(n \rightarrow \infty)} 0 \quad \text{falls } h \leq 1;$$

und für $x < 1$, also $h < 0$, ist $\xi_n \in (1 - |h|, 1)$, somit $|\xi_n| \geq 1 - |h|$, somit

$$|\tilde{R}_{n,x^*}(h)| \leq \underbrace{\frac{1}{n+1}}_{\rightarrow 0} \cdot \left(\frac{|h|}{1 - |h|} \right)^{n+1} \xrightarrow{(n \rightarrow \infty)} 0 \quad \text{falls } h \geq -\frac{1}{2}. \quad (3.17)$$

Bem.: Beachten Sie, dass die Stelle ξ i.a. von n abhängt; wir haben sie deshalb mit ξ_n bezeichnet. Wenn man diese Abhängigkeit des ξ von n vergisst, können u.U. Fehler passieren. So könnte man z.B. – losgelöst von obiger Rechnung – bei der Untersuchung z.B. eines Terms ξ_n^{n+1} , wobei $|\xi_n| < 1$ gelte, *fälschlicherweise* mit Ignorierung des Index n Schlüsse ziehen wie ” $|\xi| < 1$, also $|\xi|^{n+1} \xrightarrow{(n \rightarrow \infty)} 0$ ” machen; der letzte Schluss ist jedoch bei n -Abhängigkeit von ξ nicht erlaubt: Z.B. im Fall $\xi = \xi_n = 1 - \frac{1}{n+1}$ wäre zwar $|\xi_n| < 1$, aber mit (3.12) ist $(1 - \frac{1}{n+1})^{n+1} \xrightarrow{(n \rightarrow \infty)} e^{-1} \neq 0$. Dieser Effekt ist bei der Restglieduntersuchung $\lim_{n \rightarrow 0} \tilde{R}_{n,x^*}(h)$ eine potenzielle Fehlerquelle (nicht jedoch wenn $n = \text{fest}$).

Wir haben somit in der Taylor-Reihe eine *Reihenentwicklung des Logarithmus*, die **Logarithmusreihe** gefunden, auf dem Intervall $\frac{1}{2} < x \leq 2$ (man kann zeigen, dass diese Reihendarstellung sogar auf dem gesamten Konvergenzbereich der Reihe, also $(0, 2]$, gilt; dazu fehlen uns jedoch an dieser Stelle die Mittel):

$$\ln x = \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} (x-1)^k \quad (0 < x \leq 2)$$

bzw. äquivalent

$$\ln(1+h) = \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} h^k \quad (-1 < h \leq 1). \quad (3.18)$$

Insbesondere für $x=2$ bekommen wir den Grenzwert der alternierenden harmonischen Reihe (s.S. 22):

$$\sum_{k=1}^{\infty} \frac{(-1)^k}{k} = -\ln 2.$$

Wir haben uns in dem Kapitel über Taylor-Entwicklung damit beschäftigt, wie man zu einer gegebenen Funktion eine Reihenentwicklung findet (nämlich durch Aufstellen der Taylor-Reihe, wobei, wenn man Pech hat, die Taylor-Reihe nicht oder nur auf einem kleinen Bereich gegen die gegebene Funktion f konvergiert). Was passiert, wenn wir für Funktionen, die wir mit Hilfe einer Potenzreihe definiert haben (\exp , \sin , \cos , ...), die zugehörige Taylor-Reihe aufstellen? Eine naheliegende Vermutung ist wohl, dass die

Reihenentwicklung, über die diese Funktionen definiert sind, gleich den Taylor-Reihen dieser Funktionen sind. Dies ist in der Tat der Fall, wie man mit Hilfe des gliedweisen Ableitens leicht zeigen kann. Ist f als Potenzreihe $f(x) := \sum_{k=0}^{\infty} a_k (x-x_*)^k$ gegeben, so bekommt man durch gliedweises Ableiten und Einsetzen von $x = x_*$, dass $f(x_*) = a_0$, $f'(x_*) = a_1$, $f''(x_*) = 2a_2, \dots$, $f^{(k)}(x_*) = k!a_k$; die Taylor-Formel ergibt also als Taylor-Reihe $\sum_{k=0}^{\infty} \frac{f^{(k)}(x_*)}{k!} (x-x_*)^k$ genau die Ausgangsreihe.

Kurz: Die Taylor-Reihe einer (konvergenten) Potenzreihe ist die Potenzreihe selbst (identische Entwicklungspunkte vorausgesetzt).

Sucht man also die Taylor-Reihe von $f = \exp$ um den Entwicklungspunkt $x_* = 0$, so lautet die Antwort $\sum_{k=0}^{\infty} \frac{x^k}{k!}$; das gleiche gilt für Taylor-Polynome; so ist z.B. das Taylor-Polynom 4. Grades von $f = \cos$ zum Entwicklungspunkt $x_* = 0$ einfach der Beginn der den Kosinus definierenden Reihenentwicklung $1 - \frac{x^2}{2!} + \frac{x^4}{4!}$; das Restglied $R_{4,0}(x)$ ist $R_{4,0}(x) = \frac{\cos^{(5)}(\xi)}{5!} x^5 = -\frac{\sin \xi}{5!} x^5$, kann somit als $|R_{4,0}(x)| \leq \frac{1}{120} |x|^5$ abgeschätzt werden; dies ist eine einfachere Vorgehensweise als z.B. analog zu S. 56 den Rest in Form von unendlich vielen Summanden (aus der Kosinusreihe) darzustellen und dann abzuschätzen. ⁵¹

Beachte: Auch für einige verkettete Funktionen kann man leicht das Taylor-Polynom finden, z.B. für $f(x) := \sin(2x)$ bekommt man $T_5(x) = \frac{(2x)^1}{1!} - \frac{(2x)^3}{3!} + \frac{(2x)^5}{5!}$, sowie auch für $f(x) := \exp(x^2)$ bekommt man $T_6(x) = 1 + \frac{(x^2)^1}{1!} + \frac{(x^2)^2}{2!} + \frac{(x^2)^3}{3!} = 1 + x^2 + \frac{1}{2} x^4 + \frac{1}{6} x^6$.

Aufstellen der Taylor-Reihe für eine Umkehrfunktion.

Auf S. 57 sowie in Aufgabe P25 wurde erklärt, wie man für die Umkehrfunktion f^{-1} einer gegebenen Funktion f die Ableitungen an einer Stelle y_* berechnen kann – und zwar selbst dann, wenn man die Umkehrfunktion selbst gar nicht explizit aufstellen kann. Die Kenntnis der Ableitungswerte $(f^{-1})^{(k)}(y_*)$ an einer festen Stelle y_* aber reicht, wie wir wissen, völlig aus, um das Taylor-Polynom für f^{-1} aufzustellen (\rightarrow Übung). Damit kann man die Schwierigkeit, dass man Umkehrfunktionen oft nicht explizit angeben kann, ein wenig 'aushebeln': Wir können immerhin das Taylor-Polynom, das häufig eine gute Näherung von f^{-1} darstellt, ausrechnen.

Anwendung der Taylor-Entwicklung zur Herleitung eines hinreichenden Kriteriums für lokale Maximal- und Minimalstellen:

Laut Satz von Taylor ist

$$f(x) = f(x_*) + \underbrace{f'(x_*)}_{=0} (x-x_*) + \underbrace{\frac{f''(\xi)}{2} (x-x_*)^2}_{>0? \text{ od. } <0?}$$

⁵¹Sucht man bei \sin, \cos, \exp dagegen die Taylor-Entwicklung um einen anderen Entwicklungspunkt als $x_* = 0$, so muss man rechnen.

für ein ξ zwischen x und x_* . Ist

$$f \in \mathcal{C}^2 \text{ und } f'(x_*) = 0 \text{ und } f''(x_*) > 0 \text{ (bzw. } f''(x_*) < 0),$$

so ist auch $f''(\xi) > 0$ (bzw. $f''(\xi) < 0$) für ξ hinreichend nahe bei x_* (d.h. für x hinreichend nahe bei x_*). Es folgt $f(x) > f(x_*)$ (bzw. $f(x) < f(x_*)$) für alle x hinreichend nahe bei x_* , also: x_* ist eine lokale Minimal- (bzw.: Maximal-)Stelle von f .

Ganz analog zeigt man, indem man Taylor-Polynome höherer Ordnung benutzt:

Ist

$$f \in \mathcal{C}^3, \quad f'(x_*) = f''(x_*) = 0, \quad f'''(x_*) \neq 0,$$

so hat man

$$f(x) = f(x_*) + \underbrace{f'(x_*)}_{=0} (x-x_*) + \underbrace{\frac{f''(x_*)}{2}}_{=0} (x-x_*)^2 + \underbrace{\frac{f'''(\xi)}{6}}_{\neq 0} (x-x_*)^3,$$

und somit ist dann x_* weder eine lokale Maximal- noch Minimalstelle (sondern ein sog. Sattelpunkt), denn in einer kleinen Umgebung von x_* hat dann $f'''(\xi)$ das gleiche Vorzeichen wie $f'''(x_*)$ (ist insbesondere $\neq 0$), aber unabhängig vom Vorzeichen von $f'''(x_*)$ findet man in jeder noch so kleinen Umgebung von x_* immer Stellen x , an denen das Restglied positiv oder negativ ist, indem man x größer bzw. kleiner als x_* wählt.

Z.Ü.: Wie lauten allgemein Kriterien im Fall $f'(x_*) = \dots = f^{(n)}(x_*) = 0, f^{(n+1)} \neq 0$?

3.6.5 Das Newton-Verfahren (und das Bisektionsverfahren)

In diesem Kapitel geht es darum, Nullstellen zu finden für nichtlineare Gleichungen, wie z.B. $f(x) := e^x - x \stackrel{!}{=} 0$. Wenn wir hierfür Verfahren finden, können wir dieses natürlich auch anwenden für das Finden von Lösungen allgemeiner nichtlinearer Gleichungen, denn solche Probleme lassen sich offensichtlich als Nullstellenproblem umschreiben ('alles auf eine Seite bringen').

Für das allgemeine **Nullstellenproblem**

$$\text{Löse } f(x) = 0.$$

wobei f eine nichtlineare gegebene Funktion ist, gibt es kein allgemeines Verfahren, das die exakte Lösung liefert.⁵²

Wir suchen daher nach einem Verfahren, das uns eine Näherungslösung liefert.

⁵²In Spezialfällen gibt es natürlich Verfahren, z.B. für $f(x) = ax + b, a \neq 0$, ist $x = -\frac{b}{a}$ Lösung; im Fall $f(x) = x^2 + px + q$ sind $x_{1,2} = -\frac{p}{2} \pm \sqrt{D}$ mit $D = \frac{p^2}{4} - q$ Lösungen, sofern $D \geq 0$. Man kann übrigens auch für Polynome vom Grad 3 und vom Grad 4 allgemeine Lösungsformeln formulieren; diese sind aber sehr kompliziert und daher eher von geringer Bedeutung. Man kann interessanterweise beweisen (\rightarrow Algebra), dass es keine allgemeine Lösungsformel für die Nullstellen von Polynomen vom Grad größer oder gleich 5 geben kann.

Ein ganz einfaches Verfahren ist das *Intervallhalbierungsverfahren* (= *Bisektionsverfahren*), das lediglich die Stetigkeit von f voraussetzt. Man sucht einen Punkt a mit $f(a) > 0$ sowie ein b mit $f(b) < 0$. O.B.d.A. sei $a < b$ für die folgende Darstellung. Nach dem Zwischenwertsatz gibt es dann eine Nullstelle $x_* \in (a, b)$. Um die Lage einzugrenzen, berechnet man $f(\frac{a+b}{2})$. Ist der Wert positiv, so liegt (mindestens) eine Nullstelle im Intervall $(\frac{a+b}{2}, b)$. Ist der Wert negativ, so liegt eine Nullstelle im Intervall $(a, \frac{a+b}{2})$. Diese Intervall halbiert man weiter und weiter, bis die gesuchte Genauigkeit erreicht ist, bis also die Intervalllänge kleiner als eine vorgegebene Fehlertoleranz (Fehlerschranke) ist.

Ein anderes Verfahren zur Nullstellensuche ist das berühmte Newton-Verfahren. Die Grundlegende Idee ist: Wenn wir die gesuchte Nullstelle x_* einer Funktion $f : \mathbb{R} \rightarrow \mathbb{R}$ nicht exakt berechnen können, so können wir stattdessen eine *Tangente* an f , gebildet an einem Punkt x_0 , betrachten und stattdessen die Nullstelle der Tangente berechnen. Nach der Taylor-Formel ist die Tangente ja eine Näherung an die Funktion f , zumindest in der Nähe des Punktes x_0 . Wenn also x_0 in der Nähe der gesuchten Nullstelle x_* liegt, dann sollte die Nullstelle der Tangente eine Näherung an die Nullstelle von f darstellen. (Skizze!). Wir rechnen: Die Tangentengleichung an f im Punkt x_0 lautet (siehe z.B. Taylor-Formel mit $n=1$)

$$T(x) = f(x_0) + f'(x_0) \cdot (x - x_0).$$

Die Nullstelle der Tangentengleichung ist offensichtlich $x = x_0 - \frac{f(x_0)}{f'(x_0)}$. Also: Unter der Annahme, dass x_0 eine "Näherung" an die gesuchte Nullstelle x_* ist, bekommen wir eine neue (vermutlich bessere) Näherung $x_1 := x_0 - \frac{f(x_0)}{f'(x_0)}$. Mit dieser Näherung können wir die Berechnung wiederholen ('iterieren') und ein $x_2 := x_1 - \frac{f(x_1)}{f'(x_1)}$ berechnen, allgemein:

$$\boxed{x_{n+1} := x_n - \frac{f(x_n)}{f'(x_n)}}$$

sofern $f'(x_n) \neq 0$. Dies ist das sog. *Newton-Verfahren* zur iterativen Berechnung einer Nullstelle von f .

Beispiel. Wir wollen eine Nullstelle der Funktion $f(x) = x^2 - 5$ mit dem Newton-Verfahren berechnen. Beachte: Das exakte Ergebnis ist $\pm\sqrt{5}$.

Es ist $f'(x) = 2x$. Bei einem vorgegebenen x_0 lautet also die Iterationsvorschrift des Newton-Verfahrens

$$x_{n+1} = x_n - \frac{x_n^2 - 5}{2x_n} = \frac{1}{2}x_n + \frac{5}{2x_n}.$$

Wir bekommen, bei Startwert $x_0 := 1$, die Folge von Näherungen

n	Näherung x_n	Fehler $x_n - \sqrt{5}$
0	1.00000000000000000000000000000000	1.23606797749978969640917366873
1	3.00000000000000000000000000000000	0.76393202250021030359082633127
2	2.33333333333333333333333333333333	0.09726535583354363692415966460
3	2.23809523809523809523809523809523809	0.00202726059544839882892156936
4	2.23606889564336372847011144883	$9.1814357 \cdot 10^{-7}$
5	2.23606797749997819409459355859	$1.8849768 \cdot 10^{-13}$
6	2.23606797749978969640917367668	$7.9450575 \cdot 10^{-27}$

Da wir in diesem Fall die exakte Lösung $\sqrt{5} = 2.23606797749978969640917366873\dots$ kennen, können wir in der Tabelle auch den Fehler $x_n - x_*$ angeben und sehen, wie außerordentlich schnell das Verfahren konvergiert. Es hat den Anschein, dass sich der Fehler von Iterationsschritt zu Iterationsschritt in etwa quadriert, d.h. sind in einem Schritt bereits m Dezimalstellen korrekt, so sind im darauf folgenden Iterationsschritt bereits ungefähr $2m$ Dezimalstellen korrekt. So ein Verhalten sollte man erwarten, wenn der Fehler im $n+1$ -ten Iterationsschritt in etwa das Quadrat des Fehlers im n -ten Iterationsschritt ist.

Wir wollen keinen strengen Konvergenzbeweis führen in dem Sinne, dass wir uns genau überlegen, unter welchen Voraussetzungen die Folge der Newton-Iterierten konvergiert; aber die folgende Überlegung soll die gerade beobachtete Quadrierung des Fehlers von Schritt zu Schritt plausibel machen:

Sei $e_n := x_n - x_*$ der Fehler im n -ten Iterationsschritt. Nach dem Satz von Taylor (zum Entwicklungspunkt x_n) gibt es für den Fehler zwischen f und der Tangente T (angelegt an der Stelle x_n) die Darstellung

$$f(x) = T(x) + \frac{f''(\xi)}{2} (x - x_n)^2;$$

ξ zwischen x_n und x . Speziell für $x = x_*$ bekommen wir, wegen $f(x_*) = 0$:

$$T(x_*) = -\frac{f''(\xi)}{2} e_n^2. \tag{3.19}$$

Die Steigung von T ist $f'(x_n)$, d.h. es gilt ('Steigungsdreieck', Skizze!)

$$f'(x_n) = \frac{\overbrace{T(x_{n+1}) - T(x_*)}^{=0}}{x_{n+1} - x_*} = -\frac{T(x_*)}{e_{n+1}} \tag{3.20}$$

Gleichungen (3.19) und (3.20) kombiniert ergeben

$$e_{n+1} = \frac{f''(\xi)}{2 f'(x_n)} e_n^2.$$

Dies ist eine Gleichung, die angibt, wie sich der Fehler von Schritt zu Schritt verändert. Wir folgern: Wenn wir einige zusätzliche Voraussetzungen an f annehmen, wie z.B. dass

f' (zumindest in einer Umgebung U von x_*) 'von der Null wegbleibt', also $\min_U |f'(x)| \geq c > 0$, und f zweimal differenzierbar ist mit beschränkter zweiter Ableitung (zumindest in U), dann kann man

$$|e_{n+1}| \leq \text{const} \cdot |e_n|^2 \quad (3.21)$$

folgern (dabei ist die Konstante $\frac{\max_{x \in U} |f''(x)|}{2 \min_{x \in U} |f'(x)|}$), was das in der Tabelle beobachtete Konvergenzverhalten erklärt. Allerdings, wenn der Startwert x_0 zu weit entfernt von der Lösung x_* liegt, (e_0 also zu groß ist), kann es passieren, dass die Newton-Iteration nicht konvergiert! (Beispiele: siehe Skizze und siehe Übung). Wenn der Anfangsfehler jedoch hinreichend klein ist, z.B. $|e_0| \leq \frac{1}{2 \text{const}}$ (mit const aus (3.21)), dann ist die Konvergenz sicher, denn dann ist

$$\begin{aligned} |e_1| &\leq \underbrace{\text{const} \cdot |e_0|}_{\leq \frac{1}{2}} \cdot |e_0| \leq \frac{1}{2} |e_0|, \\ |e_2| &\leq \underbrace{\text{const} \cdot |e_1|}_{\leq \frac{1}{4}} \cdot |e_1| \leq \frac{1}{8} |e_0| \end{aligned}$$

usw., d.h. die Fehler $|e_n|$ bilden eine monoton fallende Nullfolge.

Beide Eigenschaften, die Tatsache, dass die Folge (i.a. nur) für Startwert hinreichend nahe bei der Nullstelle konvergiert, und dass der Fehler wie durch (3.21) fällt, fasst man zusammen, indem man sagt:

Das Newton-Verfahren ist lokal quadratisch konvergent.

Beim Bisektionsverfahren wird der Fehler dagegen in jedem Iterationsschritt lediglich halbiert.⁵³

Was kann man tun, wenn das Newton-Verfahren nicht konvergiert, oder nicht gegen die gewünschte Nullstelle konvergiert? Wenn auch das Herumprobieren mit anderen Startwerten nicht hilft, kann man zunächst mit dem Bisektionsverfahren einige Schritte durchführen, bis man die Lage der Nullstelle hinreichend genau eingegrenzt hat, und dann mit einem Wert aus dem erhaltenen Intervall als Startwert startet man die Newton-Iteration.

Verhalten bei mehreren Nullstellen. Falls f mehrere Nullstellen hat, so stellt sich die Frage, gegen welche Nullstelle die Newton-Iteration abhängt. sicher ist: Startet man 'hinreichend nahe' bei einer der Nullstellen, so wird die Newton-Iteration gegen diese Nullstelle konvergieren (lokal quadratische Konvergenz). Für Startwerte, wie 'weit' weg von allen Nullstellen liegen, ist es i.a. unvorhersehbar, gegen welche der Nullstellen die Newton-Iteration konvergiert.

Die Anwendung führt zu Fraktalen: Auch wenn wir uns im Kapitel über Differenzierbarkeit

⁵³Beim Bisektionsverfahren müsste man als erstes darüber nachdenken, was man unter 'Fehler' versteht. Z.B. $e_n := \max_{x \in [a_n, b_n]} |x - x_*|$ ist mögliche Definition des Fehlers.

auf Funktionen $f : \mathbb{R} \rightarrow \mathbb{R}$, konzentriert haben, so kann man die Differenzierbarkeit analog auch für Funktionen $f : \mathbb{C} \rightarrow \mathbb{C}$ definieren; auch die üblichen Rechenregeln fürs Ableiten bleiben erhalten. Für ein Polynom $f(z) := z^3 - 1$, $z \in \mathbb{C}$, haben wir somit $f'(z) = 3z^2$, somit können wir für jeden beliebigen Startwert $z_0 \in \mathbb{C} \setminus \{0\}$ die Folge der Newton-Iterierten $z_{n+1} = z_n - \frac{f(z_n)}{f'(z_n)}$ verfolgen. Wir wissen aus dem 1. Semester, dass die Funktion f drei Nullstellen $z_{*,1} = 1$, $z_{*,2} = -\frac{1}{2} + \frac{i}{2}\sqrt{3}$, $z_{*,3} = -\frac{1}{2} - \frac{i}{2}\sqrt{3}$ hat. Man kann nun für *jeden* Startwert die Folge der Newton-Iterierten verfolgen, und dann in der komplexen Ebene den Startwert z_0 in einer von drei Farben einfärben, je nachdem gegen welche der drei Nullstellen die Folge konvergiert ist. Was man auf diese Weise bekommt ist ein wunderschönes *Fraktal!* In der Praxis wird man als Konvergenztest $|z_n - z_{*,i}| \leq \epsilon$ für ein kleines $\epsilon > 0$ verwenden. Nimmt man eine andere Funktion (Polynom) mit mehreren Nullstellen, so bekommt man ein anderes Fraktal.

Numerische Mathematik. Mit den Suchen nach und Analysieren von Algorithmen, mit denen man mathematische Probleme näherungsweise (jedoch oft: "beliebig genau", d.h. mit Fehler kleiner als eine beliebige vorgegebene Fehlerschranke) lösen kann, und die sich auf Computern implementieren lassen, beschäftigt sich die sog. *Numerische Mathematik* (kurz: *Numerik*). Wichtige Aspekte: Rechenaufwand von Verfahren (Anzahl der nötigen Rechenoperationen in Abhängigkeit von z.B. der gewünschten Genauigkeit oder der 'Problemgröße' (unter Problemgröße könnte man z.B. bei Verfahren zum Lösen von linearen Gleichungssystemen die Anzahl der Gleichungen/Unbekannten verstehen), Konvergenzbeweise,...

Newton-Verfahren und Heron-Verfahren. Wir haben im ersten Semester das Heron-Verfahren zur näherungsweisen Berechnung von Wurzeln \sqrt{a} kennengelernt. Wie Ihnen vielleicht aufgefallen ist, ist das Newton-Verfahren, angewandt auf die Funktion $f(x) = x^2 - a$, völlig identisch mit dem Heron-Verfahren zum Finden von \sqrt{a} , d.h. das Heron-Verfahren ist ein Spezialfall des Newton-Verfahrens. Heron hatte nicht unsere Möglichkeiten der Fehlerabschätzung, die bei uns auf dem Satz von Taylor beruhen; er hat geometrisch argumentiert, um die Konvergenz des Verfahrens (einigermaßen) zu begründen. Das allerdings schon fast 2000 Jahre vor Newton und Taylor.

3.7 Integration

3.7.1 Stammfunktionen

Das Suchen nach Stammfunktionen ist nichts anderes als die Umkehrung des Differenzierens. Die grundlegende Bedeutung von Stammfunktionen ist darin zu sehen, dass, wie wir später in Kap. 3.7.2 erkennen werden, es Stammfunktionen uns ermöglichen, den Flächeninhalt unter dem Graphen einer Funktion zu berechnen).

Def. (Stammfunktion) Sei $D := (a, b)$, wobei $-\infty \leq a < b \leq +\infty$. Sei $f : D \rightarrow \mathbb{R}$ stetig.

Eine Funktion $F : D \rightarrow \mathbb{R}$ heißt *Stammfunktion* von f , wenn $F' = f$ gilt auf D .

Ist F eine Stammfunktion von f , so schreiben wir

$$F = \int f(x) dx$$

Bsp. Die Funktion $f(x) := x e^{x^2}$ hat als Stammfunktion $F(x) = \frac{1}{2} e^{x^2}$, denn $\frac{d}{dx} \frac{1}{2} e^{x^2} = x e^{x^2}$.

Aber auch jede Funktion der Art $F(x) = \frac{1}{2} e^{x^2} + c$, $c \in \mathbb{R}$, ist offensichtlich eine Stammfunktion von f , denn beim differenzieren fällt die Konstante weg:

$$\int x e^{x^2} dx = \frac{1}{2} e^{x^2} + c, \quad c \in \mathbb{R}$$

Allgemein gilt: Kennt man zu f eine Stammfunktion F , so haben alle Stammfunktionen von f die Form $F + \text{const}$. Anders ausgedrückt: Die Differenz zweier Stammfunktionen von f ist immer eine Konstante.

Unser Wissen über Ableitungen ausnutzend, können wir sofort eine Liste einiger Stammfunktionen angeben (machen Sie jeweils die 'Probe', also z.B. $\frac{d}{dx}(\frac{1}{n+1}x^{n+1} + c) = x^n$ usw.):

$$\begin{aligned} \int x^n dx &= \frac{1}{n+1} x^{n+1} + c \quad \forall n \in \mathbb{Z} \setminus \{-1\} \\ \int \frac{1}{x} dx &= \left\{ \begin{array}{l} \ln x + c \text{ auf } (0, \infty) \\ \ln(-x) + c \text{ auf } (-\infty, 0) \end{array} \right\} = \ln|x| + c \text{ auf } \mathbb{R} \setminus \{0\} \\ \int e^{ax} dx &= \frac{1}{a} e^{ax} + c \quad \forall a \neq 0 \\ \int \sin x dx &= -\cos x + c \\ \int \cos x dx &= \sin x + c \\ \int \frac{1}{1+x^2} dx &= \arctan x + c \end{aligned}$$

$$\begin{aligned} \int \sinh x \, dx &= \cosh x + c \\ \int \cosh x \, dx &= \sinh x + c \\ \int \frac{1}{\sqrt{1+x^2}} \, dx &= \operatorname{Arsinh} x + c, \quad -1 < x < 1 \\ \int \frac{1}{\sqrt{x^2-1}} \, dx &= \operatorname{Arcosh} x + c, \quad |x| > 1 \\ \int \frac{1}{\sqrt{1-x^2}} \, dx &= \arcsin x + c = -\arccos x + \tilde{c}, \quad -1 < x < 1 \end{aligned}$$

Für komplizierte Funktionen (zusammengesetzt aus einfacheren Funktionen) braucht man i.a. *Rechenregeln*, nämlich die Regel von der *Partiellen Integration*, die *Substitutionsregel*, und die *Partialbruchzerlegung*. Diese sind leider etwas schwieriger anzuwenden als die Rechenregeln der Differentiation, da man i.a. nachdenken muss, welche der Regeln man anwenden muss, um die Stammfunktion(en) zu finden.

Die Regel von der Partiellen Integration. Es gilt

$$\int f(x)g'(x) \, dx = f(x)g(x) - \int f'(x)g(x) \, dx$$

Beweis: Differenzieren der Gleichung, Produktregel der Differentiation.

Beispiele:

Es ist

$$\begin{aligned} \int x e^{2x} \, dx &= x \frac{1}{2} e^{2x} - \int 1 \cdot \frac{1}{2} e^{2x} \, dx = \left(\frac{1}{2}x - \frac{1}{4} \right) e^{2x} + c \\ \int \ln x \, dx &= \int 1 \cdot \ln x \, dx = x \ln x - \int x \cdot \frac{1}{x} \, dx = x \ln x - \int 1 \cdot dx = x \ln x - x + c \end{aligned}$$

Die Substitutionsregel. Es gilt

$$\int f(g(x)) g'(x) \, dx = \int f(y) \, dy + c,$$

wobei die Substitution $y = g(x)$ am Ende rückgängig gemacht werden muss ("Rücksubstitution"). Merkhilfe zur Transformation des Differentials: $\frac{dy}{dx} = g'(x)$ auflösen nach dy (was auch immer dies sein mag): $dy = g'(x) \, dx$.

Spezialfälle der Substitutionsregel sind, mit $f(y) := \frac{1}{y}$,

$$\int \frac{g'(x)}{g(x)} \, dx = \int \frac{1}{y} \, dy = \ln |y| + c = \ln |g(x)| + c,$$

sowie mit $g(x) := \alpha x$, $\alpha \neq 0$,

$$\int f(\alpha x) dx = \frac{1}{\alpha} \int f(y) dy,$$

wobei immer am Ende zurücks substituiert werden muss.

Beispiele:

$$\int \underbrace{(\sin x)^2}_{=:g(x)} \underbrace{\cos x}_{=:g'(x)} dx = \int y^2 dy = \frac{1}{3} y^3 + c = \frac{1}{3} \sin^3 x + c \quad |y := g(x) = \sin(x)$$

$$\int \frac{6x^2 - 5}{2x^3 - 5x + 1} dx = \ln |2x^3 - 5x + 1| + c$$

$$\int \cos(5x) dx = \frac{1}{5} \sin(5x) + c$$

$$\int \frac{(\ln x)^5}{2x} dx = \frac{1}{2} \int y^5 dy = \frac{1}{2} \cdot \frac{1}{6} y^6 + c = \frac{1}{12} (\ln x)^6 + c \quad |y := \ln x$$

$$\int \tan x dx = \int \frac{\sin x}{\cos x} dx = \int \frac{-1}{y} dy = -\ln |y| + c = -\ln |\cos x| + c$$

Bemerkung: Beim Integrieren ist es sinnvoll, am Ende die "Probe" zu machen (differenzieren ist leicht!)

Eine weitere Technik, die sog. **Partialbruchzerlegung**, siehe Kap. 3.7.4.

3.7.2 Das Riemann-Integral

In diesem Kapitel geht es darum, eine Methode zu entwickeln, den Flächeninhalt zwischen dem Graphen einer Funktion $y = f(x)$ und der x -Achse in den Grenzen $x = a$ und $x = b$, $a < b$ zu berechnen. Neben der Beschränktheit des Intervalls $[a, b]$ setzen wir zunächst einmal lediglich voraus, dass die Funktion f auf $[a, b]$ beschränkt sei.

wir folgen der Idee von Bernhard Riemann (1826-1866): Wir zerlegen das Intervall in $n \in \mathbb{N}$ Teilintervalle durch Einfügen von Punkten

$$a = x_0 < x_1 < x_2 < \dots < x_n = b.$$

Ein solches Tupel $Z := (x_0, \dots, x_n)$ nennen wir eine *Zerlegung* des Intervalls $[a, b]$.

Wir betrachten die zu dieser Zerlegung gehörende sog. *Untersumme*

$$\underline{S} = \underline{S}_Z := \sum_{i=1}^n (x_i - x_{i-1}) \cdot \inf_{x \in [x_{i-1}, x_i]} f(x)$$

und die sog. *Obersumme*

$$\overline{S} = \overline{S}_Z := \sum_{i=1}^n (x_i - x_{i-1}) \cdot \sup_{x \in [x_{i-1}, x_i]} f(x)$$

Für den gesuchten (noch zu definierenden) Flächeninhalt A ist plausibel:

1. Es ist $\underline{S} \leq A \leq \overline{S}$ für jede beliebige Zerlegung.
2. \underline{S} und \overline{S} stellen Näherungen an A dar, zumindest sofern die Zerlegung 'fein genug' gewählt wurde.

Als Maß für die Feinheit einer Zerlegung $Z := (x_0, x_1, \dots, x_n)$ setzen wir $|Z| := \max_{i=1, \dots, n} x_i - x_{i-1}$.

Wir definieren daher, Riemann folgend:

Def. (Riemann-Integral) Falls für beliebige Folgen von Zerlegungen, deren Feinheit gegen null geht, die Obersumme und die Untersumme gegen ein und denselben Wert konvergiert, so bezeichnen wir diesen Grenzwert als das Riemann-Integral von f in den Grenzen von a bis b (anschaulich: Flächeninhalt unter dem Graphen); wir bezeichnen das Riemann-Integral mit

$$\int_a^b f(x) dx := \lim_{|Z| \rightarrow 0} \underline{S}(Z) = \lim_{|Z| \rightarrow 0} \overline{S}(Z).$$

Existiert das Riemann-Integral, so heißt f *Riemann-integrierbar*.

Bsp. Wir wollen das Riemann-Integral von $f(x) := x^2$, in den Grenzen von 0 bis b berechnen. Wir betrachten zur Vereinfachung lediglich äquidistante Zerlegungen $Z = Z_n = (0, \frac{b}{n}, \frac{2b}{n}, \dots, \frac{nb}{n})$; deren Feinheit ist $|Z_n| = \frac{b}{n} \xrightarrow{(n \rightarrow \infty)} 0$. (Eigentlich müssten wir beliebige Folgen von Zerlegungen, deren Feinheit gegen null geht, betrachten!) Wir bekommen als Ober- und als Untersumme

$$\begin{aligned} \overline{S}(Z_n) &= \sum_{i=1}^n \frac{b}{n} \cdot \left(\frac{bi}{n}\right)^2 = \frac{b^3}{n^3} \cdot \sum_{i=1}^n i^2 = \frac{b^3}{n^3} \cdot \frac{n}{6} (n+1)(2n+1) = \frac{b^3}{6} \left(1 + \frac{1}{n}\right) \left(2 + \frac{1}{n}\right), \\ \underline{S}(Z_n) &= \sum_{i=0}^{n-1} \frac{b}{n} \cdot \left(\frac{bi}{n}\right)^2 = \frac{b^3}{n^3} \cdot \sum_{i=1}^{n-1} i^2 = \frac{b^3}{n^3} \cdot \frac{n-1}{6} n(2n-1) = \frac{b^3}{6} \left(1 - \frac{1}{n}\right) \left(2 - \frac{1}{n}\right). \end{aligned}$$

Lassen wir n gegen unendlich gehen, konvergieren sowohl Ober- als auch Untersumme gegen $\frac{1}{3}b^3$, also

$$\int_0^b x^2 dx = \frac{1}{3} b^3.$$

Das Ergebnis ist offenbar gleich der Stammfunktion F von f , evaluiert an der Stelle b (minus der Stammfunktion von f , evaluiert an der Stelle 0, was aber gleich null ist):

$$\underbrace{\int_a^b f(x) dx}_{\text{Flächeninhalt}} = \underbrace{F(b)}_{\text{Stammfkt.}} - \underbrace{F(a)}_{\text{Stammfkt.}}$$

Ist das Zufall, oder ein allgemeines Gesetz? Dass ein grundsätzlicher Zusammenhang besteht zwischen der Stammfunktion F einer Funktion f einerseits und dem Flächeninhalt unter dem Graphen von f andererseits, besagt der folgende Satz:

Satz (Erster Hauptsatz der Differential- und Integralrechnung) Sei f auf $[a, b]$ Riemann-integrierbar und sei F eine Stammfunktion von f (also $F' = f$ bzw. $F = \int f dx$). Dann gilt

$$\boxed{\int_a^b f(x) dx = F(b) - F(a)},$$

d.h. man kann den Flächeninhalt unter dem Graphen durch Auswertung einer (beliebigen) Stammfunktion von f an den Endpunkten des Intervalls als Differenz dieser Werte berechnen.

Bemerkung: Beachten Sie, dass man unter Einsetzen von $F' = f$ die obige Formel so lesen kann, dass sie Auskunft gibt, was man erhält wenn man nacheinander differenziert und integriert: $\int_a^b F'(x) dx = F(b) - F(a)$.

Beweis: Wir bilden eine Zerlegung des Intervalls $[a, b]$ und schreiben

$$F(b) - F(a) = \sum_{i=1}^n F(x_i) - F(x_{i-1}) \stackrel{(*)}{=} \sum_{i=1}^n f(\xi_i) (x_i - x_{i-1}), \quad (3.22)$$

wobei wir an der Stelle (*) den Mittelwertsatz auf jedem Teilintervall benutzt haben und $\xi_i \in (x_{i-1}, x_i)$. Der (von der Wahl der Zerlegung unabhängige) Wert (3.22) liegt offensichtlich zwischen \underline{S} und \overline{S} . Da wir die Riemann-Integrierbarkeit von f vorausgesetzt haben, wissen wir, dass für Feinheit der Zerlegung gegen null gehend \underline{S} und \overline{S} gegen $\int_a^b f(x) dx$ gehen. Somit muss auch (3.22) gleich $\int_a^b f(x) dx$ sein. \square

Anmerkungen.

- Statt $F(b) - F(a)$ schreibt man auch $F(x) \Big|_a^b$.
- Die Berechnung einer Stammfunktion $\int f(x) dx$ bezeichnet man auch als *unbestimmtes Integral*; die Berechnung eines Flächeninhalts $\int_a^b f(x) dx$ als *bestimmtes Integral*.

Mit dem obigen Hauptsatz haben wir ein Instrumentarium, um Integrale/Flächeninhalte unter Funktionen zu berechnen, indem wir Stammfunktionen berechnen. Insbesondere übertragen sich die Regeln zur Berechnung von Stammfunktionen auf die Berechnung von Integralen. So gibt es für die Partielle Integration sowie für die Substitutionsregel neben der Version für unbestimmte Integrale (s.o.) auch eine Version für *bestimmte* Integrale:

$$\boxed{\int_a^b f(x)g'(x) dx = f(x)g(x) \Big|_a^b - \int_a^b f'(x)g(x) dx}$$

$$\int_a^b f(g(x))g'(x) dx = \int_{g(a)}^{g(b)} f(y) dy$$

Einige Eigenschaften⁵⁴ des Riemann-Integrals:

$$\begin{aligned} \int_a^b f(x) dx + \int_b^c f(x) dx &= \int_a^c f(x) dx \\ \int_a^b [f(x) + g(x)] dx &= \int_a^b f(x) dx + \int_a^b g(x) dx \\ \int_a^b \alpha f(x) dx &= \alpha \int_a^b f(x) dx \\ f(x) \leq g(x) \quad \forall x \in [a, b] &\implies \int_a^b f(x) dx \leq \int_a^b g(x) dx \end{aligned}$$

Diese Eigenschaften kann man leicht mittels Ober- und Untersummen zeigen. (Sie sind so zu lesen: "Sind $f, g : [a, b] \rightarrow \mathbb{R}$ Riemann-integrierbar, dann ist auch $f+g$ Riemann-integrierbar, und es gilt $\dots=\dots$ "). Ein Spezialfall der letzten der obigen Eigenschaften ist offensichtlich die folgende Abschätzung eines Integrals:

$$(b-a) \cdot \min_{x \in [a, b]} f(x) \leq \int_a^b f(x) dx \leq (b-a) \cdot \max_{x \in [a, b]} f(x)$$

Man definiert außerdem

$$\int_b^a f(x) dx := - \int_a^b f(x) dx$$

d.h. Integrale sind auch dann definiert, wenn die 'untere' Grenze größer als die 'obere' Grenze ist; in dem Fall ist das Integral über eine positive (negative) Funktion also negativ (positiv).

Welche Funktionen sind überhaupt Riemann-integrierbar? Eine ganze Menge:

Satz (Riemann-Integrierbarkeit).

- Jede auf $[a, b]$ stetige Funktion ist Riemann-integrierbar.
- Jede auf $[a, b]$ monotone Funktion ist Riemann-integrierbar.
- Jede Funktion auf $[a, b]$, für die sich $[a, b]$ derart in endlich viele disjunkte Teilintervalle zerlegen lässt, so dass f auf jedem Teilintervall stetig oder monoton ist, ist Riemann-integrierbar.

⁵⁴Die 2. und die 3. Eigenschaft beschreiben die *Linearität* des R-Integrals (genauer: Die Abbildung $I : f \mapsto \int_a^b f(x) dx, \mathcal{C}^1([a, b]) \rightarrow \mathbb{R}$ ist linear); die 4. Eigenschaft bezeichnet man als *Monotonie* des R-Integrals.

Zum Beweis. Wir überspringen hier den Beweis der Tatsache, dass jede stetige Funktion Riemann-integrierbar ist.⁵⁵

Für *monotone* Funktionen ist jedoch ist es ganz leicht zumindest zu zeigen, dass der Abstand von Ober- und Untersumme gegen null geht (was strenggenommen aber noch nicht beweist, dass Ober- und Untersumme konvergieren): Man macht sich anhand einer Skizze sofort klar: Bei monoton wachsendem f ist für jede beliebige Zerlegung die Ober- und die Untersumme

$$\bar{S} = \sum_{i=1}^n (x_i - x_{i-1}) \cdot f(x_i), \quad \underline{S} = \sum_{i=1}^n (x_i - x_{i-1}) \cdot f(x_{i-1}),$$

somit die Differenz

$$\bar{S} - \underline{S} = \sum_{i=1}^n (x_i - x_{i-1}) \cdot (f(x_i) - f(x_{i-1})) \leq \sum_{i=1}^n |Z| (f(x_i) - f(x_{i-1})) = |Z| \cdot (f(b) - f(a)),$$

was offensichtlich gegen null geht für Zerlegungsfeinheit $|Z|$ gegen null. \square

Kann man auch Funktionen angeben, die nicht Riemann-integrierbar sind? Das bekannteste Beispiel für eine solche Funktion ist

$$f(x) := \begin{cases} 0, & x \in \mathbb{Q} \\ 1, & x \in \mathbb{R} \setminus \mathbb{Q} \end{cases},$$

mit Grenzen $a < b$ beliebig: Offensichtlich ist für jede beliebige Zerlegung von $[a, b]$ die Untersumme immer gleich null und die Obersumme immer gleich $b - a$; beide konvergieren also nicht gegen einen gemeinsamen Wert.⁵⁶

Wir formulieren nun einen Mittelwertsatz speziell für Integrale:

⁵⁵Der Beweis dieser Tatsache ist eher technisch; er verwendet letztendlich die ϵ - δ -Charakterisierung der Stetigkeit von f . Man muss jedoch zunächst als Hilfsaussage beweisen, dass das δ , das zunächst von ϵ und x abhängen kann, bei einer *kompakten* Menge $[a, b]$ unabhängig von x gewählt werden kann; dies nennt man auch *gleichmäßige Stetigkeit*; d.h. man muss zunächst den Hilfssatz beweisen, dass auf einer kompakten Menge jede stetige Funktion gleichmäßig stetig ist. Wenn man diesen Hilfssatz hat, dann kann man sich überlegen, dass man für jedes vorgegebene $\epsilon > 0$ die Feinheit der Zerlegung nur kleiner als $\delta > 0$ wählen muss, um die Differenz von Ober- und Untersumme unter $(b-a) \cdot \epsilon$ zu drücken. Daraus, dass dies für *jedes* $\epsilon > 0$ möglich ist, folgert man die Behauptung, ϵ gegen null gehen lassend.

⁵⁶Um diesem Mangel zu beheben wurde später das sog. *Lebesgue-Integral* entwickelt (Henri Lebesgue, 1902). Seine Definition ist zunächst mal komplizierter als beim Riemann-Integral; deswegen verzichten wir hier darauf. Es hat jedoch viele Vorteile: Neben der Tatsache, dass man damit 'mehr' Funktionen integrieren kann, kommt es in vielen mathematischen Sätzen vor, die, mit dem Riemann-Integral deutlich schwieriger zu formulieren wären. Es spielt daher heutzutage eine sehr große Rolle in vielen Gebieten, wie der Funktionalanalysis, Maßtheorie, Stochastik (=Wahrscheinlichkeitsrechnung), angewandten Mathematik, Physik. Beide Integrationsbegriffe sind konsistent in dem Sinn, dass wenn das Riemann-Integral einer Funktion existiert, immer auch das Lebesgue-Integral existiert und beide den gleichen Wert haben.

Satz (Mittelwertsatz der Integralrechnung). Sei $f : [a, b] \rightarrow \mathbb{R}$ stetig. Dann gibt es ein $\xi \in (a, b)$, so dass

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

Beweis. Da $[a, b]$ kompakt und f stetig, gibt es nach dem Satz vom Minimum/Maximum

$$M := \max_{x \in [a, b]} f(x), \quad m := \min_{x \in [a, b]} f(x),$$

sowie $x_1, x_2 \in [a, b]$ mit $f(x_1) = M$, $f(x_2) = m$. Offensichtlich liegt der Mittelwert η von f auf $[a, b]$, $\eta := \frac{1}{b-a} \int_a^b f(x) dx$ zwischen $\frac{1}{b-a} \int_a^b m dx = m$ und $\frac{1}{b-a} \int_a^b M dx = M$. Nach dem Zwischenwertsatz, angewendet auf f , gibt es also ein ξ zwischen x_1 und x_2 (somit in (a, b)), so dass $\eta = f(\xi)$; also $f(\xi) = \frac{1}{b-a} \int_a^b f(x) dx$. Multiplikation mit $b-a$ liefert die Behauptung. \square

Der nächste Satz besagt, wie man Stammfunktionen mittels eines Riemann-Integrals finden/darstellen kann (oder anders ausgedrückt: was passiert, wenn man ein Integral bzgl. seiner Integrationsgrenze differenziert):

Satz (Zweiter Hauptsatz der Differential- und Integralrechnung) Sei f stetig auf $[a, b]$. Dann ist die Funktion $F : [a, b] \rightarrow \mathbb{R}$, definiert durch

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

eine Stammfunktion von f ; d.h. es ist

$$\frac{d}{dx} \int_a^x f(t) dt = f(x). \tag{3.23}$$

Beweis. Dass f wegen seiner Stetigkeit Riemann-integrierbar ist, haben wir schon weiter oben festgehalten. Wir zeigen nun noch (3.23), indem wir den Differenzenquotienten untersuchen:

$$\lim_{h \rightarrow 0} \frac{\int_a^{x+h} f(t) dt - \int_a^x f(t) dt}{h} = \lim_{h \rightarrow 0} \frac{1}{h} \int_x^{x+h} f(t) dt \stackrel{(*)}{=} \lim_{h \rightarrow 0} \frac{1}{h} \cdot h \cdot f(\xi)$$

wobei wir an der Stelle (*) den Mittelwertsatz der Integration angewendet haben, und ξ zwischen x und $x+h$ liegt. Somit, für $h \rightarrow 0$, geht ξ gegen x , und somit erhalten wir wegen der Stetigkeit von f oben den Grenzwert $f(x)$. \square

Anwendungen des zweiten Hauptsatzes:

- Berechnen Sie die Ableitung von $f(x) := \int_0^{g(x)} h(t) dt$ (Hinweis: Kettenregel; H sei Stammfunktion von h)
- Berechnen Sie die Ableitung von $f(x) := \int_x^b h(t) dt$ (H sei Stammfunktion von h)
und von $\tilde{f}(x) = \int_x^{2x} h(t) dt$
- Stellen Sie das Taylor-Polynom dritter Ordnung für $f(x) := \int_0^x e^{t \cos t} dt$ auf zum Entwicklungspunkt $x_* = 0$ auf.
- (Klausuraufgabe 2009): Stellen Sie die Taylor-Polynome T_1, T_2 erster und zweiter Ordnung für $f(x) = \cos x \int_{\pi}^x t \cos t dt$ auf zum Entwicklungspunkt $x_* = \pi$. Stellen Sie für T_1 das Restglied R_1 auf und finden Sie unter der Annahme $x \in [0, \pi]$ eine Abschätzung der Form $|R_1(x)| \leq c|x - \pi|^2$ (d.h. geben Sie ein c an).

3.7.3 Uneigentliche Riemann-Integrale

Für die Definition des Riemann-Integrals im vorangegangenen Kapitel hatten wir vorausgesetzt,

1. dass das Integrationsintervall $[a, b]$ beschränkt ist,
2. dass die zu integrierende Funktion auf $[a, b]$ beschränkt ist.

Sobald man eine der beiden Bedingungen fallen lässt, sind nämlich Ober- und Untersumme nicht mehr zu definieren (d.h. können unendlich sein).

Eine Möglichkeit, auch für unbeschränkte Integrationsgebiete und/oder unbeschränkte Integranden ein Integral zu definieren ist (neben dem im vorangegangenen Kapitel kurz erwähnten Lebesgue-Integral) das sog. *uneigentliche Riemann-Integral*. Eine allgemeine Darstellung ist schwierig; es ist sinnvoll, ein paar Fälle/Beispiele darzustellen:

Bsp. 1: Integrationsbereich ist $I = [a, \infty)$; Integrand ist auf $[a, \infty)$ beschränkt:

Man definiert dann

$$\int_a^{\infty} f(x) dx := \lim_{b \rightarrow \infty} \int_a^b f(x) dx$$

sofern dieser Grenzwert existiert.

Analog im Fall $I = (-\infty, b]$.

Bsp. 2: Integrationsbereich ist $I = \mathbb{R}$; Integrand ist auf \mathbb{R} beschränkt:

Man definiert

$$\int_{-\infty}^{\infty} f(x) dx := \lim_{a \rightarrow -\infty} \int_a^m f(x) dx + \lim_{b \rightarrow \infty} \int_m^b f(x) dx$$

falls diese beiden Grenzwerte existieren. Die Wahl von $m \in \mathbb{R}$ ist dabei beliebig; das Ergebnis hängt nicht von der Wahl von m ab.

Bsp. 3: Integrationsbereich $I = [a, b)$ ist beschränkt, aber $f(x)$ ist (nur) für $x \rightarrow b$ unbeschränkt:

Man definiert dann

$$\int_a^b f(x) dx := \lim_{\epsilon \rightarrow 0} \int_a^{b-\epsilon} f(x) dx$$

falls dieser Grenzwert existiert.

Analog geht man vor, wenn $f(x)$ (nur) für $x \rightarrow a$ unbeschränkt ist.

Bsp. 4: Integrationsbereich $I = (a, b)$ ist beschränkt, aber $f(x)$ ist sowohl für $x \rightarrow a$ als auch für $x \rightarrow b$ unbeschränkt:

Man definiert dann

$$\int_a^b f(x) dx := \lim_{\epsilon \rightarrow 0} \int_{a+\epsilon}^m f(x) dx + \lim_{\epsilon \rightarrow 0} \int_m^{b-\epsilon} f(x) dx$$

falls diese beiden Grenzwerte existieren. Dabei ist die Wahl von $m \in (a, b)$ beliebig.

Darüber hinaus gibt es noch Kombinationsmöglichkeiten wie $I = [a, \infty)$, $f(x)$ für $x \rightarrow a$ unbeschränkt etc; diese behandelt man analog zu oben Fällen 2 und 4 durch Unterteilung an einer Stelle m .

Beispiel. Untersuche die Existenz von $\int_I \frac{1}{x^\alpha} dx$ für $\alpha > 0$, $I = [a, \infty)$, $I = (0, b]$, $I = (0, \infty)$:

Eine Stammfunktion F von f ist

$$F(x) = \begin{cases} \frac{1}{1-\alpha} x^{1-\alpha} & , \alpha \neq 1 \\ \ln x & , \alpha = 1 \end{cases}$$

also

$$\int_a^b f(x) dx = F(x) \Big|_a^b = \begin{cases} \frac{1}{1-\alpha} (b^{1-\alpha} - a^{1-\alpha}) & , \alpha \neq 1 \\ \ln b - \ln a & , \alpha = 1 \end{cases}$$

Wir erhalten

$$\int_a^\infty \frac{dx}{x^\alpha} = \lim_{b \rightarrow \infty} \int_a^b \frac{dx}{x^\alpha} = \begin{cases} \frac{a^{1-\alpha}}{\alpha-1} & , \alpha > 1 \\ \infty \text{ (ex. nicht)} & , \alpha \leq 1 \end{cases}$$

und

$$\int_0^b \frac{dx}{x^\alpha} = \lim_{a \rightarrow 0} \int_a^b \frac{dx}{x^\alpha} = \begin{cases} \infty \text{ (ex. nicht)} & , \alpha \geq 1 \\ \frac{b^{1-\alpha}}{1-\alpha} & , \alpha < 1 \end{cases}$$

und

$$\int_0^{\infty} \frac{dx}{x^\alpha} = \int_0^a \frac{dx}{x^\alpha} + \int_a^{\infty} \frac{dx}{x^\alpha} \text{ existiert nicht } \forall \alpha > 0.$$

Merke:

$\int_{x_0}^{\infty} \frac{1}{x^\alpha} dx$ existiert nur für $\alpha > 1$, und

$\int_0^{x_0} \frac{1}{x^\alpha} dx$ existiert nur für $\alpha < 1$

(α und x_0 positiv vorausgesetzt); also z.B. $\int_1^{\infty} \frac{1}{x^2} dx$ und $\int_0^1 \frac{1}{\sqrt{x}} dx$ existieren. ⁵⁷

3.7.4 Partialbruchzerlegung

Die Partialbruchzerlegung ist eine Methode, um gebrochen-rationale Funktionen

$$f(x) = \frac{\sum_{k=0}^n \alpha_k x^k}{\sum_{k=0}^m \beta_k x^k}, \quad \alpha_k, \beta_k \in \mathbb{R}, \quad (3.24)$$

in Form einer Summe von (einfacheren) Termen/Brüchen zu schreiben. Die Partialbruchzerlegung ist insbesondere nützlich, um gebrochen-rationale Funktionen zu integrieren, denn die einzelnen Summanden lassen sich leichter integrieren.

Erster Schritt: Abspalten des "ganzen" Anteils.

Ist $n \geq m$, so kann man einen polynomiellen Anteil (mit Polynomgrad $n-m$) abspalten derart, dass ein gebrochen-rationaler Anteil mit Zählergrad echt kleiner als Nennergrad zurückbleibt.

Beispiel: Für $f(x) := \frac{x^4+x^3+3x^2+5x-3}{x^2+x-1}$ liefert eine Polynomdivision x^2+4 mit Rest $x+1$, also

$$f(x) = x^2+4 + \frac{x+1}{x^2+x-1}.$$

Der "ganze" Anteil ist als Polynom leicht zu integrieren. Wir können also im Folgenden o.B.d.A. voraussetzen, dass wir es grundsätzlich mit $n < m$ zu tun haben.

Zweiter Schritt: Bestimme alle (auch komplexen) Nullstellen des Nenners sowie deren Vielfachheit

Hierzu gibt es (bei großem m) keinen Algorithmus; man kann Nullstellen raten und dann durch Polynomdivision den Polynomgrad reduzieren; bei ganzzahligen Koeffizienten sollte man alle ganzzahligen Teiler des absoluten Gliedes durchprobieren. Wenn wir Glück haben ist der Nenner bereits in faktorisierte Form gegeben.

⁵⁷Kurz, um sich's zu merken: Für großes x und großes α wird $\frac{1}{x^\alpha}$ klein, d.h. hinreichend klein, so dass das Integral existiert.

Beachte: Aus dem 1. Semester wissen wir, dass nicht-reelle Nullstellen von reellen Polynomen immer paarweise konjugiert komplex als $z_0 = a+bi$ und $\bar{z}_0 = a-bi$ auftreten; multipliziert man die zugehörigen Linearfaktorenpaare aus, so bekommt man immer ein *reelles* Polynom zweiten Grades: $(x-z_0)(x-\bar{z}_0) = x^2 - (z_0+\bar{z}_0)x + z_0\bar{z}_0 = x^2 - 2a x + (a^2+b^2)$.

Der sog. **Satz über die Partialbruchzerlegung** besagt: Für jede Funktion der Form (3.24) mit $n < m$ gibt es eine Darstellung

$$f(x) = \sum_{i=1}^k \sum_{j=1}^{r_i} \frac{a_{ij}}{(x-x_i)^j} + \sum_{i=1}^l \sum_{j=1}^{s_i} \frac{b_{ij}x + c_{ij}}{((x-z_i)(x-\bar{z}_i))^j}, \quad a_{ij}, b_{ij} \in \mathbb{R} \quad (3.25)$$

Dabei sind die x_i die reellen Nullstellen mit Vielfachheit r_i , und die z_i, \bar{z}_i sind die nicht-reellen Nullstellen mit Vielfachheit s_i .

Beispiel: $f(x) = \frac{1}{(x-3)(x+4)^2(x^2+6)}$ hat eine Partialbruchzerlegung

$$f(x) = \frac{a_1}{x-3} + \frac{a_2}{x+4} + \frac{a_3}{(x+4)^2} + \frac{a_4x+a_5}{x^2+6}$$

Schritt 3: Berechnen der Koeffizienten der Partialbruchzerlegung. Die gängigste Methode ist: Die Gleichung "f(x) = Partialbruchzerlegung" wird mit dem Hauptnenner durchmultipliziert; dies gibt auf beiden Seiten der Gleichung ein Polynom; die Unbekannten werden mittels Koeffizientenvergleich ermittelt. (Eine andere Methode ist es, jeweils nur mit einem der Nenner $(x-x_i)^{r_i}$ durchzumultiplizieren und dann $x := x_i$ einzusetzen; insbesondere falls alle Nullstellen einfach und reell sind, kommt man so unmittelbar an alle Koeffizienten; andernfalls, also falls die so erhaltenen Gleichungen noch zu wenige sind, kann man weitere konkrete x -Werte (die keine Nenner-Nullstellen sind) einsetzen, um weitere Gleichungen zu erhalten.)

Schritt 4: Integrieren der Partialbrüche. Lautet die Aufgabe, eine Stammfunktion oder ein bestimmtes Integral von f (zwischen zwei Polstellen) zu bestimmen, so suchen wir nun Stammfunktionen der Partialbrüche, also der Summanden, wie sie in (3.25) vorkommen.

Im Fall, dass alle Nullstellen reell sind, ist das einfach, denn wir kennen die Stammfunktion von Termen der Form $\frac{a_{ij}}{(x-x_i)^j}$.

Falls auch nicht-reelle Nullstellen vorkommen, so wollen wir uns hier auf den Fall beschränken, dass deren Vielfachheit eins ist, d.h. wir wollen Terme der Form $\frac{ax+b}{x^2+cx+d}$ bestimmen. Eine mögliche Vorgehensweise:

Falls $a \neq 0$, 'identifiziert' man zunächst die Ableitung des Nenners im Zähler, d.h. man schreibt $\frac{ax+b}{x^2+cx+d} = \frac{\frac{a}{2}(2x+c) + b - \frac{ac}{2}}{x^2+cx+d} = \frac{a}{2} \frac{2x+c}{x^2+cx+d} + (b - \frac{ac}{2}) \frac{1}{x^2+cx+d}$. Den ersten der beiden Summanden kann man sofort mittels Substitution integrieren; man erhält einen logarithmischen Term. Es reicht also, sich auf Integranden der Form $\frac{1}{x^2+cx+d}$ zu konzentrieren. Für solche schreibt man den Nenner, quadratische Ergänzung benutzend,

um in die Form $(x+\alpha)^2 + \beta$; dabei ist $\beta > 0$, denn andernfalls hätte der Term reelle Nullstelle(n). Dann teilt man sowohl Zähler als auch Nenner durch β und führt anschließend eine affin-lineare Substitution $y := (x+\alpha)/\sqrt{\beta}$ durch, so dass ein Bruch der Form $\gamma \cdot \frac{1}{y^2+1}$ zu integrieren bleibt. Dessen Stammfunktion ist der Arcustangens.

4 Grundlagen der Analysis im \mathbb{R}^n

Beispiele und Veranschaulichungen von/für Funktionen $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$.

1. Die Topografie einer Landschaft (Höhenrelief) kann man als Abbildung $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ oder $f : D \rightarrow \mathbb{R}$, $D \subset \mathbb{R}^2$ auffassen. Die Visualisierung kann mittels Farbtönen oder Höhenlinien (mathematisch: =:Niveaulinien) erfolgen (\rightarrow Landkarten); der Graf einer solchen Abbildung, also die Gesamtheit der Punkte $(x, f(x))$, ist eine Teilmenge des \mathbb{R}^3 .
2. Die physikalischen Größen *Druck*, *Dichte*, *Temperatur* in einem Gas oder einem Objekt oder einer Flüssigkeit kann man als Funktion $f : \mathbb{R}^3 \rightarrow \mathbb{R}$ auffassen oder als $f : D \rightarrow \mathbb{R}$, wobei $D \subset \mathbb{R}^3$. An jedem Punkt des \mathbb{R}^3 oder von $D \subset \mathbb{R}^3$ ist ein Funktionswert 'angeheftet'. Eine Visualisierung ist i.a. nichttrivial.
3. Bei strömenden Gasen oder Flüssigkeiten (Luftströmung) kann man an jedem Punkt des \mathbb{R}^3 (oder jedem Punkt aus $D \subset \mathbb{R}^3$) sich einen Geschwindigkeits'pfeil' (-vektor) vorstellen; Das Geschwindigkeitsfeld einer Strömung ist also eine Abbildung $F : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ oder $f : D \rightarrow \mathbb{R}^3$, $D \subset \mathbb{R}^3$.

Viele Naturvorgänge, insbesondere Strömungsvorgänge, werden durch sog. *partielle Differentialgleichungen* beschrieben; das sind Gleichungen, die einen Zusammenhang herstellen zwischen verschiedenen sog. *partiellen Ableitungen* (siehe Kap. 4.3.1) solcher Größen wie Temperatur, Druck, Geschwindigkeitsfeldern. Ein einfaches Beispiel ist die *Wärmeleitungsgleichung* $\partial_t T(t, x, y, z) - k \cdot [\partial_x \partial_x T(t, x, y, z) + \partial_y \partial_y T(t, x, y, z) + \partial_z \partial_z T(t, x, y, z)] = 0$; gesucht ist eine *Funktion*(!) T =Temperatur, die diese Gleichung erfüllt (sowie außerdem gewisse Anfangsbedingungen zu einem Zeitpunkt $t = t_0$ und gewisse Randbedingungen am Rande des zu beschreibenden Objekts); k ist eine gegebene Größe, die die Wärmeleitfähigkeit beschreibt. In manchen speziellen Situation kann man die exakte Lösung von (einfachen) partiellen Differentialgleichungen berechnen; meist aber muss man mit Hilfe von Computern und der Numerik nach Näherungslösungen suchen.

4. Es gibt Kurven, die man nicht ohne weiteres als Graf einer Funktion darstellen kann, z.B. $K := \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 = 1\}$.

Für solche Kurven ist es oft sinnvoll, eine sog. *Parametrisierung* zu betrachten; im Beispiel: $K = \{(\cos t, \sin t) \mid t \in [0, 2\pi)\}$. Das bedeutet: Kurven im \mathbb{R}^2 kann man beschreiben, indem man eine Abbildung $f : D \rightarrow \mathbb{R}^2$ (die sog. Parametrisierung) findet, wobei $D \subseteq \mathbb{R}$, so dass $K = \{f(t) \mid t \in D\}$; also: Die Kurve K ist als Bild der Abbildung f zustande gekommen.

Analoges geht auch im \mathbb{R}^3 . (Was für eine Kurve beschreibt die Parametrisierung $f(t) := (\cos t, \sin t, at)$, $t \in \mathbb{R}$?)

5. Als Verallgemeinerung von 4. kann man (i.a. gekrümmte) *Flächen* im \mathbb{R}^3 mittels einer Parametrisierung $f : D \rightarrow \mathbb{R}^3$, wobei $D \subset \mathbb{R}^2$, darstellen.

Welche Fläche wird z.B. parametrisiert durch $f(t, \varphi) := (r \cos \varphi, r \sin \varphi, t)$, wobei $t \in [0, H]$, $\varphi \in [0, 2\pi)$, und $r, H > 0$ fest? Die Mantelfläche eines Zylinders.

Und durch $f(\varphi, \nu) = (r \cos \varphi \cos \nu, r \sin \varphi \cos \nu, r \sin \nu)$ mit $\varphi \in [0, 2\pi)$, $\nu \in [-\frac{\pi}{2}, \frac{\pi}{2}]$, $r > 0$, wird die Oberfläche einer Kugel mit Radius r parametrisiert; φ entspricht der geografischen Länge, ν der geographischen Breite, d.h. $\nu = 0$ entspricht dem Äquator...

Solche Parametrisierungen von Flächen liefern später u.a. Formeln zur Berechnung des Flächeninhalts und des Schwerpunktes von Flächen (d.h. des Gewichts und des Schwerpunktes von Blechen etc.)

6. Die bisherigen Beispiele handelten von Funktionen die von einem (n -dimensionalen) \mathbb{R}^n in einen m -dimensionalen Vektorraum \mathbb{R}^m abbilden. Man kann darüberhinaus auch Abbildungen zwischen unendlichdimensionalen Vektorräumen (z.B. sog. Funktionenräumen) betrachten. Beispiel: Die Abbildung \mathcal{D} (für 'Differenzieren'), die jeder differenzierbaren Funktion f ihre Ableitung f' zuordnet, also etwa $\mathcal{D} : \mathcal{C}^1(a, b) \rightarrow \mathcal{C}^0(a, b)$, $f \mapsto f'$ (also: $\mathcal{D}(f) := f'$). Ein weiteres Beispiel ist die Abbildung $I : \mathcal{C}([a, b]) \rightarrow \mathbb{R}$, $f \mapsto \int_a^b f(x) dx$.

Wir werden uns in dieser Vorlesung jedoch auf Funktionen zwischen endlichdimensionalen Vektorräumen konzentrieren; die Untersuchung von Abbildung zwischen unendlichdimensionalen Vektorräumen führt in das Gebiet der sog. *Functionalanalysis*.

Die Vielfältigkeit der Anwendungsmöglichkeiten motiviert, dass es sinnvoll ist, möglichst viele der in Kap. 2 definierten Begriffe (Stetigkeit, Differenzierbarkeit,...) auf Funktionen $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ bzw. $f : D \rightarrow \mathbb{R}^m$ mit $D \subseteq \mathbb{R}^n$ zu verallgemeinern. Da diese Begriffe i.a. mittels *Folgen* erklärt werden, müssen wir uns zunächst mit Folgen im \mathbb{R}^n (oder allgemeiner: in einem Vektorraum) beschäftigen und festlegen, was wir unter *Konvergenz einer Folge* in einem Vektorraum verstehen wollen.

4.1 Folgen im \mathbb{R}^n

Um in \mathbb{R} zu definieren, wann eine Folge (x_n) konvergent ist gegen ein $x \in \mathbb{R}$, haben wir den 'Abstand' von x_n und x betrachtet. Diesen Abstand haben wir gemessen mit Hilfe des absoluten Betrages als $|x - x_n|$ und haben gefordert, dass dieser $\leq \epsilon$ wird. Was ist eine sinnvolle Übertragung auf den \mathbb{R}^n , oder allgemeiner, auf Vektorräume? Der absolute Betrag hat die Eigenschaften (1.) $|x| \geq 0$, (2) $|x| = 0 \Leftrightarrow x = 0$, (3) $|\alpha x| = |\alpha| |x|$, (4) $|x + y| \leq |x| + |y|$ für alle $x, y, \alpha \in \mathbb{R}$. Wir erinnern uns an das erste Semester, dass *eine Norm* in einem Vektorraum genau diese Eigenschaften hat: Eine Abbildung $\|\cdot\| : V \rightarrow \mathbb{R}$ heißt *Norm* auf dem \mathbb{R} -Vektorraum V , wenn (1.) $\|\vec{x}\| \geq 0$, (2) $\|\vec{x}\| = 0 \Leftrightarrow \vec{x} = \vec{0}$, (3) $\|\alpha \vec{x}\| = |\alpha| \|\vec{x}\|$, (4) $\|\vec{x} + \vec{y}\| \leq \|\vec{x}\| + \|\vec{y}\|$ für alle $x, y \in V$, $\alpha \in \mathbb{R}$.

Ein Vektorraum V mit einer Norm $\|\cdot\|$ heißt *normierter Vektorraum*.

Beispiele für Normen im Vektorraum $V = \mathbb{R}^n$ sind (siehe 1. Sem.)

1. Die Summennorm $\|\vec{x}\|_1 := |x_1| + \dots + |x_n|$,

2. die Euklidische Norm $\|\vec{x}\|_2 := (|x_1|^2 + \dots + |x_n|^2)^{\frac{1}{2}}$

3. die Maximumnorm $\|\vec{x}\|_\infty := \max\{|x_1|, \dots, |x_n|\}$

Den Abstand von zwei Punkten \vec{x}, \vec{y} in einem normierten Vektorraum $(V, \|\cdot\|)$ kann man also messen mittels $\|\vec{x} - \vec{y}\|$.

Einfach indem wir den absoluten Betrag durch die Norm ersetzen, bekommen wir eine Definition des Begriffs 'Konvergenz von Folgen' auf (normierte) Vektorräume:

Def. (Konvergenz einer Folge). Sei $(\vec{x})_{n \in \mathbb{N}}$ eine Folge im \mathbb{R} -Vektorraum V und $\|\cdot\|$ eine Norm auf V . Die Folge heißt konvergent gegen $\vec{x} \in V$, wenn

$$\forall \epsilon > 0 \exists n_0 \in \mathbb{N} \forall n \geq n_0 : \|\vec{x}_n - \vec{x}\| \leq \epsilon$$

analog überträgt man die *Beschränktheit* von Folgen:

Def. (Beschränktheit von Folgen). Eine Folge (\vec{x}_n) in V heißt beschränkt bzgl. der Norm $\|\cdot\|$, wenn es ein $c \in \mathbb{R}$ gibt so dass

$$\forall n \in \mathbb{N} : \|\vec{x}_n\| \leq c$$

Eine naheliegende Frage ist folgende: Im \mathbb{R}^m gibt es verschiedene Normen. Ist es relevant, *welche* Norm ich benutze zur Konvergenzuntersuchung oder zur Untersuchung der Beschränktheit? D.h. kann es sein, dass eine Folge bezüglich einer Norm gegen ein $\vec{x} \in \mathbb{R}^m$ konvergiert und bezüglich einer anderen Norm divergent ist (oder gegen einen anderen Wert konvergiert)? Die Antwort lautet, dass dies im \mathbb{R}^m nicht möglich ist: **Ist eine Folge im \mathbb{R}^m bezüglich einer Norm konvergent, so ist sie auch bezüglich jeder anderen Norm konvergent, und zwar gegen den selben Grenzwert. Ist sie bezüglich einer Norm beschränkt, so ist sie auch bezüglich jeder anderen Norm beschränkt.** Man kann sich zur Untersuchung der Konvergenz oder der Beschränktheit also die Norm aussuchen, die einem am einfachsten dazu geeignet erscheint.

(Aber Vorsicht: Diese Aussagen gelten nur im \mathbb{R}^m und sind in unendlichdimensionalen Vektorräumen i.a. falsch (was das Arbeiten in solchen Vektorräumen, z.B. also Funktionenräumen, erheblich schwieriger macht im Vergleich zum \mathbb{R}^m .)

Wie in \mathbb{R} gilt auch in $(V, \|\cdot\|)$: **Aus Konvergenz folgt Beschränktheit.**

Und ganz analog wie bei der obigen Definition der Konvergenz kann man auch das **Cauchy-Kriterium** auf normierte Vektorräume verallgemeinern, indem man $|\cdot|$ durch die Norm ersetzt. Im \mathbb{R}^m gilt, wie schon in \mathbb{R} , dass eine Folge genau dann konvergent ist, wenn sie das Cauchy-Kriterium erfüllt (auch dies ist in unendlich-dimensionalen Vektorräumen i.a. leider falsch).

Auch Teilfolgen sind genau so zu definieren wie bei Folgen in \mathbb{R} ; der Satz von Bolzano-Weierstraß, dass jede beschränkte Folge eine konvergente Teilfolge hat, gilt

auch im \mathbb{R}^m , ist jedoch in unendlichdimensionalen Vektorräumen i.a. falsch.

Es bleibt noch der Begriff der *komponentenweisen Konvergenz*: Bei einer Folge (\vec{x}_n) im \mathbb{R}^m ist jedes \vec{x}_n ein Objekt im \mathbb{R}^m , kann mittels seiner Komponenten $\vec{x}_n = (x_{n,1}, \dots, x_{n,m})^T$ geschrieben werden. Hat die Konvergenz (Beschränktheit) der Folge $(\vec{x}_n)_{n \in \mathbb{N}}$ im \mathbb{R}^m etwas mit der Konvergenz (Beschränktheit) der m Komponentenfolgen $(x_{n,k})_{n \in \mathbb{N}}$, $k = 1, \dots, m$, zu tun? In der Tat:

Satz (Konvergenz, Komponentenfolgen). Eine Folge im \mathbb{R}^m ist genau dann konvergent (bzw.: beschränkt), wenn alle ihre m Komponentenfolgen in \mathbb{R} konvergent (bzw.: beschränkt) sind.

Das sieht man, wenn man die komponentenweise Konvergenz mit der Konvergenz in der $\|\cdot\|_\infty$ -Norm vergleicht: Es ist offensichtlich

$$\begin{aligned} \|\vec{x}\|_\infty \leq c &\iff \forall j = 1, \dots, m : |x_{n,j}| \leq c && \text{für Beschränktheit, bzw.} \\ \|\vec{x} - \vec{x}_*\|_\infty \leq \epsilon &\iff \forall j = 1, \dots, m : |x_{n,j} - x_{*,j}| \leq \epsilon && \text{für Konvergenz. } \square \end{aligned}$$

Bemerkung (Rechenregel für Summen konvergenter Folgen): Es gilt, wie zu erwarten war: Sind (\vec{x}_n) und (\vec{y}_n) konvergente Folgen in $(V, \|\cdot\|)$, so ist auch die Folge $(\vec{x}_n + \vec{y}_n)$ konvergent, und es gilt $\lim_{n \rightarrow \infty} \vec{x}_n + \vec{y}_n = \lim_{n \rightarrow \infty} \vec{x}_n + \lim_{n \rightarrow \infty} \vec{y}_n$.

(Beachten Sie: Eine entsprechende Regel für "Produkte" wäre sinnlos, da das Produkt von Elementen eines Vektorraums i.a. nicht definiert ist.)

4.2 Stetigkeit

Vorbemerkung: Die Definition der *Topologischen Grundbegriffe* (offen, abgeschlossen, Abschluss, Rand), die wir für Teilmengen von \mathbb{R} definiert hatten, lassen sich direkt auf Teilmengen von normierten Vektorräumen übertragen; man ersetzt den Absolutbetrag durch die Norm.

Wir betrachten Funktionen $f : V \rightarrow W$ (oder $f : D \rightarrow W$, wobei $D \subseteq V$), wobei $(V, \|\cdot\|_V)$ und $(W, \|\cdot\|_W)$ normierte \mathbb{R} -Vektorräume sind. Zur Klarheit der Darstellung haben wir die beiden verschiedenen Normen mittels Indizes gekennzeichnet; oft wird so ein Index auch weggelassen, da i.a. aus dem Zusammenhang hervorgeht, welche Norm jeweils gemeint ist.

Wie schon bei Funktionen von \mathbb{R} nach \mathbb{R} kann man die Stetigkeit von Funktionen zwischen normierten Vektorräumen mittels Epsilon-Delta-Kriterium definieren, und zwar, indem man absolute Beträge durch Normen ersetzt, oder man kann, was äquivalent ist (sogar in unendlichdimensionalen Vektorräumen) die Stetigkeit mittels Folgen definieren (s. Satz unten):

Def. (Stetigkeit, Epsilon-Delta-Kriterium). $f : D \rightarrow W$, wobei $D \subseteq V$, und wobei $(V, \|\cdot\|_V)$ und $(W, \|\cdot\|_W)$ normierte \mathbb{R} -Vektorräume sind, heißt *stetig* an der

Stelle $x_* \in D$, wenn

$$\forall \epsilon > 0 \exists \delta > 0 \forall x \in D : \|x - x_*\|_V \leq \delta \implies \|f(x) - f(x_*)\|_W \leq \epsilon$$

f heißt *stetig*, wenn f stetig ist an allen Stellen $x \in D$.

Satz (Stetigkeit, Folgenkriterium). f ist genau dann stetig an der Stelle $x_* \in D$, wenn für jede Folge (\vec{x}_n) mit $\lim_{n \rightarrow \infty} \vec{x}_n = \vec{x}_*$ gilt: $\lim_{n \rightarrow \infty} f(\vec{x}_n) = f(\vec{x}_*)$.

Bemerkungen:

- Auch die dritte Charakterisierung der Stetigkeit, dass Urbilder offener Mengen offen sind, ist weiterhin gültig.
- Die Rechenregeln, dass **Verknüpfungen und Verkettungen stetiger Funktionen wieder stetig sind**, gelten auch hier sinngemäß.
- Ist der Bildraum $W := \mathbb{R}^m$ von f endlichdimensional, so kann man die 'Komponentenfunktionen' f_1, \dots, f_m betrachten; diese sind Abbildungen von $D \subseteq V$ nach \mathbb{R} . Es ist: $f : D \rightarrow \mathbb{R}^m$ ist genau dann stetig, wenn alle Komponentenfunktionen $f_k : D \rightarrow \mathbb{R}$ stetig sind. Also: Das Überprüfen der Stetigkeit einer Funktion mit Werten in \mathbb{R}^m kann man immer zurückführen auf das Überprüfen der Stetigkeit von 'skalaren' Funktionen (d.h. Funktionen, die nach \mathbb{R} abbilden); es reicht also, wenn man sich auf skalarwertige Funktionen konzentriert, siehe die folgenden Beispiele.

Beispiele.

(a) **(Beweis von Unstetigkeit)** Die Funktion $f : \mathbb{R}^2 \setminus \{\vec{0}\} \rightarrow \mathbb{R}^2$, $f(x, y) = \frac{x^2 - y^2}{x^2 + y^2}$ lässt sich an der Stelle $(0, 0)$ nicht stetig fortsetzen.

Beweis: Für alle Folgen (x_n, y_n) , die gegen $(0, 0)$ konvergieren, müsste, damit sich f stetig fortsetzen ließe, die Folge der Funktionswerte $(f(x_n, y_n))_{n \in \mathbb{N}}$ gegen ein und denselben Wert konvergieren.

Wählen wir eine Folge $(x_n, 0)$, so ist $f(x_n, 0) = 1$, konvergiert somit gegen 1. Wählen wir jedoch eine Folge $(0, y_n)$, so ist $f(0, y_n) = -1$, konvergiert somit gegen -1 . (Das Verhalten hängt also von der *Richtung* ab, aus der wir uns $(0, 0)$ nähern.)

(b) **(Beweis von Stetigkeit)** Die Funktion $f : D \rightarrow \mathbb{R}$, $D = \{(x, y) \in \mathbb{R}^2 \mid x \neq 0\}$,

$$f(x, y) = \begin{cases} \frac{\sin(xy)}{x} & , x \neq 0 \\ y & , x = 0 \end{cases}$$

ist stetig auf D .

Beweis:

An Stellen (x_0, y_0) mit $x_0 \neq 0$ ist die Funktion nach den Rechenregeln der Stetigkeit stetig. Es reicht also, f an allen Stellen $(0, y_0)$ zu untersuchen. Wir untersuchen also $|f(x, y) - f(0, y_0)|$,⁵⁸ wobei $x \rightarrow 0$ und $y \rightarrow y_0$:

$$|f(x, y) - f(0, y_0)| = \left| \frac{\sin(xy)}{x} - y_0 \right| = \left| \underbrace{y}_{\rightarrow y_0} \cdot \underbrace{\frac{\sin(xy)}{xy}}_{\rightarrow 1} - y_0 \right| \rightarrow 0$$

Dabei haben wir im letzten Schritt ausgenutzt, dass $xy \rightarrow 0 \cdot y_0 = 0$ konvergiert.

- (c) **Bemerkung.** Zum Beweis von Unstetigkeit ist es sinnvoll, zunächst Wege entlang der Koordinatenrichtungen zu probieren (so wie in (a) geschehen). Wenn das nicht funktioniert (d.h. wenn man entlang der Koordinatenrichtungen immer die gleichen Grenzwerte bekommt (z.B. $f(x, y) = \frac{xy}{x^2+y^2}$), kann man Wege der Art (x, ax) oder (ay, y) mit $a \in \mathbb{R}$ fest und $x \rightarrow 0$ bzw. $y \rightarrow 0$ probieren, also entlang von Geraden gegen $(0, 0)$ gehen ($(x_0, y_0) = (0, 0)$ vorausgesetzt). Es gibt jedoch auch Fälle, bei denen entlang all dieser Geraden ein Grenzwert existiert, der Grenzwert sogar immer der gleiche ist, die Funktion an der Stelle $(0, 0)$ aber trotzdem nicht stetig ist. Beispiel: $f(x, y) = \frac{x^2+y^2}{x+y}$ mit $D = \{(x, y) \in \mathbb{R}^2 \mid x+y \neq 0\}$:

Mit $y := ax$, $a \in \mathbb{R}$ fest, ist $\lim_{x \rightarrow 0} f(x, ax) = \frac{x^2+a^2x^2}{x+ax} = \frac{1+a^2}{1+a}x \xrightarrow{(x \rightarrow 0)} 0$; analog für $x = 0$, $y \rightarrow 0$ ist $f(0, y) = y \rightarrow 0 = y_0$. Jedoch auf dem krummlinigen Weg $x := \underbrace{y^2}_{\rightarrow 0} - y$, $y \rightarrow 0$ ist $\lim_{y \rightarrow 0} f(y^2 - y, y) = \lim_{y \rightarrow 0} \frac{(y^2 - y)^2 + y^2}{y^2} = \lim_{y \rightarrow 0} y^2 - 2y + 2 = 2 \neq 0$.

Manchmal muss man also auch 'krummlinige' Wege, die sich (x_0, y_0) nähern, untersuchen, um Unstetigkeit zu belegen (zu finden durch 'Herumprobieren', ggf. unterstützt durch Visualisierung von f mittels Computer)

Einige Sätze aus früheren Kapiteln über stetige Funktionen lassen sich verallgemeinern:

Satz (Satz vom Minimum/Maximum). Sei $f : D \rightarrow \mathbb{R}$ stetig und $D \subset \mathbb{R}$ kompakt (also abgeschlossen und beschränkt). Dann ist f beschränkt, und f nimmt ein Minimum/Maximum an, d.h. es gibt $\vec{x}_1, \vec{x}_2 \in D$ mit

$$\min_{\vec{x} \in D} f(\vec{x}) = \inf_{\vec{x} \in D} f(\vec{x}) = f(\vec{x}_1), \quad \max_{\vec{x} \in D} f(\vec{x}) = \sup_{\vec{x} \in D} f(\vec{x}) = f(\vec{x}_2).$$

Für den nächsten Satz brauchen wir den Begriff des *Wegzusammenhangs*:

Def. (Wegzusammenhang). Eine Menge $M \subseteq \mathbb{R}^m$ heißt (*weg-*)*zusammenhängend*, wenn es für beliebige $\vec{x}_1, \vec{x}_2 \in M$ immer einen Weg, d.h. eine stetige Abbildung $f : [0, 1] \rightarrow \mathbb{R}^m$ gibt, der ganz in M verläuft, d.h. $f([0, 1]) \subset M$, und der \vec{x}_1 und \vec{x}_2

⁵⁸Stattdessen kann man auch $f(x, y)$ betrachten und zeigen, dass es gegen $f(0, y_0)$ geht; oft ist die hier gewählte Vorgehensweise der Betrachtung der Differenzen jedoch etwas leichter.

verbindet, d.h. $f(0) = \vec{x}_1$, $f(1) = \vec{x}_2$.

Satz (Zwischenwertsatz). Sei $D \subseteq \mathbb{R}^m$ zusammenhängend und $f : D \rightarrow \mathbb{R}$ stetig; seien $\vec{x}_1, \vec{x}_2 \in D$ mit $f(\vec{x}_1) < f(\vec{x}_2)$. Dann gibt es für jedes $\eta \in [f(\vec{x}_1), f(\vec{x}_2)]$ ein $\vec{\xi} \in D$ mit $f(\vec{\xi}) = \eta$. ⁵⁹

Beweis. Sei $\gamma : [0, 1] \rightarrow D$ ein Weg (also: stetige Abbildung) mit $\gamma(0) = \vec{x}_1$ und $\gamma(1) = \vec{x}_2$ (eine solche existiert, da D wegzusammenhängend ist). Die Verkettung $f \circ \gamma : [0, 1] \rightarrow \mathbb{R}$ ist dann stetig und erfüllt $f \circ \gamma(0) = f(\vec{x}_1)$, $f \circ \gamma(1) = f(\vec{x}_2)$. Nach dem Zwischenwertsatz aus Kap. 3.3 folgt die Behauptung \square

Beachten Sie, dass die Aussagen der beiden obigen Sätze nur für *skalarwertige* Funktionen gelten; beim Satz vom Minimum/Maximum ist dies klar, da für vektorwertige Funktionen Minimum/Maximum nicht definiert sind; beim Zwischenwertsatz ist dies nicht ganz so selbstverständlich, dass er für vektorwertige Funktionen seine Gültigkeit verliert.

Die obige Beweisidee, einen Satz für den Fall " $f : \mathbb{R}^n \rightarrow \mathbb{R}$ " auf einen entsprechenden Satz für die Situation " $f : \mathbb{R} \rightarrow \mathbb{R}$ " zurückzuführen, indem man eine Kurve/Weg durch den Definitionsbereich legt und nur Funktionswerte auf dieser Kurve betrachtet, wird uns bei der Formel von Taylor (Kap. 4.3.4) wiederbegegnen.

4.3 Differenzierbarkeit

4.3.1 Richtungsableitung, partielle Ableitung, Jacobi-Matrix

Wir betrachten zunächst skalarwertige Funktionen $f : D \rightarrow \mathbb{R}$, $D \subseteq \mathbb{R}^n$, D offen, um die Begriffe *Richtungsableitung* und *partielle Ableitung* zu definieren. Die Übertragung auf vektorwertige Funktionen erfolgt dann komponentenweise.

Def. (partielle Ableitung, skalarwertiger Fall; Gradient). Seien D und f wie oben und $\vec{x}_* \in D$. Sei $\vec{e}_i = (0, \dots, 0, 1, 0, \dots, 0)^T$ der i -te Standardbasisvektor im \mathbb{R}^n . Dann heißt

$$\begin{aligned} \frac{\partial f}{\partial x_i}(\vec{x}_*) &:= \lim_{h \rightarrow 0} \frac{f(\vec{x}_* + h\vec{e}_i) - f(\vec{x}_*)}{h} \\ &= \lim_{h \rightarrow 0} \frac{f(x_{*,1}, \dots, x_{*,i-1}, x_{*,i} + h, x_{*,i+1}, \dots, x_{*,n}) - f(x_{*,1}, \dots, x_{*,i}, \dots, x_{*,n})}{h}, \quad 1 \leq i \leq n, \end{aligned}$$

(bei Existenz) *die partielle Ableitung* von f nach x_i an der Stelle x_* . f heißt an der Stelle \vec{x}_* *partiell differenzierbar*, wenn $\frac{\partial f}{\partial x_1}(\vec{x}_*), \dots, \frac{\partial f}{\partial x_n}(\vec{x}_*)$ existieren. f heißt *partiell differenzierbar*, wenn f an allen Stellen $x_* \in D$ *partiell differenzierbar* ist.

Weitere Bezeichnungen für die partielle Ableitung sind

$$\partial_{x_i} f(\vec{x}_*), \quad \partial_i f(\vec{x}_*), \quad f_{x_i}(\vec{x}_*).$$

⁵⁹Man kann den Begriff des Wegzusammenhangs als Verallgemeinerung des Begriffs 'Intervall' (der ja in der skalaren Version vom Zwischenwertsatz aus Kap. 3.3 gebraucht wird) auffassen: Im \mathbb{R}^n ist eine Menge genau dann wegzusammenhängend, wenn sie ein Intervall ist.

(Beachte: "∂" bei partiellen Ableitungen, "d", wenn es nur ein Argument gibt.)

Der Vektor, bestehend aus den n partiellen Ableitungen an einer festen Stelle x_* wird als *Gradient* von f an der Stelle x_* bezeichnet:

$$\nabla f(\vec{x}) := \begin{pmatrix} \partial_1 f(\vec{x}) \\ \vdots \\ \partial_n f(\vec{x}) \end{pmatrix} \in \mathbb{R}^n$$

Das Symbol "∇" wird "Nabla" genannt/ausgesprochen (da es angeblich an ein antikes Saiteninstrument (Harfe=nabla (griech.) = nablium (lat.)) erinnert).

Beispiel. In der Praxis ist das Berechnen von partiellen Ableitungen ganz einfach: Man betrachtet alle x_j mit $j \neq i$ als 'fest', als Parameter, und betrachtet nur x_i als Argument, nach dem man ableitet:

Für $f(x, y) := x^3y + \exp(x^2+y)$ sind die partiellen Ableitungen

$$f_x(x, y) = \frac{\partial f}{\partial x}(x, y) = 3x^2y + 2x \exp(x^2+y), \quad f_y(x, y) = \frac{\partial f}{\partial y}(x, y) = x^3 + \exp(x^2+y),$$

der Gradient ist also

$$\nabla f(x, y) = \begin{pmatrix} 3x^2y + 2x \exp(x^2+y) \\ x^3 + \exp(x^2+y) \end{pmatrix}.$$

Also: Ist f eine *skalarwertige* Funktion, deren partielle Ableitungen existieren, so kann man den Gradienten bilden. Dieser ist der (Spalten-)vektor, der aus den partiellen Ableitungen von f besteht; der Gradient, als Funktion von \vec{x} , ist eine *vektorwertige* Funktion.

Statt "in Richtung von \vec{e}_i " kann man auch in beliebige andere Richtungen ableiten; dies führt zum Begriff der *Richtungsableitung*:

Def. (Richtungsableitung). Seien D, f, \vec{x}_* wie im vorherigen Satz. Sei $\vec{r} \in \mathbb{R}^n \setminus \{\vec{0}\}$ ein beliebiger Vektor, genannt *Richtungsvektor*. Dann heißt

$$\frac{\partial f}{\partial \vec{r}}(\vec{x}_*) := \lim_{h \rightarrow 0} \frac{f(\vec{x}_* + h\vec{r}) - f(\vec{x}_*)}{h},$$

(bei Existenz) die Richtungsableitung von f an der Stelle x_* in Richtung von \vec{r} .

Setzt man $r := \vec{e}_i$, so bekommt man die partielle Ableitung nach x_i . Hat man $\|\vec{r}\|_2 = 1$, so beschreibt die Richtungsableitung die anschauliche Steigung der Funktion f in Richtung von \vec{r} ; falls $\|\vec{r}\| \neq 1$, so geht diese Bedeutung verloren; man sieht der

Definition sofort an: Bei fest vorgegebener Richtung und Orientierung ist der Wert der Richtungsableitung proportional zur Länge von \vec{r} :

$$\frac{\partial f}{\partial(\alpha\vec{r})}(\vec{x}_*) = \lim_{h \rightarrow 0} \frac{f(\vec{x}_* + h\alpha\vec{r}) - f(\vec{x}_*)}{h} = \alpha \cdot \lim_{\tilde{h} \rightarrow 0} \frac{f(\vec{x}_* + \tilde{h}\vec{r}) - f(\vec{x}_*)}{\tilde{h}} = \alpha \frac{\partial f}{\partial\vec{r}}(\vec{x}_*)$$

wobei im vorletzten Schritt $\tilde{h} := \alpha h$ gesetzt wurde.

Die Richtungsableitung mittels der Definition zu berechnen, ist i.a. mühsam. Einfacher geht das mit folgender Formel:

Satz (Formel zur Berechnung der Richtungsableitung). Falls in einer Umgebung von x_* alle partiellen Ableitungen von f existieren und stetig sind, so gilt

$$\frac{\partial f}{\partial\vec{r}}(\vec{x}_*) = \langle \nabla f(\vec{x}_*), \vec{r} \rangle \quad (4.1)$$

Beweisidee: Man schreibt in der Definition der Richtungsableitung \vec{r} als Linearkombination der \vec{e}_i ; die dann erfolgenden Umformungen erfordern die Stetigkeit der partiellen Ableitungen.

Diese Formel wird uns eine anschauliche Bedeutung des Gradienten liefern:

Frage: An einer festen Stelle \vec{x}_* , in welche Richtung \vec{r} steigt/fällt die Funktion f am steilsten; in welcher Richtung ist sie konstant? Dazu betrachten wir alle Richtungsvektoren $\vec{r} \in \mathbb{R}^n$ mit $\|\vec{r}\|_2 = 1$ und betrachten die Formel (4.1): Das Skalarprodukt auf der rechten Seite ist (s. 1. Sem.) = $\underbrace{\|\nabla f(\vec{x}_*)\|}_{\text{gegeben}} \cdot \underbrace{\|\vec{r}\|}_{=1} \cdot \underbrace{\cos(\text{Winkel})}_{\text{variabel}}$; dieses wird genau

dann am größten, wenn beide Vektoren in die gleiche Richtung zeigen (Winkel=0); sie wird am kleinsten, wenn beide Vektoren in entgegengesetzte Richtungen zeigen (Winkel= π), und die wird null, wenn die beiden Vektoren orthogonal zueinander stehen.

Daraus folgt: **Der Gradient zeigt immer in die Richtung des steilsten Anstiegs einer Funktion $f : \mathbb{R}^n \rightarrow \mathbb{R}$. Die Richtung des steilsten Abfallens ist $-\nabla f(x_*)$. Der Gradient steht immer orthogonal auf Niveaumengen von f (=Linien gleicher Funktionswerte $\{\vec{x} \mid f(\vec{x}) = c\}$, falls $n=2$). (Skizze!)**

Nun zu *vektorwertigen* Funktionen. Wir können partielle Ableitungen vektorwertiger Funktionen definieren, indem wir die einzelnen (skalarwertigen) Komponentenfunktionen betrachten. Eine Funktion $f : \mathbb{R}^n \rightarrow \mathbb{R}$ hat n partielle Ableitungen. Für eine Funktion $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ hat jede der m Komponentenfunktionen n partielle Ableitungen; somit bekommen wir für f insgesamt $n \cdot m$ viele partielle Ableitungen, die man in Form einer $m \times n$ -Matrix anordnet. Diese Matrix heißt *Jacobi-* oder *Funktionalmatrix* von f :

Def. (Jacobi-, Funktionalmatrix).⁶⁰ Sei $f : D \rightarrow \mathbb{R}^m$ mit $D \subseteq \mathbb{R}^n$ offen. Sei $x \in D$. Die Matrix

$$Jf(\vec{x}) = (\partial_j f_i)_{i=1..m, j=1..n} = \begin{pmatrix} \partial_1 f_1(\vec{x}) & \dots & \partial_n f_1(\vec{x}) \\ \vdots & & \vdots \\ \partial_1 f_m(\vec{x}) & \dots & \partial_n f_m(\vec{x}) \end{pmatrix} \in \mathbb{R}^{m \times n}$$

heißt (bei Existenz) die *Jacobi-* oder *Funktionalmatrix* von f an der Stelle x .⁶¹ Die Zeilen von $Jf(x)$ sind gerade die Gradienten der Komponentenfunktionen f_i , diese jedoch als Zeilenvektoren geschrieben:

$$Jf(x) = \begin{pmatrix} (\nabla f_1(x))^T \\ \vdots \\ (\nabla f_m(x))^T \end{pmatrix}$$

Insbesondere im Fall $m = 1$ ist die Jacobi-Matrix ein Zeilenvektor, und zwar das Transponierte des Gradienten.

Partielle Ableitungen höherer Ordnung:

Wenn $f : D \rightarrow \mathbb{R}$, $D \subseteq \mathbb{R}^n$, partiell differenzierbar ist, dann sind die partiellen Ableitungen $\partial_i f$ wiederum Funktionen von $D \subseteq \mathbb{R}^n$ nach \mathbb{R} ; man kann also jede der n partiellen Ableitungen ggf. wiederum nach n Koordinatenrichtungen ableiten und bekommt so die partiellen Ableitungen zweiter Ordnung; diese bilden eine $n \times n$ -Matrix, die sogenannte *Hesse-Matrix*.⁶²

$$Hf(\vec{x}) := (\partial_i \partial_j f)_{i=1..n, j=1..n} = \begin{pmatrix} \partial_1 \partial_1 f(\vec{x}) & \dots & \partial_1 \partial_n f(\vec{x}) \\ \vdots & & \vdots \\ \partial_n \partial_1 f(\vec{x}) & \dots & \partial_n \partial_n f(\vec{x}) \end{pmatrix} \in \mathbb{R}^{n \times n}$$

Bei Existenz kann man Ableitungen noch höherer Stufe $\partial_i \partial_j \partial_k f(\vec{x})$ bilden usw..

Beispiel (inklusive der Verwendung verschiedener üblicher Schreibweisen für Ableitungen höherer Ordnung):

Für $f(x, y) = x^5 y^3$ ist

$$\partial_1 \partial_2 \partial_1 f(x, y) = \partial_x \partial_y \partial_x f(x, y) = f_{xyx}(x, y) = 60x^3 y^2.$$

Bildet man mehrmals die Ableitung nach der gleichen Koordinate, so schreibt man kurz

$$\partial_i^k f := \underbrace{\partial_i \dots \partial_i}_k f.$$

⁶⁰Carl Gustav Jacob Jacobi, 1804-1851)

⁶¹Wichtig: Zeilen und Spalten nicht verwechseln. Als Merkhilfe ggf.: Die Komponenten von f schreibt man, wenn man sich streng an die Konvention hält, 'übereinander' (f =Spaltenvektor). Diese Anordnung gilt auch innerhalb der Jacobi-Matrix.

⁶²Wir werden gleich noch sehen, dass diese Matrix unter gewissen Annahmen *symmetrisch* ist (\rightarrow Satz von Schwarz), dass also $\partial_i \partial_j f = \partial_j \partial_i f$.

Für partielle Ableitungen höherer Ordnung gibt es noch die sogenannte *Multiindex-Schreibweise*: Ein **Multiindex** $\vec{\alpha} = (\alpha_1, \dots, \alpha_n)^T$ ist ein Vektor, dessen Einträge α_i aus \mathbb{N}_0 sind. Man schreibt kurz

$$D^\alpha f := \partial_1^{\alpha_1} \dots \partial_n^{\alpha_n} f,$$

also z.B. $D^{(2,1,2)} f(x, y, z) = \partial_x^2 \partial_y \partial_z^2 f(x, y, z) = \partial_x \partial_x \partial_y \partial_z \partial_z f(x, y, z)$. Die Ordnung einer partiellen Ableitung in Multiindex-Schreibweise $D^\alpha f$ ergibt sich offensichtlich als

$$|\alpha| := \alpha_1 + \dots + \alpha_n$$

Satz (Satz von Schwarz, Satz über die Vertauschbarkeit von partiellen Ableitungen).⁶³ Sei $f : D \rightarrow \mathbb{R}^n$, $D \subseteq \mathbb{R}$ offen, und es sollen die partiellen Ableitungen von f bis zur Ordnung zwei existieren, und diese seien stetig. Dann gilt

$$\partial_i \partial_j f(\vec{x}) = \partial_j \partial_i f(\vec{x})$$

für $i, j \in \{1, \dots, n\}$, $\vec{x} \in D$. Insbesondere ist dann offensichtlich die Hesse-Matrix $Hf(\vec{x})$ symmetrisch.

Indem man den Satz statt auf f auf eine partielle Ableitung von f anwendet, überträgt er sich auch auf partielle Ableitungen höherer Ordnung; so ist z.B.

$$\partial_2 \partial_1 \partial_3 \partial_1 f = \partial_2 \partial_1 \partial_1 \partial_3 f = \partial_1 \partial_2 \partial_1 \partial_3 f = \partial_1^2 \partial_2 \partial_3 f,$$

sofern alle partiellen Ableitungen von f bis zur Ordnung 4 existieren und stetig sind; dabei wurde in obiger Rechnung der Satz von Schwarz zuerst auf f , dann auf $\partial_1 \partial_3 f$, und dann auf $\partial_3 f$ angewendet.

Anwendungen von partiellen Ableitungen höherer Ordnung, Hesse-Matrix (Ausblick auf 3. Semester): Partielle Ableitungen höherer Ordnung werden wir brauchen, um die *Taylor-Entwicklung* für Funktionen mehrerer Argumente aufstellen zu können (Kap. 4.3.4).

Ferner: Zum Klassifizieren lokaler Extrema hatten wir bei Funktionen $f : \mathbb{R} \rightarrow \mathbb{R}$ das Vorzeichen von $f''(x)$ zu untersuchen. Für Funktionen $f : \mathbb{R}^n \rightarrow \mathbb{R}$ übernimmt die Hesse-Matrix, also die Matrix der zweiten partiellen Ableitungen, die Rolle von f'' : Das Kriterium, das man im 3. Semester erhält, ist: Falls an einer kritischen Stelle x gilt dass $\langle Hf(\vec{x})\vec{v}, \vec{v} \rangle > 0$ (bzw.: < 0) für alle $\vec{0} \neq \vec{v} \in \mathbb{R}^n$, so ist \vec{x} eine Minimalstelle (bzw.: Maximalstelle). Die Bedingung kann man äquivalent umformen zu: Alle *Eigenwerte* (s. 1. Semester!) von H sind strikt positiv (bzw.: strikt negativ). Das für Extremstellen notwendige Kriterium $f'(x) = 0$ wird im mehrdimensionalen übrigens zu $\nabla f(\vec{x}) = \vec{0}$. Siehe 3. Semester.

4.3.2 Totale Ableitung

Der Begriff der partiellen Differenzierbarkeit ist leider mit diversen "Mängeln" behaftet: Für Funktionen $f : \mathbb{R} \rightarrow \mathbb{R}$ waren wir gewohnt, dass die Differenzierbarkeit eine recht

⁶³Hermann Amandus Schwarz, 1843-1921

'starke' Eigenschaft ist: Aus der Differenzierbarkeit folgt die Stetigkeit, und mit Hilfe der Ableitung kann man (siehe Taylor-Formel oder Mittelwertsatz) eine Funktion lokal gut approximieren (durch die Tangente). Wie steht es um diese Eigenschaften im Fall $f : \mathbb{R}^n \rightarrow \mathbb{R}$? Dazu das Beispiel

$$f(x, y) := \begin{cases} \frac{xy}{x^2+y^2} & , (x, y) \neq (0, 0) \\ 0 & , (x, y) = (0, 0). \end{cases}$$

An der Stelle $(0, 0)$ berechnen wir (mit der Definition, d.h. mit dem Differenzenquotienten) die partiellen Ableitungen. Sie sind $\partial_1 f(0, 0) = 0$ und $\partial_2 f(0, 0) = 0$; f ist also an dieser Stelle partiell differenzierbar. Doch ist f stetig? Nähern wir uns mit einer Folge (x_n, x_n) der Stelle $(0, 0)$, so haben wir $f(x_n, x_n) = \frac{x_n^2}{x_n^2+x_n^2} = \frac{1}{2}$, jedoch z.B. $f(x_n, -x_n) = -\frac{1}{2}$; für $x_n \rightarrow 0$ bekommen wir so verschiedene Grenzwerte für die Funktionswerte, d.h. f ist nicht stetig! *Aus der Existenz der partiellen Ableitungen folgt also nicht die Stetigkeit der Funktion!* (Anschaulich: Die Berechnung der partiellen Ableitungen an der Stelle $(0, 0)$ verwendet ausschließlich Information über die Funktion auf den Koordinatenachsen. Die Funktionswerte, die abseits der Koordinatenachsen liegen, gehen nicht ein, können daher auch bei partieller Diff'barkeit beliebig stark von $f(0, 0)$ abweichen, f somit dort unstetig sein.)

f ist sogar nicht nur an der Stelle $(0, 0)$, sondern überall partiell differenzierbar, denn an allen Stellen $(x, y) \neq (0, 0)$ ist f als Verknüpfung differenzierbarer Funktionen offensichtlich partiell differenzierbar. Also: Selbst wenn f in einer ganzen Umgebung von \vec{x}_* partiell diff'bar ist, folgt daraus immer noch nicht die Stetigkeit der Funktion an der Stelle \vec{x}_* (was anschaulich nicht ganz offensichtlich/einsichtig ist).

Dieses Beispiel zeigt außerdem, dass man, wenn man $f(0, 0)$ und $\partial_1 f(0, 0)$, $\partial_2 f(0, 0)$ kennt und damit f in der Nähe von $(0, 0)$ näherungsweise durch eine affin-lineare Funktion (geometrisch: Tangentialebene(!) an den Grafen=Relief legen) unter Verwendung der drei Werte darstellen möchte, man keine gute Näherung erwarten kann, da die Funktionswerte – trotz der partiellen Diff'barkeit – in beliebig kleinen Umgebungen von $(0, 0)$ sehr stark schwanken (springen).

Man kann sogar noch weiter gehen: Beispiele belegen, dass selbst aus der Existenz aller *Richtungsableitungen* sogar in einer ganzen Umgebung des betrachteten Punktes noch nicht die Stetigkeit der Funktion am betrachteten Punkt folgt! Die Anschauung, dass die Richtungs-differenzierbarkeit eine stärkere Eigenschaft ist als die Stetigkeit, täuscht also.

Aus diesen Gründen hat man noch einen weiteren Differenzierbarkeitsbegriff, den Begriff der *totalen Differenzierbarkeit/totalen Ableitung* eingeführt, und zwar derart, dass aus der totalen Diff'barkeit von f immer die Stetigkeit von f folgt. Den Begriff der totalen Ableitung können wir herleiten, indem wir uns nochmal die Situation $f : \mathbb{R} \rightarrow \mathbb{R}$ vergegenwärtigen und überlegen, wie wir die dortige Definition der Ableitung auf den vektoriellen Fall übertragen können:

Für Funktionen $f : \mathbb{R} \rightarrow \mathbb{R}$ hatten wir Differenzierbarkeit als Existenz des Grenzwertes

$$f'(x_*) := \lim_{h \rightarrow 0} \frac{f(x_*+h) - f(x_*)}{h} \quad (4.2)$$

definiert. Eine unmittelbare Übertragung dieser Definition auf Funktionen $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ scheitert daran, dass dann $\vec{h} \in \mathbb{R}^n$ ist und wir nicht durch \vec{h} dividieren können. Es ist auch nicht möglich, \vec{h} durch $\|\vec{h}\|$ zu ersetzen, denn schon in (4.2) ist es unzulässig, im Nenner $|h|$ zu schreiben (was man sich leicht an einfachen Beispielen, wie $f(x) := x$, überlegen kann). Wir formen daher die Bedingung (4.2) soweit um, bis eine Form erscheint, die man auf den mehrdimensionalen Fall übertragen kann:

Bedingung (4.2) ist gleichbedeutend dazu, dass es eine Zahl $f'(x_*)$ gibt, so dass der 'Rest' $R_{x_*}(h) := \frac{f(x_*+h) - f(x_*)}{h} - f'(x_*)$ gegen null geht für $h \rightarrow 0$. Multiplikation mit h ergibt:

$$\exists f'(x_*) \in \mathbb{R} : R_{x_*}(h) \cdot h + f'(x_*) \cdot h + f(x_*) - f(x_*+h) = 0, \quad \text{wobei } R_{x_*}(h) \xrightarrow{(h \rightarrow 0)} 0$$

und somit, unter Setzung $\tilde{R}_{x_*}(h) := R_{x_*}(h) \cdot h$ und Auflösen nach $f(x_*+h)$:

$$\exists f'(x_*) \in \mathbb{R} : f(x_*+h) = f(x_*) + f'(x_*) \cdot h + \tilde{R}_{x_*}(h), \quad \text{wobei } \frac{\tilde{R}_{x_*}(h)}{h} \xrightarrow{(h \rightarrow 0)} 0$$

Die Konvergenz gegen null(!) ist gleichbedeutend mit Konvergenz von Beträgen gegen null; man kann die Bedingung " $\frac{\tilde{R}_{x_*}(h)}{h} \xrightarrow{(h \rightarrow 0)} 0$ " ersetzen durch " $\left| \frac{\tilde{R}_{x_*}(h)}{h} \right| \xrightarrow{(h \rightarrow 0)} 0$ " oder auch durch " $\frac{\tilde{R}_{x_*}(h)}{|h|} \xrightarrow{(h \rightarrow 0)} 0$ ". Dies führt auf folgende Definition der Ableitung:

Definition (Differenzierbarkeit, totale Ableitung). Sei $f : D \rightarrow \mathbb{R}^m$, $D \subseteq \mathbb{R}^n$, D offen. f heißt *total differenzierbar* an der Stelle x , falls es eine Matrix(!) $f'(\vec{x}) \in \mathbb{R}^{m \times n}$ gibt, so dass

$$f(\vec{x}_* + \vec{h}) = \underbrace{f(\vec{x}_*) + f'(\vec{x}_*) \vec{h}}_{\text{aff.-lin. Approx. an } f} + \tilde{R}_{x_*}(\vec{h}), \quad \text{wobei } \frac{\tilde{R}_{x_*}(\vec{h})}{\|\vec{h}\|} \xrightarrow{(\vec{h} \rightarrow \vec{0})} \vec{0} \quad (4.3)$$

Die Matrix $f'(\vec{x}_*)$ heißt dann die *totale Ableitung* von f an der Stelle \vec{x}_* .

f heißt *total differenzierbar*, wenn f an allen Stellen $\vec{x} \in D$ total differenzierbar ist.

Deutung: Totale Differenzierbarkeit bedeutet, dass sich die Funktion $\vec{h} \mapsto f(\vec{x}_* + \vec{h})$ 'hinreichend gut' durch eine affin-lineare Funktion $f(\vec{x}_*) + f'(\vec{x}_*) \vec{h}$ approximieren lässt; was dadurch ausgedrückt wird, dass das Restglied "hinreichend schnell" (insbes. schneller als linear) gegen null geht. Man vergleiche auch die Analogie zur Taylor-Entwicklung (Polynomgrad=1). Die in (4.3) affin-lineare Approximation an f in der Nähe von \vec{x} ist im Fall $m=1$ nichts anderes als die Ebene, die sich an den Grafen von f anschmiegt, also die *Tangentialebene* an den Grafen von f im Punkt x .

Wie findet/berechnet man die totale Ableitung? Hätte man nur obige Definition zur Verfügung, so müsste man f' offenbar *raten* und anschließend durch Untersuchung von \tilde{R}_x zeigen, dass es die gesuchte Ableitung ist. Der folgende Satz vermeidet, dass man raten muss:

Falls an einer Stelle \vec{x} die totale Ableitung von f existiert, so ist f an dieser Stelle auch partiell diff'bar und stetig, und die totale Ableitung ist gleich der Jacobi-Matrix $f'(\vec{x}) = Jf(\vec{x})$; ferner ist f dann stetig an der Stelle \vec{x} . Wenn man also die totale Diff'barkeit von f an einer Stelle \vec{x} prüfen will, so muss man nicht $f'(\vec{x})$ raten, sondern man berechnet $Jf(\vec{x})$ und prüft dann, ob der Rest im Sinne von (4.3) hinreichend schnell gegen null geht.

Jedoch auch diese Überprüfung, ob der Rest hinreichend schnell gegen null geht, kann man sich ebenfalls ersparen, wenn man verwendet:

Falls alle partiellen Ableitungen in einer Umgebung von \vec{x} existieren und stetig(!) sind, dann ist f an der Stelle \vec{x} total differenzierbar. Das bedeutet: *Wenn man die partiellen Ableitungen in einer Umgebung von \vec{x} berechnen kann und diese stetig sind, dann ist f in x total diff'bar mit $f'(\vec{x}) = Jf(\vec{x})$; das Verhalten des Restes muss nicht weiter geprüft werden.*

Geometrisch bedeutet die totale Differenzierbarkeit, dass (jede Komponentenfunktion) eine Tangentialebene hat, die sich im Punkt x an den Grafen anschmiegt.

Bemerkungen:

- Natürlich gibt es per Substitution $\vec{x} := \vec{x}_* + \vec{h}$, die Darstellung

$$f(\vec{x}) = \underbrace{f(\vec{x}_*) + f'(\vec{x}_*)(\vec{x} - \vec{x}_*)}_{\text{lin. Approx. an } f} + \tilde{R}_{x_*}(\vec{x} - \vec{x}_*) = \vec{0}, \quad \text{wobei } \frac{\tilde{R}_{x_*}(\vec{x} - \vec{x}_*)}{\|\vec{x} - \vec{x}_*\|} \xrightarrow{(\vec{x} \rightarrow \vec{x}_*)} \vec{0}$$

- Im Spezialfall $m = 1$ (d.h. f skalarwertig) wird die Vektorgleichung (4.3) zur skalaren Gleichung

$$f(\vec{x}_* + \vec{h}) = \underbrace{f(\vec{x}_*) + \langle \nabla f(\vec{x}_*), \vec{h} \rangle}_{\text{aff.-lin. Approx. an } f} + \tilde{R}_{x_*}(\vec{h}) = 0, \quad \text{wobei } \frac{\tilde{R}_{x_*}(\vec{h})}{\|\vec{h}\|} \xrightarrow{(\vec{h} \rightarrow \vec{0})} 0$$

bzw.

$$f(\vec{x}) = \underbrace{f(\vec{x}_*) + \langle \nabla f(\vec{x}_*), \vec{x} - \vec{x}_* \rangle}_{\text{aff.-lin. Approx. an } f} + \tilde{R}_{x_*}(\vec{x} - \vec{x}_*) = 0, \quad \text{wobei } \frac{\tilde{R}_{x_*}(\vec{x} - \vec{x}_*)}{\|\vec{x} - \vec{x}_*\|} \xrightarrow{(\vec{x} \rightarrow \vec{x}_*)} 0.$$

- Man kann diese Definition der Differenzierbarkeit auch auf allgemeine normierte Vektorräume V, W übertragen. Ist $f : V \rightarrow W$, so ist (bei Existenz) $f'(x) \in \text{Lin}(V, W)$ (so wie eine Matrix

$f'(x) \in \mathbb{R}^{m \times n}$ sich als Element von $\text{Lin}(\mathbb{R}^n, \mathbb{R}^m)$ auffassen lässt), d.h. für jedes feste $x \in V$ ist $f'(x)$ eine lineare Abbildung von V nach W . $f' : x \mapsto f'(x)$ ist dann eine Abbildung von V nach $\text{Lin}(V, W)$, also $f' \in \text{Abb}(V, \text{Lin}(V, W))$. Dies ist zugegebenermaßen ziemlich abstrakt, daher verfolgen wir diesen allgemeinen Fall hier nicht weiter.

- Wie schon bei der Stetigkeit gilt auch bei der totalen Differenzierbarkeit:
Eine Funktion $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ ist genau dann differenzierbar an der Stelle x , wenn jede der m Komponentenfunktionen f_1, \dots, f_m an der Stelle x total differenzierbar ist; es gilt dann der Zusammenhang, dass die m Zeilen der Matrix $f'(x)$ gerade die Zeilenvektoren $f'_i(x)$ sind:

$$f'(x) = \begin{pmatrix} f'_1(x) \\ \vdots \\ f'_m(x) \end{pmatrix} = \begin{pmatrix} (\nabla f_1(x))^T \\ \vdots \\ (\nabla f_m(x))^T \end{pmatrix}$$

- Für das Abklingverhalten von R_x gibt es eine (insbes. in der Informatik und der Angewandten Mathematik) übliche Schreibweise: Man sagt $f = o(g)$ für $x \rightarrow x_*$, falls $\lim_{x \rightarrow x_*} \frac{f(x)}{g(x)} = 0$. In (4.3) also: $R_{x_*}(h) = o(\|h\|)$.

Diese Schreibweise ist nicht zu verwechseln mit $f = O(g)$ für $x \rightarrow x_*$, was bedeutet, dass es eine Konstante c gibt, so dass $|f(x)| \leq c \cdot g(x)$ für alle x in einer Umgebung von x_* (oft: $x_* = \infty$).
Bemerkung: Eigentlich wäre die Schreibweise " $f \in o(g)$ " besser als " $f = o(g)$ ", denn mit " $o(g)$ " wird ja eine Klasse (Menge) von Funktionen bezeichnet; außerdem folgt aus $f = o(g)$ nicht dass $g = o(f)$ ist; und aus z.B. $f_1, f_2 = o(g)$ folgt $f_1 + f_2 = o(g)$ — alles Eigenschaften, die man intuitiv nur schwer mit dem Gleichheitszeichen in Verbindung bringen kann.

Satz (Kettenregel) Seien $g : D_g \rightarrow D_f$, $f : D_f \rightarrow \mathbb{R}^l$; $D_g \subseteq \mathbb{R}^n$ und $D_f \subseteq \mathbb{R}^m$ offen. Seien f und g total differenzierbar. Dann gilt die Matrix-Gleichung

$$(f \circ g)'(\vec{x}) = f'(g(\vec{x})) g'(\vec{x})$$

oder gleichbedeutend

$$\underbrace{J(f \circ g)(\vec{x})}_{\in \mathbb{R}^{l \times n}} = \underbrace{Jf(g(\vec{x}))}_{\in \mathbb{R}^{l \times m}} \underbrace{Jg(\vec{x})}_{\in \mathbb{R}^{m \times n}}.$$

Im Spezialfall $n = l = 1$ (also $f \circ g : D_g \rightarrow \mathbb{R}$ mit $D_g \subseteq \mathbb{R}$) sollte die Ableitung, wie wir wissen, eine Zahl sein. Und in der Tat, in obiger Formel steht auf der rechten Seite das Produkt aus einer Matrix mit einer Zeile und einer Matrix mit nur einer Spalte; das Ergebnis ist ein Skalarprodukt, d.h. eine Zahl.

Eine Anwendung: Berechnen Sie die Ableitung von $u(x) := \int_0^x e^{-t^2} dt$. Wir haben in Kap. 3 keine Rechenregel zum Differenzieren eines solchen Ausdrucks; auch können wir die Stammfunktion nicht explizit ausrechnen. Wir gehen stattdessen wie folgt vor: Wir führen die Hilfsfunktionen $f(y, z) := \int_0^y e^{-zt^2} dt$ und $g(x) := (x, x)^T$ ein. Es ist offensichtlich $u(x) = f(x, x) = f(g(x))$. Somit können wir u' mit der obigen Kettenregel

berechnen. Es ist $g'(x) = (1, 1)^T$ und $f'(y, z) = (e^{-zy^2}, -\int_0^y t^2 e^{-zt^2} dt)$ und $f'(g(x)) = f'(x, x) = (e^{-x^3}, -\int_0^x t^2 e^{-xt^2} dt)$, also, mit Kettenregel,

$$u'(x) = f'(g(x)) g'(x) = \left(e^{-x^3}, -\int_0^x t^2 e^{-xt^2} dt \right) \begin{pmatrix} 1 \\ 1 \end{pmatrix} = e^{-x^3} - \int_0^x t^2 e^{-xt^2} dt$$

Allgemein bekommt man so

$$\frac{d}{dx} \int_a^{h(x)} f(t, x) dt = f(h(x), x) h'(x) + \int_a^{h(x)} \partial_x f(t, x) dt$$

für hinreichend glatte f, h .

4.3.3 Weitere Differentialoperatoren und Anwendungen

Die Divergenz eines Vektorfeldes

Wir kennen bereits als grundlegende Differentialoperatoren:

Für Skalarfeld $f : \mathbb{R}^n \rightarrow \mathbb{R}$: Den Gradienten $\nabla f : \mathbb{R}^n \rightarrow \mathbb{R}^n$

Für Vektorfeld $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$: Die Jacobi-Matrix $Jf : \mathbb{R}^n \rightarrow \mathbb{R}^{n \times n}$.

Ferner definiert man die sogenannte *Divergenz*⁶⁴ eines Vektorfeldes $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ als

$$\operatorname{div} f : \mathbb{R}^n \rightarrow \mathbb{R}, \quad \operatorname{div} f := \sum_{i=1}^n \frac{\partial f_i}{\partial x_i} = \sum_{i=1}^n \partial_i f_i$$

Zur Übung: Berechnen Sie die Divergenz des Vektorfeldes

$$\vec{v}(x_1, x_2, x_3) = \left(-v_0 x_2 / (x_1^2 + x_2^2)^{\frac{\alpha}{2}}, -v_0 x_1 / (x_1^2 + x_2^2)^{\frac{\alpha}{2}}, 0 \right)^T, \quad (4.4)$$

wobei $v_0, \alpha > 0$. Machen Sie sich klar, dass dieses Vektorfeld, als Strömungsfeld aufgefasst, einen 'Wirbel' um die z -Achse beschreibt.

Anwendungen der Divergenz: In der Strömungsmechanik/Kontinuumsmechanik: Ist $v : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ das Geschwindigkeitsfeld eines Fluids mit konstanter Dichte (z.B. Wasser gilt als inkompressibel, d.h. die Dichte ist unabhängig vom Druck; bei konstanter Temperatur hat es somit konstante Dichte), so gilt

$$\operatorname{div} \vec{v} = 0. \quad (4.5)$$

Dabei kann \vec{v} die Argumente x_1, x_2, x_3 haben, oder, wenn das Strömungsfeld zeitlich variabel ist ('instationär'), die vier Argumente t, x_1, x_2, x_3 . Vorsicht: In *beiden* Fällen jedoch soll sich 'div' nur auf die *räumlichen* Koordinaten x_1, x_2, x_3 beziehen, nicht auf t .

⁶⁴Diese hat rein gar nichts zu tun mit dem Begriff 'Divergenz' im Sinne von Abwesenheit von Konvergenz.

Begründung für Gleichung (4.5): Wir betrachten einen kleinen Würfel $Q_\epsilon := [0, \epsilon] \times [0, \epsilon] \times [0, \epsilon]$, den wir o.B.d.A. an der Stelle $\vec{x} = \vec{0}$ angeheftet haben. Wir betrachten die Massenbilanz für ein Fluid mit konstanter Dichte ρ und Geschwindigkeitsfeld \vec{v} : "So viel wie in einem Zeitintervall in Q_ϵ hineinfließt, muss im gleichen Zeitintervall auch wieder hinausfließen." Die Summe der Massenflüsse (Masse pro Zeiteinheit) über die 6 Seitenflächen des Würfels muss also null ergeben. Man beachte, dass für den Fluss durch eine Seitenfläche nur diejenige Geschwindigkeitskomponente v_i relevant ist, die senkrecht zur Fläche steht; tangentielle Geschwindigkeitskomponenten sind irrelevant. Die Massenbilanz lautet also⁶⁵

$$\begin{aligned} & \int_0^\epsilon \int_0^\epsilon \rho v_1(0, y, z) dy dz - \int_0^\epsilon \int_0^\epsilon \rho v_1(\epsilon, y, z) dy dz \\ & + \int_0^\epsilon \int_0^\epsilon \rho v_2(x, 0, z) dx dz - \int_0^\epsilon \int_0^\epsilon \rho v_2(x, \epsilon, z) dx dz \\ & + \int_0^\epsilon \int_0^\epsilon \rho v_3(x, y, 0) dx dy - \int_0^\epsilon \int_0^\epsilon \rho v_3(x, y, \epsilon) dx dy = 0. \end{aligned} \quad (4.6)$$

Um einen Zusammenhang zu (4.5) herzustellen, müssen wir 1. die Integralzeichen loswerden und 2. Differenzenquotienten in (4.6) erzeugen. Wir schreiben zunächst (4.6) um, indem wir jeweils gegenüberliegende Seitenflächen zusammenfassen:

$$\begin{aligned} & \rho \int_0^\epsilon \int_0^\epsilon [v_1(0, y, z) - v_1(\epsilon, y, z)] dy dz \\ & + \rho \int_0^\epsilon \int_0^\epsilon [v_2(x, 0, z) - v_2(x, \epsilon, z)] dx dz \\ & + \rho \int_0^\epsilon \int_0^\epsilon [v_3(x, y, 0) - v_3(x, y, \epsilon)] dx dy = 0. \end{aligned}$$

Nun teilen wir durch ρ und wenden den Mittelwertsatz der Integralrechnung an; zunächst jeweils für das äußere, dann für das innere Integral. Das ergibt

$$\begin{aligned} & \epsilon^2 [v_1(0, \eta_1, \zeta_1) - v_1(\epsilon, \eta_1, \zeta_1)] \\ & + \epsilon^2 [v_2(\xi_1, 0, \zeta_2) - v_2(\xi_1, \epsilon, \zeta_2)] \\ & + \epsilon^2 [v_3(\xi_2, \eta_2, 0) - v_3(\xi_2, \eta_2, \epsilon)] = 0 \end{aligned}$$

für gewisse $\xi_1, \xi_2, \eta_1, \eta_2, \zeta_1, \zeta_2 \in (0, \epsilon)$.

Nun teilen wir durch $-\epsilon^3$ und bekommen Differenzenquotienten:

$$\frac{v_1(\epsilon, \eta_1, \zeta_1) - v_1(0, \eta_1, \zeta_1)}{\epsilon} + \frac{v_2(\xi_1, \epsilon, \zeta_2) - v_2(\xi_1, 0, \zeta_2)}{\epsilon} + \frac{v_3(\xi_2, \eta_2, \epsilon) - v_3(\xi_2, \eta_2, 0)}{\epsilon} = 0$$

Nun bilden wir $\lim_{\epsilon \rightarrow 0}$. Setzen wir dabei voraus, dass die v_i hinreichend 'glatt' sind (z.B. total diff'bar), und nutzen aus, dass $\xi_1, \xi_2, \eta_1, \eta_2, \zeta_1, \zeta_2$ (die von ϵ abhängen, in $[0, \epsilon]$ liegen, und somit gegen null gehen müssen, bekommen wir im Grenzübergang

$$\frac{\partial v_1(0, 0, 0)}{\partial x_1} + \frac{\partial v_2(0, 0, 0)}{\partial x_2} + \frac{\partial v_3(0, 0, 0)}{\partial x_3} = 0.$$

Da die Rechnung auch dann funktioniert, wenn Q_ϵ nicht am Nullpunkt liegt, sondern an einem beliebigen anderen Punkt, folgt (4.5).

Nun eine Verallgemeinerung von (4.5): Ein Fluid, dessen Dichte ρ (Skalarfeld) zeitlich/räumlich *variabel* ist, erfüllt die Gleichung

$$\partial_t \rho + \operatorname{div}(\rho \vec{v}) = 0.$$

⁶⁵Während die Integrale jeweils einen Massenfluss beschreiben (Einheit: Masse pro Zeit), werden die Integranden als *Massenflussdichten* bezeichnet (Einheit: Masse pro Zeit pro Fläche = Dichte mal Geschwindigkeit).

Diese *Partielle Differentialgleichung* beschreibt den *Massenerhalt* eines Fluids.

Weitere Notation: Da die Definition der Divergenz an ein Skalarprodukt erinnert, findet man häufig die Notation $\vec{\nabla} \bullet \vec{f}$. Die Gleichung des Massenerhalts wird so zu $\partial_t \rho + \vec{\nabla} \bullet (\rho \vec{v}) = 0$.

Die Rotation eines Vektorfeldes.

Man definiert die sogenannte *Rotation* eines Vektorfeldes $f : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ als

$$\text{rot } f : \mathbb{R}^3 \rightarrow \mathbb{R}^3, \quad \text{rot } f := \begin{pmatrix} \partial_2 f_3 - \partial_3 f_2 \\ \partial_3 f_1 - \partial_1 f_3 \\ \partial_1 f_2 - \partial_2 f_1 \end{pmatrix}$$

Merkhilfe: $\text{rot } f = \det \begin{pmatrix} \vec{e}_1 & \partial_1 & f_1 \\ \vec{e}_2 & \partial_2 & f_2 \\ \vec{e}_3 & \partial_3 & f_3 \end{pmatrix}$ (nur formal)

Weitere Notation: Obige Merkhilfe erinnert an die Merkhilfe fürs *Vektorprodukt*. Deshalb schreibt man auch gelegentlich " $\vec{\nabla} \times \vec{f}$ " für $\text{rot } \vec{f}$.

Anwendungen der Rotation: In der Strömungsmechanik/Kontinuumsmechanik:

Ist $v : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ das Geschwindigkeitsfeld eines Fluids, so gibt der Betrag von $\text{rot } f(\vec{x})$ die 'Wirbelstärke' an der Stelle \vec{x} an, und die Richtung von $\text{rot } f(\vec{x})$ gibt die Richtung der Rotationsachse an.

Zur Übung: Berechnen Sie die Rotation des obigen Vektorfeldes (4.4).

Anwendungen der Rotation: In der Elektrotechnik:

Die sog. *Maxwell'schen Gleichungen* beschreiben den Zusammenhang zwischen Magnetfeldern \vec{B} und Elektrischen Feldern \vec{E} und Elektrischer Stromstärkendichte \vec{I} und Ladungsdichte ρ :

$$\begin{aligned} \text{div } (\epsilon \vec{E}) &= \rho \\ \text{div } \vec{B} &= 0 \\ \partial_t \vec{B} + \text{rot } \vec{E} &= \vec{0} \\ \epsilon \mu \partial_t \vec{E} - \text{rot } \vec{B} + \mu \vec{I} &= \vec{0} \end{aligned}$$

Der Laplace-Operator. Man definiert den sog. *Laplace-Operator* Δ auf Skalarfeldern $f : \mathbb{R}^n \rightarrow \mathbb{R}$ als

$$\Delta f : \mathbb{R}^n \rightarrow \mathbb{R}, \quad \Delta f := \sum_{i=1}^n \frac{\partial^2 f}{(\partial x_i)^2} = \sum_{i=1}^n \partial_i^2 f$$

Anwendung des Laplace-Operators: Z.B. in der Strömungsmechanik:

Diffusionsgleichung: Die Konzentration c (Skalarfeld) eines Stoffes, der in einem ruhenden Fluid gelöst ist, verbreitet sich durch Diffusion (Brown'sche Bewegung) per

$$\partial_t c - k \Delta c = 0 \quad (\text{Diffusionsgleichung})$$

Mathematisch analog: Die Temperaturverteilung T (Skalarfeld) in einem ruhenden Medium verhält sich gemäß

$$\partial_t T - k \Delta T = 0 \quad (\text{Wärmeleitungsgleichung}).$$

Einige einfache Rechenregeln:

- $\operatorname{div} \nabla = \Delta$
- $\operatorname{div} (\rho \vec{v}) = \vec{\nabla} \rho \bullet \vec{v} + \rho \operatorname{div} \vec{v}$ (eine Art 'Produktregel')

Der Beweis ist elementar; man setzt die Definition der Differentialoperatoren ein.

4.3.4 Die Formel von Taylor im \mathbb{R}^n

Wie schon für Funktionen von *einem* Argument wollen wir nun auch Funktionen mit *mehreren* Argumenten durch polynomielle Ausdrücke (Taylor-Polynome) lokal in einer Umgebung eines Entwicklungspunktes approximieren. Beachten Sie, dass wir hier zunächst nur *skalarwertige* Funktionen betrachten. Im Fall mehrerer Argumente wird auch das zu konstruierende Taylor-Polynom von mehreren Argumenten abhängen. Was ist ein Polynom in mehreren Argumenten?

Def. (Polynom) Eine Abbildung $f : \mathbb{R}^n \rightarrow \mathbb{R}$ der Form

$$f(\vec{x}) = \sum_{k=0}^m \sum_{\substack{\alpha_1, \dots, \alpha_n \in \mathbb{N}_0 \\ \alpha_1 + \dots + \alpha_n = k}} a_{\alpha_1, \dots, \alpha_n} x_1^{\alpha_1} \cdot \dots \cdot x_n^{\alpha_n},$$

wobei die $a_{\alpha_1, \dots, \alpha_n} \in \mathbb{R}$, heißt *reelles Polynom in n Variablen*. Sein Grad ist m , sofern es einen von null verschiedenen Koeffizienten $a_{\alpha_1, \dots, \alpha_n}$ gibt mit $\alpha_1 + \dots + \alpha_n = m$. Eine kürzere Schreibweise ergibt sich unter Verwendung von Multiindizes:

$$f(\vec{x}) = \sum_{k=0}^m \sum_{|\vec{\alpha}|=k} \vec{a}_{\vec{\alpha}} \vec{x}^{\vec{\alpha}};$$

dabei ist $\vec{x}^{\vec{\alpha}} := x_1^{\alpha_1} \cdot \dots \cdot x_n^{\alpha_n}$.

Beispiele. $f(x, y, z) = 3 + 2x - 5y + z - 3x^2 + 3xyz$ ist ein Polynom dritten Grades. Polynome dritten Grades in *zwei* Variablen haben die allgemeine Form $f(x, y) = a_1 + \underbrace{a_2x + a_3y}_{\text{Terme 1. Ordnung}} + \underbrace{a_4x^2 + a_5xy + a_6y^2}_{\text{Terme 2. Ordnung}} + \underbrace{a_7x^3 + a_8x^2y + a_9xy^2 + a_{10}y^3}_{\text{Terme 3. Ordnung}}$.

Nun zur Formel von Taylor. Wir betrachten $f : D \rightarrow \mathbb{R}$ mit $D \subseteq \mathbb{R}^n$ offen. Sei $\vec{x}_* \in D$ fest (der sog. Entwicklungspunkt). Es sei f $m+1$ mal partiell differenzierbar auf D , und all diese partiellen Ableitungen seien stetig. Wir wollen f in einer Umgebung von

\vec{x}_* durch ein Polynom approximieren. Sei ein Richtungsvektor $\vec{h} \in \mathbb{R}^n$ so gewählt, dass nicht nur $\vec{x}_* + \vec{h}$, sondern die gesamte (geradlinige) Strecke von \vec{x}_* nach $\vec{x}_* + \vec{h}$ in D liegt: $\{\vec{x}_* + t\vec{h} \mid t \in [0, 1]\} \subset D$.

Um die Situation auf die von Kap. 3.6.4 zurückzuführen, betrachten wir die Abbildung $g(t) := f(\vec{x}_* + t\vec{h})$, $[0, 1] \rightarrow \mathbb{R}$, d.h. wir betrachten die Funktionswerte von f nur entlang der Strecke, die von \vec{x}_* nach $\vec{x}_* + \vec{h}$ führt (vergleiche die Ähnlichkeit zum Beweis des Zwischenwertsatzes S. 101). Da g nur ein skalares Argument hat, können wir auf g den Satz von Taylor aus Kap. 3.6.4 anwenden. Wir wählen als Entwicklungspunkt $t_* := 0$, da $g(0) = f(x_*)$. Wir erhalten die übliche Formel

$$g(t) = g(0) + g'(0)t + \frac{g''(0)}{2!}t^2 + \dots + \frac{g^{(m)}(0)}{m!}t^m + \frac{g^{(m+1)}(\tau)}{(m+1)!}t^{m+1} \quad (4.7)$$

mit τ zwischen 0 und t . Wir setzen hier $t = 1$ ein, denn dann erhalten wir auf der linken Seite $g(1) = f(\vec{x}_* + \vec{h})$. Wir müssen diese Formel nun nur 'rückübersetzen', d.h. die rechte Seite mittels f ausdrücken:

Es ist $g(0) = f(\vec{x}_*)$. Die Ableitungen von g rechnen wir mit der Kettenregel aus; es liegt offensichtlich der auf S. 109 diskutierte Fall $n=l=1$ vor: Es ist

$$g'(t) = \underbrace{f'(\vec{x}_* + t\vec{h})}_{\text{Zeilenvektor}} \cdot \underbrace{\vec{h}}_{\text{Spalten-V.}} = \langle \nabla f(\vec{x}_* + t\vec{h}), \vec{h} \rangle = \sum_{i=1}^n h_i \partial_i f(\vec{x}_* + t\vec{h}) \quad (4.8)$$

und somit

$$g'(0) = \langle \nabla f(\vec{x}_*), \vec{h} \rangle = \sum_{k=1}^n \partial_k f(\vec{x}_*) \cdot h_k.$$

Indem wir (4.8) mit der Kettenregel differenzieren, bekommen wir

$$g''(t) = \frac{d}{dt} \sum_{i=1}^n h_i \partial_i f(\vec{x}_* + t\vec{h}) = \sum_{i=1}^n \sum_{j=1}^n h_i h_j \partial_i \partial_j f(\vec{x}_* + t\vec{h}) = \langle Hf(\vec{x}_* + t\vec{h}) \vec{h}, \vec{h} \rangle$$

und somit

$$g''(0) = \sum_{i=1}^n \sum_{j=1}^n h_i h_j \partial_i \partial_j f(\vec{x}_*) = \langle Hf(\vec{x}_*) \vec{h}, \vec{h} \rangle.$$

Der nächste Term lautet

$$g'''(0) = \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n h_i h_j h_k \partial_i \partial_j \partial_k f(\vec{x}_*)$$

und allgemein

$$g^{(m)}(0) = \sum_{i_1=1}^n \dots \sum_{i_m=1}^n h_{i_1} \cdot \dots \cdot h_{i_m} \cdot \partial_{i_1} \dots \partial_{i_m} f(x_*),$$

sowie für das Restglied

$$g^{(m+1)}(\tau) = \sum_{i_1=1}^n \dots \sum_{i_{m+1}=1}^n h_{i_1} \cdot \dots \cdot h_{i_{m+1}} \cdot \partial_{i_1} \dots \partial_{i_{m+1}} f(x_*).$$

Wenn wir all dies in (4.7) einsetzen haben wir unsere Taylor-Formel für die Funktion f . Jedoch ist es sinnvoll, die mühsame Schreibweise (" $\sum_{i_1=1}^n \dots \sum_{i_m=1}^n$ ") durch eine knappere zu ersetzen. Hierzu sind die oben eingeführten Multi-Indizes nützlich. Wir wollen der Einfachheit halber die folgenden Überlegungen nur für den Fall $n = 2$ durchführen: Z.B. für den Term 2. Ordnung haben wir (s.o.)

$$g''(0) = h_1 h_1 \partial_1 \partial_1 f(x_*) + h_1 h_2 \partial_1 \partial_2 f(x_*) + h_2 h_1 \partial_2 \partial_1 f(x_*) + h_2 h_2 \partial_2 \partial_2 f(x_*),$$

was sich unter Verwendung der Symmetrie (Satz von Schwarz) als

$$g''(0) = h_1^2 \partial_1^2 f(x_*) + 2h_1 h_2 \partial_1 \partial_2 f(x_*) + h_2^2 \partial_2^2 f(x_*)$$

schreiben lässt. Und analog für den Term 3. Ordnung bekommen wir durch Zusammenfassen gleicher Terme:

$$g'''(0) = h_1^3 \partial_1^3 f(x_*) + 3h_1^2 h_2 \partial_1^2 \partial_2 f(x_*) + 3h_1 h_2^2 \partial_1 \partial_2^2 f(x_*) + h_2^3 \partial_2^3 f(x_*).$$

Um nun die Multi-Indizes ins Spiel zu bringen, definieren wir für Multi-Indizes $\alpha \in \mathbb{N}_0^n$ die Fakultät

$$\alpha! := \alpha_1! \cdot \dots \cdot \alpha_n!$$

und schreiben wir obiges als

$$\begin{aligned} g''(0) &= (h_1, h_2)^{(2,0)} D^{(2,0)} f(x_*) + 2(h_1, h_2)^{(1,1)} D^{(1,1)} f(x_*) + (h_1, h_2)^{(0,2)} D^{(0,2)} f(x_*) \\ &= 2! \cdot \left[\frac{(h_1, h_2)^{(2,0)}}{2!0!} D^{(2,0)} f(x_*) + \frac{(h_1, h_2)^{(1,1)}}{1!1!} D^{(1,1)} f(x_*) + \frac{(h_1, h_2)^{(0,2)}}{0!2!} D^{(0,2)} f(x_*) \right] \\ &= 2! \sum_{|\alpha|=2} \frac{\vec{h}^\alpha}{\alpha!} D^\alpha f(x_*) \end{aligned}$$

und

$$\begin{aligned} g'''(0) &= (h_1, h_2)^{(3,0)} D^{(3,0)} f(x_*) + 3(h_1, h_2)^{(2,1)} D^{(2,1)} f(x_*) \\ &\quad + 3(h_1, h_2)^{(1,2)} D^{(1,2)} f(x_*) + (h_1, h_2)^{(0,3)} D^{(0,3)} f(x_*) \\ &= 3! \cdot \left[\frac{(h_1, h_2)^{(3,0)}}{3!0!} D^{(3,0)} f(x_*) + \frac{(h_1, h_2)^{(2,1)}}{2!1!} D^{(2,1)} f(x_*) + \frac{(h_1, h_2)^{(1,2)}}{1!2!} D^{(1,2)} f(x_*) + \frac{(h_1, h_2)^{(0,3)}}{0!3!} D^{(0,3)} f(x_*) \right] \\ &= 3! \sum_{|\alpha|=3} \frac{\vec{h}^\alpha}{\alpha!} D^\alpha f(x_*) \end{aligned}$$

Man kann zeigen, dass allgemein für $n \in \mathbb{N}$ und $m \in \mathbb{N}$ der Term m -ter Ordnung die Form

$$g^{(m)}(0) = \sum_{i_1=1}^n \dots \sum_{i_m=1}^n h_{i_1} \cdot \dots \cdot h_{i_m} \cdot \partial_{i_1} \dots \partial_{i_m} f(x_*) = m! \sum_{|\alpha|=m} \frac{D^\alpha f(\vec{x}_*)}{\alpha!} \vec{h}^\alpha$$

hat. Also lautet die Taylor-Formel in der 'kompakten' Darstellung:

$$\begin{aligned}
 f(\vec{x}_* + \vec{h}) &= \overbrace{f(\vec{x}_*) + \sum_{|\alpha|=1} \frac{D^\alpha f(\vec{x}_*)}{\alpha!} \vec{h}^\alpha + \dots + \sum_{|\alpha|=m} \frac{D^\alpha f(\vec{x}_*)}{\alpha!} \vec{h}^\alpha}^{\text{Taylor-Polynom}} \\
 &\quad + \underbrace{\sum_{|\alpha|=m+1} \frac{D^\alpha f(\vec{x}_* + \tau \vec{h})}{\alpha!} \vec{h}^\alpha}_{\text{Restglied}} \\
 &= \underbrace{\sum_{|\alpha| \leq m} \frac{D^\alpha f(\vec{x}_*)}{\alpha!} \vec{h}^\alpha}_{\text{Taylor-Polynom}} + \underbrace{\sum_{|\alpha|=m+1} \frac{D^\alpha f(\vec{x}_* + \tau \vec{h})}{\alpha!} \vec{h}^\alpha}_{\text{Restglied}}, \quad \tau \in (0, 1)
 \end{aligned} \tag{4.9}$$

Die 'Zwischenstelle' $\vec{\xi} := \vec{x}_* + \tau \vec{h}$ (mit $\tau \in (0, 1)$) ist ein Punkt auf der Verbindungslinie zwischen \vec{x}_* und $\vec{x}_* + \vec{h}$.

Auch hier gibt es natürlich auch wieder die Möglichkeit, durch Substitution $\vec{x} := \vec{x}_* + \vec{h}$ das \vec{h} zu eliminieren.

Um die abstrakte Schreibweise (4.9) etwas konkreter zu machen, hier das Taylor-Polynom *zweiter* Ordnung:

$$\begin{aligned}
 f(\vec{x}_* + \vec{h}) &= f(\vec{x}_*) + \langle \nabla f(\vec{x}_*), \vec{h} \rangle + \frac{1}{2} \langle H f(\vec{x}_*) \vec{h}, \vec{h} \rangle \\
 &= f(x_*) + \sum_{i=1}^n \partial_i f(\vec{x}_*) \cdot h_i + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \partial_i \partial_j f(\vec{x}_*) \cdot h_i h_j
 \end{aligned}$$

Beispiel. Berechne das Taylor-Polynom erster und zweiter Ordnung von $f(x, y) = \exp(x^2 + x + y)$ zum Entwicklungspunkt $(x_*, y_*) = (0, 0)$.

Es ist

$$\begin{aligned}
 f_x(x, y) &= (2x+1) \exp(x^2 + x + y), & f_x(0, 0) &= 1 \\
 f_y(x, y) &= \exp(x^2 + x + y), & f_y(0, 0) &= 1 \\
 f_{xx}(x, y) &= [2 + (2x+1)^2] \exp(x^2 + x + y), & f_{xx}(0, 0) &= 3 \\
 f_{xy}(x, y) &= (2x+1) \exp(x^2 + x + y) = f_{yx}(x, y), & f_{xy}(0, 0) &= 1 = f_{yx}(0, 0) \\
 f_{yy}(x, y) &= \exp(x^2 + x + y), & f_{yy}(0, 0) &= 1
 \end{aligned}$$

somit

$$f(h_1, h_2) \approx T_1(h_1, h_2) = 1 + h_1 + h_2;$$

T_1 beschreibt die *Tangentialebene* an den Grafen von f im Punkt $(0, 0)$; und die Approximation höherer Ordnung (sich besser an den Grafen von f anschmiegend)

$$f(h_1, h_2) \approx T_2(h_1, h_2) = 1 + h_1 + h_2 + \frac{3}{2} h_1^2 + h_1 h_2 + \frac{1}{2} h_2^2.$$

Indem man den Satz von Schwarz ausnutzt, stimmen, insbesondere bei Termen hoher Ordnung, viele der Summanden überein. So bekommt man z.B. als Terme dritter Ordnung

$$\frac{1}{3!} [f_{xxx}(0,0) h_1^3 + 3f_{xxy}(0,0) h_1^2 h_2 + 3f_{xyy}(0,0) h_1 h_2^2 + f_{yyy}(0,0) h_2^3].$$

Bemerkung: Vorsicht bei der Anwendung auf *vektorwertige* Funktionen: Für solche Funktionen kann man ebenfalls eine Taylor-Entwicklung angeben, einfach indem man jede einzelne Komponentenfunktion Taylor-entwickelt. Die Stelle $\vec{\xi} = \vec{x}_* + \tau \vec{h}$ ist jedoch für jede der Komponentenfunktionen i.a. eine andere; es gibt keinen Grund anzunehmen (und es gibt auch Gegenbeispiele) dass man ein einheitliches $\vec{\xi}$ findet für alle Komponenten.

Ende des 2. Semesters.

Themen des 3. Semesters:

- Fortsetzung der Vektoranalysis (Extremwertaufgaben, auch mit Nebenbedingungen) (\rightarrow Optimierung)
- Gewöhnliche Differentialgleichungen
- Algebra/Gruppentheorie und Anwendungen in der Kryptographie (Verschlüsselung)