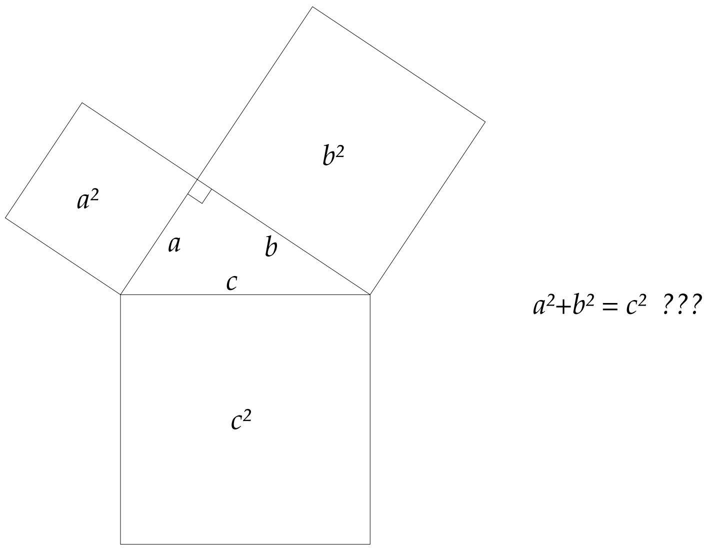
Fächen- und Rauminhalt, Integration

J.-H. Eschenburg, Universität Augsburg, SS 05, 07, 09, 11, 13

Vorbemerkung

In den "Bildungsstandards für den mittleren Schulabschluss" der Kultusminister (2003) werden die folgenden mathematischen Leitideen genannt: Zahl, Messen, Raum und Form, Funktionaler Zusammenhang, Daten und Zufall. Die Vorlesung nimmt diese Vorgabe auf. Sie ist Teil eines Zyklus von 4 Semestern, der die fachlichen Grundlagen für das nichtvertiefte Lehramtsstudium der Mathematik bereitstellen soll. Die vier Teile sind folgende: Zahl und Funktion (§55(1)1 LPO), Flächen- und Rauminhalt, Integration (§55(1)1 LPO), Linearität (§55(1)2 LPO), Mehrere Variable (§55(1)2 LPO). Mit normalen Schulkenntnissen sollte man den Zyklus mit jeder dieser Vorlesungen beginnen können.

1. Der Satz des Pythagoras



Die Mathematik hat die Aufgabe, das Verborgene auf Offensichtliches zurückzuführen. Ein schönes Beispiel für diesen Prozess ist der Lehrsatz des *Pythagoras*,[1] einer der grundlegenden Sätze der Geometrie:

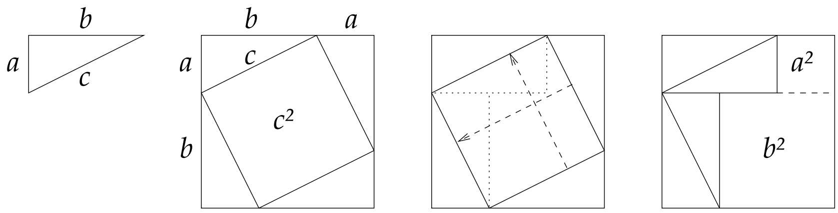
[1] Pythagoras von Samos, 569 (Samos) - 475 v. Chr.

In einem rechtwinkligen Dreieck mit Katheten a, b und Hypothenuse c gilt:[2]

$$a^2 + b^2 = c^2.$$

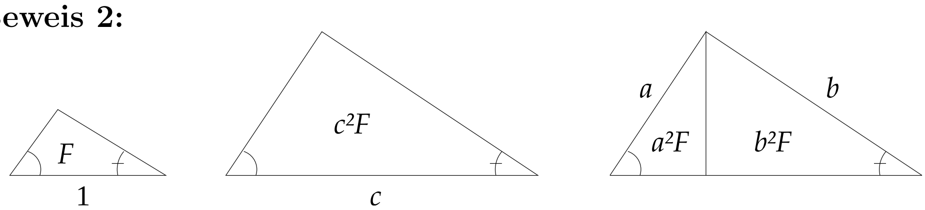
Aus der Figur ist zunächst in keiner Weise zu entnehmen, warum die beiden kleineren Quadrate zusammen den gleichen Flächeninhalt wie das große Quadrat haben sollten. Wir müssen dazu eine Brücke bauen zwischen dem Behaupteten und dem Offensichtlichen; diese Brücke nennt man *Konstruktion*; wir finden sie in jedem mathematischen Beweis wieder. Wir wollen nicht nur einen, sondern gleich drei Beweise dieses Satzes vorstellen; sie beruhen alle auf "offensichtlichen" Eigenschaften des *Flächeninhalts*, die wir anschließend diskutieren wollen.

Beweis 1:



Die Konstruktion besteht darin, das Dreieck in ein Quadrat mit Kantenlänge $b + a$ hineinzuplazieren. Dieses besteht aus vier Kopien des Dreiecks, die ein Quadrat über c aussparen. Wenn wir zwei der Dreiecke verschieben (rechte Figur), so sparen die vier Dreiecke stattdessen je ein Quadrat über a und b aus, deshalb muss $c^2 = a^2 + b^2$ gelten.

Beweis 2:



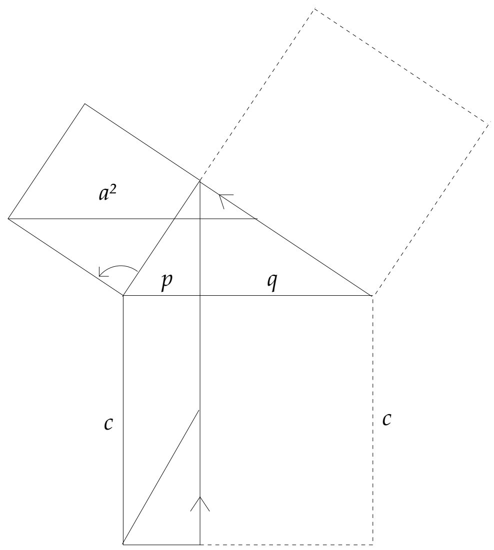
Die beiden rechtwinkligen Teildreiecke in der rechten Figur sind ähnlich zum großen Dreieck, mit dem sie jeweils einen Winkel gemeinsam haben.[3] Der Flächeninhalt ähnlicher Dreiecke wächst mit den Quadrat der

[2] In einem rechtwinkligen Dreieck sind die *Katheten* die beiden Seiten, die den rechten Winkel umschießen; die dritte, dem rechten Winkel gegenüberliegende Seite heißt *Hypothenuse*.

[3] Zwei Figuren sind *ähnlich*, wenn sie die gleiche Form bei möglicherweise unterscheidlicher Größe haben. Bei Dreiecken bedeutet dies einfach Gleichheit der Winkel. Zwei *rechtwinklige* Dreiecke sind ähnlich, wenn sie einen gemeinsamen Winkel

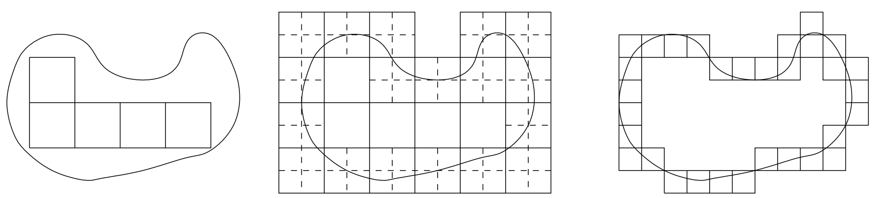
Länge. Wenn F den Flächeninhalt eines ähnlichen Dreiecks mit Hypothenuse 1 bezeichnet, so ist der Flächeninhalt des großen Dreiecks c^2F und der beiden Teildreiecke a^2F und b^2F, also ist $c^2F = a^2F + b^2F$.

Beweis 3:



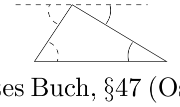
Das ist der Beweis von *Euklid*.[4] Die Hypothenuse c wird durch den Höhenfußpunkt in zwei Abschnitte p und q zerlegt. Damit wird das Quadrat über c in zwei Rechtecke pc und qc zerlegt. Wenn wir $pc = a^2$ und entsprechend $qc = b^2$ zeigen können (*Kathetensatz*), dann folgt $a^2 + b^2 = pc + qc = c^2$. Die Flächengleichheit $pc = a^2$ ergibt sich in drei Schritten: Das Rechteck pc wird durch *Scherung* in ein flächengleiches Parallelogramm verwandelt, das um 90^o gedreht wird und dann wieder durch Scherung (entlang der mit einem Pfeil bezeichneten Kante) in das Quadrat über der Seite a übergeht.

2. Flächeninhalt ebener Figuren



---

haben; der dritte Winkel ist bestimmt,



weil die Winkelsumme stets 180^o beträgt.

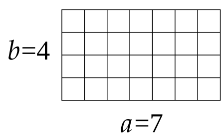
[4]Euklid (ca. 360 - 290 v.Chr.): Die Elemente, Erstes Buch, §47 (Oswalds Klassiker der exakten Wissenschaften, Bd. 235, S.32)

Wollen wir den Flächeninhalt F einer ebenen Figur messen, so zählen wir die Quadrate eines Gitters (Rechenkästchen), die ganz in der Figur Platz haben; deren Anzahl N_1 ist eine untere Schranke für den Flächeninhalt, wenn wir den Flächeninhalt F_o jedes Kästchens kennen. Zählen wir stattdessen die Kästchen, die die Figur nur treffen, so erhalten wir eine obere Schranke $N_2 F_o$:

$$N_1 F_0 \leq F \leq N_2 F_0. \tag{$*$}$$

In unserer Beispielfigur ist $N_1 = 5$ und $N_2 = 23$; der Flächeninhalt liegt also zwischen $5F_o$ und $23F_o$. Will man genauer messen, unterteilt man die Quadrate feiner, wie in der mittleren Figur angedeutet (natürlich braucht man nur die Randquadrate zu unterteilen) und kann damit die Differenz zwischen oberer und unterer Schranke verkleinern, in der Figur von 18 Quadraten mit Gesamtfläche $10F_0$ auf 30 Viertelquadrate mit Gesamtfläche $\frac{30}{4}F_0 = 7{,}5\,F_o$. Auf diese Weise lässt sich der Flächeninhalt beliebig genau messen.

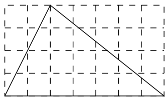
Mit dieser Definition ist zwar der Flächeninhalt aller Figuren bestimmt, aber durch eine Formel ausdrücken können wir ihn zunächst nur für achsenparallele Rechtecke: Bei Seitenlängen a und b ist der Flächeninhalt ab. Sind a und b nämlich ganzzahlig, ist der Flächeninhalt die Anzahl der Kästchen (Figur). Im Fall rationaler Kantenlängen $a = \frac{n}{p}$ und $b = \frac{m}{p}$ muss man die Kästchen in p^2 kleinere Kästchen mit Kantenlänge $1/p$ unterteilen. Es gibt $n \cdot m$ kleine Kästen, jedes mit Flächeninhalt $1/p^2$, somit erhält man den Flächeninhalt $\frac{n \cdot m}{p^2} = a \cdot b$. Irrationale Kanntenlängen a, b werden durch rationale angenähert.



Ebenso sehen wir, dass eine Streckung um den Faktor a in horizontaler und b in vertikaler Richtung den Flächeninhalt um den Faktor ab vergrößert.

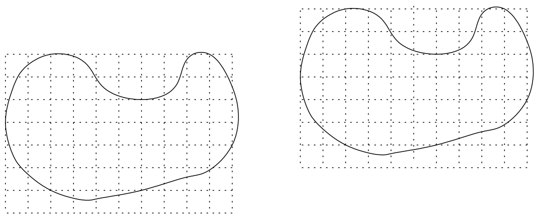
Wenn wir Formeln für den Flächeninhalt anderer einfacher Figuren[5] suchen, ist diese Definition wenig hilfreich. Bereits der Flächeninhalt eines Dreiecks würde uns Mühe machen:

[5] Das Wort *Figur* ist absichtlich etwas unpräzise gewählt. Wir meinen damit eine Teilmenge der Ebene mit "gutartigem" Rand. Der Fehler (obere minus untere Schranke) bei dieser Bestimmung des Flächeninhalts ist ja der Flächeninhalt aller Kästchen, die den Rand treffen, und dieser soll für genügend kleine Kästchen beliebig klein sein. Solche Teilmengen der Ebene nennt man *messbar*.

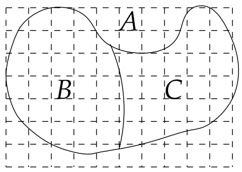


Wichtiger als die Definition des Flächeninhaltes sind deshalb die daraus folgenden Eigenschaften. Es genügen zwei; alle anderen lassen sich aus ihnen ableiten.

(1) *Der Flächeninhalt einer Figur bleibt ungeändert bei Verschiebungen.*

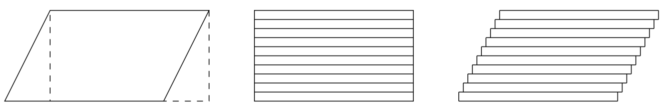


(2) *Zerlegt man eine Figur A in zwei Teilfiguren B und C, so ist der Flächeninhalt von A die Summe der Flächeninhalte von B und C.*



In der Tat kann mam aus diesen beiden Eigenschaften die Definition $(*)$ auf S. 4 zurückgewinnen: Die Kästchen eines Quadratgitters haben alle gleichen Flächeninhalt (Verschiebungsinvarinanz). Die Figur ist in der Vereinigung aller sie treffenden Kästchen enthalten und hat deshalb kleineren Flächeninhalt als diese Vereinigung (Zerlegungsinvarinanz: Diese Vereinigung lässt sich zerlegen in die Figur und eine Restmenge), andererseits enthält die Figur die Vereinigung aller Kästchen, die ganz darin liegen, und hat deshalb größeren Flächeninhalt als die Vereinigung dieser Kästchen. Damit erhalten wir $(*)$.

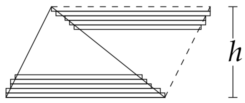
Mit den beiden Eigenschaften kann man zum Beispiel sofort sehen, dass der Flächeninhalt bei *Scherungen* derselbe bleibt:



Die Figuren geben zwei unterschiedliche Beweise dafür, dass das Rechteck denselben Flächeninhalt hat wie das Parallelogramm: In der linken Figur entsteht das Parallelogramm aus dem Rechteck, indem rechts ein

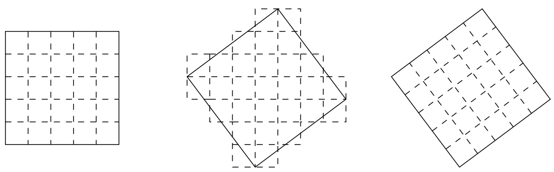
Dreieck abgeschnitten und links wieder angesetzt wird. In der mittleren Figur wird das Rechteck in schmale Streifen zerlegt, die anders angeordnet wieder das Parallelogramm ergeben (rechte Figur). Eigentlich müssen die Streifen dafür "unendlich schmal" sein, damit die Zackenlinie eine Gerade wird; das Argument gehört damit streng genommen bereits in den Bereich der *Infinitesimalrechnung*. Hier ist es noch entbehrlich, aber bei Volumina wird es unabdingbar (*Cavalierisches Prinzip*, siehe nächster Abschnitt).

Damit ergeben sich die Flächenformeln *Grundseite* · *Höhe* für das Parallelogramm und $\frac{1}{2}$(*Grundseite* · *Höhe*) für das Dreieck, denn dieses lässt sich zu einem Parallelogramm verdoppeln.



Die Flächengleichheit der beiden Dreiecke sehen wir durch Zerlegung in Streifen (Cavalieri-Prinzip). Das auf den Kopf gestellte (um 180 Grad gedrehte) Dreieck hat deshalb den gleichen Flächeninhalt.

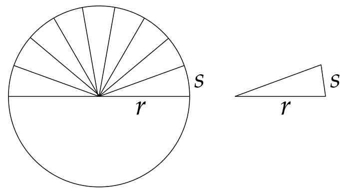
Wir können aus den beiden Eigenschaften aber sofort beweisen, dass der Flächeninhalt sogar unter beliebigen Drehungen invariant bleibt. Das ist nicht ganz offensichtlich, wie die folgenden Figuren zeigen:



Dass die Zahl der überdeckenden Kästchen im linken und im mittleren Bild dieselbe ist, sieht man nicht ohne weiteres. Anders wäre es, wenn man mit anderen, nicht mehr achsenparallelen Kästchen überdecken dürfte (rechte Figur). Aber haben diese gedrehten Kästchen den gleichen Flächeninhalt wie die ungedrehten? Das ist ja genau unsere Frage!

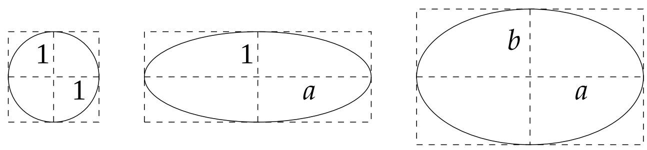
Und doch ist es so. Wir können mit den Kästchen eines um einen bestimmten Winkel α gedrehten Quadratgitters genauso wie mit den achsenparallelen jede Figur überdecken, und mit dem gleichen Argument wie vorher gilt für die gedrehten Kästchen die Beziehung $(*)$, wobei der Flächeninhalt F_0 des achsenparallelen durch den des gedrehten Kästchens ersetzt werden muss; diesen wollen wir F_α nennen. Es gibt aber eine Figur, die bei jeder Drehung unverändert bleibt: der Kreis. Da wir ihn durch die gleiche Anzahl achsenparalleler wie gedrehter Kästchen überdecken können, muss $F_\alpha = F_o$ gelten.

Der Flächeninhalt eines *Kreises* mit Radius r ist bekanntlich πr^2, wobei π als das Verhältnis von Umfang u und Durchmesser $2r$ des Kreises definiert ist.[6] Diese Formel stammt wie viele andere von *Archimedes*,[7] der sie durch eine Zerlegung des Kreises in "unendlich viele" schmale Dreiecke gewann:



Jedes Teildreieck hat Höhe r und Grundseite s,[8] also Flächeninhalt $\frac{1}{2}rs$, und durch Aufsummieren aller Teildreiecke ergibt sich die Kreisfläche zu $F = \frac{1}{2}ru$. Da $u = 2\pi r$, ist $F = \pi r^2$.

Die *Ellipse* mit Halbachsen a und b entsteht aus dem Kreis mit Radius 1 und Flächeninhalt π durch horizontale Streckung um den Faktor a und vertikale Streckung um b; damit hat sie den Flächeninhalt πab.



3. Rauminhalte

Rauminhalte oder *Volumina* werden ganz analog wie Flächeninhalte definiert, wobei die Einheitsquadrate durch *Einheitswürfel* (Würfel der Kantenlänge 1) zu ersetzen sind. Als Konsequenz ergeben sich die gleichen Regeln wie vorher und eine weitere, die die Beziehung zum Flächeninhalt beschreibt:

(1) *Das Volumen eines Körpers bleibt ungeändert bei Verschiebungen.*
(2) *Zerlegt man einen Körper A in zwei Teilkörper B und C, so ist das Volumen von A die Summe der Volumina von B und C.*

---

[6] Dieses Verhältnis ist für alle Kreise gleich, denn je zwei Kreise gehen durch Verschiebung und zentrische Streckung ineinander über. Verschiebungen verändern Längen gar nicht, zentrische Streckungen verändern sie um einen konstanten Faktor; das *Verhältnis* von zwei Längen bleibt also gleich.

[7] Archimedes von Syrakus, 287 - 212 v.Chr., Syrakus, Sizilien

[8] Das stimmt nicht ganz genau, aber der Fehler wird im Grenzübergang beliebig klein.

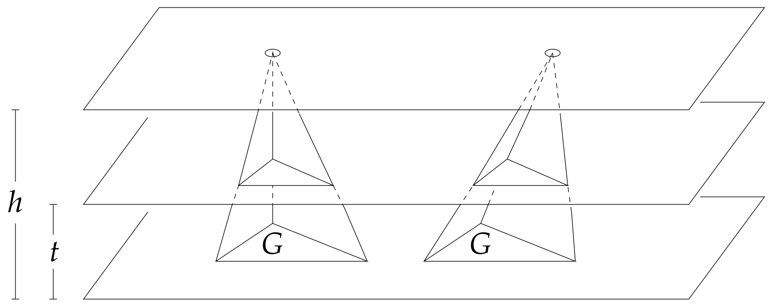
(3) *Das Volumen einer Platte oder Säule mit Höhe h über einem Flächenstück der Größe G ist Gh.*

Ein weiteres wichtiges Hilfsmittel zur Volumenberechnung ist das *Prinzip von Cavalieri*,[9] das in dieser Form bereits auf Archimedes zurückgeht:

> *Zwei Körper haben gleiches Volumen, wenn ihre Schnitte mit jeder horizontalen Ebene den gleichen Flächeninhalt haben.*

In der Tat können wir uns die Körper aus dünnen Scheiben über solchen horizontalen Schnitten zusammengesetzt denken; da deren Volumina übereinstimmen, haben die Körper insgesamt gleiches Volumen.

Wir wollen als erstes den Rauminhalt eines Kegels oder einer Pyramide von der Höhe h über einem Flächenstück G berechnen.

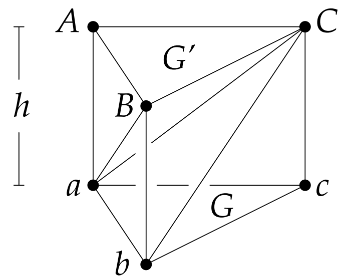


Nach dem *Cavalieri'schen Prinzip* hängt das Volumen nur von der Höhe h und der Grundfläche G ab: Wenn zwei Kegel mit gleicher Grundfläche und gleicher Höhe gegeben sind, dann sind die Schnitte mit einer Ebene parallel zur Grundfläche bei beiden Kegeln dieselben.[10] Damit sind auch die Volumina gleich, weil wir uns die beiden Kegel aus volumengleichen flachen Scheiben über diesen Querschnitten aufgebaut denken können (siehe Figur zu den Scherungen im vorigen Abschnitt).

Nun können wir das Kegelvolumen berechnen. Dazu betrachten wir eine Scheibe oder Säule mit Höhe h und dreieckiger Grundfläche G, ein *Prisma*. Die unteren Eckpunkte mögen a, b, c genannt werden, die oberen A, B, C. Dieses Prisma zerlegen wir in die drei *Tetraeder* (Kegel über einem Dreieck), die jeweils durch die vier Eckpunkte $(abcC)$, $(abBC)$ und $(aABC)$ gegeben sind.

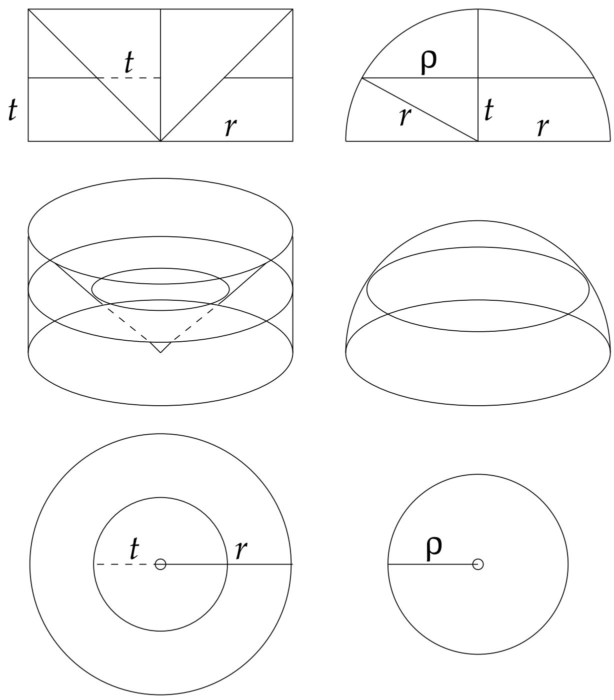
[9] Bonaventura Francesco Cavalieri, 1598 (Mailand) - 1647 (Bologna); vgl. aber "Mehrere Variable", Satz 15.4, Seite 65f.

[10] Der Schnitt mit einer Ebene in Höhe t ist eine Verkleinerung der Grundfläche G um den Faktor $(h-t)^2/h^2$, denn alle Längen werden um den Faktor $(h-t)/h$ verkleinert.



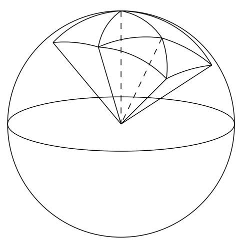
Alle drei Tetraeder haben das gleiche Volumen; für $(abcC)$ und $(aABC)$ ist das unmittelbar klar, denn es sind Kegel der Höhe h über der Grundfläche $G = (abc)$ und der Deckfläche $G' = (ABC)$, die gleich groß sind. Um die noch fehlende Gleichheit zu sehen, können wir z.B. die Tetraeder $(abBC)$ und $(aABC)$ ansehen als Kegel mit Spitze C über den Dreiecken (abB) und (aAB). Diese teilen das Rechteck $(abBA)$ diagonal in zwei Hälften, haben also gleichen Flächeninhalt, und da sie in derselben Ebene liegen, ist auch ihre Höhe die gleiche, also sind auch die Volumina von $(abBC)$ und $(aABC)$ gleich.

Damit ist das Volumen von jedem der drei Tetraeder ein Drittel des Prismavolumens Gh. Der Tetraeder $(abcC)$ mit Grundfläche G und Höhe h hat also das Volumen $\frac{1}{3}Gh$. Da sich jede Fläche in Dreiecke zerlegen lässt, gilt dieselbe Formel für Kegel über beliebigen Grundflächen.



Die Kegelformel war *Archimedes* bereits bekannt, und er benutzte sie zur Berechnung von Volumen und Oberfläche der *Kugel*. Dabei kommt wieder das *Cavalierische Prinzip* zur Geltung.[11] Die Halbkugel vom Radius r wird verglichen mit einem Kreiszylinder mit Radius r und Höhe r, aus dem ein auf der Spitze stehender Kreiskegel mit Radius und Höhe r ausgeschnitten ist. Schneidet man beide Körper mit einer horizontalen Ebene der Höhe t, so ist der Schnitt mit dem ausgebohrten Kreiszylinder ein Kreisring mit innerem Radius t und äußerem r, und der Schnitt mit der Halbkugel ist nach Pythagoras ein Kreis mit Radius $\rho = \sqrt{r^2 - t^2}$. Der Kreisring hat den Flächeninhalt $\pi r^2 - \pi t^2$ und der Kreis $\pi \rho^2 = \pi(r^2 - t^2)$; der Flächeninhalt ist also derselbe. Nach dem Cavalierischen Prinzip hat die Halbkugel daher dasselbe Volumen wie der ausgebohrte Zylinder, nämlich $\pi r^2 \cdot r - \frac{1}{3}\pi r^2 \cdot r = \frac{2}{3}\pi r^3$. Das Volumen der ganzen Kugel ist demnach $\frac{4}{3}\pi r^3$.

Dieselbe Überlegung, mit der Archimedes den Zusammenhang zwischen Umfang und Flächeninhalt des Kreises erkannte, führte ihn auch auf die Beziehung zwischen Oberfläche und Volumen der Kugel:



Wenn man sich die Kugelfläche aus kleinen, fast ebenen Flächenstücken (z.B. Dreiecken) zusammengesetzt denkt, dann setzt sich die Kugel aus den Kegeln mit Höhe r über diesen Flächenstücken zusammen, und nach der Kegelformel ist damit das Kugelvolumen gleich $\frac{1}{3} \cdot F \cdot r$, wobei F die Gesamtoberfläche der Kugel bezeichnet. Also ist $\frac{4}{3}\pi r^3 = \frac{1}{3} F r$ und daher $F = 4\pi r^2$. Die Kugelfläche ist also genau viermal so groß wie die Kreisfläche mit demselben Radius!

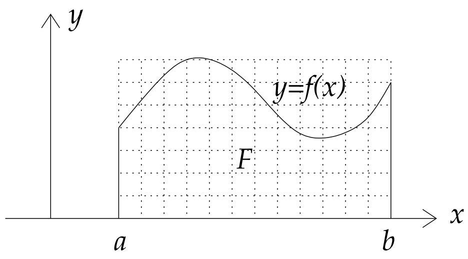
4. Die Fläche unter einer Kurve

Die Flächen- und Volumenberechnungen von Archimedes sind wunderschön, aber doch auf wenige sehr spezielle Beispiele beschränkt. Erst im 17. Jahrhundert wurde von *Barrow*[12] eine allgemeine Methode zur

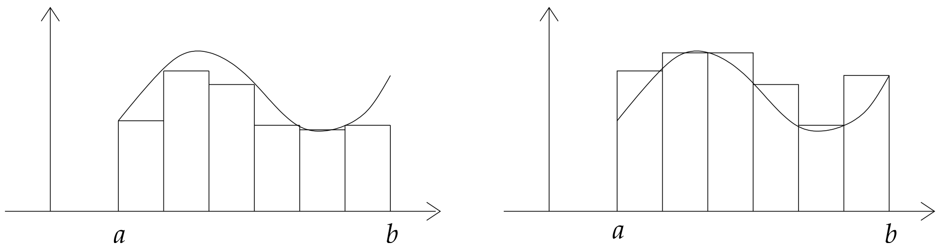
[11] Das Argument von Archimedes war etwas anders, vgl. http://www.cut-the-knot.com/pythagoras/Archimedes.shtml

[12] Isaac Barrow, 1630 - 1677 (London); Lehrer und Vorgänger von *Sir Isaac Newton* (1643 - 1727) als Professor in Cambridge. Er entdeckte, dass die Flächenberechnung die Umkehrung der Differentialrechnung ist (*Hauptsatz*

Flächenberechnung gefunden, und diese entfaltete sehr schnell Wirkungen, die weit über den ursprünglichen Zweck, die Flächenberechnungen, hinausreichten. Barrow beschränkte sich auf Flächen, die Rechtecken noch sehr ähnlich und nur an einer Seite von einer Kurve statt einer geraden Linie begrenzt sind, nämlich die Fläche unter dem *Graphen einer Funktion*[13] $f : [a, b] \to \mathbb{R}_+$.



Wir können in diesem Fall leicht untere und obere Schranken für den Flächeninhalt finden:



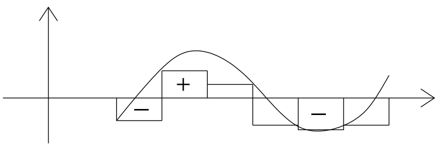
Wir müssen dazu das Intervall $[a, b]$ in kleinere Abschnitte unterteilen und die Fläche durch eine Vereinigung von Rechtecken über jedem Abschnitt ersetzen, deren Höhe der minimale oder maximale Wert der Funktion f in diesem Teilabschnitt ist. Wenn die Abschnitte immer kleiner werden, sollten die obere und die untere Schranke zusammenrücken; eine solche Funktion nennen wir *integrierbar*. Diese Eigenschaft ist erfüllt, wenn die Funktion f *stetig* ist. Wir haben den Begriff der *Stetigkeit* im letzten Semester kennengelernt; grob gesprochen bedeutet er, dass der Graph von f keine Sprünge macht.

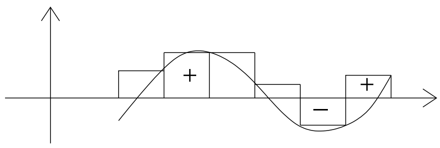
---

der Differential- und Integralrechnung), vgl. http://www.maths.uwa.edu.au/∼schultz/3M3/L18Barrow.html

[13] Eine *Funktion* ist durch eine Formel gegeben, mit der einer variablen Zahl x eine andere Zahl $y = f(x)$ zugeordnet wird, z.B. $y = x^2$. Der Bereich, aus dem die Variable x entnommen werden darf, heißt *Definitionsbereich* D von f, und der y-Bereich heißt *Wertebereich* W; wir schreiben dann $f : D \to W$. Der *Graph* von f ist die Menge $\{(x, y) \in D \times W;\ y = f(x)\}$. Im vorliegenden Fall ist der Definitionsbereich das *abgeschlossene Intervall* $[a, b] = \{x \in \mathbb{R};\ a \leq x \leq b\}$, d.h. der Bereich, der zwischen den reellen Zahlen a und b liegt, wobei $a \leq b$ vorausgesetzt ist, und der Wertebereich ist die Menge $\mathbb{R}_+ = [0, \infty) = \{y \in \mathbb{R};\ y \geq 0\}$.

Die Einschränkung auf Funktionen mit positiven Werten ist eigentlich überflüssig (wir werden sie aber der Anschaulichkeit halber in den Figuren meist beibehalten): Bei (stetigen) Funktionen $f : [a, b] \to \mathbb{R}$, die auch negative Werte annehmen dürfen, wird ebenso der Flächeninhalt zwischen der x-Achse und dem Graphen gemessen, aber die Anteile, die unterhalb der x-Achse liegen, bekommen ein negatives Vorzeichen; sie werden abgezogen.

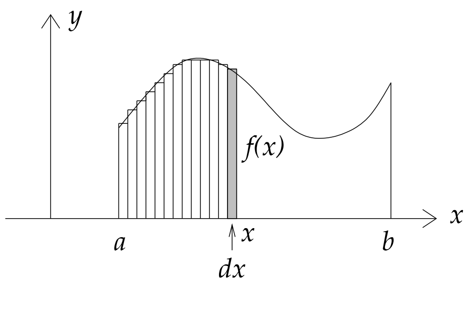




Für diesen mit *Vorzeichen* gerechneten Flächeninhalt zwischen x-Achse und Graph einer Funktion f gibt es die Bezeichnung

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

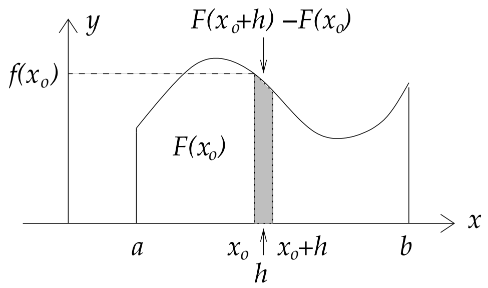
(*Integral von* a *bis* b *über* f). Das Integralsymbol $\int$ ist ein stilisiertes S, das für “*Summe*” steht; die Schreibweise $\int_a^b f(x)dx$ soll daran erinnern, dass man die Flächen von sehr schmalen Rechtecken der Höhe $f(x)$ und der Breite dx aufsummiert, wobei d für “*Differenz*” steht: Die Breite dx ist die Differenz von zwei nahe benachbarten x-Werten.



Bisher haben wir allerdings noch keinen Hinweis gefunden, wie wir die Zahl $F = \int_a^b f(x)dx$ wirklich berechnen können. Der Trick ist, nicht nur die eine Fläche F, sondern gleich eine ganze Schar von Flächen

$$F(x_o) := \int_a^{x_o} f(x)dx \tag{1}$$

für alle $x_o \in [a, b]$ zu betrachten und den Zuwachs dieser Flächen bei Vergrößerung von x_o zu studieren:



Die Fläche $F(x_o + h)$ umfasst $F(x_o)$ und den dunkel unterlegten Streifen; dieser ist also die Differenz $F(x_o + h) - F(x_o)$. Andererseits ist dieser Streifen im Wesentlichen ein Rechteck der Breite h und der Höhe $f(x_o)$, wenn man von der geringen Schwankung der Werte der (stetigen!) Funktion f in dem kleinen Intervall zwischen x_o und $x_o + h$ absieht, also erhalten wir:

$$\frac{F(x_o + h) - F(x_o)}{h} \approx f(x_o), \tag{2}$$

und die Annäherung wird umso besser, je mehr h sich der Null nähert, d.h. wir erhalten einen *Grenzwert* (über den Begriff wird im nächsten Abschnitt zu reden sein)

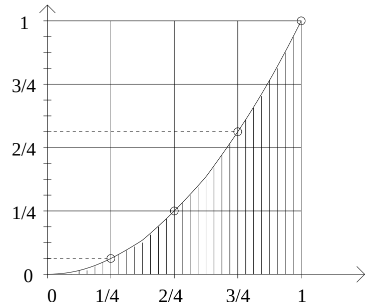
$$\lim_{h \to 0} \frac{F(x_o + h) - F(x_o)}{h} = f(x_o). \tag{3}$$

Die linke Seite beschreibt die *Veränderungsrate* oder *Ableitung* von $F(x)$ bei x_o, genannt $F'(x_o)$, und wir erhalten

$$F'(x_o) = f(x_o) \tag{4}$$

an jeder Stelle $x_o \in [a, b]$. Das ist der von Barrow gefundene *Hauptsatz der Differential- und Integralrechnung*: Die gesuchte Größe $F(x)$, die Flächenfunktion, entsteht aus der gegebenen Funktion f durch *Umkehrung des Ableitungsprozesses*, als sogenannte *Stammfunktion*. Das Ableiten (die *Differentiation*) ist aber rechnerisch gut verstanden (wir werden in den nächsten Abschnitten darüber reden), daher ist auch der Umkehrprozess in vielen Fällen nicht schwierig und führt zu den aus der Schule bekannten Ergebnissen, z.B.

$$\int_0^1 x^2 dx = \left[\frac{1}{3}x^3\right]_0^1 = \frac{1}{3}.$$



Bevor wir soweit sind, gibt es noch einige Fragen zu klären, z.B. die folgende: Selbst wenn wir wissen, dass $f(x) = x^2$ die Ableitung von $F(x) = \frac{1}{3}x^3$ ist, muss deshalb $F(1) = \frac{1}{3}$ die gesuchte Fläche sein? Könnte es nicht noch ganz andere Funktionen F mit $F'(x) = x^2$ geben? Vielleicht haben wir die falsche gewählt? Wie würden Sie diese Frage beantworten?

5. Grenzwerte

Von *Grenzwerten* können wir sprechen, wenn eine Größe y von einer anderen Größe x abhängig ist, $y = f(x)$ für eine Funktion $f : D \to W$. Wir nennen y^* den *Grenzwert* oder *Limes* von $f(x)$ für x gegen x^* und schreiben dafür

$$[x \to x^* \Rightarrow f(x) \to y^*] \quad \text{oder} \quad f(x) \xrightarrow{x \to x^*} y^* \quad \text{oder} \quad y^* = \lim_{x \to x^*} f(x),$$

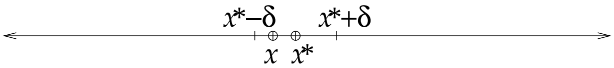
falls gilt:

> $f(x)$ kommt dem Wert y^* beliebig nahe, wenn x sich dem Wert x^* nähert.

Was bedeutet "beliebig nahe kommen"? x ist *nahe* an x^*, wenn die Differenz (die Abweichung) zwischen x und x^* "klein" ist. Aber wir wissen ja nicht, welche der beiden Zahlen die größere ist; die Differenz $x - x^*$ kann also auch negativ sein. Was bedeutet "klein" bei negativen Zahlen? Zweifellos ist -100 "kleiner" als $-1/100$ (je größer die Schulden, desto kleiner das Vermögen ...), aber das ist hier nicht gemeint: Wenn die Differenz -100 ist, sind die Zahlen sicher nicht nahe beieinander. Es ist der *Betrag*[14] *der Differenz*, $|x - x^*|$, der klein sein soll, noch kleiner als eine vorgegebenen kleine (aber positive) Zahl ("*Schranke*"), z.B. $1/100$. Diese vorgegebene kleine Zahl wird meistens mit einem kleinen griechischen Buchstaben, z.B. δ (*Delta*) oder ϵ (*Epsilon*) bezeichnet:

[14] Der *Betrag* $|x|$ einer reellen Zahl x ist gleich x, wenn $x \geq 0$ und gleich $-x$, wenn $x < 0$, oder in einem Audruck: $|x| = \sqrt{x^2}$. Bei *komplexen* Zahlen $x = u + iv$ ist der Betrag ganz ähnlich definiert: $|x| = \sqrt{x\bar{x}}$ mit $\bar{x} = u - iv$; vgl. Abschnitte 15 und 16 von "Zahl und Funktion".

$|x - x^*| < \delta$. Die Größe $|x - x^*|$ interpretieren wir auch geometrisch, nämlich als *Abstand* zwischen den Zahlen x und x^*, die wir dann als Punkte auf der *Zahlengeraden* verstehen; $|x - x^*| < \delta$ bedeutet, dass x zwischen $x^* - \delta$ und $x^* + \delta$ liegt.



1. Definition des Grenzwerts: $f(x) \overset{x \to x^*}{\longrightarrow} y^*$ oder $y^* = \lim_{x \to x^*} f(x)$ heißt:

> *Für alle* $\epsilon > 0$ *gibt es ein* $\delta > 0$ *mit der Eigenschaft, dass für alle* $x \in D$ *gilt:*
> *Wenn* $|x - x^*| < \delta$, *dann* $|f(x) - y^*| < \epsilon$.[15]

Auch mit *Folgen* kann man Annäherungen beschreiben (vgl. "Zahl und Funktion", Abschnitt 9):[16]

2. Definition des Grenzwerts: $f(x) \overset{x \to x^*}{\longrightarrow} y^*$ oder $y^* = \lim_{x \to x^*} f(x)$ heißt:

> Für jede Folge (x_n) in D mit $x_n \to x^*$ gilt $f(x_n) \to y^*$.

Eine Anwendung des Grenzwerts ist der bereits erwähnte Begriff der *Stetigkeit* (vgl. "Zahl und Funktion", Abschnitt 19): Eine Funktion $f : D \to \mathbb{R}$, heißt *stetig*, wenn für alle $x^* \in D$ gilt: $f(x^*) = \lim_{x \to x^*} f(x)$ oder (mit anderen Symbolen):

$$D \ni x \to x^* \Rightarrow f(x) \to f(x^*). \tag{5}$$

In vielen Fällen ist der Punkt x^* gar nicht selbst Element von D; er muss nur durch Punkte $x \in D$ *approximierbar* sein, d.h. zu jeder

---

[15] Mit den gebräuchlichen Abkürzungen $\forall$ = "für alle" und $\exists$ = "existiert" lautet diese Definition:

$$\forall_{\epsilon>0} \exists_{\delta>0} \forall_{x \in D} \; [\, |x - x^*| < \delta \Rightarrow |f(x) - y^*| < \epsilon \,].$$

[16] 1. Def. erfüllt $\overset{!}{\Rightarrow}$ 2. Def. erfüllt: $x_n \to x^* \Rightarrow |x_n - x^*| < \delta$ für große $n \in \mathbb{N} \overset{1.\text{Def.}}{\Rightarrow}$ $|f(x_n) - y^*| < \epsilon \Rightarrow f(x_n) \to y^*$.
1. Def. nicht erfüllt $\overset{!}{\Rightarrow}$ 2. Def. nicht erfüllt: 1. Def. nicht erfüllt $\Rightarrow$

$$\exists_{\epsilon>0} \forall_{\delta>0} \exists_{x \in D} \; [\, |x - x^*| < \delta \text{ und } |f(x) - y^*| \geq \epsilon \,].$$

Insbesondere können wir zu $\delta = \frac{1}{n}$ für beliebiges $n \in \mathbb{N}$ ein $x_n \in D$ finden mit $|x_n - x^*| < \frac{1}{n}$ und $|f(x_n) - y^*| \geq \epsilon$. Damit ist eine Folge (x_n) in D gefunden mit $x_n \to x^*$, aber $f(x_n) \not\to y^*$, somit ist die 2. Def. nicht erfüllt.

Schranke $\epsilon > 0$ muss es ein $x \in D$ mit $|x - x^*| < \epsilon$ geben.[17] Zum Beispiel kann D ein offenes Intervall (a, b) sein[18] und x^* einer der *Randpunkte*, etwa $x^* = a$. Das trifft häufig dann zu, wenn $f(x)$ ein Quotient $\frac{p(x)}{q(x)}$ ist, dessen Zähler $p(x)$ und Nenner $q(x)$ beide für $x \to a$ gegen Null (oder auch gegen Unendlich) gehen. Beispiel: $f(x) = \frac{x^2 - 2ax + a^2}{x^2 - a^2}$ mit $a \neq 0$:

$$\frac{x^2 - 2ax + a^2}{x^2 - a^2} = \frac{(x-a)^2}{(x-a)(x+a)} = \frac{x-a}{x+a} \xrightarrow{x \to a} \frac{0}{2a} = 0$$

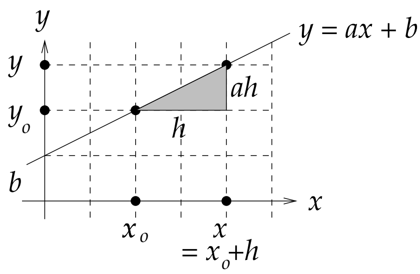
Der hier angewandte Trick ist typisch für Grenzwert-Berechnungen:

> *Zuerst Vereinfachen (z.B. Kürzen), dann zum Grenzwert übergehen!*

Wenn wir im Zähler $p(x) = x^2 - 2ax + a^2$ und im Nenner $q(x) = x^2 - a^2$ *vor* dem Kürzen zum Grenzwert $x \to a$ übergegangen wären, d.h. a anstelle von x eingesetzt hätten,[19] so hätten wir den unsinnigen Ausdruck $\frac{0}{0}$ erhalten; *nach* dem Kürzen des Faktors $(x - a)$ löst sich diese Schwierigkeit ganz von selbst auf.

6. Differenzen und Differenzieren

Die einfachsten Funktionen $y = f(x)$ sind die "linearen", deren Graph eine Gerade ist; genauer heißen sie eigentlich *inhomogen linear* oder *affin*. Für sie sind die Änderungen der x- und der y-Werte, also die Differenzen $x - x_o$ und $y - y_o$ zueinander proportional: $y - y_o = a(x - x_o)$ oder $f(x_o + h) - f(x_o) = ah$ für eine Konstante a, die von x und h (und sogar von x_o) unabhängig ist, oder anders gesagt, der *Differenzenquotient* $\frac{y - y_o}{x - x_o} = \frac{f(x+h) - f(x)}{h}$ ist konstant gleich a.



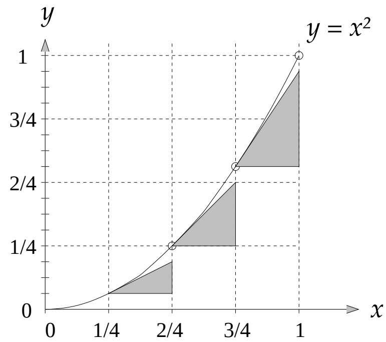
[17] Ein Sonderfall ist "$x^* = \infty$"; dabei nähert sich x keinem Zahlenwert an, sondern wächst über alle Grenzen. Die Aussage $f(x) \xrightarrow{x \to \infty} y^*$ oder $y^* = \lim_{x \to \infty} f(x)$ bedeutet: Für alle $\epsilon > 0$ gibt es eine Zahl R mit der Eigenschaft, dass für alle $x \in D$ gilt: Wenn $x \geq R$, dann $|f(x) - y^*| < \epsilon$. Oder mit der 2. Definition: Für jede Folge $x_n \to \infty$ in D gilt $f(x_n) \to y^*$.

[18] Beim *offenen Intervall* (a, b) sind die Randpunkte a, b nicht enthalten: $(a, b) = \{x \in \mathbb{R};\ a < x < b\}$.

[19] Zähler und Nenner sind stetig, daher $\lim_{x \to a} p(x) = p(a)$, $\lim_{x \to a} q(x) = q(a)$.

Aus der Figur wird die geometrische Bedeutung der Zahl a als *Steigung* der Geraden deutlich.

Dass die sehr einfache Funktion $f(x) = ax+b$ spezielle Eigenschaften hat, ist nicht weiter verwunderlich. Höchst bemerkenswert dagegen ist, dass sehr viele andere, beliebig komplizierte Funktionen eine ähnliche Eigenschaft haben:



(1) $f(x) = x^2$

$$\begin{aligned} f(x+h) - f(x) &= x^2 + 2hx + h^2 - x^2 \\ &= 2hx + h^2, \\ \frac{f(x+h) - f(x)}{h} &= 2x + h \xrightarrow{h \to 0} 2x \end{aligned}$$

(2) $f(x) = x^3$

$$\begin{aligned} f(x+h) - f(x) &= x^3 + 3hx^2 + 3h^2x + h^3 - x^3 \\ &= 3hx^2 + 3h^2x + h^3, \\ \frac{f(x+h) - f(x)}{h} &= 3x^2 + 3hx + h^2 \xrightarrow{h \to 0} 3x^2 \end{aligned}$$

(3) $f(x) = \frac{1}{x}$

$$\begin{aligned} f(x+h) - f(x) &= \frac{1}{x+h} - \frac{1}{x} = \frac{x - (x+h)}{(x+h)x} \\ &= -\frac{h}{x^2 + hx} \\ \frac{f(x+h) - f(x)}{h} &= -\frac{1}{x^2 + hx} \xrightarrow{h \to 0} -\frac{1}{x^2} \end{aligned}$$

Bemerkung Wir haben hier (wie schon in den vorigen Beispielen) benutzt, dass die vier Grundrechenarten *stetig* sind: $h \to 0 \Rightarrow hx \to 0 \Rightarrow x^2 + hx \to x^2 \Rightarrow \frac{1}{x^2+hx} \to \frac{1}{x^2}$. Wenn wir für h eine Nullfolge $h_n \to 0$ einsetzen, dann folgt dies aus Satz 10.1, S. 35 “Zahl und Funktion”.

(4) $f(x) = e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots$ (*Exponentialfunktion*)[20]

$$\begin{aligned}
f(x+h) - f(x) &= e^{x+h} - e^x = e^x e^h - e^x = e^x(e^h - 1) \\
&= e^x(h + \frac{h^2}{2!} + \frac{h^3}{3!} \dots) = e^x \cdot h(1 + \frac{h}{2!} + \frac{h^2}{3!} \dots) \\
\frac{f(x+h) - f(x)}{h} &= e^x(1 + \frac{h}{2!} + \frac{h^2}{3!} \dots) \stackrel{h \to 0}{\longrightarrow} e^x
\end{aligned}$$

Gegenüber der affinen Funktion $f(x) = ax + b$ fallen zwei Unterschiede auf:

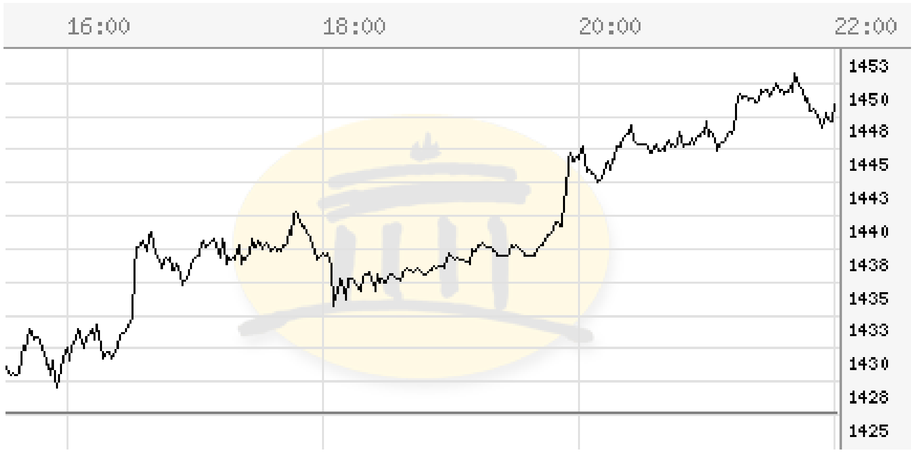
- Der Ausdruck $f(x_o + h) - f(x_o)$ ist nicht länger ein konstantes Vielfaches von h, d.h. der Differenzenquotient $\frac{f(x_o+h)-f(x_o)}{h}$ ist nicht konstant, sondern hängt von h ab. Aber dieser Ausdruck stabilisiert sich, wenn h gegen Null strebt; er hat einen (natürlich von h unabhängigen) *Grenzwert* a, wenn $h \to 0$ geht. Der Funktionsgraph sieht in der Nähe von x_o annähernd wie eine Gerade mit Steigung a aus (die *Tangente*), und wir nennen a deshalb auch die *Steigung* von f (eigentlich: der Tangente des Graphen von f) im Punkt x_o.
- Anders als die Steigung einer Geraden ist die Steigung a eines beliebigen Funktionsgraphen von x_o abhängig und definiert damit eine neue Funktion $x_o \mapsto a = f'(x_o)$, die wir als *Ableitung* f' von f bezeichnen. Den Rechenprozess, mit dem man die Ableitung f' aus der gegebenen Funktion f berechnet, nennt man *Ableiten* oder *Differenzieren.*

Definition: Es sei $D \subset \mathbb{R}$ ein offenes Intervall, $D = (a, b)$. Eine Funktion $f : D \to \mathbb{R}$ heißt *differenzierbar* in einem Punkt $x_o \in D$, wenn der Differenzenquotient $\frac{f(x_o+h)-f(x_o)}{h}$ für $h \to 0$ einen Grenzwert a besitzt. Diesen nennen wir die *Ableitung* von f in x_o und schreiben dafür $a = f'(x_o)$.

Unsere Beispiele zeigen, dass diese Eigenschaft für Funktionen gilt, die nur die vier Grundrechenarten verwenden (*rationale Funktionen*), aber auch für gewisse unendliche Summen, vgl. Beispiel (4). Der Graph

[20] Für jede natürliche Zahl n ist $n!$ („n-Fakultät") das Produkt der Zahlen von 1 bis n, also $1! = 1$, $2! = 1 \cdot 2 = 2$, $3! = 1 \cdot 2 \cdot 3 = 6$, $4! = 24$, $5! = 120$, $6! = 720$ usw. Diese Zahlen werden bald sehr groß; da sie im Nenner stehen, sind die Zahlen $\frac{x^n}{n!}$ für große n sehr nahe bei 0 und spielen bei der Addition keine Rolle mehr. Deshalb können wir diese unendlich vielen Zahlen addieren; die Summe *konvergiert*; vgl. “Zahl und Funktion”, S. 61 und 62f

differenzierbarer Funktionen ist “glatt”; er hat an jeder Stelle eine *Richtung*. Das muss keineswegs bei jeder Funktion der Fall sein, wie z.B. der zeitliche Verlauf von Aktienkursen zeigt.[21]



Eine solche Funktion ist zwar stetig (der Graph hat keine Lücken und enthält keine senkrechten Geradenstücke), aber nicht differenzierbar; sie ist das Ergebnis einer sehr großen Menge gleichzeitiger Kauf- und Verkauf-Entscheidungen von unabhängigen Akteuren mit unterschiedlichen Intentionen, so dass im Kleinen keine Richtung zustandekommt. Ein vergleichbares physikalisches Beispiel ist die von dem englischen Botaniker Robert Brown 1827 beobachtete Irrfahrt eines leichten Materieteilchen (z.B. Pollen auf einer Wasseroberfläche) auf Grund unzählig vieler Stöße von Molekülen (Brownsche Bewegung).[22] Natürlich sind auch solche Prozesse Gegenstand mathematischer Theorien (“Stochastische Prozesse”).

7. Ableitungsregeln

Um Ableitungen systematisch zu berechnen, muss man als Ausgangspunkt einige einfache Funktionen ableiten (differenzieren) können und braucht zusätzlich Rechenregeln, um kompliziertere Funktionen abzuleiten, die aus einfachen zusammengesetzt sind. Als Ausgangspunkt nehmen wir die *affinen* (inhomogen linearen) Funktionen $f(x) = ax + b$

---

[21] Die Abbildung zeigt die zeitliche Entwicklung des amerikanischen NASDAQ-Aktienindex am 4. Mai 2005; Quelle: http://finanzen.web.de/

[22] Diese Erklärung stammt von Albert Einstein (1905). Eine sehr schöne Simulation dazu findet man auf der Webseite
http://galileo.phys.virginia.edu/classes/109N/more_stuff/Applets/
unter “Einstein’s explanation of Brownian motion”.

mit ihrer konstanten Ableitung (Steigung) $f'(x) = a$. Insbesondere haben die *Konstanten* $f(x) = b = 0 \cdot x + b$ die Ableitung Null.[23] Um komplizierte Beispiele zu behandeln, müssen wir verstehen, wie Funktionen *zusammengesetzt* werden können. Es gibt grundsätzlich zwei Arten, aus zwei gegebenen Funktionen f und g eine neue zu machen:

Durch Rechnen: Sind zwei Funktionen $f, g : D \to \mathbb{R}$ gegeben, dann können wir aus ihne mit den vier Grundrechenarten neue Funktionen $f + g, f - g,\ f \cdot g,\ f/g : D \to \mathbb{R}$ (sofern g nirgends Null ist) berechnen, die folgendermaßen für alle $x \in D$ definiert sind: $(f + g)(x) := f(x) + g(x)$, $(f \cdot g)(x) := f(x) \cdot g(x)$ usw.

Durch Verketten: Sind zwei Funktionen $f : A \to B$ und $g : B \to C$ gegeben (A, B, C irgendwelche Mengen), dann ist eine neue Funktion $g \circ f : A \to C$ für alle $x \in A$ definiert durch $(g \circ f)(x) = g(f(x))$. Ist z.B. $A = B = C = \mathbb{R}$ und $f(x) = x^2 + 1$ und $g(x) = 3x - 2$, dann ist $(g \circ f)(x) = 3f(x) - 2 = 3(x^2 + 1) - 2$ und $(f \circ g)(x) = (3x - 2)^2 + 1$.

Dementsprechend gibt es zwei Sorten von Regeln, wie man zusammengesetzte Funktionen differenziert:

Satz 7.1. Summenregel, Produktregel, Quotientenregel:
Ist $D \subset \mathbb{R}$ ein offenes Intervall und $f, g : D \to \mathbb{R}$ differenzierbar, so sind $f \pm g$, $f \cdot g$, f/g (falls $g \neq 0$) wieder differenzierbar, und die Ableitung der neuen Funktion lässt sich folgendermaßen aus f und g sowie ihren Ableitungen f' und g' bestimmen:

$$(f \pm g)' = f' \pm g' \tag{6}$$

$$(f \cdot g)' = f' \cdot g + f \cdot g' \tag{7}$$

$$(f/g)' = (f' \cdot g - f \cdot g')/g^2 \tag{8}$$

Beweis. Zu jedem $x \in D$ müssen wir die Differenzenquotienten der neuen Funktionen berechnen. Für Summe und Differenz ist das sehr einfach:

$$\begin{aligned}\frac{(f \pm g)(x + h) - (f \pm g)(x)}{h} &= \frac{f(x + h) \pm g(x + h) - (f(x) \pm g(x))}{h} \\ &= \frac{f(x + h) - f(x)}{h} \pm \frac{g(x + h) - g(x)}{h} \\ &\xrightarrow{h \to 0} f'(x) \pm g'(x)\end{aligned}$$

[23] Das entspricht ja auch der Anschauung, denn der Graph der Konstanten ist eine horizontale Gerade, d.h. eine Gerade mit Steigung Null.

Beim Produkt benötigen wir einen kleinen Trick: Als Zwischenschritt von $f(x+h)g(x+h)$ zu $f(x)g(x)$ fügen wir den Null-Term

$$-f(x)\cdot g(x+h)+f(x)\cdot g(x+h)$$

hinzu (der Trick wurde bei Produkten schon öfters verwendet, siehe “Zahl und Funktion”, S. 35, Gl. (25)):[24]

$$\begin{aligned}
&\frac{(f\cdot g)(x+h)-(f\cdot g)(x)}{h}\\
=\;&\frac{f(x+h)\cdot g(x+h)-f(x)\cdot g(x)}{h}\\
=\;&\frac{f(x+h)g(x+h)-f(x)g(x+h)+f(x)g(x+h)-f(x)g(x)}{h}\\
=\;&\frac{f(x+h)-f(x)}{h}\cdot g(x+h)\;+\;f(x)\cdot\frac{g(x+h)-g(x))}{h}\\
\xrightarrow{h\to 0}\;&\qquad f'(x)\cdot g(x)\qquad + \qquad f(x)\cdot g'(x)
\end{aligned}$$

Um schließlich den Quotienten $f/g = f\cdot(1/g)$ zu differenzieren, genügt es, den Differenzenquotient von $1/g$ zu bestimmen und dann die Produktregel anzuwenden:

$$\begin{aligned}
\frac{(1/g)(x+h)-(1/g)(x)}{h} &= \frac{\frac{1}{g(x+h)}-\frac{1}{g(x)}}{h}\\
&= \frac{1}{g(x+h)g(x)}\cdot\frac{g(x)-g(x+h)}{h}\\
&\xrightarrow{h\to 0} \frac{1}{g(x)^2}\cdot(-g'(x))
\end{aligned}$$

und somit $(1/g)'=-g'/g^2$. Die Quotientenregel (8) folgt nun aus (7): Da f und $1/g$ differenzierbar, ist auch $f\cdot(1/g)$ differenzierbar mit

$$(f\cdot\frac{1}{g})' = f'\cdot\frac{1}{g}+f\cdot\left(\frac{1}{g}\right)' = \frac{f'}{g}-f\cdot\frac{g'}{g^2}=\frac{f'g-fg'}{g^2} \qquad \square$$

[24] Für den letzten Schluss der nachfolgenden Gleichungskette benötigen wir auch die *Stetigkeit* der Funktion g, d.h. $g(x+h)\xrightarrow{h\to 0} g(x)$. In der Tat ist eine differenzierbare Funktion automatisch stetig: Da der Ausdruck $\frac{g(x+h)-g(x)}{h}$ für $h\to 0$ einen Grenzwert $a=g'(x)$ besitzt, ist er beschränkt (kleiner als $a+\epsilon$, größer als $a-\epsilon$) für $|h|<\delta$, also $\frac{|g(x+h)-g(x)|}{|h|}\le C$ und damit $|g(x+h)-g(x)|\le C\cdot|h|\xrightarrow{h\to 0} 0$.

Satz 7.2. Kettenregel:
Sind $A, B \subset \mathbb{R}$ offene Intervalle und sind die Funktionen $f : A \to B$ und $g : B \to \mathbb{R}$ differenzierbar, so ist $g \circ f : A \to \mathbb{R}$ wieder differenzierbar, und die Ableitung dieser Funktion lässt sich folgendermaßen aus f und g und deren Ableitungen f' und g' bestimmen:

$$\begin{aligned} (g \circ f)' &= (g' \circ f) \cdot f' \\ (g \circ f)'(x) &= g'(f(x)) \cdot f'(x) \quad \forall_{x \in A} \end{aligned} \tag{9}$$

Beweis. Wir müssen wieder den Differenzenquotienten von $g \circ f$ berechnen, d.h. den Ausdruck $\frac{(g\circ f)(x+h)-(g\circ f)(x)}{h}$. Dabei setzen wir lieber $x+h =: \tilde{x}$, also $h = \tilde{x}-x$. Somit haben wir den Quotienten $\frac{g(f(\tilde{x}))-g(f(x))}{\tilde{x}-x}$ zu berechnen. Diesen Bruch wollen wir folgendermaßen erweitern:

$$\frac{g(f(\tilde{x})) - g(f(x))}{\tilde{x} - x} = \frac{g(f(\tilde{x})) - g(f(x))}{f(\tilde{x}) - f(x)} \cdot \frac{f(\tilde{x}) - f(x)}{\tilde{x} - x} \tag{10}$$

Jetzt stehen rechts Differenzenquotienten von g und f, und da mit $\tilde{x} - x \to 0$ auch $f(\tilde{x}) - f(x) \to 0$ (Stetigkeit von f), sollte der erste Faktor rechts gegen $g'(f(x))$ und der zweite Faktor gegen $f'(x)$ streben. Aber es gibt ein kleines Problem: Der Nenner $f(\tilde{x}) - f(x)$ des ersten Faktors auf der rechten Seite kann Null werden, auch wenn $\tilde{x} \neq x$. Doch das ist kein wirkliches Hindernis, denn dieser Faktor ist auch dann noch definiert: Für festes x und $y = f(x)$ definieren wir auf B eine neue Funktion g^* durch

$$g^*(\tilde{y}) = \begin{cases} \frac{g(\tilde{y})-g(y)}{\tilde{y}-y} & \text{falls} \quad \tilde{y} \neq y \\ \\ g'(y) & \text{falls} \quad \tilde{y} = y \end{cases} \tag{11}$$

Die Differenzierbarkeit von g in y bedeutet gerade

$$g^*(\tilde{y}) \xrightarrow{\tilde{y}\to y} g^*(y) = g'(y). \tag{12}$$

Nun folgt für *alle* $y \in B$

$$g(\tilde{y}) - g(y) = g^*(\tilde{y}) \cdot (\tilde{y} - y) \tag{13}$$

(für $\tilde{y} = y$ steht da nur $0 = 0$). Setzen wir $f(\tilde{x})$ für $\tilde{y}$ und $f(x)$ für y ein, so folgt

$$g(f(\tilde{x}) - g(f(x)) = g^*(f(x)) \cdot (f(\tilde{x}) - f(x))$$

und mit (12) erhalten wir anstelle von (10):

$$\frac{g(f(\tilde{x})) - g(f(x))}{\tilde{x} - x} = g^*(f(\tilde{x})) \cdot \frac{f(\tilde{x}) - f(x)}{\tilde{x} - x} \xrightarrow{\tilde{x}\to x} g'(f(x)) \cdot f'(x) \qquad \square$$

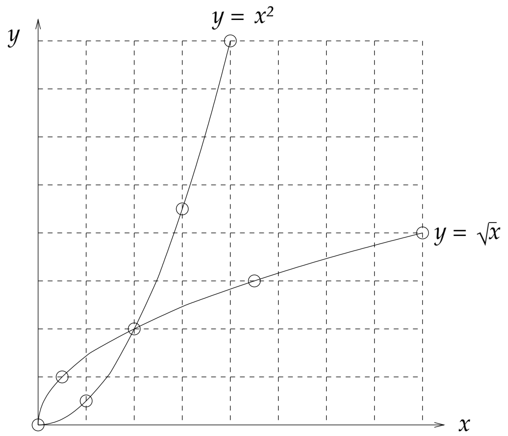
Eine Funktion $f : A \to B$ nennt man *bijektiv* oder *invertierbar*,[25] wenn sie durch eine weitere Funktion $g : B \to A$ "rückgängig" gemacht werden kann:

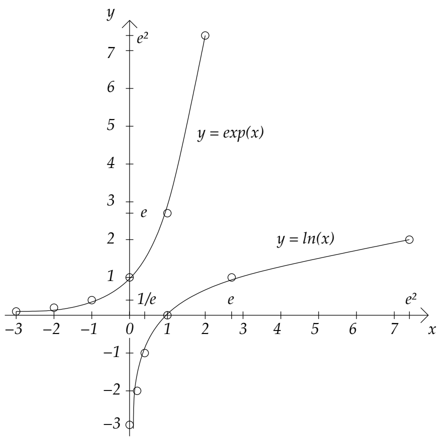
$$\forall_{x\in A} \;\; g(f(x)) = x, \;\; \forall_{y\in B} \;\; f(g(y)) = y \tag{14}$$

oder kurz

$$g \circ f = \mathrm{id}_A\,, \quad f \circ g = \mathrm{id}_B\,. \tag{15}$$

Hierbei bezeichnen id_A und id_B die *identischen Funktionen*, die gar nichts verändern: $\mathrm{id}_A(x) = x$ für alle $x \in A$ und $\mathrm{id}_B(y) = y$ für alle $y \in B$. Wir nennen g die *Inverse* oder *Umkehrfunktion* von f und schreiben dafür oft f^{-1} statt g. Wenn A und B abgeschlossene Intervalle sind und $f : A \to B$ umkehrbar und stetig ist, dann ist auch $g = f^{-1} : B \to A$ stetig ("Zahl und Funktion", Satz 21.3, S.75). Zum Beispiel ist die Umkehrfunktion der n-ten Potenz $p_n : \mathbb{R}_+ \to \mathbb{R}_+$, $p_n(x) = x^n$ die n-te Wurzel $w_n(x) = \sqrt[n]{x} = x^{1/n}$, und die Umkehrfunktion der e-Funktion $\exp : \mathbb{R} \to \mathbb{R}_+^*$ ist der *natürliche Logarithmus* $\ln : \mathbb{R}_+^* \to \mathbb{R}$.





Satz 7.3. Ableitung der Umkehrfunktion:
Sind $A, B \subset \mathbb{R}$ Intervalle und $f : A \to B$ umkehrbar und differenzierbar mit $f'(x) \neq 0$ für alle $x \in A$, dann ist auch $g = f^{-1} : B \to A$ differenzierbar mit Ableitung

$$g' = \frac{1}{f' \circ g}\,, \quad g'(y) = \frac{1}{f'(g(y))} \; \forall_{y\in B} \tag{16}$$

Beweis. Zu $y, \tilde{y} \in B$ setzen wir $x = g(y)$, $\tilde{x} = g(\tilde{y})$. Wenn $\tilde{y} \to y$, dann geht auch $\tilde{x} \to x$ (Stetigkeit von g), und

$$\frac{g(\tilde{y}) - g(y)}{\tilde{y} - y} = \frac{g(f(\tilde{x})) - g(f(x))}{f(\tilde{x}) - f(x)} = \frac{\tilde{x} - x}{f(\tilde{x}) - f(x)} \xrightarrow{\tilde{y}\to y} \frac{1}{f'(x)} \qquad \square$$

[25] Vgl. "Zahl und Funktion", Abschnitt 21, S. 72f

Beispiel 1: Die n-te Wurzel $w_n = p_n^{-1} : \mathbb{R}_+^* \to \mathbb{R}_+^*$ ist differenzierbar mit Ableitung $w_n'(x) = \frac{1}{p_n'(w_n(x))}$. Die Berechnung dieses Ausdrucks ist einfach, wenn wir die Ableitung von $p_n(x)$ wieder durch $p_n(x)$ ausdrücken: $p_n'(x) = n \cdot x^{n-1} = n \cdot (x^n)^{(n-1)/n} = n \cdot p_n(x)^{(n-1)/n} = n \cdot p_n(x)^{1-(1/n)}$ und damit $p_n'(w_n(x)) = n \cdot x^{1-(1/n)}$, also erhalten wir

$$w_n'(x) = \frac{1}{p_n'(w_n(x))} = \frac{1}{n \cdot x^{1-(1/n)}} = \frac{1}{n} \cdot x^{(1/n)-1} \tag{17}$$

Beispiel 2: Der Logarithmus $\ln : \mathbb{R}_+^* \to \mathbb{R}$ ist differenzierbar mit Ableitung $\ln'(x) = \frac{1}{\exp'(\ln(x))}$. Wir haben bereits gesehen (Beispiel (4) auf S. 17), dass die e-Funktion mit ihrer Ableitung identisch ist, $\exp' = \exp$, also erhalten wir

$$\ln'(x) = \frac{1}{\exp'(\ln(x))} = \frac{1}{\exp(\ln(x))} = \frac{1}{x} \tag{18}$$

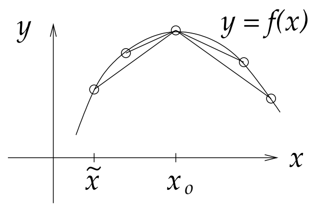
8. Maxima und Minima

Das *Maximum* einer Funktion $f : D \to \mathbb{R}$ ist der größte Wert, den f annehmen kann, d.h. der Wert $f(x_o)$ eines Punktes $x_o \in D$ mit der Eigenschaft

$$\forall_{x \in D} \quad f(x) \leq f(x_o). \tag{19}$$

Unglücklicherweise wird oft auch die Stelle x_o, an dem der Maximalwert $f(x_o)$ angenommen wird, als Maximum bezeichnet; besser sollte man x_o als *Maximalstelle* bezeichnen. Ganz analog ist das *Minimum* als der kleinste angenommene Wert definiert. Wenn D ein offenes Intervall und f differenzierbar ist, kann man Maximal- und Minimalstellen von f mit Hilfe der Ableitung f' entdecken: Sie befinden sich unter den Nullstellen von f':

Satz 8.1. *Für eine differenzierbare Funktion $f : D \to \mathbb{R}$ auf einem offenen Intervall $D \subset \mathbb{R}$ gilt: Wenn $x_o \in D$ eine Maximal- oder Minimalstelle von f ist, dann ist $f'(x_o) = 0$.*



Beweis. Wenn x_o Maximalstelle ist, so ist $f(\tilde{x}) - f(x_o) \leq 0$ für alle $\tilde{x} \in D$. Das Vorzeichen des Differenzenquotienten $\frac{f(\tilde{x})-f(x_o)}{\tilde{x}-x_o}$ hängt daher nur vom Nenner ab, der ≥ 0 ist falls $\tilde{x} < x_o$ und ≤ 0 falls $\tilde{x} > x_o$. Der Limes für $\tilde{x} \to x$ muss daher zugleich ≥ 0 und ≤ 0 sein, je nachdem, ob wir Folgen $\tilde{x}_n \to x_o$ mit $\tilde{x}_n < x_o$ für alle $n \in \mathbb{N}$ und solche mit $\tilde{x}_n > x_o$ für alle $n \in \mathbb{N}$ einsetzen. Also ist $f'(x_o) = \lim_{\tilde{x} \to x_o} \frac{f(\tilde{x})-f(x_o)}{\tilde{x}-x_o} = 0$. Für Minimalstellen führt man ein analoges Argument, oder man geht zu $-f$ über; die Minimalstellen von f sind die Maximalstellen von $-f$. □

Beispiel 1: $f : (0, \infty) \to \mathbb{R}$, $f(x) = \frac{x}{1+x^2}$. Dann ist

$$f'(x) = \frac{1 + x^2 - x \cdot 2x}{(1 + x^2)^2} = \frac{1 - x^2}{(1 + x^2)^2}$$

und die einzige Nullstelle von f' ist da, wo $1 - x^2 = 0$, also $x = 1$. Das Maximum ist demnach $f(1) = \frac{1}{2}$.[26]

Beispiel 2: $f : (0, \infty) \to \mathbb{R}$, $f(x) = x^3 - 3x$. Dann ist $f'(x) = 3x^2 - 3$, und die einzige Nullstelle von f' ist da, wo $3x^2 = 3$, also $x = 1$ ist. Das Maximum sollte $f(1) = 1 - 3 = -2$ sein. **Aber das ist nicht wahr**, denn z.B. ist $f(2) = 8 - 6 = 2 > f(1)$. Was haben wir falsch gemacht? Nach dem vorigen Satz müsste doch das Maximum unter den Nullstellen von f' zu finden sein! Aber wenn es nun gar kein Maximum gibt? Eben das ist hier der Fall: die Funktion besitzt kein Maximum; sie wächst über alle Grenzen![27]

Wie ist es möglich, dass die Wertemenge $W = \{f(x);\ x \in D\}$ kein größtes Element besitzt? Unter endlich vielen Zahlen findet man immer eine größte (man vergleiche die größere der ersten beiden Zahlen mit der dritten, die größte der ersten drei Zahlen mit der vierten usw.), doch unter unendlich vielen Zahlen braucht es keine größte zu geben (man kommt mit dem Vergleichen dann nie zum Ende). Aber es gibt einen Ersatz, das *Supremum* (vgl. "Zahl und Funktion", Abschnitt 12, S. 40 ff). Zur Erinnerung: Eine Zahl s heißt *obere Schranke* für W, wenn $s \geq w$ für alle $w \in W$; Kurzschreibweise: $s \geq W$. Wenn $W \neq \emptyset$ überhaupt eine obere Schranke besitzt, also nicht über alle Grenzen

[26] Das sieht man auch anders: $1/f(x) = (1 + x^2)/x = x + \frac{1}{x}$ ist zweimal das *arithmetische Mittel* der Zahlen x und $\frac{1}{x}$, und letzteres ist größer als ihr *geometrische Mittel* $(x \cdot \frac{1}{x})^{1/2} = 1$, wenn $x \neq \frac{1}{x}$ (vgl. "Zahl und Funktion", S. 3). Also ist $1/f(x) \geq 2$. Somit ist $f(x) \leq \frac{1}{2} = f(1)$.

[27] In einem Kriminalroman weisen alle Indizien auf einen Mann namens Egon als Mörder hin. Die Polizei kann ausschließen, dass ein anderer die Tat begangen haben kann. Und dennoch ist Egon unschuldig - woran kann das liegen? (Lösung: llafnu nie raw se, drom neniek bag se)

wächst, dann gibt es unter den oberen Schranken von W auch eine kleinste ("Zahl und Funktion", Satz 12.1); diese kleinste obere Schranke heißt *Supremum* von W, geschrieben $\sup W$.[28] Wenn W gar keine obere Schranke besitzt, also *unbeschränkt* ist, dann setzt man $\sup W = \infty$. In diesem Sinne besitzt also jede Teilmenge $W \subset \mathbb{R}$ ein Supremum in $\mathbb{R} \cup \{\infty\}$. Dieses werden wir jetzt benutzen, um in vielen Fällen doch ein Maximum zu konstruieren. (Ganz entsprechend besitzt W eine größte untere Schranke in $\mathbb{R} \cup \{-\infty\}$, das *Infimum* von W, geschrieben $\inf W$.)

Satz 8.2. *Ist $D = [a, b]$ ein beschränktes abgeschlossenes Intervall und $f : D \to \mathbb{R}$ stetig, so besitzt f ein Maximum und ein Minimum, d.h. die Wertemenge $W = f(D) = \{f(x);\ x \in D\}$ besitzt ein größtes und ein kleinstes Element.*

Beweis. Es sei $s = \sup W \in \mathbb{R} \cup \{\infty\}$. Da s kleinste obere Schranke von W ist, gibt es keine "Lücke" zwischen s und W, d.h. es gibt eine Folge (w_n) in W mit $w_n \to s$.[29] Da jedes $w_n \in W$ liegt, gibt es dazu $x_n \in D = [a, b]$ mit $w_n = f(x_n)$. Die Folge (x_n) in $[a, b]$ ist beschränkt und besitzt daher eine konvergente Teilfolge $x_{n_k} \to x_o \in [a, b]$ (Satz von Bolzano und Weierstraß, "Zahl und Funktion", Satz 11.4, S. 39). Aus der Stetigkeit von f erhalten wir $f(x_{n_k}) \to f(x_o)$, aber andererseits gilt auch $f(x_{n_k}) = w_{n_k} \to s$, also folgt $s = f(x_o) \in W$. Damit hat sich das Supremum als Maximum entpuppt; insbesondere ist $s \neq \infty$. □

Bemerkung: Der Beweis des Satzes 8.2 ist *nicht konstruktiv*; kein Computer könnte die Folge $(f(x_n))$ konstruieren und so das Maximum finden. Er könnte zwar versuchen, immer größere Werte $f(x_1) < f(x_2) < f(x_3) \dots$ zu finden, aber er kommt damit dem Maximum vielleicht niemals auch nur nahe. Finden lässt es sich aber mit Satz 8.1, indem ich mir die Werte von f in allen Nullstellen der Ableitung f' ansehe und davon den größten nehme. Satz 8.2 wiederum garantiert

[28] Zur Erinnerung: Man konstruiert $s = \sup W$ als Limes einer konvergenten Intervallschachtelung (I_k). Man beginnt mit $I_1 = [w_1, s_1]$ für ein Element $w_1 \in W$ und eine obere Schranke $s_1 \geq w$. Nun halbiert man dieses Intervall. Für den Mittelpunkt $m_1 = \frac{1}{2}(s_1 + w_1)$ gibt es zwei Möglichkeiten: Entweder $m_1 \geq W$, dann setzen wir $I_2 = [w_1, m_1]$, oder es gibt ein Element $w_2 \in W$ mit $w_2 > m_1$, dann setzen wir $I_2 = [w_2, s_2]$. In beiden Fällen ist I_2 höchstens halb so lang wie I_1. Jetzt halbiert man I_2 und definiert I_3 ganz analog, usw. Der Limes dieser Intervallschachtelung (Vollständigkeitsaxiom!) ist eine obere Schranke $s \geq W$, die gleichzeitig Limes einer Folge in W ist; damit kann es keine noch kleinere obere Schranke geben.

[29] Zu jedem $n \in \mathbb{N}$ gibt es ein $w_n \in W$ mit $w_n > s - \frac{1}{n}$, sonst wäre ja $s - \frac{1}{n}$ eine kleinere obere Schranke. Da außerdem $w_n \leq s$, folgt $|w_n - s| < \frac{1}{n} \to 0$, also $w_n \to s$.

mir, dass dieses Verfahren Erfolg hat, vgl. Beispiel 2. Die beiden Sätze bilden eine Art Symbiose; nur zusammen sind sie von Nutzen:

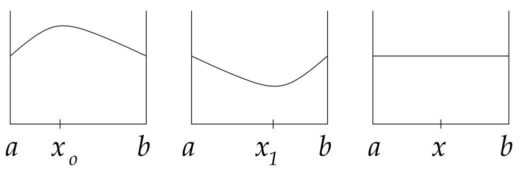
Satz 8.3. *Gegeben sei eine stetige Funktion $f : [a, b] \to \mathbb{R}$, die auf dem offenen Intervall (a, b) differenzierbar ist. Dann besitzt f ein Maximum und ein Minimum. Um diese Werte zu finden, berechnet man die Randwerte $f(a)$ und $f(b)$ sowie die Werte von f in den Nullstellen von f', d.h. man bestimmt alle $x_i \in (a, b)$, $i = 1, \ldots, k$, mit $f'(x_i) = 0$ und berechnet dort den Wert $f(x_i)$. Das Maximum ist der größte, das Minimum der kleinste der $k + 2$ Werte $f(a), f(b), f(x_i)$, $i = 1, \ldots, k$.*

Bemerkung: Der Satz gilt auch noch für $a = -\infty$ oder $b = \infty$, falls $\lim\limits_{x\to-\infty} f(x)$ oder $\lim\limits_{x\to\infty} f(x)$ existiert. Wenn die Funktion nach oben (oder nach unten) unbeschränkt ist, gibt es kein Maximum (Minimum).

9. Mittelwertsatz, Monotonie, Umkehrfunktion

Wir erinnern uns an den "Hauptsatz der Differential- und Integralrechnung", mit dem der Flächeninhalt unter den Graphen einer Funktion $f : [a, b] \to \mathbb{R}$ berechnet werden konnte: Der Flächeninhalt $F(x)$ unter dem Graphen von $f|_{[a,x]}$ war als Funktion von x differenzierbar mit Ableitung $f(x)$, d.h. $F(x)$ war eine *Stammfunktion* von $f(x)$. Um ihn wirklich anwenden zu können, müssen wir noch wissen, dass es zu gegebener Funktion f im Wesentlichen nur *eine* Stammfunktion gibt (bis auf eine additive Konstante). Dies können wir mit den folgenden ganz unschuldig anmutenden Sätzen zeigen:

Satz 9.1. Satz von Rolle:[30] *Ist $f : [a, b] \to \mathbb{R}$ stetig mit $f(a) = f(b)$, und ist f differenzierbar auf (a, b), dann gibt es einen Punkt $x \in (a, b)$ mit $f'(x) = 0$.*

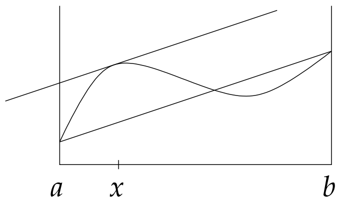


Beweis. Auf dem beschränkten abgeschlossenen Intervall $[a, b]$ nimmt f ein Maximum $f(x_o)$ und ein Minimum $f(x_1)$ an. Wenn die Maximalstelle oder die Minimalstelle im Inneren liegt, $x_o \in (a, b)$ oder $x_1 \in (a, b)$, dann gilt $f'(x_o) = 0$ oder $f'(x_1) = 0$ nach Satz 8.1 und wir sind fertig. Andernfalls liegen Maximal- und Minimalstelle beide am Rand, also $x_o, x_1 \in \{a, b\}$. Da $f(a) = f(b)$, haben dann Maximum und

[30]Michel Rolle, 1652 (Ambert, Auvergne, Frankreich) - 1719 (Paris)

Minimum den gleichen Wert und damit ist die Funktion konstant. Also gilt $f'(x) = 0$ sogar für *alle* $x \in (a, b)$. □

Satz 9.2. Mittelwertsatz: *("Sekantensteigung = Tangentensteigung")*



Ist $f : [a, b] \to \mathbb{R}$ stetig und auf (a, b) differenzierbar, dann gibt es eine Stelle $x \in (a, b)$ mit

$$f'(x) = \frac{f(b) - f(a)}{b - a} \tag{20}$$

Beweis. Im Fall $f(a) = f(b)$ ist dies offensichtlich der Satz von Rolle. Wir führen den allgemeinen Fall auf diesen Spezialfall zurück, indem wir von f die "Sekante" abziehen; das ist die affine Funktion g mit $g(a) = f(a)$ und $g(b) = f(b)$, also

$$g(x) = f(a) + \frac{f(b) - f(a)}{b - a}(x - a)$$

Die Differenzfunktion $\tilde{f} = f - g$ erfüllt $\tilde{f}(a) = \tilde{f}(b) = 0$ und damit die Voraussetzungen des Satzes von Rolle. Also gibt es ein $x \in (a, b)$ mit $\tilde{f}'(x) = 0$. Andererseits ist $\tilde{f}'(x) = f'(x) - g'(x) = f'(x) - \frac{f(b)-f(a)}{b-a}$, und damit folgt die Behauptung. □

Eine auf einem Intervall $D \subset \mathbb{R}$ definierte Funktion $f : D \to \mathbb{R}$ heißt *monoton steigend* (*fallend*), wenn f *schwache Ungleichungen* erhält (umdreht), d.h. wenn für alle $x_1, x_2 \in D$ gilt:

$$x_2 \geq x_1 \Rightarrow f(x_2) \overset{(\leq)}{\geq} f(x_1) \tag{21}$$

und *streng monoton steigend* (*fallend*), wenn f sogar *starke Ungleichungen* erhält (umdreht):

$$x_2 > x_1 \Rightarrow f(x_2) \overset{(<)}{>} f(x_1). \tag{22}$$

Satz 9.3. *Es sei D offen und $f : D \to \mathbb{R}$ differenzierbar. Dann gilt:*

a) $f' = 0 \iff f = const,$

b) $f' \overset{(\leq)}{\geq} 0 \iff f$ *monoton wachsend (fallend),*

c) $f' \overset{(<)}{>} 0 \Rightarrow f$ *streng monoton wachsend (fallend).*

Beweis. Dies folgt aus dem Mittelwertsatzes 9.2 und der Definition der Ableitung als Limes des Differenzenquotienten: Es seien $x_1, x_2 \in D$ mit $x_2 > x_1$ gegeben. Nach dem Mittelwertsatz gibt es $x \in (x_1, x_2)$ mit $\frac{f(x_2)-f(x_1)}{x_2-x_1} = f'(x)$. Damit gilt:

$$f'(x) \geq 0 \ (= 0, > 0) \Rightarrow f(x_2) - f(x_1) \geq 0 \ (= 0, > 0),$$

was die Monotonie (bzw. Konstanz bzw. strenge Monotonie) beweist. Ist umgekehrt die Monotonie (Konstanz) von f vorausgesetzt und sind $x, x+h \in D$ mit $h > 0$, so gilt $f(x+h) \geq f(x)$ $(= f(x))$ und damit $\frac{f(x+h)-f(x)}{h} \geq 0$ $(= 0)$. Da dies für *jedes* (genügend kleine) $h > 0$ gilt, folgt auch $f'(x) = \lim_{h\to 0} \frac{f(x+h)-f(x)}{h} \geq 0$ $(= 0)$. [31] □

Beispiel: Die Funktionen $\sin, \cos : \mathbb{R} \to \mathbb{R}$,

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - + \dots, \quad \cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - + \dots,$$

sind differenzierbar mit Ableitung

$$\sin' = \cos, \quad \cos' = -\sin, \tag{23}$$

denn nach den Additionstheoremen (vgl. "Zahl und Funktion", S. 70, (77) und (78)) gilt

$$\begin{aligned} \sin(x+h) &= \sin x \cos h + \cos x \sin h \\ \cos(x+h) &= \cos x \cos h - \sin x \sin h \end{aligned}$$

und damit

$$\begin{aligned} \frac{\sin(x+h) - \sin(x)}{h} &= \sin x \frac{\cos(h) - 1}{h} + \cos x \frac{\sin h}{h} \to \cos x, \\ \frac{\cos(x+h) - \cos(x)}{h} &= \cos x \frac{\cos(h) - 1}{h} - \sin x \frac{\sin h}{h} \to -\sin x, \end{aligned}$$

denn

$$\begin{aligned} \frac{\cos h - 1}{h} &= -\frac{h}{2!} + \frac{h^3}{4!} - + \dots \stackrel{h\to 0}{\longrightarrow} 0 \\ \frac{\sin h}{h} &= 1 - \frac{h^2}{3!} + \frac{h^4}{5!} - + \dots \stackrel{h\to 0}{\longrightarrow} 1 \end{aligned}$$

[31] *Schwache Ungleichungen* bleiben im Limes erhalten: Ist (a_n) eine Folge mit $a_n \to a$ und $a_n \geq 0$ für alle $n \in \mathbb{N}$, so ist auch $a \geq 0$, denn wäre $a = -\epsilon < 0$, so wäre $|a_n - a| \geq \epsilon$ für alle n im Widerspruch zur Konvergenz. Für *starke Ungleichungen* gilt dies nicht; zum Beispiel ist $\frac{1}{n} > 0$ für alle $n \in \mathbb{N}$, aber $\lim_{n\to\infty} \frac{1}{n} = 0 \not> 0$.

(konvergente Potenzreihen sind stetig, vgl. "Zahl und Funktion", S. 63, Satz 19.3). Die ersten Nullstellen[32] von cos liegen bei $\pm\frac{\pi}{2}$, also ist die *Tangensfunktion*

$$\tan x = \frac{\sin x}{\cos x}$$

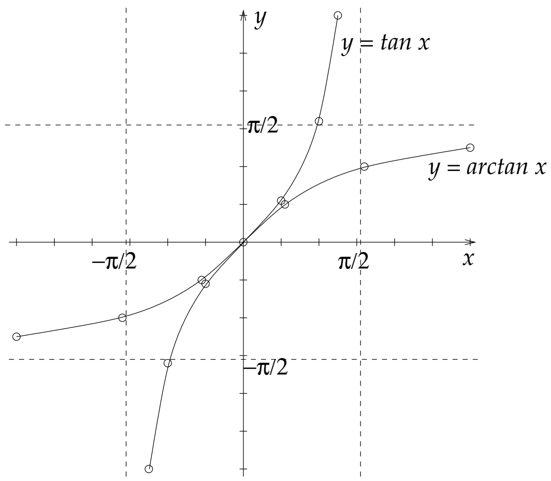
auf dem Intervall $(-\frac{\pi}{2}, \frac{\pi}{2})$ definiert. Die Ableitung ist nach Quotientenregel

$$\tan' = \left(\frac{\sin}{\cos}\right)' = \frac{\sin'\cos - \sin\cos'}{\cos^2} = \frac{\cos^2 + \sin^2}{\cos^2}$$

und der letzte Ausdruck kann auf zwei Weisen ausgedrückt werden: Entweder man nutzt $\cos^2 + \sin^2 = 1$ oder man multipliziert die Summe aus: $\frac{\cos^2+\sin^2}{\cos^2} = \frac{\cos^2}{\cos^2} + \frac{\sin^2}{\cos^2} = 1 + \tan^2$. Je nachdem erhält man

$$\tan' = \frac{1}{\cos^2}\;, \quad \tan' = 1 + \tan^2\,. \tag{24}$$

Aus beiden Gleichungen sieht man $\tan' > 0$, also ist tan streng monoton steigend. Der Wertebereich[33] ist ganz $\mathbb{R}$, denn $\lim\limits_{x\to\pi/2} \tan x = \infty$ und $\lim\limits_{x\to-\pi/2} \tan x = -\infty$, da $\sin x \to \pm 1$ und $\cos x \to 0$ für $x \to \pm\frac{\pi}{2}$. Die Funktion $\tan : (-\frac{\pi}{2}, \frac{\pi}{2}) \to \mathbb{R}$ ist also umkehrbar (bijektiv).



[32] Vgl. "Zahl und Funktion", S. 69f, besonders die Figur auf S. 70 oben: Die komplexe Zahl $\cos t + i\sin t = e^{it} \in \mathbb{C} = \mathbb{R}^2$ ist der Punkt auf dem Einheitskreis mit Winkel t zur positiven x-Achse. Beim Winkel $t = \pm\frac{\pi}{2} = \pm 90^o$ liegt e^{it} zum ersten Mal auf der y-Achse, d.h. $e^{\pm i\pi/2} = \pm i$. Für den Realteil (die x-Koordinate) $\cos t$ von e^{it} gilt also: $\cos t > 0$ für $t \in (-\frac{\pi}{2}, \frac{\pi}{2})$ und $\cos \pm\frac{\pi}{2} = 0$, während $\sin \pm\frac{\pi}{2} = \pm 1$.

[33] Um zu zeigen, dass wirklich alle Zahlen zwischen $-\infty$ und ∞ als Werte der Tangensfunktion vorkommen, verwenden wir den *Zwischenwertsatz*; vgl "Zahl und Funktion" S. 73f

Nach Satz 7.3 ist die Umkehrfunktion[34] $\tan^{-1} = \arctan$ wieder differenzierbar, da $\tan'(x) \neq 0$ für alle $x \in (-\frac{\pi}{2}, \frac{\pi}{2})$, und für die Ableitung gilt:

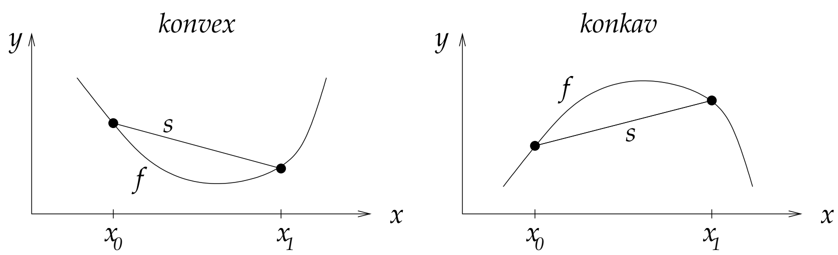
$$\arctan'(x) = \frac{1}{\tan'(\arctan x)} \overset{(24)}{=} \frac{1}{1 + \tan(\arctan x)^2} = \frac{1}{1+x^2} \tag{25}$$

Die schwierige Arcus-Tangens-Funktion hat also eine erstaunlich einfache Ableitung, oder umgekehrt: Die Stammfunktion der harmlosen rationalen Funktion $\frac{1}{1+x^2}$ ist erstaunlich schwierig!

Der Satz 9.3 hilft auch, Maxima und Minima einer Funktion zu entdecken: Wenn eine Funktion f vor einer Stelle x_o (für $x \leq x_o$) monoton steigt und danach (für $x \geq x_o$) monoton fällt, dann ist $f(x_o)$ offensichtlich ein Maximum. Wir erhalten also:

Satz 9.4. *Ist D ein offenes Intervall, $x_o \in D$ und $f : D \to \mathbb{R}$ eine differenzierbare Funktion mit $f'(x) \geq 0$ für $x \leq 0$ und $f'(x) \leq 0$ für $x \geq 0$, dann ist $f(x_o)$ ein Maximum von f.* □

10. Konvexität



Eine auf einem Intervall D definierte Funktion $f : D \to \mathbb{R}$ heißt *konvex* bzw. *konkav*, wenn ihr Graph zwischen je zwei Punkten immer unterhalb bzw. immer oberhalb seiner Sekante liegt. Die beiden x-Werte mögen x_0, x_1 heißen (mit $x_0 < x_1$). Die *Sekante* ist die Gerade durch zwei auf dem Graphen von f liegenden Punkte, $(x_0, f(x_0))$ und $(x_1, f(x_1))$; sie ist also der Graph der affinen Funktion

$$s(x) = f(x_0) + a \cdot (x - x_0), \quad a := \frac{f(x_1) - f(x_0)}{x_1 - x_0}.$$

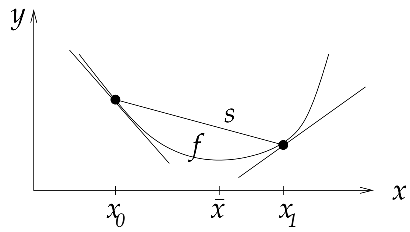
Damit ist f *konvex* genau dann, wenn $f(x) \leq s(x)$ für alle $x \in [x_0, x_1]$, d.h. für alle $x \in [x_0, x_1]$ gilt

$$f(x) \leq f(x_0) + a \cdot (x - x_0), \quad a := \frac{f(x_1) - f(x_0)}{x_1 - x_0}. \tag{26}$$

[34]Das ist die *Arcus-Tangens-Funktion*, die dem Tangens eines Winkels wieder den Winkel selbst ("arcus" = Bogen) zuordnet.

Bei konkaven Funktionen ist "$\leq$" durch "$\geq$" zu ersetzen.

Satz 10.1. *Es sei D ein offenes Intervall und $f : D \to \mathbb{R}$ differenzierbar mit Ableitung $f' : D \to \mathbb{R}$. Dann gilt:*
f ist konvex (konkav) $\iff$ f' ist monoton steigend (fallend).



Beweis. "$\Leftarrow$": $f' \nearrow \iff (f-s)' \nearrow$, denn $(f-s)' = f' - s'$ und $s' = a = const$. Weil $f-s$ an den Punkten x_0 und x_1 eine Nullstelle besitzt, gibt es nach dem Satz von Rolle zwischen x_0 und x_1 eine Nullstelle der Ableitung, also ein $\bar{x} \in (x_0, x_1)$ mit $(f-s)'(\bar{x}) = 0$. Wegen der Monotonie gilt

$$(f-s)'(x) \begin{cases} \leq 0 & \text{für} \quad x \leq \bar{x} \\ \geq 0 & \text{für} \quad x \geq \bar{x} \end{cases}$$

Damit ist $f-s$ selbst $\searrow$ auf $[x_0, \bar{x}]$ und $\nearrow$ auf $[\bar{x}, x_1]$, und damit ist $(f-s)(\bar{x})$ das Minimum von $(f-s)|_{[x_0,x_1]}$. Da $(f-s)$ an den Randpunkten x_0 und x_1 gleich Null ist, folgt $0 \geq (f-s)(x) \geq (f-s)(\bar{x})$ für alle $x \in [x_0, \bar{x}]$ und für alle $x \in [\bar{x}, x_1]$, also $(f-s) \leq 0$ auf $[x_0, x_1]$.
"$\Rightarrow$": Nach Voraussetzung ist $f-s \leq 0$ auf $[x_0, x_1]$ und $f-s = 0$ an den Randpunkten x_0, x_1. Deshalb gilt $(f-s)'(x_0) \leq 0$ und $(f-s)'(x_1) \geq 0$ (Vorzeichen des Differenzenquotienten!) und damit $f'(x_0) \leq s' \leq f'(x_1)$. Da x_0, x_1 beliebige Elemente von D mit $x_0 < x_1$ sind, haben wir damit die Monotonie von f' bewiesen.

Beispiel: Die e-Funktion $\exp : \mathbb{R} \to \mathbb{R}_+^*$, $\exp(x) = e^x$ ist konvex, denn exp ist (streng) monoton steigend, weil $\exp' = \exp > 0$, und deshalb ist auch $\exp' = \exp$ monoton steigend.

Satz 10.2. *Ist $f : D \to \mathbb{R}$ konvex und sind $x_1, \ldots, x_n \in D$ und $\lambda_1, \ldots, \lambda_n \in [0,1]$ mit $\lambda_1 + \cdots + \lambda_n = 1$, dann gilt*

$$f(\sum_{j=1}^{n} \lambda_j x_j) \leq \sum_{j=1}^{n} \lambda_j f(x_j). \tag{27}$$

Beweis. Für <u>$n = 1$</u> ist nichts zu zeigen: $f(x_1) \leq f(x_1)$.
<u>$n = 2$</u>: Jedes $x \in [x_0, x_1]$ ist von der Form

$$x = x_0 + \lambda \cdot (x_1 - x_0) = (1-\lambda)x_0 + \lambda x_1 \tag{28}$$

für ein $\lambda \in [0,1]$; für so ein x ist $x - x_0 = \lambda(x_1 - x_0)$. Also ist (26) äquivalent zu

$$f((1-\lambda)x_0 + \lambda x_1) \leq (1-\lambda)f(x_0) + \lambda f(x_1) \tag{29}$$

für alle $\lambda \in [0,1]$, denn $a \cdot (x - x_0) = \lambda \cdot (f(x_1) - f(x_0))$,

Induktionsschritt $\underline{n \to n+1}$: Gegeben seien Punkte $x_1, \ldots, x_{n+1} \in D$ und Zahlen $\lambda_1, \ldots, \lambda_{n+1} \in [0,1]$ mit $\lambda_1 + \cdots + \lambda_n + \lambda_{n+1} = 1$. Zu zeigen ist $f\left(\sum_{j=1}^{n+1} \lambda_j x_j\right) \leq \sum_{j=1}^{n+1} \lambda_j f(x_j)$. Wenn wir den letzten Punkt x_{n+1} weglassen, können wir die Induktionsvoraussetzung benutzen; allerdings ist die Summe der verbleibenden λ_j nicht Eins, sondern es gilt $\lambda_1 + \cdots + \lambda_n = 1 - \lambda_{n+1}$. Deshalb setzen wir

$$\lambda'_j := \frac{\lambda_j}{1 - \lambda_{n+1}}\,,$$

dann ist $\lambda'_1 + \cdots + \lambda'_n = 1$ und nach Induktionsvoraussetzung gilt

$$f(\sum_{j=1}^{n} \lambda'_j x_j) \leq \sum_{j=1}^{n} \lambda'_j f(x_j). \tag{30}$$

Setzen wir $\sum\limits_{j=1}^{n} \lambda'_j x_j =: x_0$ dann gilt

$$\begin{aligned} f(\sum_{j=1}^{n+1} \lambda_j x_j) &= f\left((1-\lambda_{n+1})x_0 + \lambda_{n+1}x_{n+1}\right) \\ &\overset{(29)}{\leq} (1-\lambda_{n+1})f(x_0) + \lambda_{n+1}f(x_{n+1}) \\ &\overset{(30)}{\leq} \left(\sum_{j=1}^{n} \lambda_j f(x_j)\right) + \lambda_{n+1}f(x_{n+1}) \end{aligned}$$

□

Beispiel: Die e-Funktion ist konvex, also gilt für $\lambda_j = 1/n$:

$$e^{(x_0+\cdots+x_n)/n} \leq (e^{x_0} + \cdots + e^{x_n})/n\,.$$

Andererseits gilt nach Potenzrechenregeln (Additionstheorem der e-Funktion):

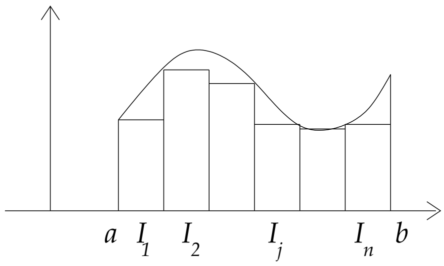
$$e^{(x_0+\cdots+x_n)/n} = (e^{x_0} \cdot \ldots \cdot e^{x_n})^{1/n}$$

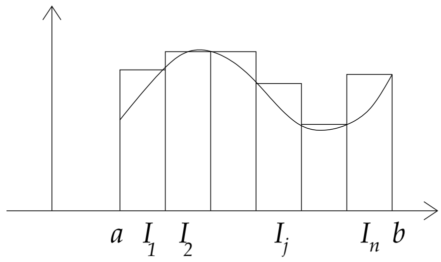
und für die Zahlen $a_j := e^{x_j} > 0$ erhalten wir die folgende allgemeine *Ungleichung zwischen Geometrischem und Arithmetischem Mittel* (vgl. “Zahl und Funktion”, S.3):

$$\sqrt[n]{a_1 \cdot \ldots \cdot a_n} \leq \frac{1}{n}(a_1 + \cdots + a_n)\,. \tag{31}$$

11. Integration

In Abschnitt 4 haben wir die Fläche unter dem Graphen einer positiven stetigen Funktion $f : [a,b] \to \mathbb{R}$ bestimmt. Diesen Flächeninhalt nennen wir das *Integral* von f und schreiben dafür $\int_a^b f$ oder $\int_a^b f(x)dx$. Das ist eine symbolische Schreibweise. Das Symbol $\int$ erinnert an "Summe", unter dx können wir uns eine kleine Differenz von x-Werten denken, unter $f(x)dx$ also den Flächeninhalt des Rechtecks der Breite dx und Höhe $f(x)$. Wir hatten diesen Flächeninhalt von oben und unten durch die *Obersumme* und die *Untersumme* eingeschachtelt:





Wir brauchen dazu eine *Zerlegung* des Intervalls $I = [a,b]$ in kleinere Intervalle $I_1, \dots, I_n$, wobei $I_j = x_j - x_{j-1}$ mit $x_0 = a < x_1 < \dots < x_n = b$. Dann sind Untersumme und Obersumme die Summe der Rechteckflächen mit Breite $h_j = x_j - x_{j-1} = \text{Länge}(I_j)$ und Höhe $\min\limits_{I_j} f$ bzw. $\max\limits_{I_j} f$ (Maximum und Minimum der Werte $f(x)$ mit $x \in I_j$) und :

$$\text{Untersumme} = \sum_{j=1}^{n} h_j \min_{I_j} f\,, \quad \text{Obersumme} = \sum_{j=1}^{n} h_j \max_{I_j} f\,.$$

Das Integral wird durch diese Einschachtelung definiert:

$$\text{Untersumme} \le \int_a^b f(x)dx \le \text{Obersumme} \tag{32}$$

Die unteren und oberen Schranken bilden eine konvergente Intervallschachtelung und definieren damit die Zahl $\int_a^b f(x)dx$, wenn ihre Differenz Obersumme $-$ Untersumme durch Verfeinerung der Unterteilung[35] beliebig klein gemacht werden kann. Eine Funktion, für die das gelingt, heißt *integrierbar*; diese Eigenschaft ist zum Beispiel dann erfüllt, wenn f stetig ist.[36]

[35] Auf einem kleineren Intervall wird das Minimus größer, das Maximum kleiner, die Werte von Unter- und Obersumme rücken also näher zusammen.

[36] Man benötigt dazu den etwas stärkeren Begriff der *gleichmäßigen Stetigkeit* von f, die aber auf dem beschränkten abgeschlossenen Intervall $[a,b]$ bereits aus

Sind zwei positive stetige Funktionen f, g auf $[a, b]$ gegeben, so gilt $\int_a^b (f+g) = \int_a^b f + \int_a^b g$ (die Rechtecke für f und g werden einfach aufeinander getürmt) Besonders einfach zu sehen ist es für $g = c = const$; dann unterscheidet sich die Fläche unter $f + c$ von der unter f durch das Rechteck mit Breite $b - a$ und Höhe c. Die gleiche Regel funktioniert auch noch für $f - c$, solange diese Funktion $f - c$ oberhalb der x-Achse liegt, also $f \geq c$ gilt (ebenso für $f - g$). Wollen wir diese Einschränkung ("solange ...") loswerden, müssen wir unsern Integralbegriff erweitern auf solche Funktionen, die nicht mehr ganz oberhalb der x-Achse liegen. Anschaulich gesprochen werden die Flächenteile unterhalb der x-Achse abgezogen. Viel einfacher ist es aber, die Definition (32) wörtlich beizubehalten; dann darf eben $\max_{I_j} f$ oder $\min_{I_j}$ auch negativ sein. Damit wird die Regel $\int_a^b (f+g) = \int_a^b f + \int_a^b g$ ohne Einschränkungen für das Vorzeichen von $f, g, f+g$ gültig. Wir sehen also, dass die Rechenregeln uns dazu zwingen, unsere Begriffe zu erweitern, damit sie unbeschränkte Gültigkeit erlangen; das war bereits bei den Zahlbereichserweiterungen so.

Der Flächeninhalt unter dem Graphen erfüllt noch eine andere Regel, die wir ebenfalls auf Integrale ausweiten wollen, die *Intervall-Additivität*: Wenn wir das Intervall $[a, b]$ unterteilen, also in zwei Intervalle $[a, c]$ und $[c, b]$ zerlegen für eine Zahl c zwischen a und b, also $c \in [a, b]$, dann gilt $\int_a^b f = \int_a^c f + \int_c^b f$ (Zerlegungsinvarianz des Flächeninhalts). Was aber passiert für $c \notin [a, b]$ (aber c noch im Definitionsbereich von f), wenn also z.B. $c < a$? In dem Fall liegt a zwischen c und b und wir haben $\int_c^b f = \int_c^a f + \int_a^b f$, mit anderen Worten $\int_a^b f = -\int_c^a f + \int_c^b f$. Wir erhalten die alte Regel zurück, wenn wir für $c < a$ setzen:

$$\int_a^c f := -\int_c^a f \tag{34}$$

---

der gewöhnlichen Stetigkeit folgt:

$$\forall_{\epsilon>0} \exists_{\delta>0} \forall_{x,\tilde{x}\in[a,b]} \; [|\tilde{x} - x| < \delta \Rightarrow |f(\tilde{x}) - f(x)| < \epsilon] \, . \tag{33}$$

(Andernfalls gäbe es nämlich $\epsilon > 0$ und Folgen (x_n) und $(\tilde{x}_n)$ in $[a, b]$ mit der Eigenschaft $|\tilde{x}_n - x_n| < 1/n \to 0$, aber $|f(\tilde{x}_n) - f(x_n)| \geq \epsilon$ für alle n. Nach Übergang zu einer Teilfolge können wir annehmen, dass (x_n) konvergiert, $x_n \to x_o$, also auch $\tilde{x}_n \to x_o$ und damit nach Stetigkeit $f(x_n), f(\tilde{x}_n) \to f(x_o)$ und insbesondere $|f(\tilde{x}_n) - f(x_n)| \to 0$, Widerspruch!)
Wenn nun jeder der Teilintervalle von $[a, b]$ eine Länge $< \delta$ hat, unterscheiden sich Maximum und Minimum von f auf dem Teilabschnitt um weniger als ϵ, und damit ist die Differenz zwischen der oberen und der unteren Schranke von F nach (33) weniger als $\epsilon \cdot (b - a)$, wobei ϵ beliebig klein ist.

(Integrieren "von rechts nach links" dreht das Vorzeichen um). Auf doppelte Weise kommen wir also von den (immer positiven) Flächeninhalten zu Zahlen mit Vorzeichen: Die Flächenteile unterhalb der x-Achse werden negativ gezählt, und bei Vertauschen der Intervallgrenzen dreht sich das Vorzeichen um.

Mit dieser Erweiterung erhalten wir die folgenden Rechenregeln für das Integral, Fortsetzungen der entsprechenden Regeln für den Flächeninhalt und sofort aus der Definition (32) ableitbar:

(1) Linearität: $\int_a^b (f+g) = \int_a^b f + \int_a^b g$, $\int_a^b cf = c\int_a^b f$,
(2) Monotonie: $f \le g \Rightarrow \int_a^b f \le \int_a^b g$,
(3) Intervall-Additivität $\int_a^b f = \int_a^c f + \int_c^b f$, $\int_a^a f = 0$,
(4) Konstanten: $\int_a^b c = c(b-a)$.

Satz 11.1. Hauptsatz der Differential- und Integralrechnung: *Ist D ein offenes Intervall, $f : D \to \mathbb{R}$ eine stetige Funktion und $a, b \in D$, so gilt*

$$\int_a^b f := \int_a^b f(x)dx = \tilde{F}(b) - \tilde{F}(a) =: \tilde{F}(x)\Big|_a^b \tag{35}$$

wobei $\tilde{F} : D \to \mathbb{R}$ eine beliebige Stammfunktion von f ist, d.h. F ist differenzierbar mit Ableitung $F' = f$.

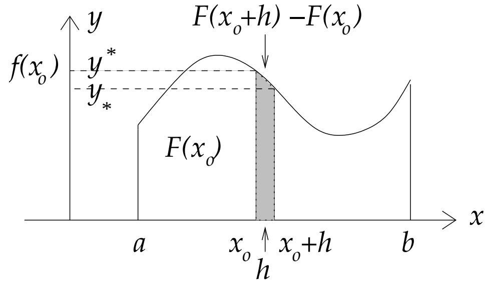
Beweis. (Vgl. S. 12) Wir definieren eine Funktion

$$F : D \to \mathbb{R}, \quad F(x_o) = \int_a^{x_o} f(x)dx.$$

Für jedes $x_o \in D$ gilt nach Intervall-Additivität (3):

$$F(x_o + h) - F(x_o) = \int_{x_o}^{x_o+h} f(x)dx,$$

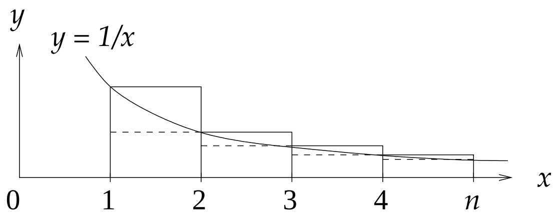
und nach (2),(4) liegt dieses Integral zwischen den Werten $y_* \cdot h$ und $y^* \cdot h$, wobei y^* das Maximum und y_* das Minimum von f im Intervall zwischen x_o und $x_o + h$ ist. Dabei darf h positiv oder negativ sein, und $|h|$ ist klein.



Somit liegt der Differenzenquotient $\frac{F(x_o+h)-F(x_o)}{h}$ zwischen y_* und y^*. Wenn wir h gegen Null gehen lassen, dann gehen sowohl y_* als auch y^* gegen den Wert $f(x_o)$. Also ist F differenzierbar mit Ableitung $F' = f$, d.h. F ist eine Stammfunktion von f. Zudem gilt $F(a) = 0$ nach (3).

Ist nun $\tilde{F}$ eine beliebige Stammfunktion von f, die gar nichts mit dem Flächeninhalt zu tun hat, sondern die wir uns irgendwie verschafft haben, so ist $(\tilde{F} - F)' = \tilde{F}' - F' = f - f = 0$ und nach Satz 9.3 folgt $\tilde{F} - F = c$ für eine Konstante $c \in \mathbb{R}$, also $\tilde{F} = F + c$. Ausgewertet für $x = a$ ergibt diese Gleichung $\tilde{F}(a) = c$, da $F(a) = 0$. Somit ist $F(x) = \tilde{F}(x) - \tilde{F}(a)$ für alle $x \in D$, und speziell für $x = b$ folgt die Behauptung. □

Beispiel: $\int_1^b \frac{1}{x}\,dx = \ln x|_1^b = \ln b$.



Dieses Integral ist eng mit der *harmonischen Reihe* $\sum \frac{1}{k} = 1+\frac{1}{2}+\frac{1}{3}+\dots$ verwandt; in der Tat zeigt die Figur

$$\frac{1}{2}+\frac{1}{3}+\dots+\frac{1}{n} \;\geq\; \int_1^n \frac{1}{x}\,dx \;\geq\; \frac{1}{1}+\frac{1}{2}+\frac{1}{3}+\dots+\frac{1}{n-1}.$$

Die linke Summe ist eine Untersumme, die rechte eine Obersumme für das Integral, wobei die unterteilenden Intervalle $I_j = [j, j+1]$ die Länge 1 haben. Weil die Differenz der beiden Summen $1 - \frac{1}{n}$ beträgt, unterscheiden sich $\sum_{k=1}^n \frac{1}{k}$ und $\int_1^n \frac{1}{x}\,dx$ um weniger als Eins. Jetzt verstehen wir, warum wir im Beweis der Divergenz der harmonischen Reihe ("Zahl und Funktion", Satz 13.1, S. 42f) die Anzahl der Summanden exponentiell wachsen lassen mussten; zum Beispiel für $b = 2^p = e^{p\cdot\ln 2}$ ist $\ln b = p \cdot \ln 2$ und daher $\sum_{k=1}^{2^p-1} \frac{1}{k} \geq \int_1^{2^p} \frac{1}{x}\,dx = p \cdot \ln 2 \to \infty$ für $p \to \infty$.

Ganz analog kann man allgemein zeigen:

Satz 11.2. Integralvergleichskriterium: *Für jede positive, monoton fallende Funktion f auf $\mathbb{R}_+$ gilt:*

$$\sum_{k=1}^{\infty} f(k) < \infty \iff \int_1^{\infty} f(x)dx < \infty \tag{36}$$

Beispiel: $\sum_{k=1}^{\infty} \frac{1}{k^\alpha}$ konvergiert genau dann, wenn $\alpha > 1$, denn die Stammfunktion von $\frac{1}{x^\alpha} = x^{-\alpha}$ für $\alpha \neq 1$ ist $\frac{1}{1-\alpha}\, x^{1-\alpha}$, und $x^{1-\alpha} = \frac{1}{x^{\alpha-1}}$ geht für $x \to \infty$ gegen 0 falls $\alpha > 1$ und gegen ∞ falls $\alpha < 1$; ebenso geht die Stammfunktion ln von $1/x$ $(\alpha = 1)$ gegen unendlich.

12. Die Kunst des Integrierens

Die Kunst des Integrierens besteht nach Satz 11.1 schlicht darin, zu einer gegebenen Funktion f eine Stammfunktion F zu finden. Wir können aus unseren bisherigen Ergebnissen bereits einen kleinen Katalog solcher Stammfunktionen zusammenstellen $(\alpha \neq -1)$:

$f(x)$	$F(x)$
x^α	$\frac{1}{\alpha+1}\, x^{\alpha+1}$
x^{-1}	$\ln x$
e^x	e^x
$\sin x$	$-\cos x$
$\cos x$	$\sin x$
$\frac{1}{1+x^2}$	$\arctan x$

Weitere Stammfunktionen gewinnen wir aus Regeln, die denen der Differentiation (Satz 7.1 und 7.2) ähneln, aber leider nicht so stark sind und keineswegs immer zum Ziel führen; wir sehen ja, dass wir bereits die Stammfunktionen von $\frac{1}{x}$ oder $\frac{1}{1+x^2}$ niemals hätten errechnen können, wenn wir nicht zufällig Funktionen mit diesen Ableitungen gekannt hätten. Viele ganz geläufige Funktionen haben Stammfunktionen, die unter den uns bisher bekannten Funktionen nicht vorkommen.[37]

Wir wollen zur Abkürzung eine Stammfunktion einer Funktion f statt mit F auch mit dem Symbol $\int f$ bezeichnen ("*unbestimmtes Integral*"). Zunächst gilt wie bei der Summenregel $\int (f+g) = \int f + \int g$ und außerdem $\int (\alpha f) = \alpha \int f$ für jede Konstante $\alpha \in \mathbb{R}$, wie man sofort durch Differenzieren sieht. Auch für die Produktregel und die Kettenregel gibt es ein Analog:

[37] Die einfachsten Funktionen dieser Art sind von der Form $f(x) = 1/\sqrt{p(x)}$, wobei $p(x)$ ein Polynom dritten oder vierten Grades mit einfachen Nullstellen ist; die zugehörigen Stammfunktionen heißen *elliptische* Funktionen. Wir erkennen hier ein allgemeines Prinzip wieder: Die Umkehroperationen erweitern den Objektbereich. Wir haben das früher schon gesehen: Die Subtraktion führte zu den negativen Zahlen, die Division zu den Brüchen, das Wurzelziehen zu den irrationalen und komplexen Zahlen. Jetzt sehen wir: Die Umkehrung der Differentiation erweitert den Bereich der Funktionen.

Satz 12.1. “Partielle Integration”: *Ist D ein offenes Intervall, $a, b \in D$ und $f, g : D \to \mathbb{R}$ differenzierbar mit stetigen Ableitungen f', g', so gilt:*

$$\begin{aligned} \int (f \cdot g') &= f \cdot g - \int (f' \cdot g), \\ \int_a^b f(x)g'(x)dx &= [f(x)g(x)]_a^b - \int_a^b f'(x)g(x)dx \end{aligned} \tag{37}$$

Beweis. Da $(f \cdot g)' = f' \cdot g + f \cdot g'$ (Produktregel), folgt $f \cdot g = \int (f \cdot g)' = \int (f' \cdot g) + \int (f \cdot g')$ und damit die Behauptung. □

Wir können mit dieser Regel leider nicht die Stammfunktion eines Produktes $f \cdot g$ aus den Stammfunktionen von f und g errechnen, wie man vielleicht erhofft hatte. Wir können lediglich die Berechnung der Stammfunktion von $f \cdot g'$ auf die von $f' \cdot g$ zurückführen; wenn dies keinen Fortschritt bringt, ist die Methode nicht anwendbar.

Beispiel 1: Was ist die Stammfunktion von $\ln x$? Wir schreiben $\ln x$ künstlich als Produkt $\ln x = (\ln x) \cdot 1$ und setzen $f(x) = \ln x$ und $g'(x) = 1$ und damit $g(x) = x$ und $f'(x) = \frac{1}{x}$. Nun ist $\int \ln = \int (fg') = fg - \int (f'g) = (\ln x) \cdot x - \int (\frac{1}{x} \cdot x) = x \cdot \ln x - \int 1 = x \cdot \ln x - x$. Probe:[38] $(x \cdot \ln x - x)' = 1 \cdot \ln x + x \cdot \frac{1}{x} - 1 = \ln x$.

Anwendung: Für $\epsilon \in (0, 1)$ ist

$$\int_\epsilon^1 \ln x \, dx = [x \cdot \ln x - x]_\epsilon^1 = -(1 - \epsilon) - \epsilon \ln \epsilon. \tag{38}$$

Was ist der Limes für $\epsilon \to 0$? Wir setzen $\epsilon = e^{-t}$ und lassen $t \to \infty$ gehen, dann ist $-\epsilon \ln \epsilon = -e^{-t} \ln(e^{-t}) = e^{-t} \cdot t = t/e^t \stackrel{t \to \infty}{\longrightarrow} 0$, denn da $e^t/t \geq \frac{t^2}{2!}/t = \frac{t}{2} \to \infty$, gilt $t/e^t \to 0$.[39] Also folgt mit (38):

$$\int_0^1 \ln x \, dx = \lim_{\epsilon \to 0} \int_\epsilon^1 \ln x \, dx = -1 \tag{39}$$

Beispiel 2: Was ist die Stammfunktion S_n von $\sin^n x := (\sin x)^n$? Wir wissen bereits $S_0 = \int \sin^0 = \int 1 = x$, $S_1 = \int \sin = -\cos$. Wenn

[38] Die Bezeichnung $(x \cdot \ln x - x)'$ ist eigentlich unüblich; stattdessen schreibt man $\frac{d}{dx}(x \cdot \ln x - x)$. Allgemein schreibt man $f'(x)$ (Ableitung der *Funktion* f an der Stelle x), aber nicht $f(x)'$, sondern stattdessen $\frac{d}{dx} f(x)$ (Differentiation des *Terms* $f(x)$ nach der Variablen x). Wir werden aber x auch als Symbol für die Funktion $\mathrm{id} : x \mapsto x$ gebrauchen; damit ist die Bezeichnung $(x \cdot \ln x - x)'$ zu rechtfertigen.

[39] Allgemeiner gilt: $e^t/t^n \to \infty$ für $t \to \infty$, weil $e^t/t^n > \frac{t^{n+1}}{(n+1)!}/t^n = \frac{t}{(n+1)!} \to \infty$. Die e-Funktion wächst in diesem Sinne schneller als jede Potenz.

wir S_n schon kennen, berechnen wir $S_{n+1} = \int(\sin \cdot \sin^n) = \int(f'g)$ mit $f' = \sin$, $f = -\cos$, $g = \sin^n$, $g' = n \cdot sin^{n-1} \cdot \cos$ wie folgt:

$$S_{n+1} = \int(f'g) = fg - \int(fg') = -\cos \cdot \sin^n + n \int(\cos^2 \cdot \sin^{n-1}).$$

Da $\cos^2 + \sin^2 = 1$, ist $\cos^2 = 1 - \sin^2$ und somit

$$\int(\cos^2 \sin^{n-1}) = \int(\sin^{n-1} - \sin^{n+1}) = S_{n-1} - S_{n+1}.$$

Daher haben wir eine Gleichung für S_{n+1} gewonnen, die wir lösen können:

$$S_{n+1} = -\cos \cdot \sin^n + nS_{n-1} - nS_{n+1}.$$

Daraus folgt $(n+1)S_{n+1} = -\cos \cdot \sin^n + nS_{n-1}$ und damit

$$S_{n+1} = \tfrac{1}{n+1}(nS_{n-1} - \cos \sin^n) \tag{40}$$

Dies ist eine Rekursionsformel, mit der wir die S_n sukzessive berechnen können: $S_0 = x$, $S_1 = -\cos$, $S_2 = \frac{1}{2}(x - \cos \sin)$ usw.

Anwendung: Wenn wir die Grenzen 0 und $\pi/2$ einsetzen, vereinfacht sich die Formel (40), da $[\cos \sin^n]_0^{\pi/2} = 0$. Für das Integral $s_n := \int_0^{\pi/2} \sin^n x \, dx = [S_n]_0^{\pi/2}$ erhalten wir aus (40) die Rekursionsformel

$$s_{n+1} = \frac{n}{n+1} s_{n-1} \tag{41}$$

mit $s_0 = \frac{\pi}{2}$ und $s_1 = [-\cos]_0^{\pi/2} = 1$, und daraus $s_2 = \frac{1}{2} \cdot a_0 = \frac{1}{2} \cdot \frac{\pi}{2}$, $s_3 = \frac{2}{3} \cdot s_1 = \frac{2}{3}$, $s_4 = \frac{3}{4} \cdot s_2 = \frac{3}{4} \cdot \frac{1}{2} \cdot \frac{\pi}{2}$, allgemein also

$$\begin{aligned} s_{2n} &= \frac{2n-1}{2n} \cdot \frac{2n-3}{2n-2} \cdot \ldots \cdot \frac{1}{2} \cdot \frac{\pi}{2} \\ s_{2n+1} &= \frac{2n}{2n+1} \cdot \frac{2n-2}{2n-1} \cdot \ldots \cdot \frac{2}{3} \end{aligned}$$

und damit

$$\frac{s_{2n+1}}{s_{2n}} = \frac{(2n)^2}{(2n+1)(2n-1)} \cdot \ldots \cdot \frac{2^2}{3 \cdot 1} \cdot \frac{2}{\pi} \tag{42}$$

Da $s_{2n+1}/s_{2n} \to 1$,[40] erhalten wir eine unendliche Produktdarstellung von $\frac{\pi}{2}$, das *Wallis-Produkt*:[41]

$$\frac{\pi}{2} = \frac{2 \cdot 2}{3 \cdot 1} \cdot \frac{4 \cdot 4}{5 \cdot 3} \cdot \frac{6 \cdot 6}{7 \cdot 5} \cdot \ldots = \frac{4}{3} \cdot \frac{16}{15} \cdot \frac{36}{35} \cdot \ldots \cdot \frac{(2n)^2}{(2n)^2 - 1} \cdot \ldots$$

[40] (s_n) ist eine monoton fallende Folge, denn für jedes feste x ist die Folge $(\sin(x)^n)_{n \in \mathbb{N}}$ monoton fallend, und $s_n = \int_0^{\pi/2} (\sin x)^n dx$. Nach (41) gilt außerdem $s_{n+1}/s_{n-1} = n/(n+1) \to 1$ und damit die Behauptung, da

$$1 \geq \frac{s_{n+1}}{s_n} \geq \frac{s_{n+1}}{s_{n-1}} \to 1.$$

[41] John Wallis, 1616 - 1703 (Oxford)

Satz 12.2. Substitutionsregel: *A und B seien offene Intervalle, $a, b \in A$ und $\phi : A \to B$ und $f : B \to \mathbb{R}$ Funktionen, ϕ differenzierbar und f stetig. Dann gilt:*

$$\int ((f \circ \phi) \cdot \phi') = \left(\int f\right) \circ \phi$$

$$\int_a^b f(\phi(x)) \cdot \phi'(x)\, dx = \int_{\phi(a)}^{\phi(b)} f(u)\, du \tag{43}$$

Beweis. $((\int f) \circ \phi)' = ((\int f)' \circ \phi) \cdot \phi' = (f \circ \phi) \cdot \phi'$, also ist $(\int f) \circ \phi$ eine Stammfunktion zu $(f \circ \phi) \cdot \phi'$. □

Dieser Satz heißt *Substitutionsregel*, weil der komplizierte Ausdruck $\phi(x)$ auf der linken Seite von (43) durch eine neue Variable u ersetzt (*substituiert*) wird. Die Regel kann man sich leicht merken, indem man formal u für $\phi(x)$ und du für $\phi'(x)dx$ substituiert.[42]

Beispiele: 1. $\int_a^b f(x+c)dx =$? Wir setzen $u = x+c$, also $\phi(x) = x+c$, dann ist $du = dx$, da $\phi' = 1$, und daher ist

$$\int_{x=a}^{x=b} f(x+c)dx = \int_{u=a+c}^{u=b+c} f(u)du. \tag{44}$$

2. $\int_a^b f(cx)dx =$? Wir setzen $u = cx$, also $u = \phi(x) = cx$, dann ist $du = c\, dx$, denn $\phi'(x) = c$, und

$$\int_a^b f(cx)dx = \frac{1}{c} \cdot \int_a^b f(cx)c\, dx = \frac{1}{c} \cdot \int_{ca}^{cb} f(u)du. \tag{45}$$

3. $\int_a^b x \cdot f(x^2)\, dx =$? Wir setzen $u = x^2$, dann ist $du = 2x\, dx$, also ist

$$\int_a^b x \cdot f(x^2)\, dx = \frac{1}{2} \int_{a^2}^{b^2} f(u)du. \tag{46}$$

4. $\int_a^b \frac{\phi'(x)}{\phi(x)}\, dx =$? Wir setzen $u = \phi(x)$ und erhalten $du = \phi'(x)dx$, also ist

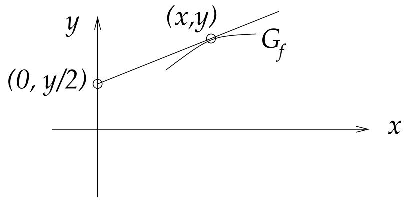
$$\int_a^b \frac{\phi'(x)}{\phi(x)}\, dx = \int_{\phi(a)}^{\phi(b)} \frac{1}{u}\, du = \ln \frac{|\phi(b)|}{|\phi(a)|} \tag{47}$$

Zum Beispiel ist

$$\int_a^b \tan(x)dx = -\int_a^b \frac{\cos'(x)}{\cos(x)}\, dx = \ln \frac{|\cos a|}{|\cos b|} \tag{48}$$

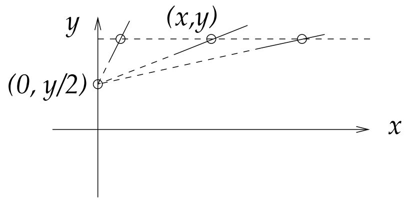
[42]Formal wird die Gleichung $\phi'(x) = \frac{d\phi}{dx} = \frac{du}{dx}$ mit dx multipliziert: $\phi'(x)dx = du$.

Aufgabe: Bestimmen Sie alle Funktionen $y = f(x)$ für $x > 0$ mit der folgenden Eigenschaft: Die Tangente an G_f = Graph f in jedem Punkt $(x, y) \in G_f$ schneidet die y-Achse im Punkt $(0, y/2)$.

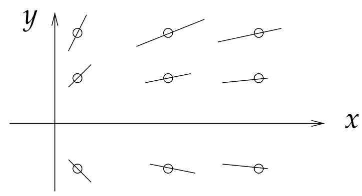


Problem: Wir wissen nicht, wo G_f liegt, weil wir f ja noch nicht kennen; wir können wir da die Angabe nutzen?[43]

Lösung: Wir nutzen die Angabe in *jedem* Punkt (x, y) der Halbebene aus und zeichnen jeweils ein Stück der Geraden durch (x, y) und $(0, y/2)$ ein:

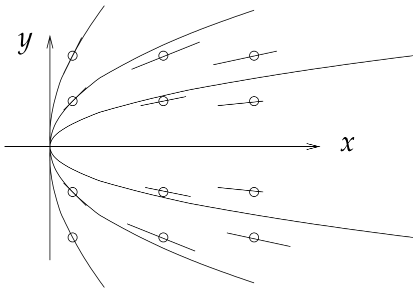


Wenn wir dies in jedem Punkt $(x, y) \in \mathbb{R}_+^* \times \mathbb{R}$ tun, erhalten wir ein *Feld* von Geradenstücken, ein *Richtungsfeld*:



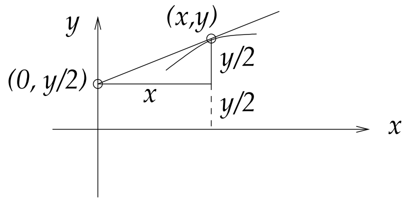
Der gesuchte Graph muss überall in dieses Richtungsfeld *passen*, d.h. an jeder Stelle die vorgegebene Richtung als Tangente haben. Damit zeichnet sich der Verlauf der gesuchten Kurven bereits ab:

[43]Eigentlich besteht dieses Problem bei *jeder* Gleichung: Wie können wir ein x mit $x^2 = x+1$ finden? Wäre uns x bekannt, dann könnten wir $x+1$ berechnen und damit $x = \sqrt{x+1}$ finden, aber wir kennen ja x nicht! Verzweiflung - oder Mathe!



Wir sehen, dass es nicht nur eine, sondern ein ganzes Feld von Lösungen gibt; durch jeden Punkt (x, y) der Halbebene genau eine; das ist typisch für Differentialgleichungen.

Analytische Lösung: Wenn (x, y) ein Punkt des gesuchten Graphen G_f ist, dann geht Tangente laut Angabe durch die Punkte (x, y) und $(0, y/2)$ und hat damit die Steigung $a = \frac{y-y/2}{x-0} = \frac{y/2}{x} = \frac{y}{2x}$.



Da die Steigung der Tangente die Ableitung in dem betreffenden Punkt ist, $a = f'(x)$, erhalten wir mit $y = f(x)$ für die unbekannte Funktion f die *Differentialgleichung*[44]

$$f'(x) = \frac{f(x)}{2x} \tag{49}$$

für alle $x \in D \subset (0, \infty)$, wobei D das (uns ebenfalls noch unbekannte) Definitionsintervall von f ist. Eine Lösung f von (49) ist $f = 0$, denn für dieses f ist $f'(x) = 0 = \frac{f(x)}{2x}$. Wenn wir dagegen $f(x) \neq 0$ für alle $x \in D$ voraussetzen, dann dürfen wir durch $f(x)$ dividieren und erhalten:

$$f' = \frac{f}{2x} \iff \frac{f'}{f} = \frac{1}{2x}$$

Auf beiden Seiten der letzteren Gleichung steht die Ableitung einer Funktion von x: Nach Kettenregel ist nämlich $\frac{f'}{f} = (\ln \circ |f|)'$ (vgl. auch (47)), und außerdem ist $\frac{1}{2x} = \frac{1}{2} \ln x$, also

[44]Man schreibt für (49) manchmal auch kurz $y' = \frac{y}{2x}$.

$$\begin{aligned} \frac{f'}{f} = \frac{1}{2x} \quad &\Longleftrightarrow \quad (\ln \circ |f|)' = \tfrac{1}{2}\ln' \\ &\Longleftrightarrow \quad \ln(|f(x)|) = \tfrac{1}{2}\ln x + c \\ &\Longleftrightarrow \quad |f(x)| = e^{\frac{1}{2}\ln x + c} = e^c \cdot e^{\frac{1}{2}\ln x} \\ &\Longleftrightarrow \quad |f(x)| = C \cdot \sqrt{x} \end{aligned}$$

mit einer Konstante $C = e^c > 0$. Wir erhalten also die Lösungen $f : (0,\infty) \to \mathbb{R}$, $f(x) = C\sqrt{x}$ für beliebige Konstanten $C \in \mathbb{R}$.[45]

Dieses Verfahren funkioniert bei allen Differentialgleichungen vom Typ

$$f'(x) = u(x)v(f(x)), \quad \text{kurz}: \quad y' = u(x)v(y) \tag{50}$$

für gegebene Funktionen u und v. Es wird *Trennung der Variablen* genannt, weil man die Gleichung $y' = u(x)v(y)$ von v so umformt, dass links nur Ausdrücke in y und rechts nur Ausdrücke in x stehen. Zunächst betrachtet man die Nullstellen von v, also die Werte y_o mit $v(y_o) = 0$. Die Konstanten $y(x) = y_o$ für alle x sind selbst Lösungen von (50), denn linke und rechte Seite sind Null. Nun betrachtet man den Fall $v(y) \neq 0$ und darf durch $v(y)$ dividieren:

$$y' = u(x)v(y) \iff \frac{y'}{v(y)} = u(x).$$

Wenn man nun Stammfunktionen $V = \int \frac{1}{v}$ von $\frac{1}{v}$ und $U = \int u$ von u kennt, dann ist $\frac{y'}{v(y)} = (V \circ y)'$ und $u = U'$ und die Gleichung wird weiter umgeformt zu

$$\begin{aligned} y' = u(x)v(y) \quad &\Longleftrightarrow \quad \frac{y'}{v(y)} = u(x) \\ &\Longleftrightarrow \quad (V \circ y)' = U' \\ &\Longleftrightarrow \quad V \circ y = U + c \\ &\Longleftrightarrow \quad y(x) = V^{-1}(U(x) + c). \end{aligned} \tag{51}$$

Beispiel 2: Lösen Sie das Anfangswertproblem

$$y' = y^2, \tag{52}$$
$$y(0) = 1, \tag{53}$$

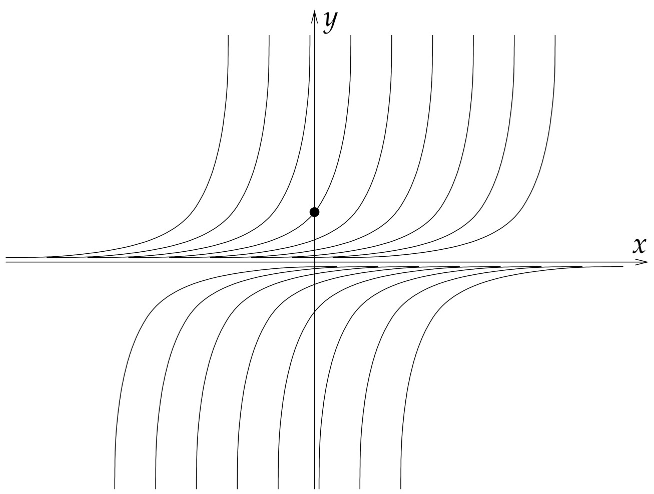
und bestimmen Sie das maximale Intervall, auf dem die Lösung definiert ist ("maximales Existenzintervall").

[45] Man müsste eigentlich noch fragen, ob es außer $f = 0$ noch andere Lösungen mit Nullstellen geben kann; dies verbietet die allgemeine Theorie: Durch jeden Punkt (x, y) der Halbebene geht genau eine Lösung.

Lösung:

$$\begin{aligned} y' = y^2 \quad &\overset{y\neq 0}{\Longleftrightarrow} \quad \frac{y'}{y^2} = 1 \\ &\Longleftrightarrow \quad \int \frac{1}{y^2}\,dy = \int 1\,dx + c \\ &\Longleftrightarrow \quad -\frac{1}{y} = x + c \\ &\Longleftrightarrow \quad y = -\frac{1}{x+c} \end{aligned}$$

Dies zusammen mit der Null-Lösung $y = 0$ nennt man die *allgemeine Lösung* der Differentialgleichung (52); es ist eine ganze Schar von Lösungen, abhängig von dem Parameter $c \in \mathbb{R}$, und durch jeden Punkt der Ebene geht genau eine Lösungskurve.



Unter allen diesen Lösungen y suchen wir nun diejenige, die auch noch (53) erfüllt, also

$$y(0) = 1 \iff -\frac{1}{0+c} = 1 \iff c = -1 \iff y = -\frac{1}{x-1}.$$

Die gesuchte Lösung ist also $y = \frac{1}{1-x}$. Sie ist überall definiert, wo $x \neq 1$, also auf $(-\infty, 1) \cup (1, \infty)$. Dieser Bereich ist eine disjunkte Vereinigung[46] der beiden offenen Intervalle $(-\infty, 1)$ und $(1, \infty)$; da 0 nach Angabe im Definitionsintervall liegen soll, muss dieses $(-\infty, 1)$ sein.

[46]Die Vereinigung $A \cup B$ von zwei Menge A, B heißt *disjunkt*, wenn A und B keinen gemeinsamen Schnitt haben, $A \cap B = \emptyset$.

14. Integration von Funktionenfolgen

Eine wichtige Klasse von Funktionen sind die *Potenzreihen*, d.h. Funktionen vom Typ

$$f(x) = a_0 + a_1x + a_2x^2 + a_3x^3 + \dots \tag{54}$$

Wir haben bereits die Funktionen exp, sin, cos als Potenzreihen kennengelernt; es gibt viele weitere. Wir werden sehen, dass wir diese unendlichen Summen in ihrem Konvergenzbereich beim Integrieren wie endliche Summen behandeln und jeden Summanden einzeln integrieren können. Die unendliche Summen $f(x)$ ist ja eigentlich der Grenzwert der Folge $f_n(x) = \sum_{k=0}^n a_kx^k$ für $n \to \infty$, und in "Zahl und Funktion" Satz 19.3, S. 63 sahen wir, dass f_n auf einem genügend kleinen Intervall $I = [-r, r]$ sogar *gleichmäßig* gegen f konvergiert:

$$f_n \stackrel{\text{glm}}{\longrightarrow} f \iff \forall_{\epsilon>0}\exists_N\forall_{n\geq N}\ |f_n - f| < \epsilon. \tag{55}$$

Satz 14.1. *Ist $f_n : [a, b] \to \mathbb{R}$ eine Folge stetiger Funktionen mit*

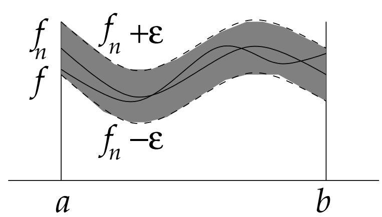
$$f_n \stackrel{\text{glm}}{\longrightarrow} f \tag{56}$$

(damit ist auch $f : [a, b] \to \mathbb{R}$ stetig, vgl. "Zahl und Funktion", Satz 19.1, S.62), dann gilt

$$\int_a^b f_n \longrightarrow \int_a^b f \tag{57}$$

Beweis.

$$\begin{aligned}
|f_n - f| \leq \epsilon \quad &\Rightarrow \quad f - \epsilon \leq f_n \leq f + \epsilon \\
&\stackrel{(4),\,S.36}{\Rightarrow} \int_a^b (f-\epsilon) \leq \int_a^b f_n \leq \int_a^b (f+\epsilon) \\
&\Rightarrow \left(\int_a^b f\right) - \epsilon(b-a) \leq \int_a^b f_n \leq \left(\int_a^b f\right) + \epsilon(b-a) \\
&\Rightarrow \left|\int_a^b f_n - \int_a^b f\right| \leq \epsilon(b-a).
\end{aligned}$$



Die Figur zeigt noch einmal das Prinzip des Arguments: Der dunkel unterlegte Bereich zwischen $f_n - \epsilon$ und $f_n + \epsilon$ hat den kleinen Flächeninhalt $2\epsilon \cdot (b-a)$, und da die Graphen von f_n und f beide darin liegen, ist der Unterschied der beiden Integrale (die Fläche zwischen den Graphen) noch kleiner. □

Folgerung 14.1. *Ist D offenes Intervall, $f_n : D \to \mathbb{R}$ stetig und gilt $f_n \xrightarrow{\text{glm}} f$ auf jedem Intervall $[a,b] \subset D$ ("lokal gleichmäßige Konvergenz", $f_n \xrightarrow{\text{lglm}} f$) und ist $F_n : D \to \mathbb{R}$ eine Stammfunktion von f_n mit $F_n(a) \to c$, dann konvergieren die Funktionen F_n punktweise gegen eine Stammfunktion F von f, kurz (mit $F_n =: \int f_n$, $F =: \int f$):*

$$f_n \xrightarrow{\text{lglm}} f,\ (\textstyle\int f_n)(a) \to c \ \Rightarrow \ \int f_n \to \int f.$$

Beweis.

$$\begin{aligned} F_n(x) &= F_n(a) + \int_a^x f_n \\ &\to \quad c \quad + \int_a^x f =: F(x). \end{aligned}$$ □

Folgerung 14.2. *Ist $F_n : D \to \mathbb{R}$ differenzierbar mit stetigen Ableitungen $F_n' = f_n$ und gilt $f_n \xrightarrow{\text{lglm}} f$ und $F_n(a) \to c$, dann konvergiert F_n und $F = \lim F_n$ ist differenzierbar mit $F' = f$, kurz:*[47]

$$F_n(a) \to c,\ F_n' \xrightarrow{\text{lglm}} f \ \Rightarrow \ (\lim F_n)' = \lim F_n'.$$ □

Folgerung 14.3. *Ist $f(x) = \sum_{k=0}^{\infty} a_k x^k$ eine Potenzreihe mit*

$$\frac{|a_k|}{|a_{k+1}|} \to R > 0,$$

so ist f auf $(-R, R)$ definiert, stetig und sogar differenzierbar mit

$$\int f = \sum_{k=0}^{\infty} \frac{a_k}{k+1} x^{k+1}, \quad f' = \sum_{k=1}^{\infty} k a_k x^{k-1}. \tag{58}$$

Beweis. Nach "Zahl und Funktion" Satz 19.3b, S.63 ist die Potenzreihe $f(x) = \sum_k a_k x^k$ auf $(-R, R)$ lokal gleichmäßig konvergent. Dasselbe gilt für die "abgeleitete Reihe" $g(x) = \sum_k b_k x^{k-1}$ mit $b_k = k a_k$, denn

$$\frac{|b_k|}{|b_{k+1}|} = \frac{k \cdot |a_k|}{(k+1) \cdot |a_{k+1}|} = \frac{k}{k+1} \cdot \frac{|a_k|}{|a_{k+1}|} \to R.$$

---

[47] Im Folgenden soll der Ausdruck $\lim F_n$ auch bedeuten, dass dieser Grenzwert existiert.

Nach den beiden voranstehenden Folgerungen gilt $g = \lim f_n' = f'$ und $\int f = \lim \int f_n$, was zu zeigen war. $\square$

Bemerkung: Die Folge $\left(\frac{|a_k|}{|a_{k+1}|}\right)_{k\in\mathbb{N}}$ in Korollar 14.3 besitzt natürlich nicht immer einen Limes. Aber der Konvergenzradius einer Potenzreihe $\sum_k a_k x^k$ ist immer definiert! Statt des Quotienten $\frac{|a_k|}{|a_{k+1}|}$ (der Nenner $|a_{k+1}|$ könnte ja Null werden!) benutzt man besser die k-te Wurzel: Wenn $\sqrt[k]{|a_k|} \le s$ für alle (genügend großen) k, also $|a_k| \le s^k$, dann ist $|a_k x^k| \le (s|x|)^k$. Wenn also $s|x| \le q$ oder $|x| \le q/s$ für eine Zahl $q \in (0,1)$, dann ist die Geometrische Reihe $\sum_k q^k$ eine konvergente Majorante für $\sum_k a_k x^k$. Die Reihe $\sum_k a_k x^k$ konvergiert daher absolut und gleichmäßig auf dem Bereich $D = \{x;\ |x| \le q/s\}$. Genauer: Ist $f_n(x) = \sum_{k=0}^n a_k x^k$ und $f(x) = \sum_{k=0}^\infty a_k x^k$, dann gilt für alle $x \in D$

$$\begin{aligned}
|f(x) - f_n(x)| &\le \sum_{k=n+1}^\infty |a_k x^k| \\
&\le \sum_{k=n+1}^\infty q^k \\
&= q^{n+1}\left(\sum_{j=0}^\infty q^j\right) \\
&= q^{n+1}/(1-q) \stackrel{n\to\infty}{\longrightarrow} 0
\end{aligned}$$

Das ist die gleichmäßige Konvergenz $f_n \stackrel{\text{glm}}{\longrightarrow} f$ auf D. Geht man zum größtmögilchen s über, $s = \lim_{k\to\infty} \sqrt[k]{|a_k|}$, so gilt immer noch $f_n \stackrel{\text{glm}}{\longrightarrow} f$ auf $\{x;\ |x| \le r\}$ für jedes $r < q/s$. Aber auch dieser Limes muss nicht existieren. Stattdessen betrachtet man den Limes Superior

$$s = \lim_{k\to\infty} \sup \sqrt[k]{|a_k|} := \lim_{k\to\infty} \sup\{\sqrt[m]{|a_m|};\ m \ge k\}$$

Dieser existiert immer, weil die Menge $\{\sqrt[m]{|a_m|};\ m \ge k\}$ mit wachsendem m immer kleiner und ihr Supremum daher auch immer kleiner wird; eine beschränkte monoton fallende Folge besitzt immer einen Limes. Man definiert den Konvergenzradius R daher als

$$R = 1/ \lim_{k\to\infty} \sup \sqrt[k]{|a_k|}; \tag{59}$$

dann gilt $f_n \stackrel{\text{glm}}{\longrightarrow} f$ auf $\{x;\ |x| \le r\}$ für alle $r < R$. Die formal abgeleitete Reihe $f'(x) = \sum_{k=1}^\infty k a_k x^{k-1}$ hat den gleichen Konvergenzradius wie $f(x) = \sum_{k=0}^\infty a_k x^k$, denn

$$\lim_{k\to\infty} \sup \sqrt[k]{k|a_k|} = \lim_{k\to\infty} \sup \sqrt[k]{|a_k|} \cdot \sqrt[k]{k} = \lim_{k\to\infty} \sup \sqrt[k]{|a_k|},$$

weil $\sqrt[k]{k} \to 1$ wegen $\ln \sqrt[k]{k} = \ln(k^{1/k}) = \frac{1}{k} \ln k \to 0$ für $k \to \infty$.[48]

Beispiel 1: $F(x) = \ln(1+x) \Rightarrow f(x) := F'(x) = \frac{1}{1+x} = \sum_{k=0}^{\infty}(-x)^k$ für $|x| < 1$ (Geometrische Reihe, "Zahl und Funktion" Satz 13.2, S.44). Damit ist $\tilde{F}(x) = \sum_{k=0}^{\infty} \frac{(-1)^k}{k+1} x^{k+1}$ eine Stammfunktion von $f(x) = \sum_{k=0}^{\infty}(-1)^k x^k$ mit $\tilde{F}(0) = 0$. Da aber auch $F(x) = \ln(1+x)$ eine Stammfunktion von f mit $F(0) = 0$ ist, muss $F = \tilde{F}$ auf $(-1,1)$ gelten. Wir erhalten demnach für alle x mit $|x| < 1$ die Formel

$$\ln(1+x) = \sum_{k=0}^{\infty}(-1)^k \frac{x^{k+1}}{k+1}. \tag{60}$$

In der Tat gilt diese Formel auch noch für $x = 1$, d.h. wir erhalten

$$\ln 2 = \sum_{k=0}^{\infty}(-1)^k \frac{1}{k+1} = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + - \dots \tag{61}$$

Das ist die uns schon bekannte *die Leibnizsche Reihe* (vgl. "Zahl und Funktion", Beispiel S. 47). Wir konnten früher nur ihre Konvergenz feststellen, jetzt haben wir ihren Wert berechnet!

Aber wie können wir (60) auf $x = 1$ ausdehnen? Wir haben

$$\ln(1+x) = s(x) = \lim s_n(x)$$

mit $s_n(x) = \sum_{k=0}^{n}(-1)^k p_k(x)$ für die Funktion $p_k(x) = \frac{x^{k+1}}{k+1}$. Für alle $x \in [0,1]$ ist $p_k(x) \geq 0$ und bildet eine monoton fallende Nullfolge. Wie wir im Beweis des Leibniz-Kriteriums ("Zahl und Funktion", S. 44) gesehen haben, bilden die "ungeraden Summen" eine untere Schranke, die "geraden Summen" eine obere Schranke für den Reihenwert:

$$\begin{array}{rcll} s_{2m+1} & = & p_0 - p_1 + p_2 - + \dots + p_{2m} - p_{2m+1} & < s \\ s_{2m} & = & p_0 - p_1 + p_2 - + \dots + p_{2m} & > s \end{array}$$

Sowohl für $n = 2m$ als auch für $n = 2m+1$ gilt daher

$$|s_n - s| < s_{2m} - s_{2m+1} = p_{2m+1} \leq \frac{1}{2m+1}$$

und somit $s_n \xrightarrow{\text{glm}} s$ auf $[0,1]$. Also ist s auf $[0,1]$ stetig, d.h.

$$s(1) = \lim_{x \nearrow 1} s(x) = \lim_{x \nearrow 1} \ln(1+x) = \ln 2$$

womit (61) bewiesen ist.

Beispiel 2: $F(x) = \arctan x \Rightarrow f(x) = F'(x) = \frac{1}{1+x^2} = \sum_{k=0}^{\infty}(-x^2)^k = \sum_{k=0}^{\infty}(-1)^k x^{2k}$ für $|x| < 1$, wieder mit der Geometrischen Reihe. Also

---

[48] $\frac{\ln x}{x} \to 0$ für $x \to \infty$, denn setzen wir $x = e^u$, so ist $\frac{\ln x}{x} = \frac{u}{e^u} \to 0$ für $u \to \infty$.

ist $\tilde{F}(x) = \sum_{k=0}^{\infty} \frac{(-1)^k}{2k+1} x^{2k+1}$ eine Stammfunktion von f mit $\tilde{F}(0) = 0$. Da auch $F(x) = \arctan x$ eine Stammfunktion von f mit $F(0) = 0$, ist $F = \tilde{F}$ auf $(-1, 1)$. Wir erhalten also für alle x mit $|x| < 1$ die Formel

$$\arctan x = \sum_{k=0}^{\infty} (-1)^k \frac{x^{2k+1}}{2k+1}. \tag{62}$$

Mit dem gleichen Argument wie in Beispiel 1 gilt diese Formel auch noch für $x = 1$. Nun ist $\arctan 1 = \frac{\pi}{4}$, denn $\tan \frac{\pi}{4} = 1$.[49] Daher erhalten wir die erstaunliche Formel

$$\frac{\pi}{4} = \sum_{k=0}^{\infty} \frac{(-1)^k}{2k+1} = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + - \dots \tag{63}$$

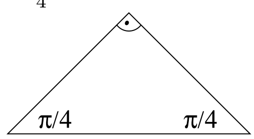
15. Höhere Ableitungen und Satz von Taylor

Gegeben sei ein offenes Intervall D und eine differenzierbare Funktion $f : D \to \mathbb{R}$. Wenn deren Ableitung $f' : D \to \mathbb{R}$ ebenfalls differenzierbar ist, so heißt f *zweimal differenzierbar* und die Ableitung der Ableitung, $f'' := (f')'$ heißt *zweite Ableitung* von f. Wenn f'' wieder differenzierbar ist, so heißt $(f'')' = f''' = f^{(3)}$ die *dritte Ableitung* usw. Rekursiv definieren wir: f ist $(n+1)$-mal differenzierbar, wenn f n-mal differenzierbar ist und die n-te Ableitung $f^{(n)}$ wieder differenzierbar ist; ihre Ableitung $(f^{(n)})' =: f^{(n+1)}$ heißt die *$(n+1)$-te Ableitung von f*. Die Funktion f heißt *n-mal stetig differenzierbar* (C^n),[50] wenn f n-mal differenzierbar und die n-te Ableitung auch noch stetig ist; die niedrigeren Ableitungen $f', \dots, f^{(n-1)}$ sind ja ohnehin stetig, weil sie sogar differenzierbar sind. Wenn *alle* Ableitungen wieder differenzierbar sind (C^n für alle $n \in \mathbb{N}$), dann heißt die Funktion *beliebig oft differenzierbar* (C^∞).

Beispiele von Funktionen mit beliebig vielen Ableitungen sind die *Polynome* $f(x) = a_0 + a_1 x + \dots a_n x^n$ (wobei alle Ableitungen $f^{(k)}$ mit $k \geq n+1$ verschwinden) und allgemeiner die Potenzreihen:

$$f(x) \;=\; a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \dots = \sum_{k=0}^{\infty} a_k x^k$$

[49] Dies gilt, weil $\tan = \frac{\sin}{\cos}$ und $\sin \frac{\pi}{4} = \cos \frac{\pi}{4}$; bei $45^o = \frac{\pi}{4}$ ist das rechtwinklige Dreieck symmetrisch, Gegenkathete = Ankathete.



[50] Für "n-mal stetig differenzierbar" gibt es die Abkürzung C^n, wobei der Buchstabe C für "continuous" (stetig) steht.

$$\begin{aligned}
f'(x) &= a_1 + 2a_2x + 3a_3x^2 + \cdots = \sum_{k=1}^{\infty} k\, a_k x^{k-1} \\
f''(x) &= 2a_2 + 6a_3x + \cdots = \sum_{k=2}^{\infty} (k-1)k\, a_k x^{k-2} \\
f^{(n)}(x) &= \sum_{k=n}^{\infty} (k-n+1)(k-n+2)\ldots(k-1)k\, a_k x^{k-n}
\end{aligned}$$

Bei $x = 0$ sind alle Summanden Null außer dem mit nullter Potenz von x (der gar kein x enthält, denn $x^0 = 1$):

$$\begin{aligned}
f(0) &= a_0, \\
f'(0) &= a_1, \\
f''(0) &= 2a_2, \\
f^{(n)}(0) &= 1 \cdot 2 \cdot \ldots \cdot (n-1) \cdot n \cdot a_n \\
&= n!\, a_n\,.
\end{aligned}$$

Wir können also die Koeffizienten der Potenzreihe $f(x)$ mit Hilfe der Ableitungen ausdrücken: $f^{(k)}(0) = k!\, a_k$ und daher $a_k = \frac{f^{(k)}(0)}{k!}$ für alle $k \in \mathbb{N}$. Damit erhalten wir:

Satz 15.1. *Ist $f(x)$ eine Potenzreihe, die für $x \in D = (-R, R)$ konvergiert für ein $R > 0$, so gilt für alle $x \in D$:*

$$f(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(0)}{k!}\, x^k\,. \tag{64}$$

□

Dieser bemerkenswerte Satz wirft eine Frage auf: Sind vielleicht *alle* beliebig oft differenzierbaren Funktionen $f : D \to \mathbb{R}$ Potenzreihen? Wenn f gegeben ist, können wir im Prinzip die Werte $f^{(k)}(0)$ berechnen; ist die Gleichung (64) dann immer erfüllt? Dies und viel mehr beantwortet der folgende *Satz von Taylor*.[51] Dabei ersetzen wir 0 gleich durch eine beliebige Stelle a, genannt *Entwicklungspunkt*, indem wir $\tilde{x} = x - a$ anstelle von x als Variable betrachten.

Satz 15.2. *Gegeben sei eine C^{n+1}-Funktion $f : D \to \mathbb{R}$ auf einem offenen Intervall D. Dann gilt für alle $a, x \in D$:*

$$f(x) = \sum_{k=0}^{n} \frac{f^k(a)}{k!}(x-a)^k + R_{n+1}(x), \tag{65}$$

$$R_{n+1}(x) = \frac{1}{n!}\int_a^x (x-t)^n f^{(n+1)}(t)\, dt. \tag{66}$$

[51] Brook Taylor, 1685 - 1731 (London)

Beweis.
$n = 0$: $f(x) = f(a) + \int_a^x f'(t)\,dt = f(a) + R_1(x)$ nach "Hauptsatz" 11.1.
$n \to n+1$: Nach Induktionsvoraussetzung ist

$$\begin{aligned} f(x) &= \sum_{k=0}^{n-1} \frac{f^k(a)}{k!}(x-a)^k + R_n(x), \qquad (67) \\ R_n(x) &= \frac{1}{(n-1)!}\int_a^x (x-t)^{n-1} f^{(n)}(t)\,dt \end{aligned}$$

Auf dieses Integral wenden wir *partielle Integration* an und setzen dazu

$$\begin{aligned} u'(t) &= (x-t)^{n-1} & v(t) &= f^{(n)}(t) \\ u(t) &= -\tfrac{1}{n}(x-t)^n & v'(t) &= f^{(n+1)}(t) \end{aligned}$$

Dann ist

$$\begin{aligned} R_n(x) &= \frac{1}{(n-1)!}\int_a^x u'(t)v(t)\,dt \\ &= \frac{1}{(n-1)!}\left([u(t)v(t)]_a^x - \int_a^x u(t)v'(t)\,dt \right) \\ &= \frac{1}{n!}\left(\left[-(x-t)^n f^{(n)}(t)\right]_a^x + \int_a^x (x-t)^n f^{(n+1)}(t)\,dt \right) \\ &= \frac{1}{n!}\left((x-a)^n f^{(n)}(a) + \int_a^x (x-t)^n f^{(n+1)}(t)\,dt \right) \end{aligned}$$

Setzen wir dies in (67) ein, so ergibt sich die Behauptung. □

Nach diesem Satz ist $f(x)$ für alle x nahe bei a im Wesentlichen durch das Polynom

$$f_n(x) = \sum_{k=0}^{n} \frac{f^k(a)}{k!}(x-a)^k \qquad (68)$$

gegeben, das sogenannte *Taylorpolynom*; der Ausdruck $R_{n+1}(x)$ (das *Taylorsche Restglied*) sollte nur eine kleine Störung sein, d.h. kleinen Betrag haben. Um dies besser zu sehen, schreibt man $R_{n+1}(x)$ noch in einer anderen Form, die auf *Lagrange*[52] zurückgeht. Dazu benötigen wir zunächst einen Hilfssatz aus der Integralrechnung:

Lemma 15.1. Mittelwertsatz der Integralrechnung: *Es seien $f, g : D \to \mathbb{R}$ stetige Funktionen mit $g \geq 0$ oder $g \leq 0$. Dann gibt es für je zwei Zahlen $a, b \in D$ ein ξ zwischen a und b mit*

$$\int_a^b (fg) = f(\xi)\int_a^b g. \qquad (69)$$

[52]Joseph Louis Lagrange, 1736 - 1813 (Paris)

Beweis. Wir dürfen $g \geq 0$ annehmen, sonst gehen wir zu $-g$ über. Wir nehmen außerdem zunächst $a < b$ an. Mit M bezeichnen wir das Maximum und mit m das Minimum von $f|_{[a,b]}$. Dann gilt auf $[a,b]$:

$$\begin{aligned} m \leq f \leq M \quad \overset{g\geq 0}{\Longrightarrow} \quad & mg \leq fg \leq Mg \\ \Rightarrow \quad & \int_a^b (mg) \leq \int_a^b (fg) \leq \int_a^b Mg \\ \Rightarrow \quad & m \int_a^b g \leq \int_a^b (fg) \leq M \int_a^b g \\ \Rightarrow \quad & m \leq \frac{\int_a^b (fg)}{\int_a^b g} \leq M \end{aligned}$$

Nach dem Zwischenwertsatz (“Zahl und Funktion”, Satz 21.1, S.73) ist jede Zahl zwischen den Werten m und M von f selbst ein Wert von f; es gibt daher ein $\xi \in (a,b)$ mit

$$\frac{\int_a^b (fg)}{\int_a^b g} = f(\xi)$$

und damit $\int_a^b (fg) = f(\xi) \int_a^b g$. Der Fall $b < a$ folgt ebenso, wobei wir überall $\int_a^b$ durch $\int_b^a = -\int_a^b$ ersetzen müssen. □

Satz 15.3. *Gegeben sei eine C^{n+1}-Funktion $f : D \to \mathbb{R}$ auf einem offenen Intervall D. Dann gilt für alle $a, x \in D$:*

$$\begin{aligned} f(x) &= \sum_{k=0}^{n} \frac{f^k(a)}{k!}(x-a)^k + R_{n+1}(x), \\ R_{n+1}(x) &= \frac{f^{(n+1)}(\xi)}{(n+1)!}\,(x-a)^{n+1} \end{aligned} \tag{70}$$

für ein ξ zwischen a und x (Lagrange'sches Restglied).

Beweis. Wir müssen nur das Restglied $R_{n+1}(x)$ aus Satz 15.2 umformen: Dazu setzen wir

$$g(t) = (x-t)^n.$$

Für gerade n ist $g(t) \geq 0$ für alle t zwischen a und x, und für ungerade n ist $g(t) \geq 0$, falls $x > a$ (und somit $x \geq t$), und $g(t) \leq 0$ falls $x < a$. Nach (66) und dem vorstehenden Lemma gilt also:

$$R_{n+1}(x) = \frac{1}{n!} \int_a^x f^{(n+1)}(t) g(t)\, dt = \frac{1}{n!}\, f^{(n+1)}(\xi) \int_a^x g$$

für ein ξ zwischen a und x. Das verbleibende Integral über g können wir berechnen:
$$\int_a^x g = \int_a^x (x-t)^n\, dt = \left[-\frac{1}{n+1}(x-t)^{n+1}\right]_{t=a}^{t=x} = \frac{1}{n+1}\,(x-a)^{n+1}.$$
Dies in die vorangehende Gleichung für $R_{n+1}(x)$ eingesetzt ergibt die Behauptung. $\square$

Leider können wir auch das Lagrange-Restglied nicht wirklich berechnen,[53] z.B. weil wir ξ gar nicht kennen. Es genügt uns aber, eine kleine obere Schranke für $|R_{n+1}(x)|$ zu finden; dann wissen wir, welchen Fehler wir höchstens machen, wenn wir $f(x)$ durch das Taylorpolynom $f_n(x)$ ersetzen. Dazu müssen wir eine obere Schranke für $|f^{(n+1)}(\xi)|$ finden. Wir wissen, dass ξ zwischen a und x liegt; wenn die Funktion $|f^{(n+1)}|$ monoton ist, können wir dann $|f^{(n+1)}(\xi)|$ durch $|f^{(n+1)}(a)|$ oder $|f^{(n+1)}(x)|$ abschätzen.

Beispiel 1: $f:\mathbb{R}\to\mathbb{R}$, $f(x)=e^x$ und $a=0$. Dann ist $f=f'=f''=\cdots=f^{(k)}$ für alle k und damit $f^{(k)}(0)=f(0)=e^0=1$. Für das Taylorpolynom ergibt sich $f_n(x)=\sum_{k=0}^n \frac{1}{k!}\,x^k$ und für das Restglied $R_{n+1}(x)=\frac{e^\xi}{(n+1)!}\,x^{n+1}$ für ein ξ zwischen 0 und x. Weil die Funktion $\xi\mapsto e^\xi$ monoton wachsend ist, folgt $e^\xi<e^0=1$ für $x<0$ und $e^\xi<e^x$ für $x>0$. Also erhalten wir:
$$\begin{aligned} x<0 &\;\Rightarrow\; |R_{n+1}(x)| \;\le\; \tfrac{1}{(n+1)!}\,|x|^{n+1},\\ x>0 &\;\Rightarrow\; |R_{n+1}(x)| \;\le\; \tfrac{e^x}{(n+1)!}\,|x|^{n+1}.\end{aligned}$$
Die erste Abschätzung (für $x<0$) ist besonders einfach: Für $x=-\frac12$ und $n=3$ zum Beispiel bekommen wir $\frac{1}{\sqrt{e}}=e^{-1/2}=f_3(-\frac12)+R_4(-\frac12)$ mit $f_3(-\frac12)=1-\frac12+\frac12\cdot\frac14-\frac16\cdot\frac18=\frac{29}{48}$ und $|R_4(-\frac12)|<\frac{1}{24}\cdot\frac{1}{16}=\frac{1}{384}$. Die zweite Abschätzung (für $x>0$) ist etwas ungünstiger; um sie verwenden zu können, brauchen wir bereits eine grobe obere Schranke für e^x, etwa $e<3$ und damit $e^x<3^n$ falls $x\le n$.[54]

---

[53] Das Taylor-Restglied in (66) ist erst recht nicht berechenbar; wir können nicht einmal den Integranden $(x-t)^n f^{(n+1)}(t)$ an jeder Stelle t berechnen.

[54] Für $x>0$ gibt es allerdings bessere Abschätzungen für $R_{n+1}(x)$ durch die geometrische Reihe:
$$\begin{aligned} R_{n+1}(x) &= \sum_{k=0}^{\infty}\frac{x^k}{k!}-\sum_{k=0}^{n}\frac{x^k}{k!}=\sum_{k=n+1}^{\infty}\frac{x^k}{k!}\\ &= \frac{x^{n+1}}{(n+1)!}+\frac{x^{n+2}}{(n+1)!(n+2)}+\frac{x^{n+3}}{(n+1)!(n+2)(n+3)}+\dots\\ &= \frac{x^{n+1}}{(n+1)!}\left(1+\frac{x}{n+2}+\frac{x^2}{(n+2)(n+3)}+\dots\right)\end{aligned}$$

Beispiel 2: $f : \mathbb{R} \to \mathbb{R}$, $f(x) = \sin x$ und $a = 0$: Hier ist das Restglied besonders einfach abzuschätzen, weil alle Ableitungen von $\sin x$ entweder $\pm \sin x$ oder $\pm \cos x$ sind, und deren Beträge sind ≤ 1. Damit ist $|R_n(x)| \leq \frac{|x|^n}{n!}$. Für $n = 6$ zum Beispiel[55] ist $\sin x = f_6(x) + R_7(x)$ mit $f_6(x) = x - \frac{x^3}{6} + \frac{x^5}{60}$ und $|R_7(x)| \leq \frac{|x|^7}{720}$.

Beispiel 3: $f : (0, \infty) \to \mathbb{R}$, $f(x) = x^\alpha$ für ein $\alpha \in \mathbb{R}$ und $a = 1$: Hier ist

$$\begin{aligned} f(x) &= x^\alpha \\ f'(x) &= \alpha\, x^{\alpha-1} \\ f''(x) &= (\alpha-1)\alpha\, x^{\alpha-2} \\ f^{(3)}(x) &= (\alpha-2)(\alpha-1)\alpha\, x^{\alpha-3} \\ f^{(k)}(x) &= (\alpha-k+1)\dots(\alpha-1)\alpha\, x^{\alpha-k} \end{aligned}$$

Somit ist

$$\frac{f^{(k)}(x)}{k!} = \binom{\alpha}{k} x^{\alpha-k}$$

wobei die *verallgemeinerten Binomialkoeffizienten* $\binom{\alpha}{k}$ für beliebige $\alpha \in \mathbb{R}$ ganz analog zu den gewöhnlichen Binomialkoeffizienten $\binom{n}{k}$ (vgl. “Zahl und Funktion” S.16) definiert sind: Wir setzen $\binom{\alpha}{0} = 1$ und

$$\binom{\alpha}{k} = \frac{\alpha(\alpha-1)\dots(\alpha-k+1)}{k!} = \prod_{j=1}^{k} \frac{\alpha-j+1}{j} \tag{71}$$

für $k \in \mathbb{N}$. Im Unterschied zu $\binom{n}{k}$ sind diese Zahlen $\binom{\alpha}{k}$ ungleich Null für alle $k \in \mathbb{N}$, sobald $\alpha \notin \{0, 1, 2, 3, \dots\}$.

Aus der Taylorformel (65) erhalten wir $f(x) = f_n(x) + R_{n+1}(x)$ mit

$$\begin{aligned} f_n(x) &= \sum_{k=0}^{n} \frac{f^{(k)}(1)}{k!}(x-1)^k = \sum_{k=0}^{n} \binom{\alpha}{k}(x-1)^k \\ R_{n+1}(x) &= \frac{f^{n+1}(\xi)}{(n+1)!}(x-1)^{n+1} = \binom{\alpha}{n+1} \xi^{\alpha-(n+1)}(x-1)^{n+1} \end{aligned}$$

$$\begin{aligned} &\leq \frac{x^{n+1}}{(n+1)!}\left(1 + \frac{x}{n+2} + \frac{x^2}{(n+2)^2} + \dots\right) \\ &\leq \frac{x^{n+1}}{(n+1)!} \cdot \frac{1}{1 - \frac{x}{n+2}} \\ &= \frac{x^{n+1}}{(n+1)!} \cdot \frac{n+2}{n+2-x} \end{aligned}$$

Zum Beispiel für $x = 1$ und $n = 2$ folgt $R_3(1) \leq \frac{1}{6} \cdot \frac{4}{3} = \frac{2}{9}$, also ist $e = 1+1+\frac{1}{2}+R_3(1)$ mit $0 < R_3(1) < \frac{2}{9}$, also $2{,}5 < e < 2{,}7222\dots$.

[55] Beim Sinus ist es günstig, n gerade zu wählen, denn z.B. ist $f_5 = f_6$, aber $|R_7|$ gibt eine bessere Fehlerabschätzung als $|R_6|$.

für ein ξ zwischen 1 und x. Für $n+1 > \alpha$ ist die ξ-Potenz in $R_{n+1}(x)$ negativ, also monoton fallend und nimmt damit ihr Maximum am linken Intervallrand an:

$$\begin{array}{lll} \xi^{\alpha-(n+1)} & \leq 1 & \text{falls } x \geq 1 \\ \xi^{\alpha-(n+1)} & \leq x^{\alpha-(n+1)} & \text{falls } x \leq 1 \end{array}$$

Somit folgt

$$|R_{n+1}(x)| \leq \begin{cases} \binom{\alpha}{n+1}|x-1|^{n+1} & \text{falls } x > 1 \\ \\ \binom{\alpha}{n+1}|x-1|^{n+1}x^{\alpha-(n+1)} & \text{falls } x < 1 \end{cases} \tag{72}$$

Für $\alpha = 1/2$ zum Beispiel haben wir

$$\begin{aligned} \binom{1/2}{1} &= \frac{1}{2} \\ \binom{1/2}{2} &= \frac{-\frac{1}{2}\cdot\frac{1}{2}}{2} = -\frac{1}{8} \\ \binom{1/2}{3} &= \frac{(-\frac{3}{2})(-\frac{1}{2})\frac{1}{2}}{6} = \frac{1}{16} \end{aligned}$$

Für $n = 2$ und $x = 1 + t$ oder $x - 1 = t$ mit $t > 0$ erhalten wir damit

$$\begin{aligned} \sqrt{1+t} &= 1 + \frac{1}{2}t - \frac{1}{8}t^2 + R_3(x), \qquad (73) \\ |R_3(x)| &\leq \frac{1}{16}t^3 \end{aligned}$$

Zum Beispiel für $t = \frac{1}{4}$ ergibt sich $\frac{1}{2}\sqrt{5} = \sqrt{1+\frac{1}{4}} = 1 + \frac{1}{8} - \frac{1}{128} \pm \epsilon = \frac{143}{128} \pm \epsilon$ mit $\epsilon = |R_3(\frac{1}{4})| \leq \frac{1}{1024}$; die Zahl $\frac{143}{128} = 1{,}1171875$ stimmt also bis auf drei Stellen hinter dem Komma mit $\frac{1}{2}\sqrt{5}$ überein. So können wir auch jede andere Quadratwurzel ziehen, indem wir die Differenz zur nächsten Quadratzahl betrachten: Zum Beispiel ist $\sqrt{10} = \sqrt{9+1} = 3\sqrt{1+t}$ mit $t = \frac{1}{9}$ und damit $\sqrt{10} = 3(1+\frac{1}{2}t-\frac{1}{8}t^2+R_3(t)) \approx 3+\frac{1}{6}-\frac{1}{8\cdot 27}$ mit einem Fehler $3|R_3(\frac{1}{9})| \leq \frac{1}{16\cdot 243} < \frac{1}{3500}$.

Lemma 15.2. *Verhalten des Restgliedes $R_{n+1}(x)$ für $n \to \infty$: Für alle $x \in (\frac{1}{2}, 2)$ gilt $R_{n+1}(x) \to 0$ für $n \to \infty$.*[56]

Beweis. Wir setzen $u = |x-1|$ falls $x \in (1,2)$ und $u = \frac{|x-1|}{x}$ falls $x \in (\frac{1}{2}, 1)$; in beiden Fällen ist $u \in (0,1)$. Mit dieser Abkürzung ist $|R_{n+1}(x)| \leq \binom{\alpha}{n+1} u^{n+1}$ im ersten bzw. $|R_{n+1}(x)| \leq \left|\binom{\alpha}{n+1}\right| u^{n+1} \cdot x^\alpha$ im

[56]Die Aussage gilt auch noch für alle $x \in (0, \frac{1}{2}]$, wie man mit Hilfe des Taylor-Restgliedes (66) zeigen kann, vgl. O. Forster, Analysis I im Kapitel über Taylorreihen.

zweiten Fall. Die Majorante $b_{n+1} = \left|\binom{\alpha}{n+1}\right| u^{n+1}$ ist eine (sogar summierbare) Nullfolge, vgl. "Zahl und Funktion", Satz 13.5, S.46, denn

$$\frac{b_{n+1}}{b_n} = \left|\frac{\binom{\alpha}{n+1}}{\binom{\alpha}{n}}\right| \cdot u = \frac{|\alpha - n|}{n+1} \cdot u \;\to\; u < 1$$

weil $\frac{n-\alpha}{n+1} = \frac{n}{n+1} - \frac{\alpha}{n+1} \to 1$. Damit ist auch $|R_{n+1}(x)|$ eine Nullfolge. □

Aus $\lim_{n\to\infty} R_{n+1}(x) = 0$ für eine Funktion f mit Entwicklungspunkt a folgt aber sofort

$$f(x) = \lim_{n\to\infty} f_n(x) = \sum_{n=0}^{\infty} \frac{f^{(k)}(a)}{k!}(x-a)^k,$$

d.h. die Funktion $f(x)$ ist eine Potenzreihe in der verschobenen Variablen $x - a$. In unserem Beispiel ergibt sich damit (mit der neuen Variablen $t = x - 1$):

Satz 15.4. Binomische Reihe: *Für alle* $t \in (-\frac{1}{2}, 1)$, *sogar für alle* $t \in (-1, 1)$, *und für alle* $\alpha \in \mathbb{R}$ *gilt:*

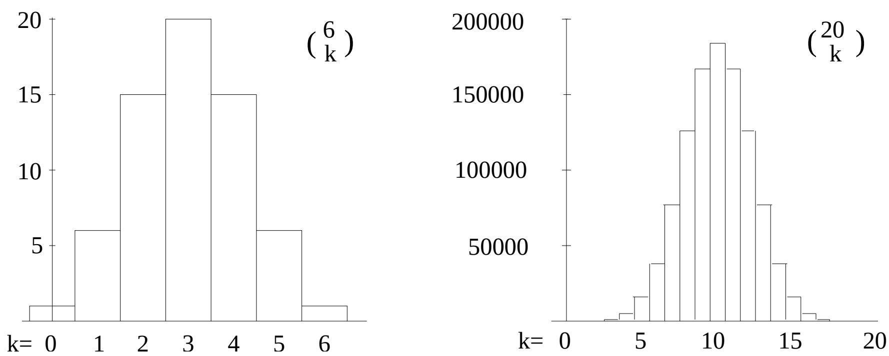
$$(1+t)^\alpha = \sum_{k=0}^{\infty} \binom{\alpha}{k} t^k \tag{74}$$

Der Satz verallgemeinert die bekannte *binomische Formel*

$$(1+t)^n = \sum_{k=0}^{n} \binom{n}{k} t^k \tag{75}$$

(vgl. "Zahl und Funktion", Satz 5.1, S.17 für $a = 1$, $b = t$).

16. Wahrscheinlichkeiten: Vom Zählen zum Messen



Die Wahrscheinlichkeit, bei einem Glücksspiel zu gewinnen, ist das Verhältnis der Anzahlen der *günstigen* und der *möglichen* Fälle. Als Beispiel betrachten wir die Wette, dass bei n Münzwürfen („Zahl" oder

„Wappen") genau k-mal „Zahl" fällt.[57] Jedes mögliche Ergebnis des Spiels ist eine Wurffolge, z.B.

Zahl - Zahl - Wappen - Zahl - Wappen - Wappen,

und die Nummern der Würfe, bei denen „Zahl" fällt, bilden eine beliebige Teilmenge der Menge $\{1, \dots, n\}$, hier z.B. $\{1, 2, 4\} \subset \{1, \dots, 6\}$. Die Menge der *möglichen* Ergebnisse ist also die Menge aller Teilmengen von $\{1, \dots, n\}$, deren Anzahl 2^n ist. Die *günstigen* Ergebnisse sind die Wurffolgen, bei denen genau k-mal „Zahl" gefallen ist, also die k-elementigen Teilmengen von $\{1, \dots, n\}$, deren Anzahl

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n \cdot (n-1) \cdot \ldots \cdot (n-k+1)}{k \cdot (k-1) \cdot \ldots \cdot 1} \tag{76}$$

beträgt. Die Wahrscheinlichkeit, die Wette zu gewinnen, ist also

$$p_n(k) = \binom{n}{k} / 2^n. \tag{77}$$

Zum Beispiel ist $p_6(3) = 20/64 = 31{,}25\%$ und $p_{20}(10) = 184756/(1024)^2 \approx 17{,}62\%$. Die durch (77) definierte Wahrscheinlichkeitsverteilung $p_n(k)$ heißt *Binomialverteilung*.[58]

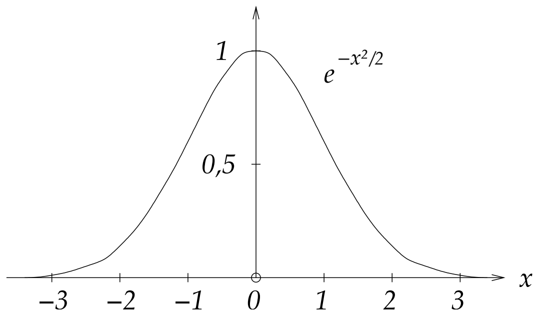
Die Wahrscheinlichkeit hat also mit *Zählen* zu tun, mit Anzahlen von Elementen von Mengen. Zählen kann man nur endlich viele unterscheidbare Objekte, wie Münzwurf-Folgen. Aber es gibt auch Wahrscheinlichkeiten bei Ereignissen, deren mögliche Ergebnisse eine unendliche Menge bilden, ein Kontinuum. Wenn Sie zum Beispiel mit Wurf-Pfeilen (wie bei Dart) das Zentrum einer Zielscheibe treffen wollen, so wird Ihnen das oft nicht gelingen, sondern Sie werden wahrscheinlich irgendwo knapp daneben liegen. Der Einfachheit halber wollen wir annehmen, dass nur die Abweichung nach rechts oder links zählt; die möglichen Ergebnisse Ihres Wurfs sind also auf der reellen Achse angeordnet, wobei der Nullpunkt das anvisierte Ziel sein soll. Die Wahrscheinlichkeit, in ein vorgegebenes Intervall $[a, b]$ zu treffen, hat natürlich etwas mit der Größe (Länge) des Intervalls zu tun, wobei der Maßstab (Maßeinheit), mit dem die Größe gemessen wird, von Ihrer individuellen Geschicklichkeit beim Werfen abhängt: Ein guter Werfer wird ein kleines Intervall eher treffen als ein ungeübter. Aber ein Intervall, das weit von der Null entfernt ist, werden Sie vermutlich nicht treffen, auch wenn es groß ist,

---

[57] vgl. „Zahl und Funktion", S. 16

[58] Man kann auch Münzwürfe betrachen, die „nicht fair" sind, bei denen also die Wahrscheinlichkeiten für „Zahl" und „Wappen" unterschiedlich ausfallen, etwa p und $q = 1 - p$. Dann ist $p_n(k) = \binom{n}{k} p^k q^{n-k}$; für $p = q = 1/2$ ergibt sich der alte Wert (77). Diese etwas allgemeinere Verteilungsfunktion wird gewöhnlich als Binomialverteilung bezeichnet, denn sie hat mit der binomischen Formel $(p+q)^n = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k}$ zu tun.

weil Sie ja ganz woanders hinzielen. Zur Bestimmung der Wahrscheinlichkeit, das Intervall $[a, b]$ zu treffen, wird also *gemessen* (die Größe ermittelt) und zugleich *gewichtet*; weit vom Nullpunkt entfernte Intervalle werden mit geringerer Wahrscheinlichkeit getroffen als solche, die den Nullpunkt enthalten.



Das mathematische Gesetz, nach dem dies geschieht, kennen Sie vermutlich: Es ist die berühmte *Gaußsche Verteilungsfunktion*, die zusammen mit einem Bild von C.F. Gauß[59] auf dem letzten 10-Mark-Schein abgebildet war:

$$g(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \tag{78}$$

Die Wahrscheinlichkeit $p([a, b])$, das Intervall $[a, b]$ zu treffen, ist nicht einfach proportional zur Länge von $[a, b]$, sondern sie ist der Flächeninhalt unter dem Graphen von $g|_{[a,b]}$, also $\int_a^b g(x)dx$. Diese Größe kann man als eine *gewichtete Länge* des Intervalls ansehen: Die Länge jedes kleinen Teilstücks wird mit dem Wert von g an der betreffenden Stelle multipliziert (gewichtet) und dann aufsummiert. Da Sie ja ganz sicher *irgendwo* hintreffen, ist die Wahrscheinlichkeit für die ganze Zahlengerade gleich 100 Prozent, und in der Tat[60] ist $p(\mathbb{R}) = \int_{-\infty}^{\infty} g(x)dx = 1$.

Der Grund für die universelle Gültigkeit dieses Gesetzes ist der „*Zentrale Grenzwertsatz*“ der Wahrscheinlichkeitsrechnung:

> *Sind $X_1, \ldots, X_n$ unabhängige Zufallsgrößen, die alle nach ein und demselben Gesetz verteilt („identisch verteilt“) sind, so ist ihre Summe $X_1 + \cdots + X_n$ nach Verschieben und Maßstabsänderung im Grenzübergang $n \to \infty$ stets nach dem Gaußschen Gesetz (78) verteilt.*

---

[59] Johann Carl Friedrich Gauß, 1777 (Braunschweig) - 1855 (Göttingen)

[60] vgl. „Mehrere Variable“, S. 61f.

Der Satz wurde zuerst im Spezialfall des Münzwurfes nachgewiesen (*Satz von Moivre und Laplace*).[61] In diesem Fall gehört zu einem einzelnen Münzwurf eine Zufallsgröße X mit $X = 1$ für „Zahl“ und $X = 0$ für „Wappen“. Zu n Münzwürfen gehören damit n identisch verteilte Zufallsgrößen $X_1, \dots, X_n$. Wenn dabei genau k-mal „Zahl“ fällt, dann ist $X_1 + \dots + X_n = k$. Gemäß dem zentralen Grenzwertsatz sollen diese Werte für große n nach dem Gaußschen Gesetz um den Mittelwert $m = n/2$ herum streuen. Andererseits wissen wir bereits, dass die Verteilung der Werte k durch die Funktion $p_n(k)$ gemäß (77) gegeben sind. Damit sollten die p_n (nach Verschieben und Maßstabsänderung) gegen g konvergierten. Dieser Fall ist deshalb besonders anschaulich, weil ja bereits unsere Figuren auf diese Konvergenz hindeuten: Die Ähnlichkeit der Graphen von p_n und g ist unübersehbar. Diese Konvergenz wollen wir nun präzise nachweisen.

Was haben die Funktionen g und p_n gemeinsam? Ich behaupte, dass ihre *Veränderungen* (Zu- und Abnahmen der Werte bei Änderung der Variablen x bzw. k) das gleiche Gesetz erfüllen. Für g ist dies die Differentialgleichung

$$g'(x) = -xg(x), \tag{79}$$

die man sofort durch Ableiten von g nach der Kettenregel gewinnt, und wir werden in gewissem Sinne das gleiche Gesetz für p_n, also für die Binimialkoeffizienten $\binom{n}{k}$ nachweisen.[62] Doch zuerst müssen wir diese sozusagen in die Mitte rücken, denn $\binom{n}{k}$ nimmt sein Maximum nicht wie g bei 0 an, sondern bei

$$m = n/2. \tag{80}$$

Deshalb betrachten wir nicht mehr $k \in \{0, \dots, n\}$ als Variable, sondern $l_k = k - m \in \{-m, \dots, m\}$. Wir betrachten also die *zentrierten Binomialkoeffizienten*

$$b_n(l_k) = \binom{n}{k} = \binom{n}{l_k + m}. \tag{81}$$

Diese erweitern wir durch lineare Interpolation zu einer Funktion auf ganz $\mathbb{R}$:

$$b_n(l_k + s) = s b_n(l_{k+1}) + (1 - s) b_n(l_k) \tag{82}$$

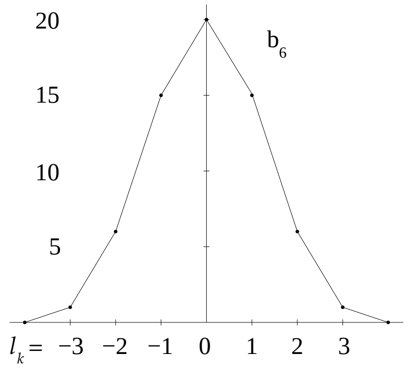
für alle $s \in [0, 1]$ und $l_k = k - m$, $k \in \mathbb{Z}$, wobei wir $b_n(l_k) = 0$ setzen für $k \in \mathbb{Z} \setminus \{0, \dots, n\}$. Insbesondere gilt für $s = \frac{1}{2}$:

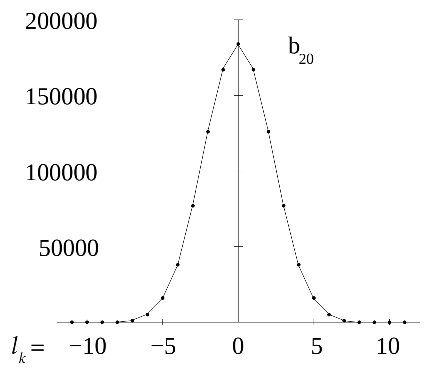
$$b_n(l_k + \tfrac{1}{2}) = \tfrac{1}{2}\left(b_n(l_k) + b_n(l_{k+1})\right). \tag{83}$$

---

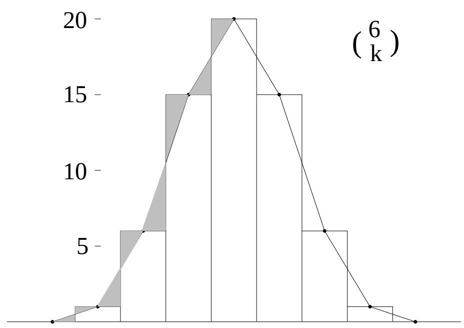
[61] Abraham de Moivre, 1667 (Vitry-le-François) – 1754 (London),
Pierre-Simon (Marquis de) Laplace, 1749 (Beaumont-en-Auge) – 1827 (Paris)

[62] Die Idee verdanken wir dem Buch
A. Büchter, H.-W. Henn: *Elementare Stochastik*, Springer 2005, S. 280 - 282.





Bei der Interpolation wurde die Fläche unter dem Graphen nicht verändert, wie die folgende Zeichnung deutlich macht; die kleinen Dreiecke ober- und unterhalb des interpolierten Graphen (b_6) haben den gleichen Flächeninhalt.



Die Ableitung von b_n ist zwischen benachbarten Stützpunkten l_k und $l_{k+1} = l_k + 1$ ist die Steigung der Geraden $b_n|_{[l_k, l_k+1]}$, also gilt für alle $x \in [l_k, l_{k+1}]$:

$$b'_n(x) = b_n(l_{k+1}) - b_n(l_k). \tag{84}$$

Das folgende Lemma zeigt, dass genau in der Mitte zwischen den Stützpunkten (bis auf einen konstanten Faktor) das gleiche Gesetz wie in (79) gilt:

Lemma 16.1. *Für jedes* $x = l_k + \frac{1}{2}$ *mit* $k \in \{0, \dots, n-1\}$ *gilt*

$$b'_n(x) = -\tfrac{4}{n+1}\, x\, b_n(x) \tag{85}$$

Beweis. Beim Übergang von $\binom{n}{k}$ zu $\binom{n}{k+1}$ tritt im Zähler der Faktor $(n-k)$, im Nenner der Faktor $(k+1)$ hinzu, also ist

$$\begin{aligned} \binom{n}{k+1} &= \binom{n}{k} \tfrac{n-k}{k+1}\,, \\ \binom{n}{k+1} - \binom{n}{k} &= \binom{n}{k} \tfrac{n-2k-1}{k+1}\,, && (86) \\ \binom{n}{k+1} + \binom{n}{k} &= \binom{n}{k} \tfrac{n+1}{k+1}\,. && (87) \end{aligned}$$

Damit folgt für $x = l_k + \frac{1}{2} = k - m + \frac{1}{2} = -\frac{1}{2}(n - 2k - 1)$:

$$\begin{aligned} b_n'(x) &= b_n(l_{k+1}) - b_n(l_k) &\overset{(86)}{=}& -2x\, b_n(l_k)\, \tfrac{1}{k+1}\,, \\ b_n(x) &= \tfrac{1}{2}\,(b_n(l_{k+1}) + b_n(l_k)) &\overset{(87)}{=}& \tfrac{1}{2}\, b_n(l_k)\, \tfrac{n+1}{k+1}\,, \end{aligned}$$

woraus sich (85) ergibt.[63] □

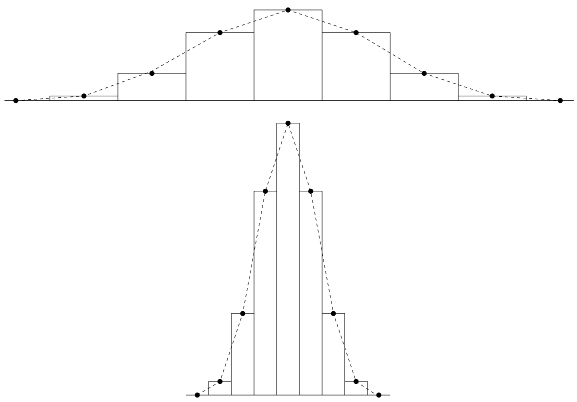
Unsere Wahrscheinlichkeitsverteilung $p_n(k) = \binom{n}{k}/2^n$ werden wir ebenso verschieben und interpolieren wie die Binomialkoeffizienten:

$$\tilde{p}_n := b_n/2^n. \tag{88}$$

Weil $\tilde{p}_n$ nur ein konstantes Vielfaches von b_n ist, erfüllt es ebenso wie b_n das Gesetz (85). Ein Problem sind allerdings die unterschiedlichen Maßstäbe. Der Maximalwert $p_n(m) = \tilde{p}_n(0)$ geht nämlich gegen Null für $n \to \infty$, weil die Potenz 2^n im Nenner alles überwiegt. Dies müssen wir durch Änderung des „vertikalen" Maßstabes verhindern: Wir multiplizieren $\tilde{p}_n$ mit einem Faktor σ_n mit der Eigenschaft

$$\sigma_n \tilde{p}_n(0) \to g(0) = 1/\sqrt{2\pi}. \tag{89}$$

Aber damit handeln wir uns ein anderes Problem ein: Die Summe der Werte an den Stützstellen oder die Fläche unter den Graphen ist für $\sigma_n p_n$ nicht mehr Eins, sondern σ_n. Dies können wir korrigieren, indem wir die Fläche unter $\tilde{p}_n$ nicht nur um den Faktor σ_n höher, sondern gleichzeitig um $1/\sigma_n$ schmaler machen, um den Flächeninhalt zu erhalten:



[63]Die Gleichung (85) gilt sogar für alle $k \in \mathbb{Z}$. Für $k = -1$ ist $l_k = -m - 1$ und $x = -m - \frac{1}{2} = -\frac{1}{2}(n + 1)$ und $b_n(x) = \frac{1}{2}(0 + 1) = \frac{1}{2}$ und $b_n'(x) = 1 - 0 = 1$, also ist $-\frac{4}{n+1}\, x\, b_n(x) = 1 = b_n'(x)$. Ähnliches gilt auf der anderen Seite bei $k = n$. Für alle übrigen k sind beide Seiten von (85) gleich Null.

Wir gehen dabei von $\tilde{p}_n$ über zu der Funktion[64]

$$g_n(x) = \sigma_n \tilde{p}_n(\sigma_n x). \tag{90}$$

Allerdings wird dadurch die Gleichung (85), die ja auch für $\tilde{p}_n$ gilt, um den Faktor σ_n^2 verändert:

$$\begin{aligned} g_n'(x) &= \sigma_n^2 \,\tilde{p}_n'(\sigma_n x) \\ &= -\frac{4\sigma_n^2}{n+1} \sigma_n \, x \, \tilde{p}_n(\sigma_n x) \\ &= -\frac{4\sigma_n^2}{n+1} \, x \, g_n(x) \end{aligned} \tag{91}$$

Welchen Wert soll σ_n haben? Dazu müssen wir die Werte $P_n := p_n(m) = \tilde{p}_n(0)$ für große n berechnen:

Lemma 16.2. *Für $n \to \infty$ gilt*

$$\tfrac{1}{2} \sqrt{n+1} \, p_n(m) \to 1/\sqrt{2\pi}. \tag{92}$$

Beweis. Wir beschränken uns auf den Fall, dass n gerade ist, $n = 2m$. Dann ist

$$\begin{aligned} P_n = \binom{2m}{m} / 2^{2m} &= \frac{(2m)!}{m! \cdot m!} \cdot \frac{1}{2^m \cdot 2^m} \\ &= \frac{(2m)!}{2^m \cdot m! \cdot 2^m \cdot m!} \\ &= \frac{1 \cdot 2 \cdot \ldots \cdot 2m}{(2 \cdot 4 \cdot \ldots \cdot 2m)^2} \\ &= \frac{1 \cdot 3 \cdot \ldots \cdot (2m-1)}{2 \cdot 4 \cdot \ldots \cdot 2m} \end{aligned}$$

Diese Zahl ist mit dem *Wallisschen Produkt* verbunden:[65]

$$\begin{aligned} \pi/2 &= \lim_{m\to\infty} w_m, \\ w_m &= \frac{2 \cdot 2 \cdot 4 \cdot 4 \cdot \quad \ldots \quad \cdot 2m \cdot 2m}{1 \cdot 3 \cdot 3 \cdot 5 \cdot \ldots \cdot (2m-1)(2m+1)} \\ &= \frac{(2 \cdot 4 \cdot \ldots \cdot 2m)^2}{(1 \cdot 3 \cdot \ldots \cdot (2m-1))^2} \cdot \frac{1}{2m+1} \\ &= \frac{1}{(P_n)^2 \cdot (2m+1)}. \end{aligned}$$

[64]Die Streifen von g_n sind nach (90) um den Faktor $1/\sigma_n$ schmaler als die von b_n oder $\tilde{p}_n$, denn wenn x um $1/\sigma_n$ wächst, dann liegt $\tilde{x} = x + (1/\sigma_n)$ bereits im nächsten Streifen, weil $\sigma_n \tilde{x} = \sigma_n x + 1$.

[65]Siehe S. 41

Also folgt $(2m+1)P_n^2 \to 2/\pi$ und damit $\frac{1}{2}\sqrt{n+1}\cdot P_n \to \frac{1}{2}\sqrt{2/\pi} = 1/\sqrt{2\pi}$. $\square$

Wir müssen also
$$\sigma_n = \tfrac{1}{2}\sqrt{n+1} \tag{93}$$
setzen, dann gilt
$$g_n(0) = \sigma_n P_n \to 1/\sqrt{2\pi} = g(0) \tag{94}$$
für $n \to \infty$. Gleichzeitig haben wir in der Gleichung (91) den Vorfaktor $\frac{4}{n+1}$ beseitigt, und in der Mitte zwischen den neuen Stützstellen $x_k = (k-m)/\sigma_n$, d.h. bei
$$\xi_k = (k - m + \tfrac{1}{2})/\sigma_n = x_k + 1/(2\sigma_n)$$
(für beliebige $k \in \mathbb{Z}$) gilt damit genau dieselbe Gleichung für g_n wie für g (vgl. (79)):
$$g_n'(\xi_k) = -\xi_k\, g_n(\xi_k). \tag{95}$$

Satz 16.1. *Es gilt $g_n \to g$ und $g_n' \to g'$ gleichmäßig auf jedem Intervall $[-R, R]$.*

Beweis. Wir vergleichen die Ausdrüche g_n'/g_n und g'/g (*logarithmische Ableitungen* von g_n und g) auf dem Intervall $(-m-1, m+1)$, wo $g_n > 0$. An den Mittelstellen ξ_k nehmen sie nach (95) beide den gleichen Wert $-\xi_k$ an. Im Intervall (x_k, x_{k+1}) ist der Graph von g_n eine Gerade mit Steigung $a_n = g_n'(\xi_k) = -\xi_k g_n(\xi_k)$, also gilt für alle $x \in (x_k, x_{k+1})$:
$$\begin{aligned}
g_n(x) &= g_n(\xi_k) + a_n(x - \xi_k)\\
\frac{g_n(x)}{g_n'(x)} &= \frac{g_n(\xi_k)}{a_n} + (x - \xi_k)\\
&= -\xi_k^{-1} + (x - \xi_k)\\
&= -\xi_k^{-1}(1 + \xi_k(x - \xi_k))\\
\frac{g_n'(x)}{g_n(x)} &= -\xi_k\,(1 + \xi_k(x - \xi_k))^{-1}\\
&\approx -\xi_k\,(1 - \xi_k(x - \xi_k))\\
&= -\xi_k + \xi_k^2(x - \xi_k)
\end{aligned} \tag{96}$$
wobei die vorletzte Zeile aus der Taylorentwicklung der Funktion $t \mapsto (1+t)^\alpha$ kommt:
$$(1+t)^\alpha \approx 1 + \alpha t \tag{$*$}$$
wenn $|t|$ genügend klein ist.[66] Wir wenden dies an auf $\alpha = -1$ und $t = \xi_k(x - \xi_k)$ mit
$$|t| \le |\xi_k||x - \xi_k| \le |\xi_k||x_k - \xi_k| \le R/(2\sigma_n).$$

[66]Insbesondere bedeutet $(*)$, dass $|(1+t)^\alpha - 1| \le A|t|$, wobei die Konstante A etwas größer als $|\alpha|$ ist, $A = |\alpha| + \epsilon$.

Andererseits ist

$$\frac{g'(x)}{g(x)} = -x = -\xi_k - (x - \xi_k) \tag{97}$$

und mit (96) ergibt sich nun

$$\left|\frac{g_n'(x)}{g_n(x)} - \frac{g'(x)}{g(x)}\right| \approx |(\xi_k^2 + 1)(x - \xi_k)| \leq C \cdot |x - \xi_k| \tag{98}$$

für eine Konstante C, die größer ist als $|\xi_k^2 + 1|$ und dabei auch noch den Fehler der Approximation „$\approx$" schluckt. Andererseits ist $g_n'/g_n = \ln(g_n)'$ und $g'/g = \ln(g)'$. Setzen wir also

$$f_n = \ln g_n, \quad f = \ln g$$

so ergibt sich aus (98)

$$|f_n' - f'| \leq C\sigma_n$$

und durch Integration der Differenz $h_n = f_n - f$ mit $h_n(\xi_k) = 0$:

$$|h_n(x)| = \left|\int_{\xi_k}^x h_n'\right| \leq \int_{\xi_k}^x |h_n'| \leq \int_{\xi_k}^x (C\sigma_n) = C\sigma_n^2.$$

Also konvergieren $f_n \to f$ und $f_n' \to f'$ gleichmäßig in $[-R, R]$, und es folgt

$$\begin{array}{ccccccc} g_n & = & e^{f_n} & \to & e^f & = & g\,, \\ g_n' & = & f_n' e^{f_n} & \to & f'e^f & = & g' \end{array} \qquad \square$$

Index

Inhaltsverzeichnis