Also ist

$$s(t + 1\,sec) \approx s(t) + v(t) \cdot 1\,sec\,.$$

Da wir sowohl $s(0\,sec) = 0\,m$ als auch die Werte $v(t)$ für $t = 0, 1, 2, 3$ und 4 *sec* kennen, können wir $s(5\,sec)$ rekursiv berechnen als

$$\begin{aligned} s(5\,sec) &\approx v(0\,sec) \cdot 1\,sec + v(1\,sec) \cdot 1\,sec + v(2\,sec) \cdot 1\,sec \\ &\quad + v(3\,sec) \cdot 1\,sec + v(4\,sec) \cdot 1\,sec \\ &= 2{,}5\,m + 6{,}1\,m + 9{,}1\,m + 11{,}5\,m + 13{,}4\,m = 42{,}6\,m\,. \end{aligned}$$

Also hat das Fahrzeug in fünf Sekunden 42,6 m zurückgelegt?

Zumindest mit dieser Genauigkeit ist das sicherlich falsch, denn bei der Berechnung des Wegs sind wir davon ausgegangen, daß das Fahrzeug während der ersten Sekunde konstant mit einer Geschwindigkeit von 2,5 m/sec gefahren ist, exakt ab deren Ende dann aber plötzlich eine Sekunde lang mit 6,1 m/sec, *usw.* Das ist natürlich unrealistisch; tatsächlich dürfte die Geschwindigkeit etwa so verlaufen sein, wie es die Kurve in Abbildung 45 angibt, wohingegen wir mit dem ebenfalls eingezeichneten stufenförmigen Verlauf gerechnet haben. Wir haben die Geschwindigkeit somit fast durchgängig unterschätzt und damit auch einen zu kleinen Weg berechnet.

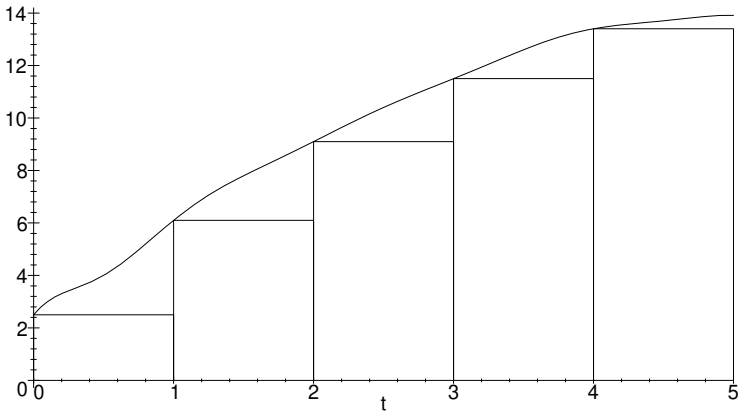


Abb. 45: Tatsächliche und unterschätzte Geschwindigkeit

Alternativ hätten wir auch einen *zu großen* Weg schätzen können, wenn wir die Geschwindigkeit während eines Sekundenintervalls jeweils auf den bekannten Wert am *Ende* dieses Intervalls gesetzt hätten; das Ergebnis wäre dann gewesen

$$\begin{aligned} s(5\,sec) &\approx v(1\,sec) \cdot 1\,sec + v(2\,sec) \cdot 1\,sec + v(3\,sec) \cdot 1\,sec \\ &\quad + v(4\,sec) \cdot 1\,sec + v(5\,sec) \cdot 1\,sec \\ &= 6{,}1\,m + 9{,}1\,m + 11{,}5\,m + 13{,}4\,m + 13{,}9\,m = 54{,}0\,m\,. \end{aligned}$$

Tatsächlich wissen wir im Augenblick also nur, daß der zurückgelegte Weg irgendwo zwischen 42,6 m und 54 m liegt.

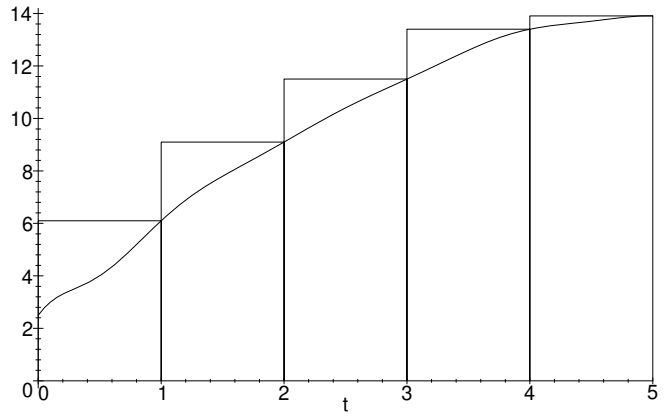


Abb. 46: Tatsächliche und überschätzte Geschwindigkeit

Eine bessere Schätzung bekommen wir, wenn wir anstelle der Sekundenintervalle Intervalle von nur einer halben Sekunde betrachtet – vorausgesetzt natürlich, wir kennen die Geschwindigkeit auch für halbzahlige Sekundenwerte. Die entsprechenden Werte seien die in der folgenden Tabelle angegebenen:

t [sec]	0	0,5	1	1,5	2	2,5	3	3,5	4	4,5	5
v [m/sec]	2,5	4,4	6,1	7,8	9,1	10,4	11,5	12,6	13,4	13,8	13,9

Jetzt ist die „unterschätzte“ Wegstrecke

$$\sum_{i=0}^{9} v\left(\frac{i}{2}\,sec\right)\cdot\frac{1}{2}\,sec = 45{,}6\,m\,,$$

und die „überschätzte“ ist

$$\sum_{i=1}^{10} v\left(\frac{i}{2}\,sec\right)\cdot\frac{1}{2}\,sec = 51{,}3\,m\,,$$

die Unsicherheit hat sich also etwa halbiert. Abbildung 47 zeigt den Verlauf von unterschätzter, tatsächlicher und überschätzter Geschwindigkeit für die Halbsekundenintervalle.

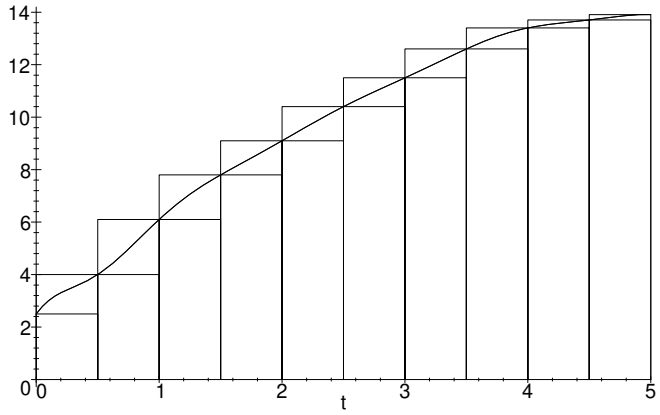


Abb. 47: Abtastung in Intervallen von einer halben Sekunde

Beide Werte wie auch ihre Differenz lassen sich anhand von Abbildung 47 leicht geometrisch veranschaulichen: Die unterschätzte Geschwindigkeit ist die Fläche der Rechtecke unterhalb der unteren Treppenfunktion, die überschätzte entsprechend die Fläche der Rechtecke unterhalb der oberen Treppenfunktion, wobei die Basis der Rechtecke jeweils auf der Zeitachse liegt. Die Differenz ist somit gleich der Fläche der Differenzrechtecke.

Diese können wir durch weitere Verfeinerung der Abtastung verkleinern. Dazu nehmen wir an, die eingezeichnete Kurve (die tatsächlich einfach eine notdürftig angepaßte Polynomfunktion darstellt) gebe den tatsächlichen Geschwindigkeitsverlauf wieder, und lassen den Computer rechnen:

Wenn wir die Geschwindigkeit viermal pro Sekunde bestimmen und den zurückgelegten Weg damit schätzen, erhalten wir als untere Schranke

$$\sum_{i=0}^{19} v\left(\frac{i}{4}\,sec\right)\cdot\frac{1}{4}\,sec \approx 47{,}14\,m$$

und als obere Schranke

$$\sum_{i=1}^{20} v\left(\frac{i}{4}\,sec\right)\cdot\frac{1}{4}\,sec \approx 49{,}99\,m\,.$$

Mit zehn Geschwindigkeitswerten pro Sekunde verbessert sich die untere Schranke auf

$$\sum_{i=0}^{49} v\left(\frac{i}{10}\,sec\right)\cdot\frac{1}{10}\,sec \approx 48{,}02\,m$$

und die obere auf

$$\sum_{i=1}^{50} v\left(\frac{i}{10}\,sec\right)\cdot\frac{1}{10}\,sec \approx 49{,}16\,m\,.$$

Wie Abbildung 48 zeigt, unterscheiden sich nun die Rechtecke der unteren Abschätzung nur noch wenig von denen der oberen, und die Fläche unter der Geschwindigkeitskurve liegt schon recht nahe bei der Fläche der Rechtecke aus jeder der beiden Abschätzungen.

Gehen wir von den Zehntelsekunden zu den Hundertsteln, sind die Rechtecke zumindest visuell in Abbildung 49 nicht mehr voneinander und von der Fläche unter der Kurve zu unterscheiden: Daß man hier keine durchweg schwarze Fläche sieht, liegt ausschließlich an sogenannten *alias*-Effekten, d.h. Diskretisierungsfehlern der Rastergraphik.

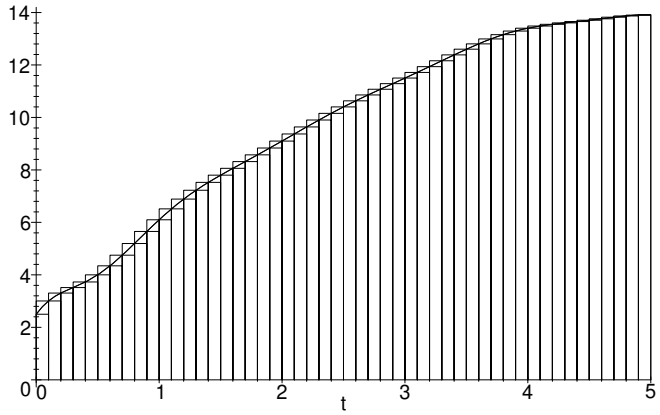


Abb. 48: Abtastung in Intervallen von einer Zehntelsekunde

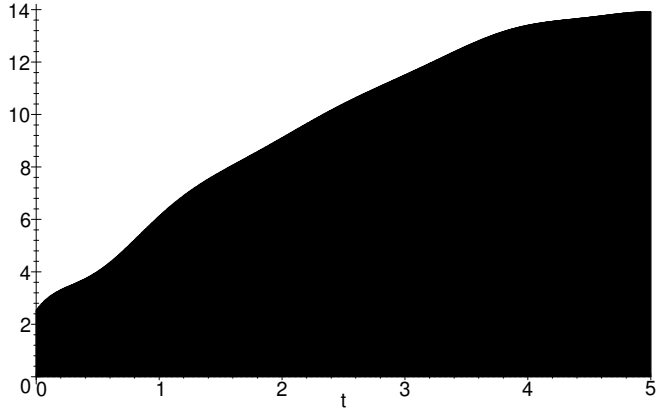


Abb. 49: Abtastung in Intervallen von einer Hundertstelsekunde

Rechnerisch gibt es noch einen kleinen Unterschied im Dezimeterbereich:

$$\sum_{i=0}^{499} v\left(\frac{i}{100}\,sec\right)\cdot\frac{1}{100}\,sec \approx 48{,}54\,m$$

und

$$\sum_{i=1}^{500} v\left(\frac{i}{100}\,sec\right)\cdot\frac{1}{100}\,sec \approx 48{,}66\,m\,.$$

Bei einer nochmaligen Verzehnfachung der Intervallanzahl ist natürlich graphisch nichts neues mehr zu sehen; die Schätzwerte verbessern sich auf

$$\sum_{i=0}^{4999} v\left(\frac{i}{1000}\,sec\right)\cdot\frac{1}{1000}\,sec \approx 48{,}595\,m$$

für die untere Schranke und

$$\sum_{i=1}^{5000} v\left(\frac{i}{1000}\,sec\right)\cdot\frac{1}{1000}\,sec \approx 48{,}607\,m$$

für die obere.

Um die eingangs gestellte Frage nach dem zurückgelegten Weg millimetergenau beantworten zu können (sofern man diese Genauigkeit wirklich als sinnvoll betrachtet), muß die Geschwindigkeitskurve 10 000 Mal pro Sekunde abgetastet werden, und wir erhalten die Schranken

$$\sum_{i=0}^{49999} v\left(\frac{i}{10000}\,sec\right)\cdot\frac{1}{10000}\,sec \approx 48{,}6006\,m$$

und

$$\sum_{i=1}^{50000} v\left(\frac{i}{10000}\right)\,sec\cdot\frac{1}{10000}\,sec \approx 48{,}6017\,m\,;$$

das Fahrzeug legte in fünf Sekunden also 48 m und 60,1 cm zurück.

2) Integration als Flächenbestimmung: Die Abbildungen 48 und vor allem 49 zeigen, daß die immer feiner werdenden Rechtecke, die wir für die obigen Abschätzungen wählten, die Fläche unter der Kurve $v = v(t)$ immer besser annähern. Wenn wir die Einheiten vergessen und t wie auch v als Längen ansehen, können wir die oben betrachteten Ausdrücke

$$\sum_{i=0}^{5n-1} v\left(\frac{i}{n}\right)\cdot\frac{1}{n} \quad \text{und} \quad \sum_{i=1}^{5n} v\left(\frac{i}{n}\right)\cdot\frac{1}{n}$$

also auch als Näherungswerte für die Fläche unter der Kurve ansehen und den Flächeninhalt als ihren gemeinsamen Grenzwert für immer größer werdendes n berechnen.

Genau genommen handelt es sich hier allerdings nicht um die *Berechnung* dieser Fläche, sondern um die *Definition* des Flächeninhalts, denn es gibt schließlich *a priori* keinen Begriff des Flächeninhalts einer krummlinig begrenzten Fläche. Die hier gewählte Definition über eine Ausschöpfung mit immer feiner werdenden Rechtecken ist zwar sehr natürlich, aber nicht zwangsläufig. Sie ist übrigens weit älter als jeder Begriff eines Integrals oder auch nur Grenzwerts: Bereits vor über zwei Jahrtausenden, um etwa 370 vor Christus, definierte der griechische Mathematiker EUDOXOS VON CNIDOS (*ca.* 408–*ca.* 355), ein Schüler und späterer Konkurrent PLATOS, Flächeninhalte auf diese Weise; später hat ARCHIMEDES VON SYRAKUS (287–212) diese Methode perfektioniert und angewandt auf Kreise und Parabeln; insbesondere berechnete er damit seine Abschätzung $223/71 < \pi < 22/7$ für π, in Dezimalzahlen ausgedrückt $3{,}1408 < \pi < 3{,}1429$, wobei diese Abschätzung allerdings nicht auf Rechtecken beruht, sondern auf der Konstruktion des regelmäßigen 96-Ecks.

3) Integration als Durchschnittsbestimmung: Kehren wir wieder zurück zum Eingangsbeispiel eines sich beschleunigenden Fahrzeugs, und fragen wir uns, wie hoch die *Durchschnittsgeschwindigkeit* war. Bei endlich vielen Werten ist der Durchschnitt natürlich einfach das arithmetische Mittel, also z.B.

$$\begin{aligned}&\frac{1}{5}\cdot\big(v(0\,sec)+v(1\,sec)+v(2\,sec)+v(3\,sec)+v(4\,sec)\big)\\ =\;&\frac{1}{5}\cdot 42{,}6\,m/sec = 8{,}52\,m/sec\,,\end{aligned}$$

wenn wir die Geschwindigkeiten vom *Anfang* jedes Sekundenintervalls nehmen, und

$$\begin{aligned}&\frac{1}{5}\cdot\big(v(1\,sec)+v(2\,sec)+v(3\,sec)+v(4\,sec)+v(5\,sec)\big)\\ =\;&\frac{1}{5}\cdot 54{,}0\,m/sec = 10{,}8\,m/sec\,,\end{aligned}$$

wenn wir die vom Ende nehmen.

Natürlich läßt sich auch hier die Abtastung immer weiter verfeinern; die entstehenden Ausdrücke

$$\frac{1}{5n}\cdot\sum_{i=0}^{5n-1} v\left(\frac{i}{n}\,sec\right) \quad \text{und} \quad \frac{1}{5n}\cdot\sum_{i=1}^{5n} v\left(\frac{i}{n}\,sec\right)$$

entsprechen bis auf einen Faktor von $\frac{1}{5\,sec}$ genau denen aus Teil *a),* wo wir den Weg abgeschätzt haben – wie es ja auch in der Tat der Fall sein muß: Die Durchschnittsgeschwindigkeit ist schließlich nichts anderes als der zurückgelegte Weg dividiert durch die zugrundeliegende Zeitspanne von fünf Sekunden.

Trotzdem ist die Interpretation eines Integrals als Durchschnitt gelegentlich von unabhängigem Interesse, da vor allem bei Anwendungen in der Quantenphysik oft zwar Durchschnitte existieren, aber keine „Zähler“ und „Nenner“, als deren Quotienten man sie interpretieren könnte.

b) Integration elementarer Funktionen

Abstrahieren wir vom Beispiel der Wegberechnung anhand einer Geschwindigkeitskurve und betrachten wir das allgemeine Problem, zu einer gegebenen reellwertigen Funktion f eine neue Funktion F zu finden derart, daß $F'(x) = f(x)$ ist.

Da jede konstante Funktion die Ableitung Null hat, ist mit F für jede reelle Zahl c auch $F + c$ eine solche Funktion; um ein konkretes F hinzuschreiben, müssen wir also einen Funktionswert von F festlegen; der Einfachheit halber sei dies, in Analogie zum obigen Beispiel, der Wert $F(0) = 0$.

Damit übernimmt f die Rolle der Geschwindigkeit v, dem Weg entspricht die Funktion F, die Variable ist nun x anstelle der Zeit t, und da uns F an einer beliebigen Stelle x interessiert, nicht nur wie im obigen Beispiel an der Stelle 5, empfiehlt es sich, n jetzt als Gesamtzahl der Intervalle zu nehmen, nicht wie oben als Anzahl der Intervalle pro Einheit (Sekunde). Die Länge eines jeden Intervalls wird dann zu x/n, und die

Analoga zu obigen Summen sind die Ausdrücke

$$\sum_{i=0}^{n-1} f\left(\frac{ix}{n}\right)\cdot\frac{x}{n} \quad \text{und} \quad \sum_{i=1}^{n} f\left(\frac{ix}{n}\right)\cdot\frac{x}{n}\,,$$

deren gemeinsamer Grenzwert für immer größer werdendes n der gesuchte Wert $F(x)$ sein sollte.

Wir *definieren* daher

$$F(x) \underset{\text{def}}{=} \lim_{n\to\infty}\sum_{i=0}^{n-1} f\left(\frac{ix}{n}\right)\cdot\frac{x}{n} \qquad (*)$$

in der Hoffnung, daß dieser Grenzwert existiert und mit dem anderen übereinstimmt – zumindest in hinreichend vielen interessanten Fällen. Bevor wir uns im nächsten Paragraphen dieser Frage zuwenden, wollen wir hier zunächst etwas mit dieser vorläufigen Definition spielen und sehen, was sie bei einfachen Funktionen liefert.

1) Die Funktion $f(x) = x^2$***:*** Hier erhalten wir

$$\begin{aligned} F(x) &\underset{\text{def}}{=} \lim_{n\to\infty}\sum_{i=0}^{n-1} f\left(\frac{ix}{n}\right)\cdot\frac{x}{n} = \lim_{n\to\infty}\sum_{i=0}^{n-1}\left(\frac{ix}{n}\right)^2\cdot\frac{x}{n} \\ &= x^3\lim_{n\to\infty}\frac{1}{n^3}\sum_{i=0}^{n-1} i^2\,. \end{aligned}$$

Wie man in seiner Formelsammlung nachschlägt oder sich von seinem Computeralgebrasystem sagen läßt, ist

$$\sum_{i=0}^{n-1} i^2 = \frac{n(2n-1)(n-1)}{6}\,,$$

eine Formel, die dann nachträglich leicht durch vollständige Induktion bewiesen werden kann:

Für $n = 1$ steht links und rechts Null, die Formel ist also korrekt.

Angenommen, wir hätten sie bewiesen für $n-1$ anstelle von n, d.h. wir wüßten, daß für ein spezielles $n > 1$

$$\sum_{i=0}^{n-2} i^2 = \frac{(n-1)(2n-3)(n-2)}{6}$$

ist. Dann folgt

$$\begin{aligned} \sum_{i=0}^{n-1} i^2 &= \frac{(n-1)(2n-3)(n-2)}{6} + (n-1)^2 \\ &= \frac{n-1}{6}\cdot\big((2n-3)(n-2)+6(n-1)\big) \\ &= \frac{n-1}{6}\cdot(2n^2-7n+6+6n-6) \\ &= \frac{n-1}{6}\cdot(2n^2-n) = \frac{(n-1)(2n-1)n}{6} \\ &= \frac{n(2n-1)(n-1)}{6}\,, \end{aligned}$$

d.h. die Formel für n selbst.

Nach dem Prinzip der vollständigen Induktion ist diese damit für *alle* Werte von n bewiesen; wir können sie in obige Rechnung einsetzen und erhalten

$$\begin{aligned} F(x) &= x^3\lim_{n\to\infty}\frac{1}{n^3}\cdot\frac{n(2n-1)(n-1)}{6} = x^3\lim_{n\to\infty}\frac{n(2n-1)(n-1)}{6n^3} \\ &= \frac{x^3}{6}\lim_{n\to\infty}\left(\frac{n}{n}\right)\left(\frac{2n-1}{n}\right)\left(\frac{n-1}{n}\right) = \frac{x^3}{6}\cdot 2 = \frac{x^3}{3}\,. \end{aligned}$$

Diese Funktion hat tatsächlich die Ableitung $F'(x) = x^2$, zumindest in diesem Beispiel funktioniert also die Definition $(*)$.

2) Die Exponentialfunktion: Versuchen wir dasselbe nochmal für die Funktion $f(x) = e^x$. Hier führt $(*)$ (wieder mit der Festlegung $F(0) = 0$)

auf den Ansatz

$$F(x) \underset{\text{def}}{=} \lim_{n\to\infty} \sum_{i=0}^{n-1} f\left(\frac{ix}{n}\right) \cdot \frac{x}{n} = \lim_{n\to\infty} \sum_{i=0}^{n-1} e^{\frac{ix}{n}} \cdot \frac{x}{n}$$
$$= x \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \left(e^{\frac{x}{n}}\right)^i .$$

Die Summe ganz rechts ist eine geometrische Reihe und kann leicht ausgerechnet werden: Setzen wir zur Abkürzung $q = e^{x/n}$, so ist

$$\begin{aligned} \sum_{i=0}^{n-1} q^i &= 1 + q + q^2 + q^3 + \cdots + q^{n-1} \\ q\sum_{i=0}^{n-1} q^i &= \quad\; q + q^2 + q^3 + \cdots + q^{n-1} + q^n \\ \hline \Rightarrow (1-q)\sum_{i=0}^{n-1} q^i &= 1 \qquad\qquad\qquad\qquad\qquad\qquad - q^n . \end{aligned}$$

Für $q \neq 1$ können wir durch $1 - q$ dividieren und erhalten die Formel

$$\sum_{i=0}^{n-1} q^i = \frac{1-q^n}{1-q} .$$

Da uns nur der Fall $x \neq 0$ interessiert, ist $q = e^{\frac{x}{n}} \neq 1$; wir können die Formel also anwenden und erhalten

$$F(x) = x \lim_{n\to\infty} \frac{1}{n} \cdot \frac{1 - \left(e^{\frac{x}{n}}\right)^n}{1 - e^{\frac{x}{n}}} = x \lim_{n\to\infty} \frac{1}{n} \cdot \frac{1 - e^x}{1 - e^{\frac{x}{n}}}$$
$$= x(1 - e^x) \lim_{n\to\infty} \frac{\frac{1}{n}}{1 - e^{\frac{x}{n}}} .$$

Im Grenzwert ganz rechts gehen für $n \to \infty$ leider sowohl der Zähler als auch der Nenner gegen Null, wir müssen also die Regel von DE L'HOSPITAL anwenden. Diese gilt zwar nur für Grenzwerte von Werten stetiger Funktionen, während wir hier den Grenzwert einer diskreten

Folge suchen, aber wenn für eine reelle Variable u der Grenzwert

$$\lim_{u\to\infty} \frac{\frac{1}{u}}{1 - e^{\frac{x}{u}}}$$

existiert, konvergiert natürlich auch die Folge der Zahlen

$$\frac{\frac{1}{n}}{1 - e^{\frac{x}{n}}} \quad \text{für } n = 1, 2, 3, \ldots$$

gegen eben diesen Grenzwert.

Nach DE L'HOSPITAL ist

$$\lim_{u\to\infty} \frac{\frac{1}{u}}{1 - e^{\frac{x}{u}}} = \lim_{u\to\infty} \frac{\frac{d}{du}\frac{1}{u}}{\frac{d}{du}(1 - e^{\frac{x}{u}})} = \lim_{u\to\infty} \frac{\frac{-1}{u^2}}{-e^{\frac{x}{u}}\left(\frac{-x}{u^2}\right)}$$
$$= \lim_{u\to\infty} \frac{-1}{xe^{\frac{x}{u}}} = \frac{-1}{x} .$$

Eingesetzt in den Ansatz für $F(x)$ führt dies auf

$$F(x) = x(1 - e^x) \cdot \frac{-1}{x} = e^x - 1 ,$$

eine Funktion, deren Ableitung in der Tat e^x ist.

3) Die Dirichletsche Sprungfunktion: Bevor wir zu übermütig werden, möchte ich als letztes Beispiel noch eine Funktion betrachten, die sich im Gegensatz zu Quadrat und Exponentialfunktion alles andere als gut verhält, die DIRICHLETsche Sprungfunktion

$$f(x) = \begin{cases} 1 & \text{für } x \in \mathbb{Q} \\ 0 & \text{für } x \notin \mathbb{Q} \end{cases} .$$

Mit $F(0) \underset{\text{def}}{=} 0$ ist wieder

$$F(x) \underset{\text{def}}{=} \lim_{n\to\infty} \sum_{i=0}^{n-1} f\left(\frac{ix}{n}\right) \cdot \frac{x}{n} .$$

Dabei ist die Zahl ix/n für $i = 0$ rational, da Null; für $i \neq 0$ ist sie genau dann rational, wenn auch x rational ist. In diesem Fall ist also

$f(ix/n) = 1$ für alle i, d.h.

$$F(x) = \lim_{n\to\infty} \sum_{i=0}^{n-1} \frac{x}{n} = \lim_{n\to\infty} n \cdot \frac{x}{n} = x$$

für rationales x. Für irrationales x dagegen sind alle ix/n außer der Null irrational, also ist $f(ix/n) = 0$ für alle $i \neq 0$. Damit bleibt von der Summe nur der erste Summand stehen, d.h.

$$F(x) = \lim_{n\to\infty} \frac{x}{n} = 0\,.$$

Also ist

$$F(x) = \begin{cases} x & \text{für } x \in \mathbb{Q} \\ 0 & \text{für } x \notin \mathbb{Q} \end{cases},$$

eine offensichtlich nicht differenzierbare Funktion, und wenn $F'(x)$ nicht einmal existiert, kann es natürlich auch nicht gleich $f(x)$ sein.

Auch unter dem Gesichtspunkt der Interpretation von $F(x)$ als Fläche unterhalb der Kurve $y = f(x)$ ist das Ergebnis unbefriedigend, denn da $f(x) \geq 0$ für alle x, sollte $F(x)$ eine monoton wachsende (oder zumindest nicht fallende) Funktion von x sein, wohingegen das gerade berechnete F bei jeder rationalen Zahl auf Null zurückfällt.

Der Ansatz $(*)$ hat also auch seine Tücken, und es wird Zeit, diese heuristische Definition durch eine bessere zu ersetzen.

c) Definition des Riemann-Integrals

Tatsächlich ist der Ansatz $(*)$ nicht so schlecht: Für die beiden gutartigen Funktionen und für das Eingangsbeispiel der Streckenbestimmung führte er schließlich zu vernünftigen Ergebnissen, und wie wir bald sehen werden, kann man Integrale über stetige Funktionen *immer* nach diesem Ansatz bestimmen; zumindest als Veranschaulichung des Integralbegriffs sollte man ihn daher durchaus im Hinterkopf behalten. Problematisch wird er erst bei „schlechten“ Funktionen wie der im letzten Beispiel.

1) Warum lohnt sich ein allgemeinerer Ansatz?: Dies legt natürlich die Frage nahe, ob man solche „schlechten“ Funktionen für Anwendungen wirklich braucht, und wenn ja, ob man sie so sehr braucht, daß sich der erhebliche technische Aufwand, den wir in diesem Paragraphen treiben müssen, lohnt.

Die DIRICHLETsche Sprungfunktion ist ein rein theoretisch konstruiertes Gegenbeispiel, das wohl keinerlei praktische Anwendung haben dürfte. Eine ihrer charakteristischen Eigenschaften taucht allerdings auch bei praktisch relevanten Funktionen auf: Auch ein periodisch wiederholter Rechteckimpuls hat abzählbar unendlich viele Sprungstellen, und zumindest als Idealisierung eines realen Signals spielt diese Funktion eine Rolle – sie ist allerdings erheblich harmloser als DIRICHLETS Beispiel, da die Menge ihrer Sprungstellen diskret ist.

Wirklich kompliziert wird die Situation dagegen beispielsweise bei der mathematischen Modellierung des Rauschens in einer elektronischen Schaltung: Hier liegt eine noch einmal deutlich schwierigere Situation vor als im obigen Beispiel, und in der Tat wird auch der in diesem Paragraphen definierte Integralbegriff nicht ausreichen, um mit diesem Problem fertig zu werden. Er ist aber eine unabdingbare Voraussetzung, um die dazu bentigten Techniken zu verstehen.

Obwohl wir im folgenden fast jeden der wichtigeren Sätze aus der Analysis I noch einmal in die Erinnerung zurückrufen und anwenden müssen, ist die nun folgende Konstruktion also kein Luxus, sondern zumindest langfristig notwendig auch für praktische Anwendungen.

2) Wo sollte der bisherige Ansatz modifiziert werden?: Ein Punkt, bei dem der bisherige Ansatz nur aufgrund der Einfachheit der betrachteten Beispiele so erfolgreich war, ist die *Abschätzung* des Grenzwerts durch die endlichen Approximationen. Beim Beispiel der Geschwindigkeit eines beschleunigenden Fahrzeugs war das problemlos, denn die Geschwindigkeit war eine monoton wachsende Funktion, die für jedes Teilintervall am linken Ende ihr Minimum und am rechten ihr Maximum annimmt.

Auch $f(x) = x^2$ ist für $x \geq 0$ monoton wachsend, $f(x) = e^x$ sogar für

alle $x \in \mathbb{R}$, so daß wir auch hier leicht untere und obere Schranken für $F(x)$ berechnen können.

Für nichtmonotone Funktionen dagegen ist die Situation völlig anders: Betrachten wir als Beispiel etwa die Sinuslinie zwischen Null und π mit einer Annäherung der Fläche durch sechs Rechtecke. In den Abbildungen 50 und 51 sind diese Rechtecke eingezeichnet, einmal bezogen auf den Funktionswert am linken Ende des Teilintervalls und einmal bezogen auf den am rechten. Wie man sieht, liegen die Rechtecke exakt spiegelsymmetrisch zueinander und haben damit insbesondere die gleiche Summe der Flächeninhalte, nämlich jeweils ungefähr 1,954. Die Fläche unter der Sinuslinie zwischen Null und π ist allerdings, wie wir bald sehen werden,

$$-\cos\pi - (-\cos 0) = -(-1) - (-1) = 1 + 1 = 2\,.$$

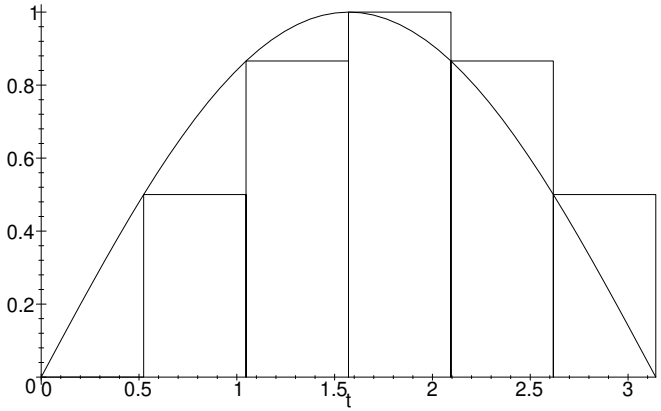


Abb. 50: Sinuslinie mit sechs Rechtecken zum jeweils linken Funktionswert

Wenn man ein (jedem Wissenschaftler zu empfehlendes) gesundes Mißtrauen gegen Computer und Taschenrechner mitbringt, könnte man versucht sein, den Wert 1,954 als eine „Taschenrechnerzwei“ zu interpretieren; da aber die Sinuswerte an den Stellen $0, \frac{\pi}{6}, \frac{\pi}{3}, \frac{\pi}{2}, \frac{4\pi}{3}, \frac{5\pi}{6}$ und 1 wohlbekannt sind (im Winkelmaß ausgedrückt sind diese Stellen gerade die Vielfachen von 30°), kann man die Summe der Rechteckflächen

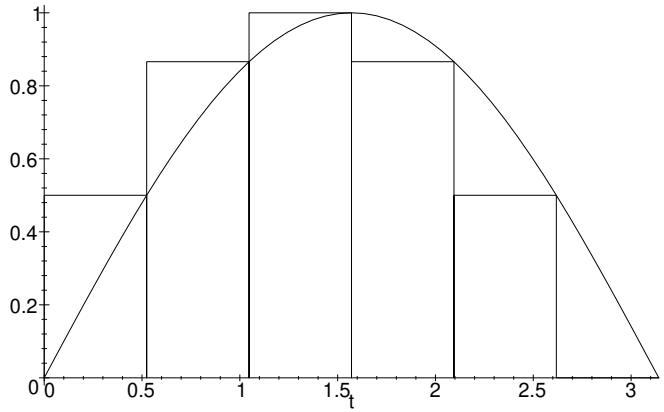


Abb. 51: Sinuslinie mit sechs Rechtecken zum jeweils rechten Funktionswert

exakt berechnen mit dem Ergebnis

$$\frac{\pi}{6}(2+\sqrt{3})\,,$$

das schon wegen der Transzendenz von π nicht gleich zwei sein kann. Somit haben wir definitiv keine obere Schranke für den korrekten Wert, und zumindest *a priori* haben wir auch keine untere, denn daß die Summe der Rechteckflächen kleiner ist als die Fläche unter der Kurve folgte je erst nachträglich durch Vergleich der beiden Werte.

Um wirklich eine untere Schranke für die Fläche zu bekommen, müßte man hier im Beispiel für die ersten drei Rechtecke den Funktionswert am linken Ende des Teilintervalls nehmen und für die letzten drei den vom rechten; für eine obere Schranke müßte man genau umgekehrt vorgehen.

Für eine beliebige Funktion gibt es aber offensichtlich keinen Grund, warum das Minimum oder Maximum überhaupt an einem Intervallende angenommen werden sollte; für einen möglichst allgemeinen Integrallbegriff sollte man also auch Punkte im Intervallinnern zur Ermittlung der Höhe des Rechtecks heranziehen: Anstelle der bislang betrachteten Unterteilung

$$a < a+\delta < a+2\delta < \cdots < a+(n-1)\delta < a+n\delta = b \quad \text{mit} \quad \delta = \frac{b-a}{n}$$

sollte also eine beliebige Unterteilung

$$a = x_0 < x_1 < x_2 < \cdots < x_{n-1} < x_n = b$$

treten.

3) Anwendung des Mittelwertsatzes: Einen weiteren Grund dafür liefert der wohl wichtigste Satz aus der Analysis I, der *Mittelwertsatz der Differentialrechnung:* Ist F stetig in einem Intervall $[u, v]$ und differenzierbar in (u, v), so gibt es einen Punkt $\xi \in (u, v)$, so daß

$$F'(\xi) = \frac{f(v) - f(u)}{v - u}$$

ist.

Wir suchen zu einer gegebenen Funktion f eine differenzierbare Funktion F mit $F' = f$. Falls wir annehmen, daß wir diese Funktion schon *hätten,* so wäre $F' = f$, und nach dem Mittelwertsatz gäbe es in jedem Intervall (x_i, x_{i+1}) ein Element ξ_i, so daß

$$\frac{F(x_{i+1}) - F(x_i)}{x_{i+1} - x_i} = F'(\xi_i) = f(\xi_i)$$

wäre, d.h.

$$F(x_{i+1}) = F(x_i) + f(\xi_i)(x_{i+1} - x_i)\,.$$

Rekursiv folgt, daß

$$\begin{aligned} F(x) = F(x_n) &= F(x_{n-1}) + f(\xi_{n-1})(x_n - x_{n-1}) \\ &= F(x_{n-2}) + f(\xi_{n-1})(x_{n-1} - x_{n-2}) + f(\xi_{n-1})(x_n - x_{n-1}) \\ &= \quad \vdots \\ &= F(x_0) + \sum_{i=0}^{n-1} f(\xi_i)(x_{i+1} - x_i)\,. \end{aligned}$$

Damit ist also die Fläche unter der Kurve $y = f(x)$ *exakt* gleich der Gesamtfläche endlich vieler geeignet gewählter Rechtecke; Abbildung 52 veranschaulicht dies anhand der ganz zu Beginn betrachteten Geschwindigkeitsfunktion. Hier ist nicht nur die Fläche unter der Kurve exakt

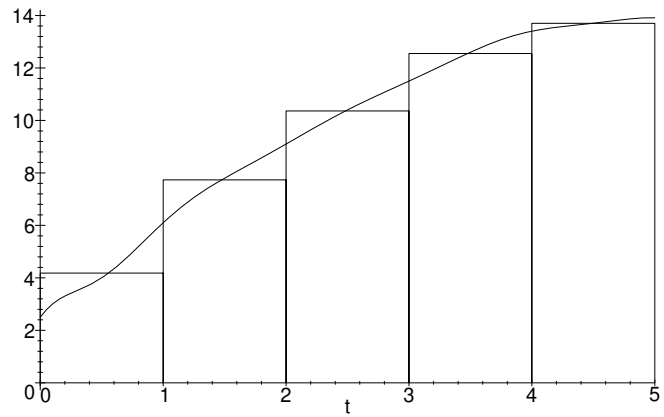


Abb. 52: Die Fläche unter der Kurve ist gleich der Fläche der fünf Rechtecke

gleich der Fläche der fünf Rechtecke, sondern auch die Fläche eines jeden Rechtecks gleich der Fläche zwischen Kurve und Grundlinie des Rechtecks, wir kommen also mit endlich vielen Rechtecken aus und brauchen nicht einmal einen Grenzwert zu berechnen.

Der Haken bei der Sache ist natürlich, daß wir die „geeignet gewählten" Rechtecke nicht kennen: Im obigen Ausdruck ist alles bekannt *außer* den Werten ξ_i, an denen die Funktion f ausgewertet wird.

Die Idee zur Definition des RIEMANN-Integrals ist nun, daß man einfach *beliebige* ξ_i zwischen x_i und x_{i+1} wählt in der Hoffnung, daß bei immer kleiner werdendem Abstand zwischen zwei aufeinanderfolgenden x_i der Grenzwert nicht mehr von der Wahl der ξ_i abhängt.

Diese Hoffnung hat natürlich nur eine Chance auf Erfüllung, wenn die Funktion f hinreichend stetig ist; ansonsten können sich die Werte von f an zwei beliebig nahe beieinanderliegenden Stellen immer noch beliebig stark unterscheiden, indem sie beispielsweise wie oben davon abhängen, ob ξ_i rational ist oder nicht.

4) Gleichmäßige Stetigkeit: Die Stetigkeit allein reicht allerdings immer noch nicht ganz aus:

Erinnern wir uns: f heißt *stetig* auf einer Teilmenge $D \subseteq \mathbb{R}$, wenn es zu jedem $\varepsilon > 0$ und zu jedem $x \in (a, b)$ ein $\delta > 0$ gibt, so daß gilt: Ist $y \in D$ und $|y - x| < \delta$, so folgt, daß $|f(y) - f(x)| < \varepsilon$ ist.

Wir möchten mehr: Wir wollen, daß sich in *jedem* Rechteck der Wert von $f(\xi)$ höchstens um $\varepsilon > 0$ ändert, falls nur die Breite des Rechtecks unter einer gewissen Schranke liegt, d.h. wir möchten, daß δ nicht von x abhängt

Dies führt auf folgende Verschärfung des Stetigkeitsbegriffs:

Definition: $D \subseteq \mathbb{R}$ sei eine Teilmenge von $\mathbb{R}$ und $f\colon D \to \mathbb{R}$ eine Funktion. f heißt *gleichmäßig stetig* in D, wenn es zu jedem $\varepsilon > 0$ ein $\delta > 0$ gibt, so daß für alle $x, y \in D$ gilt:

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

Um zu sehen, daß dies mehr ist als die bloße Stetigkeit, betrachten wir die Funktion $f(x) = 1/x$. Diese Funktion ist stetig auf der Menge aller positiver reeller Zahlen, denn für $0 < \delta < x$ und $\varepsilon > 0$ ist

$$\left|\frac{1}{x \pm \delta} - \frac{1}{x}\right| = \left|\frac{\mp\delta}{x(x \pm \delta)}\right| = \frac{\delta}{x(x \pm \delta)} < \varepsilon$$

$$\Leftrightarrow \delta < (x^2 \pm x\delta)\varepsilon \Leftrightarrow \delta \mp x\delta\varepsilon < x^2\varepsilon \Leftrightarrow \delta < \frac{x^2\varepsilon}{1 \mp x\varepsilon}\,.$$

Der Ausdruck ganz rechts ist offensichtlich kleiner, wenn im Nenner das Pluszeichen steht, also gilt: Für gegebenes $x > 0$ und $\varepsilon > 0$ gilt für jede positive reelle Zahl y

$$|y - x| < \delta \underset{\text{def}}{=} \frac{x^2\varepsilon}{1 + x\varepsilon} \Rightarrow \left|\frac{1}{y} - \frac{1}{x}\right| < \varepsilon\,.$$

Damit ist die Stetigkeit der Funktion bewiesen. Sie ist aber nicht gleichmäßig stetig, denn es gibt aber kein von x unabhängiges $\delta > 0$ mit dieser Eigenschaft: Der angegebene Ausdruck für das größtmögliche δ geht wegen des Faktors x^2 im Zähler für $x \to 0$ selbst gegen Null, fällt also unter jede vorgegebene positive Schranke. Abbildung 53 zeigt δ in Abhängigkeit von x für den speziellen Wert $\varepsilon = \frac{1}{10}$.

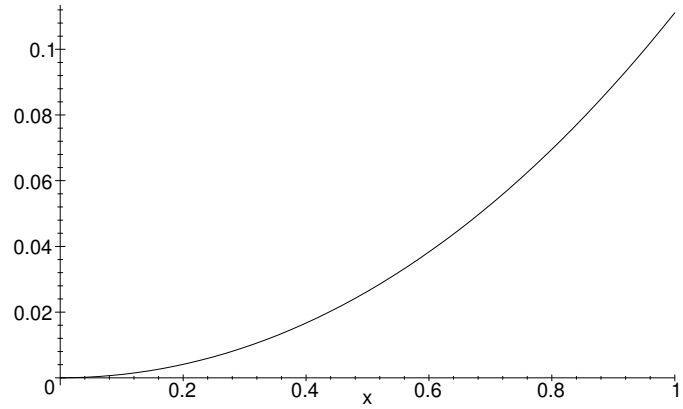


Abb. 53: δ in Abhängigkeit von x für $f(x) = 1/x$ und $\varepsilon = 0{,}1$

In einem abgeschlossenen Intervall sollte so etwas nicht möglich sein, und in der Tat gilt der (durchaus nichttriviale)

Satz: $f\colon [a, b] \to \mathbb{R}$ sei eine stetige Funktion auf dem abgeschlossenen Intervall $[a, b]$ mit $a, b \in \mathbb{R}$. Dann ist f auf $[a, b]$ gleichmäßig stetig.

Der *Beweis* ist technisch und indirekt:

Angenommen, f wäre *nicht* gleichmäßig stetig. Dann gäbe es für mindestens ein $\varepsilon > 0$ zu jedem $\delta > 0$ Punkte $x, y \in [a, b]$, so daß $|y - x| < \delta$, aber $|f(y) - f(x)| \geq \varepsilon$ wäre.

Aufgrund der Annahme, der Satz sei falsch, können wir ein solches ε fixieren und betrachten die speziellen Werte $\delta = \frac{1}{n}$: Nach dem gerade gesagten gibt es dazu Punkte $x_n, y_n \in [a, b]$, so daß $|y_n - x_n| < \frac{1}{n}$, aber $|f(y_n) - f(x_n)| \geq \varepsilon$ ist.

Nun kommt der nichttriviale Teil: Wir benötigen aus der Analysis I den

Satz von Bolzano-Weierstraß: Jede Folge (x_n) von Punkten aus dem abgeschlossenen Intervall $[a, b]$ hat (mindestens) eine konvergente Teilfolge x_{n_ν}.

Insbesondere muß also die hier betrachtete Folge (x_n) eine konvergente Teilfolge $\left(x_{n_\nu}\right)_{\nu \in \mathbb{N}}$ haben; deren Grenzwert sei $c \in [a, b]$.

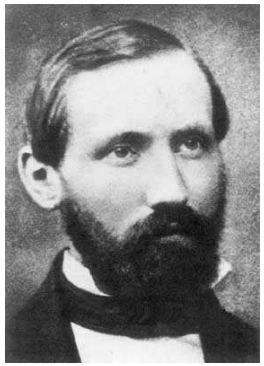
Da $\left|x_{n_\nu} - y_{n_\nu}\right| < 1/n_\nu$ ist, muß dann auch die Folge der y_{n_ν} gegen c konvergieren, und da f nach Voraussetzung eine stetige Funktion ist, gilt

$$\lim_{\nu\to\infty} f(x_{n_\nu}) = \lim_{\nu\to\infty} f(y_{n_\nu}) = f(c)\,.$$

Andererseits ist aber für jedes ν nach Konstruktion der Folgen (x_n) und (y_n) der Abstand zwischen $f(x_{n_\nu})$ und $f(y_{n_\nu})$ größer oder gleich der festgewählten Zahl ε, so daß die beiden Folgen unmöglich denselben Grenzwert haben können.

Mithin führt die Annahme, f sei *nicht* gleichmäßig stetig, auf einen Widerspruch, und damit ist der Satz bewiesen. ∎

5) Definition einer Approximation für das Integral: Nach diesen Vorarbeiten können wir ernsthaft darangehen, das Integral einer Funktion zu definieren. Es gibt in der Mathematik verschiedene Integralbegriffe; für uns reicht die Definition des deutschen Mathematikers BERNHARD RIEMANNN, das inzwischen nach ihm benannte RIEMANN-Integral.



GEORG FRIEDRICH BERNHARD RIEMANNN (1826-1866) war Sohn eines lutherischen Pastors und schrieb sich 1946 auf Anraten seines Vaters an der Universität Göttingen für das Studium der Theologie ein. Schon bald wechselte an die Philosophische Fakultät, um dort unter anderem bei GAUSS Mathematikvorlesungen zu hören. Nach Promotion 1851 und Habilitation 1854 erhielt er dort 1857 einen Lehrstuhl. Trotz seines frühen Todes initiierte er grundlegende auch noch heute fundamentale Entwicklungen in der Geometrie, der Zahlentheorie und über abelsche Funktionen. Seine Vermutung über die Nullstellen der (heute als RIEMANNsch bezeichneten) ζ-Funktion ist die berühmteste offene Vermutung der heutigen Mathematik.

Wir gehen aus von einer Funktion $f\colon [a,b] \to \mathbb{R}$ auf einem abgeschlossenen Intervall $[a,b]$; gesucht ist eine Funktion $F\colon [a,b] \to \mathbb{R}$, für die (idealerweise) $F'(x) = f(x)$ sein sollte für alle x aus dem offenen Intervall (a,b).

Wir wählen eine Unterteilung

$$a = x_0 < x_1 < x_2 < \cdots < x_{n-1} < x_n = x$$

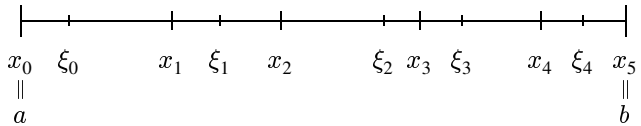
des Intervalls von a bis x. Da wir im folgenden viel mit dieser Unterteilung rechnen werden, führen wir die Abkürzung

$$\boldsymbol{x} = (x_0, x_1, \ldots, x_n)$$

dafür ein und verabreden, daß die x_i immer, wenn von $\boldsymbol{x}$ die Rede ist, die obigen Gleichungen und Ungleichungen erfüllen sollen, daß sie also tatsächlich eine Unterteilung von $[a,x]$ definieren.

Falls die gesuchte differenzierbare Funktion $F\colon [a,b] \to \mathbb{R}$ existiert, sagt uns – wie wir in Teil *c)* gesehen haben – der Mittelwertsatz der Differentialrechnung, daß für geeignete Elemente ξ_i mit

$$a = x_0 < \xi_0 < x_1 < \xi_1 < x_2 < \cdots < \xi_{n-2} < x_{n-1} < \xi_{n-1} < x_n = x$$



folgt

$$F(x) = F(a) + \sum_{i=0}^{n-1} f(\xi_i)(x_{i+1} - x_i)\,.$$

Natürlich kennen wir die Zahlen ξ_i nicht – selbst wenn wir wissen, daß F und damit die ξ_i überhaupt existieren. Deshalb betrachten wir einfach *beliebige* Zahlen ξ_i mit

$$a = x_0 < \xi_0 < x_1 < \xi_1 < x_2 < \cdots < \xi_{n-2} < x_{n-1} < \xi_{n-1} < x_n = x$$

und führen auch hier wieder die Abkürzung

$$\underline{\xi} = (\xi_0, \xi_1, \ldots, \xi_{n-1})$$

ein mit der Verabredung, daß immer, wenn ein $\underline{\xi}$ in Zusammenhang mit einem $\boldsymbol{x}$ auftritt, diese Ungleichungen erfüllt sein sollen.

Definition: Die RIEMANNsche Summe zu einem Paar $(\boldsymbol{x}, \underline{\xi})$ ist

$$I(\boldsymbol{x}, \underline{\xi}) = \sum_{i=0}^{n-1} f(\xi_i)(x_{i+1} - x_i)\,.$$

Das RIEMANN-Integral soll der Grenzwert einer Folge RIEMANNscher Summen sein, wobei die Unterteilungen $\boldsymbol{x}$ immer feiner werden. Dieses „Feinerwerden“ müssen wir natürlich auch noch erklären: Seien $\boldsymbol{x}$ und $\boldsymbol{y}$ zwei Unterteilungen des Intervalls $[a, x]$; konkret sei

$$a = x_0 < x_1 < \cdots < x_{n-1} < x_n = x$$

und

$$a = y_0 < y_1 < \cdots < y_{m-1} < y_m = x\,.$$

Dann heißt $\boldsymbol{y}$ eine *Verfeinerung* von $\boldsymbol{x}$, wenn jedes Teilintervall (y_i, y_{i+1}) von $\boldsymbol{y}$ ganz in einem der Teilintervalle (x_j, x_{j+1}) von $\boldsymbol{x}$ liegt; $\boldsymbol{y}$ entsteht also aus $\boldsymbol{x}$, indem einige der Intervalle von $\boldsymbol{x}$ noch weiter unterteilt werden. Insbesondere muß dann $m \geq n$ sein.

Nun betrachten wir eine Folge $(\boldsymbol{x}^{(\nu)})$ von Unterteilungen derart, daß

- $\boldsymbol{x}^{(\nu+1)}$ stets eine Verfeinerung von $\boldsymbol{x}^{(\nu)}$ ist *und*
- die maximale Länge der Teilintervalle von $\boldsymbol{x}^{(\nu)}$ mit wachsendem ν gegen Null geht, d.h.

$$\lim_{\nu\to\infty} \max_{i=0}^{n_\nu - 1} (x_{i+1}^{(\nu)} - x_i^{(\nu)}) = 0\,.$$

Zu jeder Unterteilung $\boldsymbol{x}^{(\nu)}$ wählen wir willkürlich eine (im Sinne obiger Konvention dazu passende) Folge $\underline{\xi}^{(\nu)}$ von Zwischenpunkten und fragen nach Existenz und gegebenenfalls Wert des Grenzwerts

$$\lim_{\nu\to\infty} I(\boldsymbol{x}^{(\nu)}, \underline{\xi}^{(\nu)})\,.$$

Dieser Grenzwert muß natürlich im allgemeinen nicht existieren, und wenn er existiert, kann er von der Wahl der Zwischenpunkte $\xi_i^{(\nu)}$ und von der Wahl einer Folge $(\boldsymbol{x}^{(\nu)})$ von Unterteilungen abhängen. Wenn er unabhängig von all diesen Wahlen existiert und immer denselben Wert hat, bezeichnen wir diesen gemeinsamen Wert als das RIEMANN-*Integral* von f zwischen a und x, in Zeichen

$$\int_a^x f(\xi)\,d\xi = \lim_{\nu\to\infty} I(\boldsymbol{x}^{(\nu)}, \underline{\xi}^{(\nu)})\,.$$

Falls dieses Integral für jedes $x \in [a, b]$ existiert, sagen wir, f sei RIEMANN-*integrierbar* auf $[a, b]$.

ξ heißt *Integrationsvariable* und übernimmt die Rolle des Summationsindex in einer Summe: Genau wie beispielsweise

$$\sum_{i=0}^{n-1} i^2 = \sum_{j=0}^{n-1} j^2 = \sum_{k=0}^{n-1} k^2 = \frac{n(2n-1)(n-1)}{6}$$

davon unabhängig ist, ob der Summationsindex i, j oder k heißt, ist auch

$$\int_a^x f(\xi)\,d\xi = \int_a^x f(u)\,du = \int_a^x f(t)\,dt$$

unabhängig davon, ob die Integrationsvariable mit ξ, u oder t bezeichnet wird.

6) Existenz des Riemann-Integrals für stetige Funktionen: Damit haben wir das RIEMANN-Integral definiert; wir wissen allerdings bisher für keine einzige Funktion, daß es existiert. In diesem Abschnitt wollen wir uns überlegen, daß es zumindest für *stetige* Funktionen immer existiert:

Satz: Jede stetige Funktion $f\colon [a, b] \to \mathbb{R}$ ist RIEMANN-integrierbar auf $[a, b]$.

Zum *Beweis* müssen wir zeigen, daß das RIEMANN-Integral

$$\int_a^x f(\xi)\,d\xi$$

für jedes $x \in [a, b]$ existiert.

Dazu fixieren wir ein beliebig vorgegebenes $x \in [a, b]$ und betrachten Folgen $(\boldsymbol{x}^{(\nu)})$ von immer feiner werdenden Unterteilungen des Intervalls $[a, x]$; dazu entsprechende Folgen $(\underline{\xi}^{(\nu)})$ von Zwischenwerten und die Grenzwerte der RIEMANN-Summen $I(\boldsymbol{x}^{(\nu)}, \underline{\xi}^{(\nu)})$.

1. Schritt: Für eine feste Folge $(\boldsymbol{x}^{(\nu)})$ von Unterteilungen existiert der Grenzwert und ist unabhängig von der Wahl der Zwischenwerte $\xi_i^{(\nu)}$:

Um das einzusehen, betrachten wir für jede Unterteilung $\boldsymbol{x}^{(\nu)}$ die RIEMANNschen Obersummen und Untersummen: Da f nach Voraussetzung stetig ist, nimmt es in jedem abgeschlossenen Intervall sowohl sein Maximum als auch sein Minimum an; insbesondere werden also im Intervall $[x_i^{(\nu)}, x_{i+1}^{(\nu)}]$ ein Maximum $M_i^{(\nu)}$ und ein Minimum $m_i^{(\nu)}$ angenommen. Damit gilt unabhängig von der Wahl des Zwischenwerts $\xi_i^{(\nu)}$

$$m_i^{(\nu)} \le f(\xi_i^{(\nu)}) \le M_i^{(\nu)} .$$

Nach Definition ist

$$I(\boldsymbol{x}^{(\nu)}, \xi^{(\nu)}) = \sum_{i=0}^{n_\nu - 1} f(\xi_i^{(\nu)})\big(x_{i+1}^{(\nu)} - x_i^{(\nu)}\big) ;$$

ersetzen wir hierin $\xi_i^{(\nu)}$ jeweils durch $m_i^{(\nu)}$, so erhalten wir eine untere Abschätzung

$$s^{(\nu)} \underset{\text{def}}{=} \sum_{i=0}^{n_\nu - 1} m_i^{(\nu)}\big(x_{i+1}^{(\nu)} - x_i^{(\nu)}\big)$$

für $I(\boldsymbol{x}^{(\nu)}, \underline{\xi}^{(\nu)})$; diese bezeichnen wir als RIEMANN*sche Untersumme* der Unterteilung $\boldsymbol{x}^{(\nu)}$. Entsprechend liefert die Ersetzung durch $M_i^{(\nu)}$ eine obere Abschätzung

$$S^{(\nu)} \underset{\text{def}}{=} \sum_{i=0}^{n_\nu - 1} M_i^{(\nu)}\big(x_{i+1}^{(\nu)} - x_i^{(\nu)}\big)$$

für $I(\boldsymbol{x}^{(\nu)}, \underline{\xi}^{(\nu)})$, die RIEMANN*sche Obersumme* der Unterteilung $\boldsymbol{x}^{(\nu)}$.

Unabhängig von der Wahl der Zwischenwerte $\xi_i^{(\nu)}$ gilt somit

$$s^{(\nu)} \le I(\boldsymbol{x}^{(\nu)}, \underline{\xi}^{(\nu)}) \le S^{(\nu)} ,$$

ein Zusammenhang, der in Abbildung 54 noch einmal graphisch dargestellt ist: Die RIEMANNsche Untersumme ist gleich der Fläche der durchgezogenen Rechtecke, für die Obersumme kommt noch der gestrichelte Anteil dazu, und für die RIEMANNsche Summe $I(\boldsymbol{x}^{(\nu)}, \underline{\xi}^{(\nu)})$ muß man bis zu den punktierten Linien gehen. Die Werte $\xi_i^{(\nu)}$ sind auf

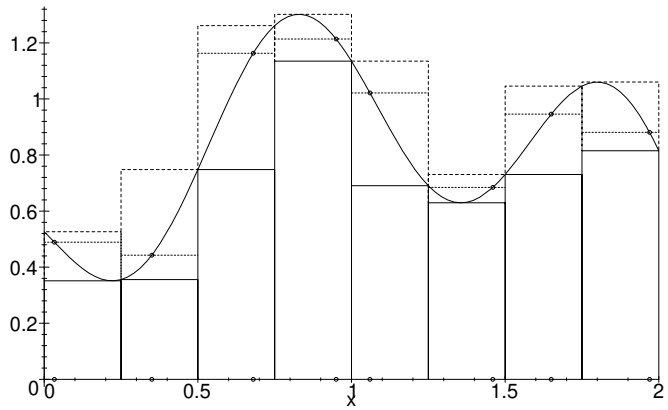


Abb. 54: Riemannsche Summen, Obersummen und Untersummen

der x-Achse durch kleine Kreise gekennzeichnet, ebenso die Punkte $\big(\xi_i^{(\nu)}, f(\xi_i^{(\nu)})\big)$ auf der Kurve.

Wenn wir jetzt noch zeigen können, daß die Folge der $s^{(\nu)}$ und die der $S^{(\nu)}$ beide gegen denselben Grenzwert konvergieren, muß wegen der Einschließung

$$s^{(\nu)} \le I(\boldsymbol{x}^{(\nu)}, \underline{\xi}^{(\nu)}) \le S^{(\nu)}$$

auch $I(\boldsymbol{x}^{(\nu)}, \underline{\xi}^{(\nu)})$ gegen diesen Grenzwert konvergieren, und die Behauptung im ersten Schritt ist bewiesen.

Vergleichen wir dazu zunächst die Untersummen $s^{(\nu)}$ und $s^{(\nu+1)}$: Die Unterteilung $\boldsymbol{x}^{(\nu+1)}$ entsteht aus $\boldsymbol{x}^{(\nu)}$ dadurch, daß einige der Intervalle weiter unterteilt werden. Ist aber $(x_i^{(\nu+1)}, x_{i+1}^{(\nu+1)})$ Teilintervall von $(x_j^{(\nu)}, x_{j+1}^{(\nu)})$, so kann das Minimum im Teilintervall natürlich nicht kleiner sein als im größeren Intervall, d.h. $s^{(\nu+1)} \ge s^{(\nu)}$. Somit ist die Folge der $s^{(\nu)}$ monoton wachsend.

Genauso folgt, daß die Folge der $S^{(\nu)}$ monoton fallend ist, und da stets $s^{(\nu)} \le S^{(\nu)}$ ist, folgt weiter, daß

$$s^{(1)} \le s^{(2)} \le s^{(3)} \le \cdots \le S^{(3)} \le S^{(2)} \le S^{(1)} .$$

Insbesondere ist also $S^{(1)}$ eine obere Schranke für die Folge der $s^{(\nu)}$ und $s^{(1)}$ eine untere Schranke für die Folge der $S^{(\nu)}$.

Damit ist also die Folge $(s^{(\nu)})$ monoton wachsend und nach oben beschränkt, während $(S^{(\nu)})$ monoton fallend und nach unten beschränkt ist. Nach einem Satz aus der Analysis I (*Jede monotone und beschränkte Folge reeller Zahlen ist konvergent.*) folgt daraus die Konvergenz beider Folgen.

Nun fehlt nur noch, daß beide denselben Grenzwert haben.

Dazu betrachten wir die Differenzen

$$\begin{aligned} S^{(\nu)} - s^{(\nu)} &= \sum_{i=0}^{n_\nu - 1} M_i^{(\nu)}(x_{i+1}^{(\nu)} - x_i^{(\nu)}) - \sum_{i=0}^{n_\nu - 1} m_i^{(\nu)}(x_{i+1}^{(\nu)} - x_i^{(\nu)}) \\ &= \sum_{i=0}^{n_\nu - 1} (M_i^{(\nu)} - m_i^{(\nu)})(x_{i+1}^{(\nu)} - x_i^{(\nu)}) \,. \end{aligned}$$

Mit

$$\Delta^{(\nu)} \underset{\text{def}}{=} \max_i \big(M_i^{(\nu)} - m_i^{(\nu)}\big)$$

ist also

$$\begin{aligned} S^{(\nu)} - s^{(\nu)} &\le \sum_{i=0}^{n_\nu - 1} \Delta^{(\nu)}(x_{i+1}^{(\nu)} - x_i^{(\nu)}) = \Delta^{(\nu)} \sum_{i=0}^{n_\nu - 1} (x_{i+1}^{(\nu)} - x_i^{(\nu)}) \\ &= \Delta^{(\nu)}(x_n^{(\nu)} - x_0^{(\nu)}) = \Delta^{(\nu)}(x - a) \,, \end{aligned}$$

und das ist eine Nullfolge, falls wir zeigen können, daß die Folge der $\Delta^{(\nu)}$ eine ist.

Hier kommt nun die in Abschnitt *c)* eingeführte gleichmäßige Stetigkeit von f zum Tragen: Danach gibt es zu jedem $\varepsilon > 0$ ein $\delta > 0$, so daß für alle Punkte $\xi_1, \xi_2 \in [a, b]$ gilt: Ist $|\xi_1 - \xi_2| < \delta$, so folgt $|f(\xi_1) - f(\xi_2)| < \varepsilon$.

Ist dabei für eine Unterteilung $\boldsymbol{x}^{(\nu)}$ jede der Differenzen $x_{i+1}^{(\nu)} - x_i^{(\nu)}$ kleiner als δ, so ist auch $M_i^{(\nu)} - m_i^{(\nu)} < \varepsilon$, denn sowohl $M_i^{(\nu)}$ als auch $m_i^{(\nu)}$ sind Funktionswerte, die irgendwo im Intervall $[x_i^{(\nu)}, x_{i+1}^{(\nu)}]$ angenommen werden.

Nun haben wir aber vorausgesetzt, daß die Unterteilungen $\boldsymbol{x}^{(\nu)}$ immer feiner werden, d.h. zu jedem $\delta > 0$ gibt es in der Tat ein ν_0, so daß $x_{i+1}^{(\nu)} - x_i^{(\nu)} < \delta$ für alle $\nu > \nu_0$ und alle i. Somit sind für $\nu > \nu_0$ alle Differenzen $M_i^{(\nu)} - m_i^{(\nu)}$ kleiner als ε, und damit ist auch $\Delta^{(\nu)} < \varepsilon$ für $\nu > \nu_0$. Also sind sowohl $(\Delta^{(\nu)})$ als auch $S^{(\nu)} - s^{(\nu)}$ Nullfolgen, d.h.

$$\lim_{\nu\to\infty} s^{(\nu)} = \lim_{\nu\to\infty} S^{(\nu)} \,.$$

Wie schon erwähnt, folgt daraus aufgrund der Einschließung

$$s^{(\nu)} \le I(\boldsymbol{x}^{(\nu)}, \underline{\xi}^{(\nu)}) \le S^{(\nu)} \,,$$

auch die Gleichung

$$\lim_{\nu\to\infty} S^{(\nu)} = \lim_{\nu\to\infty} I(\boldsymbol{x}^{(\nu)}, \underline{\xi}^{(\nu)}) = \lim_{\nu\to\infty} S^{(\nu)} \,.$$

Da die linke und die rechte Seite obiger Ungleichung nicht von den $\xi_i^{(\nu)}$ abhängen, ist auch $\lim I(\boldsymbol{x}^{(\nu)}, \underline{\xi}^{(\nu)})$ davon unabhängig, und damit ist der erste Schritt des Beweises beendet.

Der zweite ist zum Glück erheblich einfacher:

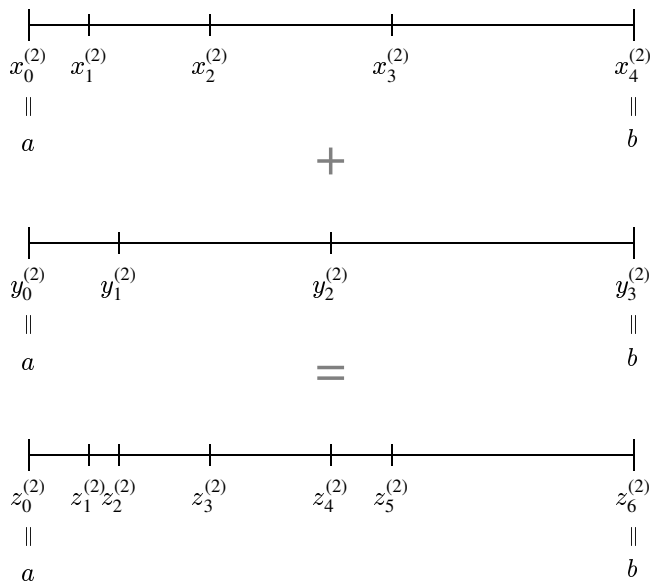
2. Schritt: $\lim\limits_{\nu\to\infty} I(\boldsymbol{x}^{(\nu)}, \underline{\xi}^{(\nu)})$ ist auch unabhängig von der Folge $(\boldsymbol{x}^{(\nu)})$:

Betrachten wir zwei Folgen $(\boldsymbol{x}^{(\nu)})$ und $(\boldsymbol{y}^{(\nu)})$ von Unterteilungen. Für jeden Index ν können wir zu den Unterteilungen $\boldsymbol{x}^{(\nu)}$ und $\boldsymbol{y}^{(\nu)}$ eine gemeinsame Verfeinerung $\boldsymbol{z}^{(\nu)}$ konstruieren, indem wir einfach die sämtlichen Zahlen $x_i^{(\nu)}$ und $y_j^{(\nu)}$ der Größe nach ordnen als $z_0^{(\nu)} = a, \ldots, z_r^{(\nu)} = x$.

Wie oben seien $s^{(\nu)}$ und $S^{(\nu)}$ die RIEMANNsche Unter- und Obersumme zur Unterteilung $\boldsymbol{x}^{(\nu)}$; entsprechend seien $\widetilde{s}^{(\nu)}$ und $\widetilde{S}^{(\nu)}$ die zur Unterteilung $\boldsymbol{z}^{(\nu)}$. Da $\boldsymbol{z}^{(\nu)}$ eine Verfeinerung von $\boldsymbol{x}^{(\nu)}$ ist, wissen wir aus dem ersten Schritt, daß

$$s^{(\nu)} \le \widetilde{s}^{(\nu)} \le \widetilde{S}^{(\nu)} \le S^{(\nu)}$$

ist; da die Folgen $(s^{(\nu)})$ und $(S^{(\nu)})$ denselben Grenzwert haben, müssen auch $(\widetilde{s}^{(\nu)})$ und $(\widetilde{S}^{(\nu)})$ gegen diesen Wert konvergieren und damit auch



die Folge der $I(\boldsymbol{z}^{(\nu)}, \underline{\tilde{\xi}}^{(\nu)})$, wobei $(\underline{\tilde{\xi}}^{(\nu)})$ irgendeine Folge von Zwischenwerten bezeichnet; wie wir bereits aus dem ersten Schritt wissen, hängt der Grenzwert nicht davon ab. Also ist

$$\lim_{\nu\to\infty} I(\boldsymbol{x}^{(\nu)}, \underline{\xi}^{(\nu)}) = \lim_{\nu\to\infty} I(\boldsymbol{z}^{(\nu)}, \underline{\tilde{\xi}}^{(\nu)}) .$$

Genauso folgt, daß auch

$$\lim_{\nu\to\infty} I(\boldsymbol{y}^{(\nu)}, \underline{\bar{\xi}}^{(\nu)}) = \lim_{\nu\to\infty} I(\boldsymbol{z}^{(\nu)}, \underline{\tilde{\xi}}^{(\nu)}) ,$$

wobei $(\underline{\bar{\xi}}^{(\nu)})$ die Folge der Zwischenwerte zu $\boldsymbol{y}^{(\nu)}$ bezeichnet, und dies zeigt schließlich die Behauptung

$$\lim_{\nu\to\infty} I(\boldsymbol{x}^{(\nu)}, \underline{\xi}^{(\nu)}) = \lim_{\nu\to\infty} I(\boldsymbol{y}^{(\nu)}, \underline{\bar{\xi}}^{(\nu)}) .$$

Damit ist der Satz vollständig bewiesen. ■

7) Stückweise stetige Funktionen: Gelegentlich möchte man die Bedingung der Stetigkeit wenigstens ein bißchen lockern, um beispielweise auch einen Rechteckimpuls, der periodisch zwischen 0 und 1 (oder -1 und 1) wechselt, behandeln zu können. Die Existenz des RIEMANN-Integral wird durch solche kleineren Abweichungen nicht beeinträchtigt:

Definition: Eine Funktion $f\colon [a,b] \to \mathbb{R}$ heißt *stückweise stetig,* wenn es eine Unterteilung

$$a = a_0 < a_1 < \cdots < a_{r-1} < a_r = b$$

des Intervalls $[a,b]$ gibt, so daß f auf jedem der offenen Intervalle (a_i, a_{i+1}) stetig ist.

f ist also überall stetig außer eventuell in endlich vielen Punkten $a_0, \ldots, a_r$. Der Wert in diesen Punkten kann, aber muß nicht mit dem linksseitigen oder rechtsseitigen Grenzwert der Funktion übereinstimmen; beispielsweise definiert man Rechteckimpulse gelegentlich auch durch die Funktion

$$f(x) = \begin{cases} 1 & \text{falls } 2n-1 < x < 2n \text{ für ein } n \in \mathbb{Z} \\ 0 & \text{falls } x \in \mathbb{Z} \\ -1 & \text{falls } 2n < x < 2n+1 \text{ für ein } n \in \mathbb{Z} \end{cases} .$$

Satz: Eine stückweise stetige Funktion $f\colon [a,b] \to \mathbb{R}$ ist RIEMANN-integrierbar.

Beweis: f sei stetig in den offenen Intervallen (a_i, a_{i+1}) mit

$$a = a_0 < a_1 < \cdots < a_{r-1} < a_r = b .$$

Wiederholt man den Beweis des entsprechenden Satzes für Funktionen, die auf ganz $[a,b]$ stetig sind, so funktioniert fast alles problemlos auch für f. Schwierigkeiten gibt es nur mit den Intervallen einer Unterteilung $\boldsymbol{x}$, die einen der Punkte a_i enthalten. In diesen Intervallen ist f nicht notwendigerweise stetig, so daß die Differenz zwischen dem Supremum und dem Infimum von f dort nicht mit Verkleinerung des Intervalls gegen Null gehen muß, sondern auch gegen einen von Null verschiedenen Wert, nämlich die „Sprunghöhe“ bei a_i, konvergieren kann.

Nun gibt es aber in jeder Unterteilung nur endlich viele Intervalle, die ein a_i enthalten, und auch die Sprunghöhen sind begrenzt; gehen also alle Intervallängen gegen Null, so geht auch wie im Beweis des Satzes die Differenz zwischen RIEMANNschen Ober- und Untersummen gegen Null. ■

8) Noch einmal die Dirichletsche Sprungfunktion: In Abschnitt *b3)* waren wir nicht zufrieden mit dem Verhalten des dort heuristisch eingeführten Integrals für die DIRICHLETsche Sprungfunktion; schauen wir, was das nun eingeführte RIEMANN-Integral daraus macht. Sei also

$$f(x) = \begin{cases} 1 & \text{für } x \in \mathbb{Q} \\ 0 & \text{für } x \notin \mathbb{Q} \end{cases};$$

wir interessieren uns für $\int\limits_0^1 f(x)\,dx$.

Dazu sei $\boldsymbol{x}$ eine Unterteilung des Intervalls $[0,1]$. Unabhängig von der Wahl dieser Unterteilung können wir stets eine Folge $\underline{\xi}$ von rationalen Zwischenwerten finden, aber auch eine Folge $\underline{\widetilde{\xi}}$ von irrationalen. Dann ist, unabhängig von der Unterteilung $\boldsymbol{x}$,

$$I(\boldsymbol{x}, \underline{\xi}) = \sum_{i=0}^{n-1} f(\xi_i)(x_{i+1} - x_i) = \sum_{i=0}^{n-1} (x_{i+1} - x_i) = 1 - 0 = 1$$

und

$$I(\boldsymbol{x}, \underline{\widetilde{\xi}}) = \sum_{i=0}^{n-1} f(\widetilde{\xi}_i)(x_{i+1} - x_i) = 0\,.$$

Damit kann kein gemeinsamer Grenzwert existieren, und somit ist die DIRICHLETsche Sprungfunktion nicht RIEMANN-integrierbar – ein sehr viel überzeugenderes Ergebnis als das aus Abschnitt *b3).*

9) Ausblick: Das Lebesgue-Integral: Wir könnten allerdings auch argumentieren, daß es „nur“ abzählbar unendlich viele rationale Zahlen, aber überabzählbar viele irrationale zwischen null und eins gibt; daher sollten letztere das Geschehen dominieren, und $\int_0^1 f(x)\,dx$ sollte verschwinden.

In der Tat kann man eine Verallgemeinerung des RIEMANN-Integrals definieren, das LEBESGUE-Integral, das für stückweise stetige Funktionen mit dem RIEMANN-Integral übereinstimmt und für die DIRICHLETsche Sprungfunktion den Wert Null liefert. Dazu geht man nach dem französischen Mathematiker HENRI LÉON LEBESGUE (1875-1941) bei den Unter- und Obersummen nicht wie bei RIEMANN von *endlich* vielen Rechtecken aus, sondern von *abzählbar unendlich* vielen. (Wie man dies im einzelnen macht, braucht uns hier nicht zu interessieren; wir werden uns im folgenden stets auf das RIEMANN-Integral beschränken.) Dieses LEBESGUE-Integral existiert *fast* immer: Man kann zwar die *Existenz* von Funktionen, für die es nicht existiert, *beweisen,* explizite Beispiele solcher Funktionen sind aber nicht bekannt.

10) Anwendung auf Flächeninhalte: Nachdem wir nun mit einen exakten Integralbegriff haben, bietet sich an, Flächen über dieses Integral zu *definieren:*

Definition: Ist $f\colon [a,b] \to \mathbb{R}$ eine nichtnegative Funktion, so bezeichnen wir die Zahl

$$\int\limits_a^b f(x)\,dx$$

als *Fläche* zwischen der Kurve $y = f(x)$ und der x-Achse zwischen den Geraden $x = a$ und $x = b$.

Was passiert, wenn f auch negative Werte annimmt? Für $f(x) = x$ etwa ist $\int\limits_{-1}^{1} f(x)\,dx = 0$, wie man sich leicht überlegt anhand von Unterteilungen, die symmetrisch zur Null liegen. Auch dies läßt sich als Aussage über Flächeninhalte interpretieren – wenn man davon ausgeht, daß die Formel

$$\textit{Fläche} = \textit{Länge} \times \textit{Breite}$$

für die Rechteckfläche auch bei *negativer* Länge und/oder Breite gelten soll. Da Integrale nicht nur zur Flächenbestimmung, sondern auch etwa zur Bestimmung der Ladung in einem Kondensator verwendet werden, ist dies die Interpretation, die man in der Mathematik in den meisten

Fällen vorzieht; für die klassische Fläche im Sinne des Tapezierens muß man mit

$$\int_a^b |f(x)|\ dx$$

arbeiten.

Im Sinne dieser negativen Längen und Breiten ist es auch sinnvoll, Integrale für $b < a$ zu definieren durch

$$\int_a^b f(x)\,dx \underset{\text{def}}{=} -\int_b^a f(x)\,dx\,.$$

Insbesondere ist dann natürlich $\int_a^a f(x)\,dx = 0$.

d) Erste Integrationsregeln

Das RIEMANN-Integral stellt die heuristischen Überlegungen vom Beginn dieses Paragraphen insofern auf eine exaktere mathematische Grundlage, als wir nun einen Integralbegriff haben, der auch bei extrem ungewöhnlichen Funktionen wie der DIRICHLETschen Sprungfunktion keine unerwarteten Ergebnisse liefert. Was allerdings den Ausgangspunkt betrifft, die Umkehrung der Differentiation, wissen wir über das neue Integral noch gar nichts, und auch sonst kennen wir noch nicht viele Regeln über den Umgang damit und insbesondere auch über seine Berechnung *ohne* die umständlichen und sehr langsam konvergierenden RIEMANN-Summen.

In diesem Abschnitt sollen die ersten (und einfachsten) solchen Regeln zusammengestellt werden; danach erweitern wir zunächst unser Instrumentarium, um dann damit auch kompliziertere und interessantere Regeln zu beweisen.

1) Monotonieregel: Eine der einfachsten, aber trotzdem oft nützlichen Regeln für den Umgang mit Integralen übersetzt Größerbeziehungen zwischen Integranden in Größenbeziehungen zwischen Integralen:

Satz: f und g seien stückweise stetige Funktionen auf dem Intervall $[a,b]$, und für alle $x \in [a,b]$ sei $f(x) \le g(x)$. Dann ist auch

$$\int_a^b f(x)\,dx \le \int_a^b g(x)\,dx\,.$$

Insbesondere ist

$$\left|\int_a^b f(x)\,dx\right| \le \int_a^b |f(x)|\ dx$$

Beweis: Wir betrachten eine Unterteilung $\boldsymbol{x}$ des Intervalls $[a,b]$ und eine dazu kompatible Sequenz $\underline{\xi}$ von Zwischenwerten; $I(\boldsymbol{x},\underline{\xi})$ sei die zugehörige RIEMANN-Summe für f und $J(\boldsymbol{x},\underline{\xi})$ die für g. Da für jedes ξ_i nach Voraussetzung $f(\xi_i) \le g(\xi_i)$ ist, folgt unmittelbar aus der Definition der RIEMANN-Summen, daß

$$I(\boldsymbol{x},\underline{\xi}) \le J(\boldsymbol{x},\underline{\xi})$$

ist. Da sich eine solche Kleinergleichbeziehung auch auf Grenzwerte überträgt, ist daher

$$\int_a^b f(x)\,dx \le \int_a^b g(x)\,dx\,,$$

wie behauptet.

Insbesondere können wir dies auch anwenden auf die Ungleichungskette

$$-|f(x)| \le f(x) \le |f(x)|$$

und erhalten

$$-\int_a^b |f(x)|\ dx \le \left|\int_a^b f(x)\,dx\right| \le \int_a^b |f(x)|\ dx\,,$$

d.h.

$$\left|\int_a^b f(x)\,dx\right| \le \int_a^b |f(x)|\ dx\,.$$

■

2) Linearität und Zusammensetzung: Genauso einfach ist die Linearitätsregel:

Satz: Für zwei stückweise stetige Funktionen $f, g\colon [a,b] \to \mathbb{R}$ und zwei reelle Zahlen α, β ist

$$\int_a^b \big(\alpha f(x) + \beta g(x)\big)\, dx = \alpha \int_a^b f(x)\, dx + \beta \int_a^b g(x)\, dx\,.$$

Beweis: Wie beim letzten Satz: Für jede RIEMANN-Summe zu einer festen Unterteilung $\boldsymbol{x}$ und jede damit kompatible Zwischenwertsequenz $\underline{\xi}$ gilt die zur Behauptung des Satzes analoge Gleichung, und damit gilt im Limes auch der Satz selbst. ∎

Für die Zusammensetzung von Integrationsintervallen gilt erwartungsgemäß

Satz: Für $a \le b \le c$ und eine stückweise stetige Funktion $f\colon [a,c] \to \mathbb{R}$ ist

$$\int_a^c f(x)\, dx = \int_a^b f(x)\, dx + \int_b^c f(x)\, dx\,.$$

Beweis: $\int_a^c f(x)\, dx$ kann als Grenzwert von RIEMANN-Summen zu einer beliebigen Folge sich verfeinernder Unterteilungen von $[a,c]$ berechnet werden; insbesondere können also Unterteilungen gewählt werden, die den Zwischenpunkt b enthalten, und dafür folgt die Behauptung unmittelbar. ∎

3) Der Mittelwertsatz der Integralrechnung: Der Mittelwertsatz der *Differentialrechnung* ist der wohl wichtigste Satz aus der Analysis I. Mit dieser überragenden Bedeutung kann der Mittelwertsatz der *Integralrechnung* nicht konkurrieren; trotzdem ist auch er eine sehr nützliche Aussage, die insbesondere auch die zu Beginn dieses Paragraphen postulierte Interpretation von Integralen als Mittelwerten für das RIEMANN-Integral auf eine solide Grundlage stellt.

Im Hinblick auf spätere Anwendungen sei der Satz zunächst etwas allgemeiner formuliert, als wir ihn im Augenblick bentigen; an der Schwierigkeit des Beweises ändert diese Verallgemeinerung nichts.

Satz: $f\colon [a,b] \to \mathbb{R}$ sei eine stetige Funktion; $g\colon [a,b] \to \mathbb{R}$ sei stückweise stetig und nirgends negativ. Dann gibt es ein $\xi \in [a,b]$, so daß

$$\int_a^b f(x)g(x)\, dx = f(\xi) \int_a^b g(x)\, dx\,.$$

Die Funktion g sollte man sich dabei als eine „Gewichtsfunktion“ vorstellen, die die verschiedenen x-Werte verschieden stark gewichtet. Am anschaulichsten ist der Spezialfall $g \equiv 1$, der deshalb noch einmal getrennt angegeben sei:

Korollar: Für eine stetige Funktion $f\colon [a,b] \to \mathbb{R}$ gibt es ein $\xi \in [a,b]$ mit

$$f(\xi) = \frac{1}{b-a} \int_a^b f(x)\, dx\,.$$

Mit anderen Worten: Der „Mittelwert“ der Funktion f im Intervall $[a,b]$ existiert nicht nur, sondern wird auch an (mindestens) einer Stelle im Intervall angenommen.

Zum *Beweis* des Satzes betrachten wir das Minimum m und das Maximum M von f auf $[a,b]$. Da g nirgends negativ wird, gilt dann auch für jedes $x \in [a,b]$

$$m\, g(x) \le f(x)g(x) \le M\, g(x)$$

und damit nach der Monotonieregel

$$m \int_a^b g(x)\, dx \le \int_a^b f(x)g(x)\, dx \le M \int_a^b g(x)\, dx\,.$$

Es gibt daher eine reelle Zahl μ zwischen m und M, den Mittelwert, so daß

$$\int_a^b f(x)g(x)\,dx = \mu \int_a^b g(x)\,dx$$

ist, und da f stetig ist, gibt es nach dem Zwischenwertsatz mindestens ein $\xi \in [a,b]$, für das $f(\xi) = \mu$ ist.

Damit ist der Satz bewiesen, und das Korollar ist natürlich einfach der Spezialfall $g \equiv 1$, für den das Integral über g gleich $b-a$ ist. ■

Man beachte, daß es hier nicht genügt, daß f nur eine *stückweise* stetige Funktion ist: Für

$$f(x) = \begin{cases} 0 & \text{für } x < 0 \\ 1 & \text{für } x \geq 0 \end{cases}$$

ist

$$\frac{1}{1-(-1)} \int_{-1}^{1} f(x)\,dx = \frac{1}{2}\int_0^1 1\,dx = \frac{1}{2}$$

nicht in der Form $f(\xi)$ darstellbar.

e) Der Hauptsatz der Differential- und Integralrechnung

Eine Motivation zur Einführung eines Integrals war die Suche nach einer Umkehroperation zur Differentiation: Genau wie die Differentiation aus einem zeitabhängigen Weg eine Geschwindigkeit macht, wollten wir ausgehend von einer zeitabhängigen Geschwindigkeit den Weg zurückbekommen. Der Hauptsatz der Differential- und Integralrechnung sagt uns nun endlich, daß wir dieses Ziel erreicht haben:

Hauptsatz der Differential- und Integralrechung: $f\colon [a,b] \to \mathbb{R}$ sei eine stetige Funktion und

$$F_a(x) = \int_a^b f(\xi)\,d\xi\,.$$

Dann ist $F_a'(x) = f(x)$ für alle $x \in (a,b)$.

Ist umgekehrt F eine differenzierbare Funktion, deren Ableitung auf dem offenen Intervall (a,b) mit f übereinstimmt, so gibt es eine reelle Zahl C derart, daß $F(x) = F_a(x) + C$ ist.

Beweis: Die Ableitung von F_a ist

$$F_a'(x) = \lim_{h\to 0} \frac{F_a(x+h) - F_a(x)}{h}\,;$$

für positives h ist

$$F_a(x+h) - F_a(x) = \int_a^{x+h} f(x)\,dx - \int_a^x f(x)\,dx = \int_x^{x+h} f(x)\,dx$$

(s. Abschnitt *b)*); für negatives h ist entsprechend

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

Nach dem Mittelwertsatz der Integralrechnung aus dem vorigen Abschnitt gibt es in jedem der beiden Fälle ein ξ zwischen x und $x+h$, so daß das Integral gleich $|h| \cdot f(\xi)$ ist, d.h.

$$\frac{F_a(x+h) - F_a(x)}{h} = f(\xi)\,.$$

Geht nun h gegen Null, so muß die zwischen x und $x+h$ liegende Zahl ξ gegen x gehen; wegen der Stetigkeit von f ist also $F_a'(x) = f(x)$

Ist F eine weitere Funktion mit Ableitung f, so hat die Differenz h von F und F_a die Ableitung Null. Nun erinnern wir uns an die Analysis I: Wenn die Ableitung einer Funktion $h(x)$ verschwindet, muß die Funktion konstant sein. In der Tat: Gäbe es im Definitionsbereich von h zwei Werte $x_1 \neq x_2$ mit $h(x_1) \neq h(x_2)$, so gäbe es nach dem Mittelwertsatz der Differentialrechnung ein $\xi \in (x_1, x_2)$, so daß

$$\frac{h(x_2) - h(x_1)}{x_2 - x_1} = h'(\xi)$$

wäre. Hier steht links eine von Null verschiedene Zahl, rechts aber Null, ein Widerspruch.

Also gibt es eine Konstante C, so daß $F(x) - F_a(x) = C$ oder

$$F(x) = F_a(x) + C$$

ist, wie behauptet. ■

Als Anwendung für die Berechnung bestimmter Integrale ergibt sich das folgende

Korollar: Ist F irgendeine differenzierbare Funktion mit der Eigenschaft, daß $F'(x) = f(x)$ ist im Intervall $[a, b]$, so gilt

$$\int_a^b f(x)\,dx = F(b) - F(a)\,.$$

Beweis: Mit obigen Bezeichnungen ist das Integral gleich $F_a(b)$. Zur Funktion F gibt es nach dem Hauptsatz eine Konstante C, so daß $F(x) = F_a(x) + C$ ist. Damit ist $F(b) - F(a) = F_a(b)$, wie gewünscht. ■

Für „einfache" Integranden ist dies im allgemeinen die beste Möglichkeit zur Berechnung eines Integrals; die Funktionen F haben daher einen Namen verdient:

Definition: Eine differenzierbare Funktion F heißt *Stammfunktion* von f, wenn $F'(x) = f(x)$ ist; wir schreiben

$$F(x) = \int f(\xi)\,d\xi + C\,,$$

wobei die *Integrationskonstante* C die Nichteindeutigkeit der Stammfunktion ausdrückt.

Diese Nichteindeutigkeit sorgt auch dafür, daß zwar nach obigem Hauptsatz die Differentiation die Integration rückgängig macht, die Integration umgekehrt aber die Differentiation nur bis auf auf eine Konstante:

Korollar: $\displaystyle\int f'(x)\,dx = f(x) + C$

Beweis: Klar, denn f ist eine Stammfunktion von f'. ■

Falls wir eine Stammfunktion von f kennen, ist die Berechnung des Integrals

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

also ganz einfach. Eine erste Auswahl von Stammfunktionen erhalten wir durch Rückwärtslesen von Differentiationsregeln, zum Beispiel

$$\frac{d}{dx}x^n = nx^{n-1} \Rightarrow \int x^n\,dx = \frac{x^{n+1}}{n+1} \quad \text{falls } n \neq -1$$

$$\frac{d}{dx}e^{ax} = ae^{ax} \Rightarrow \int e^{ax}\,dx = \frac{e^{ax}}{a} \quad \text{falls } a \neq 0\,,$$

und auch Formeln wie

$$\int \sin\omega x\,dx = -\frac{\cos\omega x}{\omega} + C \quad \text{und} \quad \int \cos\omega x\,dx = \frac{\sin\omega x}{\omega} + C$$

sind nun problemlos. Für weitere Stammfunktionen müssen wir allerdings zunächst noch einige Funktionen kennenlernen.

f) Trigonometrische Funktionen, Hyperbelfunktionen und ihre Umkehrfunktionen

Die Ableitung einer rationalen Funktion ist wieder eine rationale Funktion; für Stammfunktionen gilt allerdings nicht dergleichen, denn wie wir gerade gesehen haben, ist schon die Stammfunktion von x^{-1} nicht mehr rational, sondern der natürliche Logarithmen. Wie wir bald sehen werden, reichen (im Reellen) auch Logarithmen noch nicht aus, um alle rationalen Funktionen integrieren zu können; wir brauchen zusätzlich noch die sogenannten Arkus- und Areafunktionen.

Die Arkusfunktionen sind die Umkehrungen der trigonometrischen Funktionen: Letztere ordnen einem Winkel eine reelle Zahl zu; die Umkehrfunktionen liefern also als Ergebnis einen Winkel oder auch Bogen, auf lateinisch *arcus* genannt.

Unter den trigonometrischen Funktionen sind zumindest der Sinus und der Cosinus aus der Analysis I bekannt; die meisten kennen wohl auch

bereits deren Quotienten

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

den Tangens. Im rechtwinkligen Dreieck gibt er das Verhältnis zwischen Gegenkathete und Ankathete eines Winkels an und damit das, was man üblicherweise als *Steigung* bezeichnet: das Verhältnis der Höhendifferenz zur horizontalen Entfernung.

Da $\cos x$ für $x = \pi/2 + k\pi$ mit $k \in \mathbb{Z}$ verschwindet, ist $\tan x$ an diesen Stellen nicht definiert; in ihrer Umgebung wird er links, wo Sinus und Cosinus dasselbe Vorzeichen haben (+ für gerades k, − für ungerades) beliebig groß, rechts beliebig negativ; siehe dazu auch Abbildung 55.

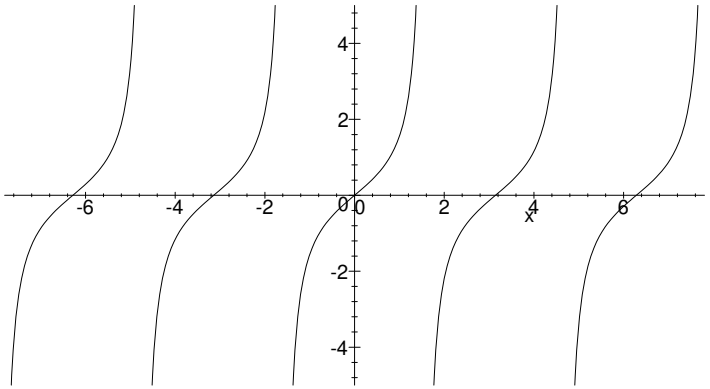


Abb. 55: Der Tangens

Aus den speziellen Werten von Sinus und Cosinus berechnet man leicht die entsprechenden Werte für Tangens:

x	0	$\frac{\pi}{6}$	$\frac{\pi}{4}$	$\frac{\pi}{3}$	$\frac{\pi}{2}$
$\tan x$	0	$\frac{\sqrt{3}}{3}$	1	$\sqrt{3}$	∞

Da alle trigonometrischen Funktionen periodisch sind, können sie keine global definierten Umkehrfunktionen haben; wir müssen sie also jeweils

auf ein Intervall einschränken, auf dem sie injektiv oder – besser noch – monoton sind.

Im Fall des Sinus bietet sich das Intervall von $-\pi/2$ bis $\pi/2$ an, in dem er monoton von -1 auf 1 ansteigt; wir definieren daher *den Arkussinus* oder – genauer ausgedrückt – den *Hauptwert* des Arkussinus für dieses Intervall:

$$\arcsin\colon [-1,1] \to \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$$

ist die Umkehrfunktion von

$$\sin\colon \left[-\frac{\pi}{2}, \frac{\pi}{2}\right] \to [-1,1]\,.$$

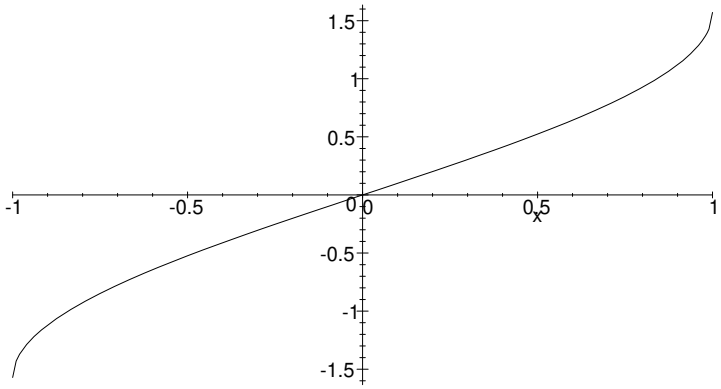


Abb. 56: Der Arkussinus

Der Cosinus fällt von seinem Maximalwert 1 an der Stelle Null monoton ab, bis er bei π den Wert -1 erreicht hat. Wir definieren den *Hauptwert* der Umkehrfunktion *Arkuscosinus* oder kurz *den* Arkuscosinus daher für dieses Intervall:

$$\arccos\colon [-1,1] \to [0,\pi]$$

ist die Umkehrfunktion von

$$\cos\colon [0,\pi] \to [-1,1]\,.$$

Aufgrund der Beziehung $\cos\varphi = \sin\left(\frac{\pi}{2} - \varphi\right)$ und der Wahl der Wertebereiche ist

$$\arccos x = \frac{\pi}{2} - \arcsin x\,.$$

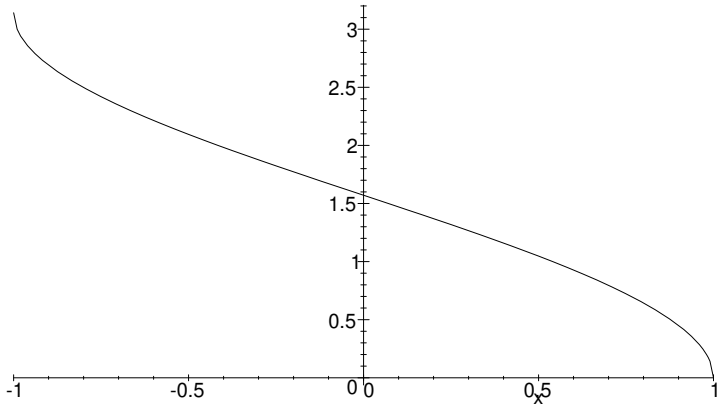


Abb. 57: Der Arkuscosinus

Der Tangens schließlich ist monoton ansteigend im offenen Intervall $(-\pi/2, \pi/2)$ und nimmt dort jeden reellen Wert an; wir definieren den *Hauptwert* der Umkehrfunktion *Arkustangens* oder kurz *den* Arkustangens daher für dieses Intervall:

$$\arctan\colon \mathbb{R} \to \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$$

ist die Umkehrfunktion von

$$\tan\colon \left(-\frac{\pi}{2}, \frac{\pi}{2}\right) \to \mathbb{R}\,.$$

Gelegentlich ganz nützlich ist der Wert an der Stelle 1: Da nach obiger Tabelle $\tan\frac{\pi}{4}$ den Wert 1 hat, ist $\arctan(1) = \frac{\pi}{4}$ oder

$$\pi = 4\,\arctan(1)\,.$$

Falls man in einem Programm den Wert π benötigt und ihn über diese Formel als Programmkonstante definiert, kann man – eine saubere Implementierung der Arkusfunktionen vorausgesetzt – sicher sein, daß π

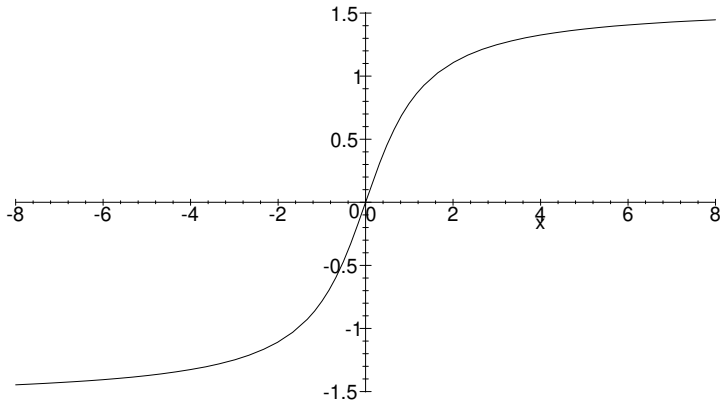


Abb. 58: Der Arkustangens

mit der maximal möglichen Stellenzahl der jeweiligen Gleitkommaarithmetik dargestellt wird – ohne daß man diese Stellenzahl oder gar π mit der entsprechenden Genauigkeit zu kennen braucht.

Abgesehen vom gerade definierten Hauptwert lassen sich natürlich noch in offensichtlicher Weise weitere Zweige definieren; wir wollen hier auf Einzelheiten verzichten. Der Leser sollte sich allerdings klarmachen, daß es beispielsweise für den Arkussinus sowohl monoton wachsende als auch monoton fallende Zweige gibt.

Mit Hilfe der Arkusfunktionen lassen sich die Polarkoordinatendarstellung und die kartesische Darstellung der komplexen Zahlen ineinander umrechnen: Da $z = re^{i\varphi}$ Realteil $x = r\cos\varphi$ und Imaginärteil $y = r\sin\varphi$ hat, sollte außer der bereits bekannten Formel

$$r = \sqrt{x^2 + y^2}$$

für $z \neq 0$ auch so etwas wie

$$\varphi = \arctan\frac{y}{x} = \arccos\frac{x}{r} = \arcsin\frac{y}{r}$$

gelten. Dies im Prinzip auch richtig, allerdings nicht immer für den Hauptwert: Da alle drei Umkehrfunktionen nur Werte in einem Intervall der Länge π annehmen, während das Argument (genauer: der

Hauptwert des Arguments) einer komplexen Zahl im doppelt so langen Intervall $(-\pi, \pi]$ liegt, kann obige Gleichung bei keiner der drei Funktionen für alle z gelten. Eine korrekte Formel mit dem Hauptwert des Arkuscosinus ist beispielsweise

$$\arg z = \begin{cases} \arccos \frac{x}{r} & \text{falls } \mathfrak{Im}\, z \geq 0 \\ -\arccos \frac{x}{r} & \text{falls } \mathfrak{Im}\, z < 0 \end{cases} .$$

Der für uns interessanteste Aspekt der Arkusfunktionen sind ihre *Ableitungen,* die uns neue, bislang unbekannte Stammfunktionen liefern sollen.

Wie in der Analysis I bei der Ableitung des Logarithmus gehen wir aus von der Formel

$$g'(x) = \frac{1}{f'\big(g(x)\big)}$$

für die Umkehrfunktion g einer Funktion f. Da die Ableitung des Sinus der Cosinus ist, erhalten wir also

$$\arcsin'(x) = \frac{1}{\cos(\arcsin x)} .$$

Wegen $\sin^2 y + \cos^2 y = 1$ ist

$$\cos^2(\arcsin x) = 1 - \sin^2(\arcsin x) = 1 - x^2 ;$$

da der Cosinus im Intervall $[-\pi/2, \pi/2]$, in dem der Arkussinus seine Werte annimmt, größer oder gleich Null ist, folgt

$$\cos(\arcsin x) = \sqrt{1 - x^2} \quad \text{und} \quad \arcsin'(x) = \frac{1}{\sqrt{1 - x^2}} .$$

Genauso ist

$$\arccos'(x) = \frac{-1}{\sin(\arccos x)} = \frac{-1}{\sqrt{1 - x^2}} ,$$

da der Sinus zwischen 0 und π keine negativen Werte annimmt.

Für den Arkustangens schließlich müssen wir zunächst Tangens selbst ableiten; nach der Quotientenregel ist

$$\tan'(x) = \frac{d}{dx} \frac{\sin x}{\cos x} = \frac{\cos^2 x + \sin^2 x}{\cos^2 x} = 1 + \tan^2 x$$

und dementsprechend

$$\arctan'(x) = \frac{1}{1 + \tan^2(\arctan x)} = \frac{1}{1 + x^2} .$$

Da sich die Ableitungen von Arkussinus und Arkuscosinus nur im Vorzeichen unterscheiden (was natürlich von vornherein klar war, da sich die Werte der beiden Funktionen stets zu $\pi/2$ ergänzen) sind als Stammfunktionen vor allem der Arkussinus und der Arkustangens interessant; wir haben die beiden neuen Formeln

$$\int \frac{dx}{1 + x^2}\, dx = \arctan x + C \quad \text{und} \quad \int \frac{dx}{\sqrt{1 - x^2}}\, dx = \arcsin x + C .$$

Als nächstes wollen wir uns überlegen, wie wir die Integrale

$$\int \frac{dx}{1 - x^2}\, dx \quad \text{und} \quad \int \frac{dx}{\sqrt{1 + x^2}}\, dx$$

mit dem jeweils anderen Vorzeichen bekommen können. Eine offensichtliche Lösung besteht darin, einfach x durch ix zu ersetzen, und das legt es nahe, in Analogie zu den EULERschen Formeln

$$\cos x = \frac{e^{ix} + e^{-ix}}{2} \quad \text{und} \quad \sin x = \frac{e^{ix} - e^{-ix}}{2i}$$

neue Funktionen

$$\cosh x = \frac{e^{x} + e^{-x}}{2} \quad \text{und} \quad \sinh x = \frac{e^{x} - e^{-x}}{2}$$

zu definieren, den *Cosinus hyperbolicus* und den *Sinus hyperbolicus.*

Der Zusatz *hyperbolicus* läßt sich leicht verstehen: Genau wie der gewöhnliche Sinus und Cosinus als *Kreisfunktionen* bezeichnet werden, weil die Punkte $(\sin t, \cos t)$ auf dem Kreis $x^2 + y^2 = 1$ liegen, haben die Hyperbelfunktionen ihren Namen daher, daß

$$\begin{aligned} \cosh^2 t - \sinh^2 t &= \left(\frac{e^t + e^{-t}}{2}\right)^2 - \left(\frac{e^t + e^{-t}}{2}\right)^2 \\ &= \frac{e^{2t} + 2 + e^{-2t}}{4} - \frac{e^{2t} - 2 + e^{-2t}}{4} = 1 \end{aligned}$$

ist, so daß die Punkte $(\cosh t, \sinh t)$ auf der Hyperbel $x^2 - y^2 = 1$ liegen, d.h

$$\cosh^2 x - \sinh^2 x = 1 \quad \text{für alle } x \in \mathbb{R}\,.$$

Damit ist insbesondere $\cosh^2 x$ stets ≥ 1; da der Cosinus hyperbolicus genau wie die beiden Exponentialfunktionen, aus denen er zusammengesetz ist, keine negativen Werte annehmen kann, folgt also

$$\cosh x \geq 1 \quad \text{für alle } x \in \mathbb{R}\,.$$

Sofort aus der Definition folgen die Ableitungsregeln

$$\cosh'(x) = \sinh x \quad \text{und} \quad \sinh' x = \cosh x\,;$$

auf Grund obiger Ungleichung ist daher der Sinus hyperbolicus monoton steigend auf ganz $\mathbb{R}$. Man sieht leicht, daß sein Vorzeichen jeweils das von x ist (Insbesondere ist also auch $\sinh 0 = 0$.), d.h. der Cosinus hyperbolicus ist monoton fallend für negative und monoton steigend für positive x.

Für große Werte von x wird e^{-x} beliebig klein, sowohl $\sinh x$ als auch $\cosh x$ unterscheiden sich für solche Werte also beliebig wenig von $\frac{1}{2}e^x$ und gehen insbesondere gegen unendlich. Für stark negative Werte wird e^x beliebig klein und das Verhalten beider Funktionen wird dominiert durch den Term $\pm e^{-x}$. Daher geht $\sinh x$ für $x \to -\infty$ gegen $-\infty$ und $\cosh x$ gegen $+\infty$. Offensichtlich ist $\sinh x$ eine ungerade, $\cosh x$ aber eine gerade Funktion. In Abbildung 59 sind gestrichelt auch noch die Funktionen $\frac{1}{2}e^x$ im positiven und $\pm\frac{1}{2}e^{-x}$ im negativen Bereich eingezeichnet; wie man sieht, ist die Übereinstimmung mit $\sinh x$ *bzw.* $\cosh x$ schon ab etwa $|x| \geq 2$ recht gut.

Der Graph des Cosinus hyperbolicus erinnert an eine durchhängende Kette, und tatsächlich kann die Variationsrechnung zeigen, daß eine Kette mit reibungsfrei gegeneinander beweglichen Gliedern genau diese Form hat.

In völliger Analogie zum Tangens definieren wir auch noch einen *Tangens hyperbolicus* durch die Vorschrift

$$\tanh x \underset{\text{def}}{=} \frac{\sinh x}{\cosh x} = \frac{e^x + e^{-x}}{e^x - e^{-x}}\,;$$

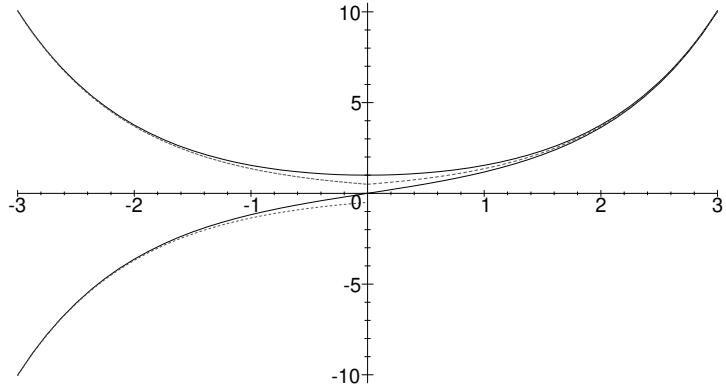


Abb. 59: Sinus hyperbolicus und Cosinus hyperbolicus

da der Cosinus hyperbolicus stets größer oder gleich eins ist, kann der Nenner nie Null werden, so daß die Funktion auf ganz $\mathbb{R}$ definiert ist. Dabei ist

$$\lim_{x\to\infty} \tanh x = \lim_{x\to\infty} \frac{e^x + e^{-x}}{e^x - e^{-x}} = \lim_{x\to\infty} \frac{e^x(1 + e^{-2x})}{e^x(1 - e^{-x})}$$
$$= \lim_{x\to\infty} \frac{1 + e^{-2x}}{1 - e^{-x}} = 1$$

und entsprechend

$$\lim_{x\to-\infty} \tanh x = \lim_{x\to-\infty} \frac{e^x + e^{-x}}{e^x - e^{-x}} = \lim_{x\to-\infty} \frac{e^{2x} + 1}{e^{2x} - 1} = -1\,.$$

Die Ableitung des Tangens hyperbolicus ist nach der Quotientenregel

$$\tanh' x = \frac{\cosh^2 x - \sinh^2 x}{\cosh^2 x} = \begin{cases} \frac{1}{\cosh^2 x} & \text{nach obiger Formel} \\ 1 - \tanh^2 x & \text{durch Ausdividieren} \end{cases},$$

wobei wie beim Tangens je nach Anwendung mal die eine, mal die andere Form des Ergebnisses nützlicher ist.

Aus beiden Ausdrücken sieht man sofort, daß die Ableitung stets positiv ist, der Tangens hyperbolicus steigt also monoton von -1 nach $+1$, wobei beide Werte nur asymptotisch angenommen werden.

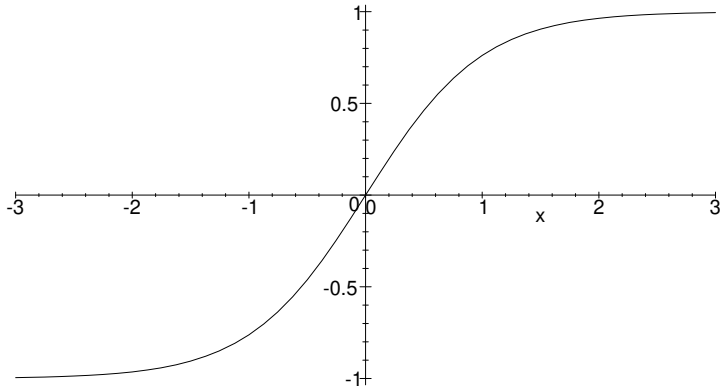


Abb. 60: Der Tangens hyperbolicus

Die Umkehrfunktionen der Hyperbelfunktionen werden als *Areafunktionen* bezeichnet und heißen *Areasinus hyperbolicus, Areacosinus hyperbolicus* und *Areatangens hyperbolicus.*

Aus obiger Diskussion folgt, daß der Areasinus hyperbolicus

$$\operatorname{arsinh}\colon \mathbb{R} \to \mathbb{R}$$

auf ganz $\mathbb{R}$ definiert ist und der Areatangens hyperbolicus

$$\operatorname{artanh}\colon (-1, 1) \to \mathbb{R}$$

nur auf dem offenen Einheitsintervall. Der Cosinus hyperbolicus hat keine eindeutig bestimmte Umkehrfunktion, da er für positive und für negative Argumente jeweils denselben Wert annimmt. Wir definieren den Areacosinus hyperbolicus

$$\operatorname{arcosh}\colon \mathbb{R}_{\geq 1} \to \mathbb{R}_{\geq 0}$$

als die Umkehrfunktion des positiven Zweigs.

Da Sinus hyperbolicus und Cosinus hyperbolicus asymptotisch wie eine Exponentialfunktion ansteigend, steigen Areasinus hyperbolicus und Areacosinus hyperbolicus wie Logarithmen, also sehr langsam.

Auch diese Funktionen sind wieder vor allem wegen ihrer Ableitungen interessant; eine weitgehend zum Fall der Arkusfunktionen analoge

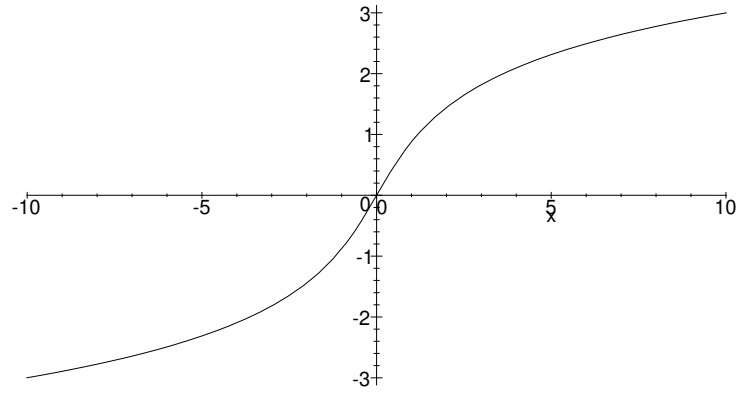


Abb. 61: Der Areasinus hyperbolicus

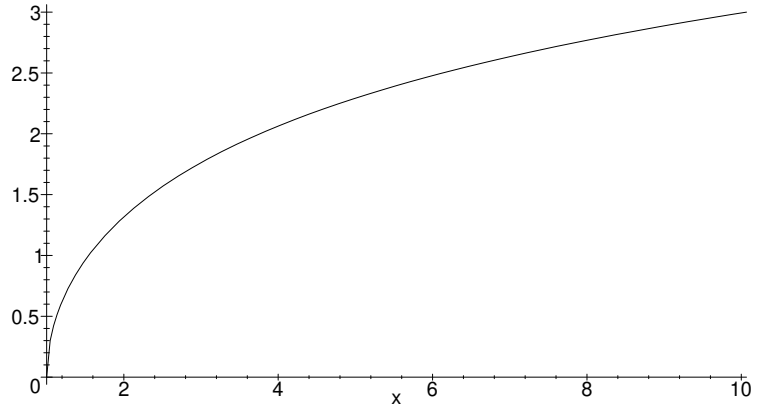


Abb. 62: Der Areacosinus hyperbolicus

Rechnung zeigt, daß

$$\operatorname{arsinh}'(x) = \frac{1}{\cosh(\operatorname{arsinh} x)} = \frac{1}{\sqrt{1+x^2}}$$

und

$$\operatorname{arcosh}'(x) = \frac{1}{\sinh(\operatorname{arcosh} x)} = \frac{1}{\sqrt{x^2-1}}$$

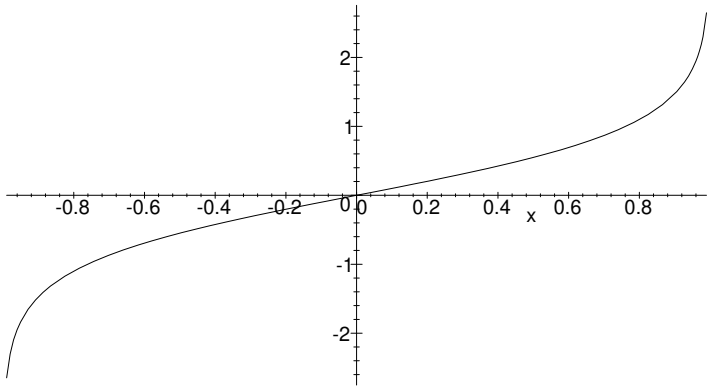


Abb. 63: Der Areatangens hyperbolicus

ist. Für den Areatangens hyperbolicus schließlich ist

$$\operatorname{artanh}'(x) = \frac{1}{1-\tanh^2(\operatorname{artanh} x)} = \frac{1}{1-x^2}\,.$$

Dies führt also auf die neuen Stammfunktionen

$$\int \frac{dx}{\sqrt{1+x^2}} = \operatorname{arsinh}(x) + C, \qquad \int \frac{dx}{\sqrt{x^2-1}} = \operatorname{arcosh}(x) + C$$

und

$$\int \frac{dx}{1-x^2} = \operatorname{artanh}(x) + C\,.$$

Die letztere wird allerdings nicht unbedingt gebraucht, denn da

$$\frac{1}{1-x^2} = \frac{1}{2}\cdot\left(\frac{1}{1-x} + \frac{1}{1+x}\right)$$

ist, können wir dieses Integral auch ausrechnen als

$$\int \frac{dx}{1-x^2} = \frac{\ln|1-x| + \ln|1+x|}{2} + C\,,$$

wobei diese Formel noch den Vorteil eines größeren Definitionsbereichs hat.

g) Partielle Integration

Eine wichtige Regel der Differentialrechnung ist die *Produktregel*

$$\frac{d}{dx}\big(f(x)g(x)\big) = f'(x)g(x) + f(x)g'(x)\,.$$

Für die Zwecke der Integralrechnung schreiben wir sie besser als

$$f'(x)g(x) = \frac{d}{dx}\big(f(x)g(x)\big) - f(x)g'(x)\,,$$

was integriert auf

$$\int f'(x)g(x)\,dx = f(x)g(x) - \int f(x)g'(x)\,dx$$

führt – die Regel der partiellen Integration. Sie ist dann nützlich, wenn sich der Integrand als Produkt schreiben läßt, wobei einer der Faktoren eine bekannte Stammfunktion hat; gelegentlich ist dann das Produkt $f(x)g'(x)$ leichter integrierbar als $f'(x)g(x)$.

Mit den wenigen Funktionen, die wir bislang kennen, lassen sich nur wenige interessante Beispiele konstruieren; die volle Kraft dieser Regel werden wir erst später kennenlernen, wenn wir unser Repertoire an Funktionen erweitert haben. Hier sei nur $\int xe^x\,dx$ betrachtet: Da die Exponentialfunktion ihre eigene Ableitung und Stammfunktion ist, empfiehlt sie sich für die Rolle der Funktion f, während $g(x) = x$ mit $g'(x) = 1$ auf Vereinfachungen auf der rechten Seite hoffen läßt. In der Tat führt partielle Integration mit

$$\int xe^x\,dx = xe^x - \int e^x\,dx = xe^x - e^x + C = (x-1)e^x + C$$

auf eine Stammfunktion. Genauso lassen sich rekursiv auch die Integrale $\int x^n e^x\,dx$ berechnen:

$$\int x^2e^x\,dx = x^2e^x - \int 2xe^x\,dx = (x^2-2x+2)e^x + C$$

und

$$\int x^ne^x\,dx = x^ne^x - n\int x^{n-1}e^x\,dx\,.$$

h) Substitutionsregel

Auch die Kettenregel

$$\frac{d}{dx} f\big(g(x)\big) = f'\big(g(x)\big) \cdot g'(x)$$

läßt sich umschreiben zu einer Integrationsregel, der *Substitutionsregel*

$$\int f\big(g(x)\big) \cdot g'(x)\, dx = F\big(g(x)\big) + C\,,$$

wobei F eine Stammfunktion von f ist.

Diese Regel dürfte wohl meist der erfolgversprechendste Versuch zum Auffinden einer Stammfunktion sein – vor allem, wenn man sie von rechts nach links liest, um für eine bekannte Funktion f die unbekannte Stammfunktion F zu berechnen. Als Substitution g wählt man hier eine „geeignete" bijektive Funktion, die zu einer Vereinfachung auf der linken Seite führt.

1) Der Spezialfall logarithmischer Ableitungen: Für $f(x) = 1/x$ führt die Substitutionsregel

$$\int f\big(g(x)\big) g'(x)\, dx = F\big(g(x)\big) + C \quad \text{mit} \quad F'(x) = f(x)$$

zur Integrationsregel

$$\int \frac{g'(x)}{g(x)}\, dx = \ln|g(x)| + C\,.$$

Als erste Anwendung betrachten wir

$$\tan x = \frac{\sin x}{\cos x} = -\frac{g'(x)}{g(x)}$$

mit $g(x) = \cos x$. Nach der gerade bewiesenen Regel ist daher

$$\int \tan x\, dx = -\ln|\cos(x)| + C\,,$$

eine Funktion, die genau wie der Tangens selbst an den Nullstellen der Cosinusfunktion nicht definiert ist.

Beim Tangens hyperbolicus gibt es keine Probleme mit dem Vorzeichen; hier ist ganz entsprechend

$$\int \tanh x\, dx = \ln|\cosh(x)| + C = \ln\cosh(x) + C\,,$$

da der Cosinus hyperbolicus nur positive Werte annimmt.

Auch Integrale wie

$$\int \frac{x}{1+x^2}\, dx$$

lassen sich nach dieser Regel ausrechnen: Da die Ableitung des Nenners $2x$ ist, folgt

$$\int \frac{x}{1+x^2}\, dx = \tfrac{1}{2} \ln\left|1+x^2\right| + C = \tfrac{1}{2}\ln(1+x^2) + C\,.$$

Damit können wir dann beispielsweise den Arkustangens integrieren: Partielle Integration zeigt, daß

$$\begin{aligned}\int 1 \cdot \arctan x\, dx &= x \arctan x - \int \frac{x}{1+x^2}\, dx + C \\ &= x \arctan x - \tfrac{1}{2}\ln(1+x^2) + C\end{aligned}$$

ist.

2) Substitutionen mit linearen Funktionen: Eine der elementarsten Anwendungen der Substitutionsregel, auf die man üblicherweise auch ohne diese Regel kommt, ist die Substitution mit linearen Funktionen $g(x) = ax + b$; hier besagt die Substitutionsregel, daß

$$\int f(ax+b) \cdot a\, dx = F(ax+b) \quad \text{mit} \quad F'(x) = f(x)$$

ist, oder besser

$$\int f(ax+b)\, dx = \frac{F(ax+b)}{a} \quad \text{mit} \quad F'(x) = f(x)\,.$$

Somit ist beispielsweise

$$\int \sin(\omega t + \varphi)\, dt = \frac{-\cos(\omega t + \varphi)}{\omega} + C$$

oder

$$\int \frac{dx}{x^2+a^2} = \int \frac{\frac{dx}{a^2}}{\left(\frac{x}{a}\right)^2+1} = \frac{1}{a}\arctan\left(\frac{x}{a}\right) + C\,.$$

Entsprechend berechnet man auch

$$\int \frac{dx}{x^2-a^2} = -\frac{1}{a}\operatorname{artanh}\left(\frac{x}{a}\right) + C$$

und sogar ganz allgemein $\displaystyle\int \frac{dx}{ax^2+bx+c}$:

Da der Fall $a = 0$ uninteressant ist (*bzw.* eine unbedingt notwendige Übungsaufgabe für alle, die nicht sofort einsehen warum), schließen wir dies aus und schreiben

$$ax^2+bx+c = a\left(x+\frac{b}{2a}\right)^2 + c - \frac{b^2}{4a}\,.$$

Mit

$$y = x + \frac{b}{2a} \quad \text{und} \quad d = \frac{c}{a} - \frac{b^2}{4a^2}$$

ist dann

$$\int \frac{dx}{ax^2+bx+c} = \frac{1}{a}\int \frac{dy}{y^2+d}$$
$$= \frac{1}{a}\begin{cases} \frac{1}{\sqrt{d}}\arctan\left(\frac{y}{\sqrt{d}}\right) + C & \text{falls } d > 0 \\ \frac{-1}{y} & \text{falls } d = 0 \\ \frac{1}{\sqrt{-d}}\operatorname{artanh}\left(\frac{y}{\sqrt{-d}}\right) + C & \text{falls } d < 0\,, \end{cases}$$

und je nach Vorzeichen von d kann dies entweder über eine der obigen Formeln oder (für $d = 0$) elementar ausgerechnet werden. Insgesamt erhalten wir mit $\Delta = b^2 - 4ac$ (und damit $d = -\Delta/4a^2$)

$$\int \frac{dx}{ax^2+bx+c} = \begin{cases} \frac{2}{\sqrt{-\Delta}}\arctan\left(\frac{2ax+b}{\sqrt{-\Delta}}\right) + C & \text{falls } \Delta < 0 \\ \frac{-2}{2ax+b} + C & \text{falls } \Delta = 0 \\ \frac{-2}{\sqrt{\Delta}}\operatorname{artanh}\left(\frac{2ax+b}{\sqrt{\Delta}}\right) + C & \text{falls } \Delta > 0\,. \end{cases}$$

3) Substitutionen mit trigonometrischen und Hyperbelfunktionen: Wegen der Beziehungen

$$\sin^2 x + \cos^2 x = 1 \quad \text{und} \quad \cosh^2 - \sinh^2 x = 1$$

bieten sich Substitutionen mit trigonometrischen Funktionen und Hyperbelfunktionen an bei Integralen, in denen Ausdrücke der Form $\sqrt{1 \pm x^2}$ und ähnliche vorkommen.

Betrachten wir als einfachstes Beispiel die Stammfunktion von $\sqrt{1-x^2}$ selbst. Mit der Substitution $x = \sin t$ mit $-\frac{\pi}{2} \le t \le \frac{\pi}{2}$ erhalten wir (durch Rückwärtslesen der Substitutionsregel)

$$\int \sqrt{1-x^2}\,dx = \int \sqrt{1-\sin^2 t}\cos t\,dt\,,$$

und dies ist $\int \cos^2 t\,dt$, da Cosinus zwischen $-\frac{\pi}{2}$ und $\frac{\pi}{2}$ nur nichtnegative Werte annimmt.

Dieses Integral kennen wir von Aufgabe *1b)* des dritten Übungsblatts:

$$\int \cos^2 t\,dt = \tfrac{1}{2}(\sin t\cos t + t) + C\,.$$

Mit der Rücksubstitution $t = \arcsin x$ erhalten wir schließlich nach der kurzen Nebenrechnung

$$\cos(\arcsin x) = \sqrt{1-\sin^2(\arcsin x)} = \sqrt{1-x^2}$$

das Ergebnis

$$\int \sqrt{1-x^2}\,dx = \tfrac{1}{2}(x\sqrt{1-x^2} + \arcsin x) + C\,.$$

Aus ähnliche Weise läßt sich auch die Stammfunktion von $\sqrt{1+x^2}$ bestimmen: Hier bietet sich die Substitution $x = \sinh t$ an und wir erhalten

$$\int \sqrt{1+x^2}\,dx = \int \sqrt{1-\sinh^2 t}\cosh t\,dt = \int \cosh^2 t\,dt\,,$$

da $\cosh t$ nur positive Werte annimmt.

Leider kennen wir das rechtsstehende Integral noch nicht; angesichts der großen Ähnlichkeiten zwischen trigonometrischen Funktionen und

Hyperbelfunktionen lohnt es sich aber sicherlich, als ersten Ansatz vor einer partiellen Integration zu schauen, ob vielleicht etwas analoges gilt wie oben. In der Tat ist nach der Produktregel

$$\frac{d}{dt}\tfrac{1}{2}(\sinh t\cosh t + t) = \cosh^2 t + C\,,$$

also

$$\int \cosh^2 t\,dt = \tfrac{1}{2}(\sinh t\cosh t + t) + C\,.$$

Somit ist

$$\int \sqrt{1+x^2}\,dx = \tfrac{1}{2}(x\sqrt{1+x^2} + \operatorname{arsinh} x) + C\,.$$

Als letztes Beispiel, in dem x einmal nicht quadratisch vorkommt, betrachten wir noch

$$\int \sqrt{\frac{1-x}{1+x}}\,dx\,.$$

Da es in diesem Abschnitt um Substitutionen mit trigonometrischen Funktionen geht, versuchen wir es wieder mit dem Ansatz $x = \sin t$ für $-\frac{\pi}{2} \le t \le \frac{\pi}{2}$ und erhalten

$$\int \sqrt{\frac{1-x}{1+x}}\,dx = \int \sqrt{\frac{1-\sin t}{1+\sin t}}\cos t\,dt\,.$$

Hier hilft nun ein Trick weiter: Erweitern wir den Bruch unter der Quadratwurzel mit $1-\sin t$, so wird das Integral zu

$$\int \sqrt{\frac{(1-\sin t)^2}{1-\sin^2 t}}\cos t\,dt = \int \frac{1-\sin t}{\cos t}\cos t\,dt = \int (1-\sin t)\,dt$$
$$= t + \cos t + C\,.$$

Damit ist

$$\int \sqrt{\frac{1-x}{1+x}}\,dx = \arcsin x + \cos(\arcsin x) + C$$
$$= \arcsin x + \sqrt{1-x^2} + C\,.$$

Bei einem so komplizierten Integral empfiehlt es sich, das Ergebnis durch Differentiation zu überprüfen; wir erhalten

$$\frac{d}{dx}\left(\arcsin x + \sqrt{1-x^2}\right) = \frac{1}{\sqrt{1-x^2}} - \frac{x}{\sqrt{1-x^2}}\,,$$

was zunächst so aussieht, als sei irgend etwas falsch gelaufen. Beachtet man aber, daß

$$\sqrt{1+x}\cdot\sqrt{1-x} = \sqrt{1-x^2}$$

ist, so ist

$$\frac{1}{\sqrt{1-x^2}} - \frac{x}{\sqrt{1-x^2}} = \frac{1-x}{\sqrt{1-x^2}} = \frac{\sqrt{1-x}}{\sqrt{1+x}}\,,$$

was erheblich vertrauenserweckender aussieht.

4) Integrale der Form $\int h(e^{ax})\,dx$: Bei solchen Integralen führt oft die Substitution $u = e^{ax}$ oder, anders ausgedrückt, $x = g(u) = \frac{1}{a}\ln u$ zu einer Vereinfachung. Die Substitutionsregel

$$\int f\big(g(u)\big)g'(u)\,du = \int f(x)\,dx$$

wird wegen $g'(u) = \dfrac{dx}{du} = \dfrac{1}{au}$ zu

$$\int h(u)\,\frac{du}{au} = \int h(e^{ax})\,dx\,;$$

insbesondere wird das Integral also bei rationalem h auf ein Integral einer rationalen Funktion zurückgeführt, und dieses kann nach dem im nächsten Abschnitt skizzierten Verfahren der Partialbruchzerlegung berechnet werden.

Im allgemeinen schreibt man eine Substitution wie die obige kurz in der Form

$$u = e^{ax}, \quad dx = \frac{du}{au}\,,$$

wobei man letztere Beziehung auch aus

$$\frac{du}{dx} = ae^{ax} = au \quad \text{oder} \quad du = audx$$

herleiten kann.

Betrachten wir als Beispiel das Integral $\displaystyle\int \frac{e^{3x}}{e^{2x}-1}\,dx$: Da sowohl e^{2x} als auch e^{3x} im Integranden vorkommen, empfiehlt sich die Substitution $u = e^x$ mit $du = u\,dx$; damit wird

$$\int \frac{e^{3x}}{e^{2x}-1}\,dx = \int \frac{u^3}{u^2-1}\,\frac{du}{u} = \int \frac{u^2}{u^2-1}\,du\,,$$

und wir sind beim Integral einer rationalen Funktion angelangt. Elementare Bruchrechnung zeigt, daß

$$\frac{u^2}{u^2-1} = 1 + \frac{1}{u^2-1} = 1 + \frac{\frac{1}{2}}{u-1} - \frac{\frac{1}{2}}{u+1}\,,$$

d.h.

$$\frac{u^2}{u^2-1}\,du = \int du + \tfrac{1}{2}\left(\int \frac{du}{u-1} - \int \frac{du}{u+1}\right) = u + \ln\left|\frac{u-1}{u+1}\right| + C\,.$$

(Für komplizierte Funktionen werden wir im nächsten Abschnitt ein Verfahren kennenlernen, daß rationale Funktionen in Summen aus einfachen Summanden zerlegt.)

Nun muß nur noch die Substitution $u = e^x$ rückgängig gemacht werden, und wir erhalten als Ergebnis

$$\int \frac{e^{3x}}{e^{2x}-1}\,dx = e^x + \tfrac{1}{2}\ln\left|\frac{e^x-1}{e^x+1}\right| + C\,.$$

Dieselbe Technik funktioniert natürlich auch bei Integralen über Hyperbelfunktionen, da man diese genausogut als Ausdrücke in Exponentialfunktionen schreiben kann. Beispielsweise ist

$$\int \frac{dx}{\sinh x} = 2\int \frac{dx}{e^x - e^{-x}} = 2\int \frac{du}{u\left(u-\frac{1}{u}\right)} = 2\int \frac{du}{u^2-1}$$
$$= \ln\left|\frac{u-1}{u+1}\right| + C = \ln\left|\frac{e^x-1}{e^x+1}\right| + C\,.$$

5) Integrale der Form $\int h(\sin x, \cos x)\,dx$***:*** Als letzte Anwendung der Substitutionsregel in diesem Abschnitt wollen wir noch Integrale betrachten, die von trigonometrischen Funktionen abhängen. Hier führt oft die Substitution

$$x = 2\arctan t, \qquad dx = \frac{2\,dt}{t^2+1}$$

zum Erfolg; sie macht $\sin x$ zu $\sin(2\arctan t)$ und $\cos x$ zu $\cos(2\arctan t)$, was wir wie folgt ausrechnen können: Aus

$$\tan^2 x = \frac{\sin^2 x}{\cos^2 x} = \frac{1-\cos^2 x}{\cos^2 x} = \frac{1}{\cos^2 x} - 1$$

folgt, daß

$$\cos(2x) = 2\cos^2(x) - 1 = \frac{2}{1+\tan^2 x} - 1 = \frac{1-\tan^2 x}{1+\tan^2 x}$$

ist und somit

$$\cos(2\arctan t) = \frac{1-t^2}{1+t^2}\,.$$

Da der Arkustangens nur Werte zwischen $-\frac{\pi}{2}$ und $\frac{\pi}{2}$ annimmt, liegt $2\arctan t$ zwischen $-\pi$ und π, ein Intervall, in dem der Sinus dasselbe Vorzeichen hat wie sein Argument. Daher ist

$$\sin(2\arctan t) = \sqrt{1-\cos^2(2\arctan t)} = \sqrt{1 - \frac{(1-t^2)^2}{(1+t^2)^2}} = \frac{2t}{1+t^2}\,.$$

Als Beispiel betrachten wir

$$\int \frac{dx}{\cos x} = \int \frac{1+t^2}{1-t^2}\,\frac{2\,dt}{1+t^2} = -2\int \frac{dt}{t^2-1}\,.$$

Das letzte Integral kennen wir:

$$\int \frac{dt}{t^2-1} = \frac{1}{2}\ln\left|\frac{1-t}{1+t}\right| + C\,.$$

Also ist

$$\int \frac{dx}{\cos x} = -\frac{2}{2}\ln\left|\frac{1-t}{1+t}\right| + C = \ln\left|\frac{1+t}{1-t}\right| + C = \ln\left|\frac{1-\tan\frac{x}{2}}{1+\tan\frac{x}{2}}\right| + C\,.$$

i) Integration rationaler Funktionen

Bekanntlich ist ein Polynom in x ein Ausdruck der Form

$$f(x) = \sum_{\nu=0}^{n} a_\nu x^\nu = a_n x^n + \cdots + a_1 x + a_0 \,;$$

ist $a_n \neq 0$, so bezeichnet man n als den *Grad* des Polynoms, in Zeichen

$$n = \deg f \,.$$

Polynome vom Grad eins, zwei, drei *bzw.* vier heißen linear, quadratisch, kubisch *bzw.* biquadratisch.

Unter einer *rationalen Funktion* versteht man eine Funktion der Form

$$x \mapsto \frac{f(x)}{g(x)} \,,$$

wobei $f(x)$ und $g(x)$ Polynome sind. Es ist klar, daß diese Funktion nur in jenen Punkten $x \in \mathbb{R}$ definiert ist, in denen $g(x) \neq 0$ ist.

Die Integrationstheorie rationaler Funktionen ist seit langem bekannt; im Prinzip kann man zu jeder rationalen Funktion eine Stammfunktion angeben, indem man wie folgt vorgeht: (Auf Einzelheiten und Beweise muß hier weitgehend verzichtet werden, da wir nicht über das notwendige mathematische Instrumentarium verfügen.)

1. Schritt: Man zerlege den Nenner in ein Produkt aus linearen und quadratischen Polynomen. Die ist zumindest grundsätzlich möglich, denn nach dem *Fundamentalsatz der Algebra* läßt sich ein Polynom vom Grad n (mir reellen oder komplexen Koeffizienten) stets in der Form

$$g(x) = a_n(x - z_1)(x - z_2) \cdot \cdots \cdot (x - z_n)$$

schreiben, wobei die *komplexen* Zahlen z_i die Nullstellen von g sind. Wie man sich leicht überlegt, folgt für ein reelles Polynom g aus $g(z) = 0$ sofort, daß auch $g(\overline{z}) = 0$ ist, d.h. die nichtreellen Nullstellen treten als Paare konjugiert komplexer Zahlen auf, die auch jeweils die gleiche Vielfachheit haben, da letztere über das Verschwinden von Ableitungen definiert werden kann.

Das Produkt der Linearfaktoren zu zwei konjugiert komplexen Zahlen ist

$$(x - z)(x - \overline{z}) = x^2 - (z + \overline{z})x + z\overline{z} = x^2 - 2\,\mathfrak{Re}\, z \cdot x + |z|^2 \,,$$

ein quadratisches Polynom mit reellen Koeffizienten. Also läßt sich $g(x)$ als Produkt linearer und quadratischer Polynome mit reellen Koeffizienten schreiben. Faßt man gleiche Faktoren zusammen, so erhält man demnach eine Darstellung

$$g(x) = a_n \ell_1(x)^{e_1} \cdot \cdots \cdot \ell_r(x)^{e_r} \cdot q_1(x)^{f_1} \cdot \cdots \cdot q_s(x)^{f_s}$$

mit linearen Polynomen ℓ_i, quadratischen Polynomen q_j und natürlichen Zahlen e_i, f_j. Dabei sind sowohl die ℓ_i als auch die q_j paarweise voneinander verschieden.

Diese Zerlegung kann algorithmisch selbst bei Polynomen mit rationalen Koeffizienten algorithmisch sehr aufwendig sein; daher kommt der obige Zusatz „im Prinzip“ für die Durchführbarkeit der Integration rationaler Funktionen.

2. Schritt: Man schreibe die Funktion $f(x)/g(x)$ als Summe eines Polynoms und von *Partialbrüchen* der Form

$$\frac{\alpha_{ik}}{\ell_i(x)^k} \quad \text{und} \quad \frac{\beta_{jk}x + \gamma_{jk}}{q_j(x)^k} \,,$$

wobei k von eins bis e_i *bzw.* f_j läuft und α_{ik}, β_{jk} und γ_{jk} reelle Zahlen sind, d.h.

$$\frac{f(x)}{g(x)} = Q(x) \sum_{i=1}^{r} \sum_{k=1}^{e_i} \frac{\alpha_{ik}}{\ell_i(x)^k} + \sum_{j=1}^{s} \sum_{k=1}^{f_j} \frac{\beta_{jk}x + \gamma_{jk}}{q_j(x)^k} \,.$$

Wir müssen uns zunächst überlegen, warum das möglich ist:

Beginnen wir mit zwei teilerfremden Polynomen $p(x)$ und $q(x)$. Der erweiterte EUKLIDische Algorithmus aus Kapitel I, §2*e)* liefert dazu neue Polynome $\alpha(x), \beta(x)$, so daß

$$\alpha(x)p(x) + \beta(x)q(x) = 1$$

ist und damit

$$\frac{1}{p(x)q(x)} = \frac{\alpha(x)p(x)+\beta(x)q(x)}{p(x)q(x)} = \frac{\alpha(x)}{q(x)} + \frac{\beta(x)}{p(x)}\,.$$

Ist ein drittes Polynom $r(x)$ teilerfremd sowohl zu $p(x)$ als auch zu $q(x)$, sind insbesondere auch $p(x)q(x)$ und $r(x)$ teilerfremd; es gibt also eine Darstellung

$$\frac{1}{p(x)q(x)r(x)} = \frac{\phi(x)}{p(x)q(x)} + \frac{\psi(x)}{r(x)} = \frac{\phi(x)\alpha(x)}{q(x)} + \frac{\phi(x)\beta(x)}{p(x)} + \frac{\psi(x)}{r(x)}$$

mit zwei neuen Polynomen $\phi(x)$ und $\psi(x)$, insgesamt also wieder eine Summe dreier Brüche mit den drei Faktoren $p(x), q(x)$ und $r(x)$ als Nennern.

Induktiv folgt auf diese Weise, daß es zu n paarweise teilerfremden Polynomen $p_1(x), \dots, p_n(x)$ stets Polynome $\alpha_1(x), \dots, \alpha_n(x)$ gibt, so daß

$$\frac{1}{p_1(x)\cdots p_n(x)} = \frac{\alpha_1(x)}{p_1(x)} + \cdots + \frac{\alpha_n(x)}{p_n(x)}$$

ist. Für ein beliebiges weiteres Polynom ist daher auch

$$\frac{f(x)}{p_1(x)\cdots p_n(x)} = \frac{\alpha_1(x)f(x)}{p_1(x)} + \cdots + \frac{\alpha_n(x)f(x)}{p_n(x)}\,.$$

Dies wenden wir an auf die obige Zerlegung des Nenners g: Da die $\ell_i(x)$ und die $q_j(x)$ paarweise teilerfremd sind, gilt dasselbe auch für ihre Potenzprodukte, es gibt also Polynome $\varphi_i(x), \psi_j(x)$, so daß gilt

$$\frac{f(x)}{g(x)} = \sum_{i=1}^{r} \frac{\varphi_i(x)}{\ell_i(x)^{e_i}} + \sum_{j=1}^{s} \frac{\psi_j(x)}{q_j(x)^{f_j}}\,.$$

Um den Grad dieser Polynome zu reduzieren, dividieren wir für jeden Bruch den Zähler mit Rest durch den Nenner und fassen die Quotientenpolynome zusammen zu einem einzigen Polynom $Q(x)$. Damit erhalten wir eine neue Darstellung

$$\frac{f(x)}{g(x)} = Q(x) + \sum_{i=1}^{r} \frac{\widetilde{\varphi}_i(x)}{\ell_i(x)^{e_i}} + \sum_{j=1}^{s} \frac{\widetilde{\psi}_j(x)}{q_j(x)^{f_j}}\,,$$

in der jeder Zähler $\widetilde{\varphi}_i(x)$ und $\widetilde{\psi}_j(x)$ kleineren Grad hat als der zugehörige Nenner.

Zur weiteren Zerlegung der Zähler überlegen wir uns zunächst, daß es zu je zwei Polynomen $\varphi(x)$ und $g(x)$ eine Darstellung

$$\varphi(x) = \alpha_0(x) + \alpha_1(x)g(x) + \alpha_2 g(x)^2 + \cdots$$

gibt mit $\deg \alpha_i(x) < \deg g(x)$. (Für ein lineares Polynom $g(x) = x - a$ ist das gerade die TAYLOR-Entwicklung von φ im Punkt a.) Für $\alpha_0(x)$ nehmen wir dazu den Divisionsrest von $\varphi(x)$ durch $g(x)$; ist $q_0(x)$ der Quotient, gilt dann

$$g(x) = \alpha_0(x) + g(x)q_0(x)\,.$$

Ist nun q_1 der Quotient und α_1 der Rest bei der Division von q_0 durch g, so erhalten wir die neue Darstellung

$$g(x) = \alpha_0(x)+g(x)\big(\alpha_1(x)+g(x)q_1(x)\big) = \alpha_0(x)+\alpha_1(x)g(x)+q_1(x)g(x)^2\,.$$

$\alpha_2(x)$ ist nun entsprechend der Divisionsrest von $q_1(x)$ durch $g(x)$ und so weiter, bis ein Quotient $q_\nu(x)$ kleineren Grad als $g(x)$ hat und damit der letzte Koeffizient $\alpha_{\nu+1}(x)$ ist.

Ist $g(x)$ ein lineares Polynom, haben alle $\alpha_\nu(x)$ kleineren Grad als eins, sind also Konstanten. Da $\deg \widetilde{\varphi}(x) < \deg \ell_i^{e_i} = e_i$ ist, können wir also schreiben

$$\widetilde{\varphi}(x) = \alpha_0 + \alpha_1\ell_i(x) + \cdots + \alpha_{e_i-1}\ell_i(x)^{e_i-1}$$

mit reellen Zahlen α_ν, und

$$\begin{aligned}\frac{\widetilde{\varphi}(x)}{\ell_i(x)^{e_i}} &= \frac{\alpha_0 + \alpha_1\ell_i(x) + \cdots + \alpha_{e_i-1}\ell_i(x)^{e_i-1}}{\ell_i(x)^{e_i-1}} \\ &= \frac{\alpha_0}{\ell_i(x)^{e_i}} + \frac{\alpha_1}{\ell_i(x)^{e_i-1}} + \cdots + \frac{\alpha_{e_i-1}}{\ell_i(x)}\,.\end{aligned}$$

Entsprechend erhalten wir die Darstellungen

$$\frac{\widetilde{\psi}(x)}{q_i(x)^{f_i}} = \frac{\beta_0 x + \gamma_0}{q_i(x)^{f_i}} + \frac{\beta_1 x}{q_i(x)^{f_i-1}} + \cdots + \frac{\beta_{f_i-1}x + \gamma_{f_i-1}}{q_i(x)}\,.$$

Faßt man alles zusammen und gibt den Zählern die richtigen Indizes, erhält man die oben angegebene Zerlegung, die sogenannte *Partialbruchzerlegung.*

Der gerade durchgeführte Beweis ist zwar (bei einer gegebenen Zerlegung des Nenners) konstruktiv, das verwendete Verfahren ist allerdings nicht besonders effizient. Nachdem wir wissen, daß eine Zerlegung wie oben existiert und auch alle Nenner kennen, ist es beispielsweise erheblich einfacher, von der obigen Zerlegung auszugehen (mit unbestimmten Zählern) und die Partialbrüche nach den üblichen Regeln der Bruchrechnung zu addieren. Der Nenner des entstehenden Bruchs ist $g(x)$, der Zähler $f(x) - Q(x)g(x)$, wobei $Q(x)$ der Quotient bei der Division von $f(x)$ durch $g(x)$ ist. Der durch Addition der Partialbrüche berechnete Zähler ist eine lineare Funktion in den Koeffizienten α, β, γ; Koeffizientenvergleich mit $f(x) - Q(x)q(x)$ liefert ein lineares Gleichungssystem für die Koeffizienten, das wir schnell und einfach nach GAUSS lösen können.

Für Linearfaktoren $\ell_i(x)$, die nur in der ersten Potenz vorkommen, gibt es sogar ein noch einfacheres Verfahren: Ist x_i die (einzige) Nullstelle von $\ell_i(x)$, so überlegt man sich leicht (durch Addition der Partialbrüche und des polynomialen Anteils), daß

$$\alpha_i = \lim_{x \to x_i} \frac{(x - x_i) f(x)}{g(x)}$$

ist. Da Zähler und Nenner für $x = x_i$ verschwinden, muß dieser Grenzwert nach DE L'HOSPITAL berechnet werden; wir erhalten

$$\alpha_i = \lim_{x \to x_i} \frac{(x - x_i) f'(x) - f(x)}{g'(x)} = -\frac{f(x_i)}{g'(x_i)} ,$$

wobei hier der Nenner nicht verschwinden kann, da x_i nur eine einfache Nullstelle von g ist.

3. Schritt: Integration der Partialbrüche: Die Integrale der Funktionen $\frac{\alpha_{ik}}{\ell_i(x)^k}$ lassen sich über die Substitution $u = \ell_i(x)$ sofort auf $\int \frac{du}{u^k}$ zurückführen und sind somit problemlos. Weiter ist

$$\int \frac{\beta x + \gamma}{ax^2 + bx + c}\, dx = \frac{\beta}{2a} \int \frac{2ax + b}{ax^2 + bx + c}\, dx + \int \frac{\gamma - \frac{\beta b}{2a}}{ax^2 + bx + c}\, dx ,$$

wobei das erste Integral nach der Regel über die logarithmische Ableitung aus Abschnitt *a)* berechnet werden kann, während wir das zweite in Abschnitt *b)* behandelt haben.

Bleiben schließlich noch die Integrale mit Potenzen eines quadratischen Polynoms im Nenner; hier wird k durch partielle Integration rekursiv erniedrigt.

Als ganz einfaches Beispiel, in dem alles explizit durchführbar ist, betrachten wir

$$\int \frac{dx}{x^2 - 5x + 6} .$$

Da hier sechs das Produkt und fünf die Summe der Nullstellen des Nenners ist, errät man leicht, daß $x^2 - 5x + 6 = (x - 2)(x - 3)$ ist. Die Partialbruchzerlegung bestimmen wir nun durch probieren: Da

$$\frac{1}{x-2} - \frac{1}{x-3} = \frac{(x-3) - (x-2)}{(x-2)(x-3)} = \frac{-1}{(x-2)(x-3)}$$

ist, folgt

$$\int \frac{dx}{(x-2)(x-3)} = -\int \frac{dx}{x-2} + \int \frac{dx}{x-3} = \ln \left| \frac{x-3}{x-2} \right| + C .$$

j) Symmetrie

Bei bestimmten Integralen kann man sich gelegentlich einen Großteil der Mühe des Ausrechnens ersparen, indem man Symmetrien berücksichtigt: Ist etwa die Funktion f spiegelsymmetrisch zur Achse $x = c$, d.h. $f(c + x) = f(c - x)$ für alle x, für die beide Seiten definiert sind, so ist auch

$$\int_c^{c+a} f(x)\, dx = \int_{c-a}^{c} f(x)\, dx ,$$

d.h.

$$\int_{c-a}^{c+a} f(x)\, dx = 2 \int_c^{c+a} f(x)\, dx .$$

Viel interessanter wird es, wenn f *punktsymmetrisch* zu einem Punkt $(c, 0)$ auf der x-Achse ist, d.h. $f(c+x) = -f(c-x)$. Dann ist

$$\int_c^{c+a} f(x)\,dx = -\int_{c-a}^{c} f(x)\,dx \quad \text{und} \quad \int_{c-a}^{c+a} f(x)\,dx = 0\,,$$

ohne daß man irgendetwas über die Stammfunktion von f wissen müßte. So ist beispielsweise

$$\int_{-1}^{1} \frac{\sin(x)}{|x|}\,dx = 0\,,$$

aber ein Leser, der mit den uns bislang bekannten Methoden nach einer Stammfunktion suchen möchte, dürfte eine sehr harte Zeit und wenig Erfolg haben.

k) Einige nicht elementar integrierbare Funktionen

Tatsächlich ist die Stammfunktion von $\dfrac{\sin x}{x}$ nicht durch elementare transzendenten Funktionen wie die Exponentialfunktion, die trigonometrischen Funktionen sowie deren Umkehrfunktionen und/oder Wurzeln und Grundrechenarten darstellbar, und sie teilt diese Eigenschaft mit einer ganzen Reihe weiterer Stammfunktionen.

Da einige dieser Stammfunktionen trotzdem für Anwendungen wichtig sind, behilft man sich damit, daß man einigen speziellen dieser Funktionen Namen gibt und versucht, alles darauf zurückzuführen. Die so definierten Funktionen können dann, genau wie der Sinus und die Exponentialfunktion, in Tabellenwerke und Unterprogrammbibliotheken aufgenommen werden, so daß man mit den dadurch abgedeckten Integralen genauso rechnen kann wie mit den elementaren transzendenten Funktionen.

Einige wichtige dieser neuen Funktionen sind

1) Der Integralsinus: Da $\lim\limits_{t\to 0} \dfrac{\sin t}{t} = 1$ ist, ist das bestimmte Integral

$$\mathrm{Si}(x) = \int_0^x \frac{\sin t}{t}\,dt$$

für alle $x \in \mathbb{R}$ wohldefiniert und eine stetige Funktion von x, der *Integralsinus.* Anwendungen finder er beispielsweise in der Elektortechnik, wo man für gewisse Filter Faltungsintegrale mit $\frac{\sin x}{x}$ benötigt.

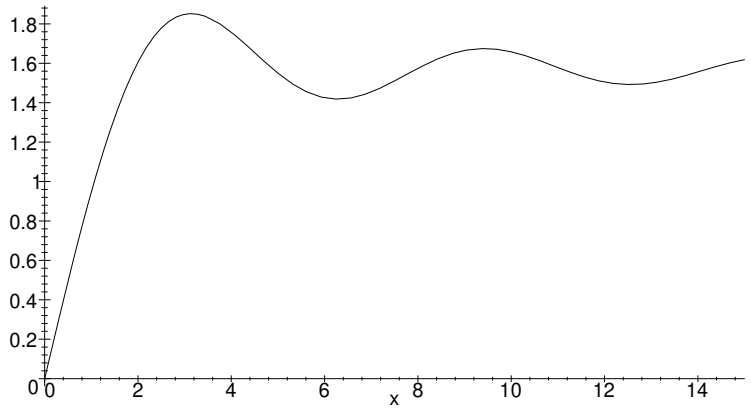


Abb. 64: Der Integralsinus

2) Die Fehlerfunktion: Auch die Funktion e^{-x^2} hat keine elementar angebbare Stammfunktion; hier definiert man als eine Stammfunktion die *Fehlerfunktion* oder *error function* durch

$$\mathrm{Erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2}\,dt\,.$$

(Der Vorfaktor sorgt dafür, daß $\lim_{x\to\infty} \mathrm{Erf}(x) = 1$ ist, was wir im Augenblick allerdings noch nicht beweisen können.)

Sie ist wichtig für die Statistik, denn für eine normalverteilte Zufallsvariable (dazu gehören die meisten Meßgrößen) mit Mittelwert $\overline{x}$ und

Standardabweichung σ ist die Wahrscheinlichkeit dafür, daß ein Meßwert zwischen a und b liegt, gleich

$$\frac{1}{2\pi}\int_a^b e^{\frac{1}{2}\left(\frac{x-\overline{x}}{\sigma}\right)^2}\,dx\,,$$

ein Integral, von dem man sich leicht überlegt, daß es mit Hilfe von Erf berechnet werden kann als

$$\frac{1}{2}\,\mathrm{Erf}\left(\frac{\sqrt{2}}{2}\,\frac{\overline{x}-a}{\sigma}\right)-\frac{1}{2}\,\mathrm{Erf}\left(\frac{\sqrt{2}}{2}\,\frac{\overline{x}-b}{\sigma}\right)\,.$$

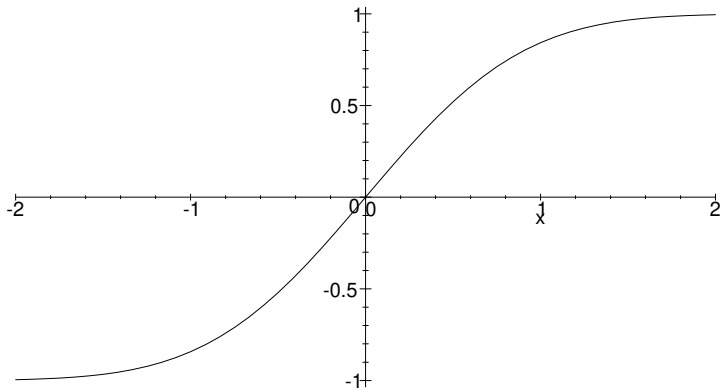


Abb. 65: Die Fehlerfunktion

3) Elliptische Integrale: Bei der Berechnung der Bogenlänge eines Ellipsensegments kommt man, je nach Art des Ansatzes und der Substitution, auf Integrale der Form

$$\int \frac{dx}{\sqrt{x^3+ax^2+bx+c}}\,,$$

wobei das Polynom im Nenner keine mehrfachen Nullstellen hat, oder auf Integrale der Form

$$\int \frac{dx}{\sqrt{1-k^2\sin^2 x}} \quad \text{oder} \quad \int \sqrt{1-k^2\sin^2 x}\,dx$$

mit $0<k<1$. Alle diese Integrale heißen *elliptische Integrale* und sind nicht elementar ausdrückbar. Eine hauptsächlich auf ADRIEN MARIE LEGENDRE (1752–1833) und KARL THEODOR WILHELM WEIERSTRASS (1815–1897) zurückgehende Theorie der elliptischen Funktionen stellt das Instrumentarium bereit, mit dem man diese vor allem in der Geodäsie, Kartographie und im Maschinenbau wichtigen Integrale berechnen kann.

4) Algebraische Integrale: Die erste Form der elliptischen Integrale ist ein Spezialfall eines sogenannen algebraischen Integrals; das sind Integrale, deren Integranden durch Wurzeln und Grundrechenarten (oder allgemeiner implizit durch Lösungen von Polynomgleichungen) gegeben sind. Die komplizierteren dieser Integrale sind fast alle nicht elementar ausdrückbar; es gibt inzwischen algorithmische Verfahren, die entscheiden, wann ein solches Integral elementar ausdrückbar ist, und die dann auch eine Stammfunktion finden können. Natürlich gibt es auch hier spezielle Funktionen, mit denen sich weitere dieser Integrale ausdrücken lassen.

l) Uneigentliche Integrale

Sei $a>0$ und $b>a$. Dann ist für eine reelle Zahl $r\neq 1$

$$\int_a^b \frac{dx}{x^r}=\frac{-1}{(r-1)x^{r-1}}\Bigg|_a^b=\frac{1}{r-1}\left(\frac{1}{a^{r-1}}-\frac{1}{b^{r-1}}\right)\,.$$

Falls $r>1$ ist, können wir hiervon den Grenzwert für b gegen unendlich betrachten und es liegt nahe, diesen als Wert des Integrals von a bis unendlich zu bezeichnen:

$$\int_a^\infty \frac{dx}{x^r}\underset{\text{def}}{=}\frac{1}{r-1}\,\frac{1}{a^{r-1}}\quad \text{für}\quad r>1\,.$$

Auch wenn wir nicht bis unendlich integrieren wollen, gibt es Beispiele von Integralen, denen wir via Grenzwertbetrachtung einen sinnvollen

Wert zuordnen können, ohne daß das Integral im Sinne unserer bisherigen Definitionen existieren würde: Beispielsweise ist für $a \le c < b$ und eine relle Zahl $0 < r < 1$

$$\int_a^c \frac{dx}{(b-x)^r} = \frac{(b-x)^{1-r}}{r-1}\Bigg|_a^c = \frac{(b-a)^{1-r}}{r-1} - \frac{(b-c)^{1-r}}{r-1}\,,$$

und auch hier liegt es nahe, den Grenzwert für $c \to b$ als Wert des Integrals von a bis b zu bezeichnen. Wir müssen hier allerdings vorsichtig sein mit dem Grenzübergang, denn die obige Formel gilt natürlich nur für $c < b$; für $c > b$ ist das Integral undefiniert. In einer solchen Situation können wir daher nicht den gewöhnlichen Limes betrachten, sondern brauchen einen eingeschränkten Grenzwertbegriff, der nur Folgen von einer der beiden Seiten berücksichtigt: Allgemein schreiben wir

$$y = \lim_{x \to b-} f(x)\,,$$

wenn es für jedes $\varepsilon > 0$ ein $\delta > 0$ gibt, so daß $|f(x) - y| < \varepsilon$ für alle x mit $b - \delta < x < b$. (Zur Erinnerung: Beim gewöhnlichen Grenzwert fordert man dies für alle x mit $|x - b| < \delta$, d.h. $b - \delta < x < b + \delta$.)

Völlig analog läßt sich natürlich auch ein rechtsseitiger Grenzwert

$$y = \lim_{x \to a+} f(x)$$

definieren durch die Bedingung, daß es für jedes $\varepsilon > 0$ ein $\delta > 0$ gibt, so daß $|f(x) - y| < \varepsilon$ für alle x mit $b < x < b + \delta$.

Schließlich sollten wir, wenn wir schon beim Definieren sind, zur Bezeichnungsökonomie noch vereinbaren, daß halboffene Intervalle wie

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

auch für unendliche b definiert sein sollen durch

$$[a, \infty) \underset{\text{def}}{=} \left\{ x \in \mathbb{R} \;\middle|\; a \le x \right\};$$

entsprechend auch

$$(-\infty, b] \underset{\text{def}}{=} \left\{ x \in \mathbb{R} \;\middle|\; x \le b \right\} \quad \text{und} \quad (-\infty, \infty) = \mathbb{R}\,.$$

Für unendliche Intervallgrenzen sind die Definitionen für links- und rechtsseitige Grenzwerte nicht sinnvoll anwendbar; wir vereinbaren daher, daß für $b = \infty$ der Ausdruck $\lim\limits_{c \to b-} \int_a^c f(x)\,dx$ für den gewöhnlichen Grenzwert $\lim\limits_{c \to \infty}$ stehen soll; entsprechend bei $\lim\limits_{x \to a+} f(x)$ für $a = -\infty$.

Mit all diesen Definitionen können wir dann eine Funktion f betrachten, die in einem halboffenen Intervall $[a, b)$ mit $b \in \mathbb{R} \cup \{\infty\}$ definiert und stückweise stetig ist; für diese definieren wir das rechtsseitig uneigentliche Integral

$$\int_a^b f(x)\,dx = \lim_{c \to b-} \int_a^c f(x)\,dx\,,$$

falls dieser Grenzwert existiert; andernfalls sagen wir, das Integral sei *divergent.*

Völlig analog definieren wir linksseitige uneigentliche Integrale: f sei stückweise stetig auf dem halboffenen Intervall $(a, b]$ mit $a \in \mathbb{R} \cup \{-\infty\}$; dann ist

$$\int_a^b f(x)\,dx = \lim_{c \to a+} \int_c^b f(x)\,dx\,,$$

falls dieser Grenzwert existiert; andernfalls sagen wir, das Integral sei *divergent.*

Diese Definitionen sind immer noch nicht allgemein genug: Eine Funktion könnte auch an *beiden* Enden eines Intervalls (a, b) undefiniert sein, wobei wir auch die Sonderfälle $a = -\infty$ und/oder $b = \infty$ zulassen wollen, und zusätzlich könnte sie auch noch Undefiniertheitsstellen $c_1 < \cdots < c_r$ im Intervallinnern haben.

In diesem Fall läßt sich das Intervall so in Teilintervalle zerlegen, daß f in jedem Teilintervall höchstens an *einem* der beiden Intervallenden uneigentlich ist: Falls es keine Undefiniertheitsstellen im Intervallinnern gibt, wählen wir willkürlich ein c_0 zwischen a und b und betrachten die beiden Intervalle $(a, c]$ und $[c, b)$. Im anderen Fall können zwei Zusatzpunkte notwendig sein: ein Punkt c_0 zwischen a und c_1 sowie ein Punkt c_{r+1} zwischen c_r und b.

Wir sagen dann, das uneigentliche Integral $\int_a^b f(x)\,dx$ konvergiere, wenn *jedes* der Integrale

$$\int_a^{c_0} f(x)\,dx, \quad \int_{c_0}^{c_1} f(x)\,dx, \quad \ldots, \quad \int_{c_{r+1}}^{b} f(x)\,dx$$

konvergiert; die Summe ihrer Werte bezeichnen wir als den Wert des Integrals von a bis b.

Als Beispiel betrachten wir das an beiden Grenzen uneigentliche Integral

$$\int_{-\infty}^{\infty} \frac{dx}{1+x^2}\,.$$

Da der Integrand eine gerade Funktion ist, empfiehlt es sich, das Integrationsintervall, das hier aus ganz $\mathbb{R}$ besteht, bei Null zu unterteilen, und wir erhalten

$$\int_{-\infty}^{\infty} \frac{dx}{1+x^2} = \int_{-\infty}^{0} \frac{dx}{1+x^2} + \int_{0}^{\infty} \frac{dx}{1+x^2} = 2\int_{0}^{\infty} \frac{dx}{1+x^2}$$

$$= 2\lim_{c\to\infty} \int_0^c \frac{dx}{1+x^2} = 2(\lim_{c\to\infty} \arctan c - \arctan 0) = \pi\,.$$

Obige Forderung, daß *jedes* der Teilintegrale einzeln konvergieren soll, ist gelegentlich etwas sehr restriktiv. Für das bei Null uneigentliche Integral

$$\int_{-2}^{4} \frac{dx}{x^3}$$

könnte man etwa argumentieren, daß der Integrand eine ungerade Funktion ist, so daß aus Symmetriegründen das Integral von -2 bis 2 verschwinden sollte und

$$\int_{-2}^{4} \frac{dx}{x^3} = \int_{-2}^{2} \frac{dx}{x^3} + \int_{2}^{4} \frac{dx}{x^3} = 0 + \frac{3}{32} = \frac{3}{32}$$

sein sollte. Diese Art der Argumentation ist durch das, was wir bislang gelernt haben, nicht gedeckt, und es gibt auch gute Gründe, sie zu verbieten: Schließlich geht die Stammfunktion $-\frac{1}{2}x^{-2}$ für $x \to 0$ gegen minus unendlich, und wenn eine Größe mit Relevanz in der realen Welt unendlich groß wird, hat dies im Allgemeinen zu gravierende Konsequenzen, als daß man einfach durch diesen Punkt hindurch weitergehen könnte.

Andererseits sind aber viele der mathematischen Formeln, die in den Naturwissenschaften und der Technik angewandt werden, nur näherungsweise gültig: Mathematische Modelle sind praktisch immer *vereinfachende* Modelle der Wirklichkeit, beispielsweise gilt das OHMsche Gesetz sicher nicht mehr, wenn man einen 5 Ω-Widerstand aus einer auf 5 V Spannung ausgelegten Schaltung im Hochspannungslabor mit 100 kV belastet, und es gilt auch nicht mehr ohne Korrekturterme, wenn man einen Wechselspannung mit 500 MHz anlegt.

Entsprechend gibt es durchaus Situationen, in denen das mathematische Modell einen unendlich großen Wert vorhersagt, wohingegen in der Realität limitierende Faktoren, die für Werte im „üblichen“ Größenbereich noch keine nennenswerte Rolle spielen, für eine Begrenzung sorgen. Falls man in einer solchen Situation sicher sein kann, daß auch in der realen Situation noch die Symmetrie zum Nullpunkt erhalten bleibt, kann man so wie oben argumentieren; falls allerdings die Symmetrie *nicht* erhalten bleibt, können durch die Begrenzung der Funktion beliebig große Abweichungen erzeugt werden, über die man mit dem vereinfachten mathematischen Modell nichts aussagen kann.

Da somit alles von der Anwendung abhängt, kann die Mathematik hier nicht mehr bieten als eine *Definition:* Falls für die Funktion f, die auf $[a,c) \cup (c,b]$ definiert ist, der Grenzwert

$$\lim_{h\to 0} \int_a^{c-h} f(x)\,dx + \int_{c-h}^{b} f(x)\,dx$$

existiert, bezeichnen wir ihn als CAUCHYschen Hauptwert von $\int_a^b f(x)\,dx$ nach dem aus der Analysis I bekannten französischen Mathematiker

Baron AUGUSTIN LOUIS CAUCHY (1789–1857). Entsprechend reden wir auch in komplizierteren Situationen mit mehreren Unstetigkeitsstellen vom CAUCHYschen Hauptwert, falls sich eine Aufteilung in Teilintervalle finden läßt, so daß für jedes Teilintervall der CAUCHYsche Hauptwert existiert. Im obigen Beispiel wäre also $\frac{3}{32}$ der CAUCHYsche Hauptwert des Integrals, wohingegen das Integral selbst undefiniert ist.

Die Frage, wann der CAUCHYsche Hauptwert für ein eigentlich divergentes Integral verwendet werden sollte, ist keine mathematische Frage: Unter rein mathematischen Gesichtspunkten gibt es **nie** eine Rechtfertigung für die Verwendung des CAUCHYschen Hauptwerts. Der CAUCHYsche Hauptwert ist nur dann sinnvoll anwendbar, wenn man davon ausgeht, daß ein mathematisches Modell eine Situation nur für nicht zu große Funktionswerte (ungefähr) korrekt beschreibt, und wenn man gleichzeitig sicher ist, daß die Unendlichkeitsstelle des mathematischen Modells für die Anwendung unproblematisch ist und gleichzeitig die Symmetrie, die der Berechnung des CAUCHYschen Hauptwerts zugrundeliegt, auch in der realen Anwendung gilt.

Der CAUCHYsche Hauptwert darf auch *nie* als eine Rechtfertigung dafür verstanden werden, daß man unbesonnen

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

setzt, wobei F eine zwar in den Punkten a und b, nicht aber auch für jeden Zwischenwert $a \le x \le b$ definierte Stammfunktion von f ist: So etwas kann zu Ergebnissen wie

$$\int_{-2}^{2} \frac{dx}{x^2} = -\frac{1}{x}\Big|_{-2}^{2} = \frac{-1}{2} - \frac{-1}{-2} = -1$$

führen, und natürlich ist keine Anwendung denkbar, in der eine negative Zahl in sinnvoller Weise als Integral über eine überall positive Funktion angesehen werden kann. In der Tat existiert im obigen Beispiel weder das Integral noch dessen CAUCHYscher Hauptwert, da sich die Unendlichkeiten links und rechts der Null hier nicht wegheben, sondern verstärken.

Zu einer etwas systematischeren Untersuchung uneigentlicher Integrale empfiehlt es sich, zunächst die Potenzen zu betrachten. Für positive Exponenten r ist x^r überall definiert, so daß Integrale über einen *endlichen* Bereich unproblematisch sind; für Integrale über einen *unendlichen* Bereich rechnet man leicht nach, daß sie immer divergieren. Für $r = 0$ ändert sich nichts an dieser Situation; interessant ist also nur der Fall $r < 0$, wo sowohl der Wert $x = 0$ als auch unendliche obere und/oder untere Grenzen zu Problemen führen können. Für negatives r ist $x^r = 1/x^{-r}$, wir interessieren uns also für

$$\int_a^b \frac{dx}{x^r} = \frac{1}{(1-r)x^{r-1}}\Big|_a^b \quad \text{für} \quad r > 0, \quad r \neq 1\,.$$

Für $a > 0$ und $b \to \infty$ existiert ein Grenzwert genau dann, wenn $r - 1 > 0$, also $r > 1$ ist; alsdann ist

$$\int_a^\infty \frac{dx}{x^r} = \lim_{b\to\infty} \int_a^b \frac{dx}{x^r} = \frac{1}{(r-1)a^{r-1}}\,.$$

Für $b > 0$ und $a \to \infty$ existiert ein Grenzwert genau dann, wenn $r - 1 < 0$, also $r < 1$ ist; alsdann ist

$$\int_0^b \frac{dx}{x^r} = \lim_{a\to 0} \int_a^b \frac{dx}{x^r} = \frac{1}{1-r} b^{1-r}\,.$$

Zusammen mit der Monotonieeigenschaft des RIEMANN-Integrals aus §3*a)* ergeben sich hieraus zwei allgemeine Kriterien für die Konvergenz uneigentlicher Integrale:

Satz: Die Funktion f sei stetig für $x \ge a$ und g sei stetig für $0 < x \le b$. Dann gilt:

1.) Falls es eine reelle Zahl K und eine reelle Zahl $r > 1$ gibt, so daß $|f(x)| \le \frac{K}{x^r}$ ist, konvergiert $\int_a^\infty f(x)\,dx$.
2.) Falls es eine reelle Zahl K und eine reelle Zahl $0 < r < 1$ gibt, so daß $|g(x)| \le \frac{K}{x^r}$ ist, konvergiert $\int_0^b g(x)\,dx$.

Beweis: 1.) Nach der Monotonieregel ist für jedes $b \geq a$

$$\int_a^b |f(x)|\, dx \leq K \int_a^b \frac{dx}{x^r}\,,$$

und letzteres Integral konvergiert, wie wir gerade nachgerechnet haben, unter den angenommenen Voraussetzungen. Das linksstehende Integral ist somit beschränkt durch eine von b unabhängige Konstante; da es wegen der Nichtnegativität des Betrags zusätzlich eine monoton wachsende Funktion von b ist, existiert daher der Grenzwert nach dem bekannten, schon für die Existenz des RIEMANN-Integrals verwendeten Satz aus der Analysis I, wonach jede monotone und beschränkte Folge reeller Zahlen konvergent ist.

Genau dasselbe gilt für die ebenfalls nichtnegative Funktion $f(x)+|f(x)|$, die durch $\frac{2K}{x^r}$ beschränkt ist; also existieren die uneigentlichen Integrale

$$\int_a^\infty \big(f(x) + |f(x)|\big)\, dx \quad \text{und} \quad \int_a^\infty |f(x)|\, dx\,,$$

und damit auch das gesuchte Integral als ihre Differenz.

2.) geht völlig analog. ∎

Als Beispiel hierfür betrachten wir die Strahlung eines schwarzen Körpers: Nach dem RAYLEIGHschen Strahlungsgesetz, wonach alle mit der Geometrie des Körpers verträglichen Wellenzahlvektoren gleichwahrscheinlich sind, wäre die Energiedichte proportional zum Quadrat der Frequenz, die Gesamtenergie also proportional zu

$$\int_0^\infty \nu^2\, d\nu\,,$$

einem offensichtlich divergenten Integral: Das ist die sogenannte „UV-Katastrophe“, die dieses Modell zum Widerspruch führt.

Die Abhilfe besteht bekanntlich darin, daß man die Gleichverteilung der Frequenzen durch eine BOSE-EINSTEIN-Statistik ersetzt und als neue

Energiedichte nun nach dem PLANCKschen Strahlungsgesetz

$$\frac{8\pi h}{c^2}\,\frac{\nu^3}{e^{\frac{h\nu}{kT}} - 1}$$

erhält, wobei h das PLANCKsche Wirkungsquantum, k die BOLTZMANN-Konstante, c die Lichtgeschwindigkeit und T die absolute Temperatur bezeichnet. Hier ist die Gesamtenergie

$$\frac{8\pi h}{c^2} \int_0^\infty \frac{\nu^3}{e^{\frac{h\nu}{kT}} - 1}\, d\nu\,,$$

und die UV-Katastrophe wird genau dann vermieden, wenn dieses Integral konvergiert.

Mit der Substitution $x = \dfrac{h\nu}{kT}$ erhalten wir

$$\frac{8\pi h}{c^2} \int_0^\infty \frac{\nu^3}{e^{\frac{h\nu}{kT}} - 1}\, d\nu = \frac{8\pi h}{c^2} \left(\frac{kT}{h}\right)^4 \int_0^\infty \frac{x^3}{e^x - 1}\, dx\,.$$

Dieses Integral ist an beiden Grenzen uneigentlich; betrachten wir also eine positive Konstante a und getrennt die Integrale von 0 bis a und von a bis ∞.

Letzteres Integral existiert nach dem gerade bewiesenen Satz, wenn wir K, r finden können, so daß

$$\frac{x^3}{e^x - 1} \leq \frac{K}{x^r} \quad \text{für alle } x \geq a\,.$$

Diese Ungleichung ist äquivalent zu

$$\frac{1}{e^x - 1} \leq \frac{K}{x^{r+3}} \quad \text{oder} \quad e^x \geq 1 + \frac{x^{r+3}}{K}\,,$$

und dies gilt mit geeignetem a, K und (beispielsweise) $r = 2$, da die Exponentialfunktion schneller wächst als jedes Polynom.

Entsprechend gilt für $0 \leq x \leq 1$ die Ungleichung etwa mit $r = \frac{1}{2}$, da dort $e^x - 1$ stärker wächst als jede x-Potenz. Also konvergiert das Integral sowohl an seiner unteren als auch seiner oberen Grenze, und

dazwischen ist ohnehin alles unproblematisch, da der Integrand stetig ist. ■

Als (vorerst) letztes Beispiel eines uneigentlichen Integrals sei noch

$$\Gamma(x) \underset{\text{def}}{=} \int_0^\infty e^{-t} t^{x-1}\, dt \quad \text{für } x > 0$$

betrachtet. Dieses Integral ist natürlich immer uneigentlich an der oberen Grenze; für $x < 1$ zusätzlich auch noch an der unteren.

Diese untere Grenze ist völlig harmlos, denn für $0 < x < 1$ ist

$$e^{-t} t^{x-1} = \frac{e^{-t}}{t^{1-x}} \le \frac{1}{t^{1-x}} \quad \text{mit} \quad 0 < 1 - x < 1\,,$$

so daß obiger Satz sofort die Konvergenz zeigt.

Auch die obere Grenze ist unproblematisch, denn die dazu notwendige Abschätzung

$$e^{-t} t^{x-1} \le \frac{K}{t^r} \iff e^t \ge t^{r+x-1}$$

folgt wieder, da die Exponentialfunktion stärker wächst als jede Potenz.

Die somit für alle $x > 0$ definierte Funktion $x \mapsto \Gamma(x)$ heißt EULERsche *Gamma-Funktion.*

LEONHARD EULER (1707–1783) wurde in Basel geboren und ging auch dort zur Schule und, im Alter von 14 Jahren, zur Universität. Dort legte er zwei Jahre später die Magisterprüfung in Philosophie ab und begann mit dem Studium der Theologie; daneben hatte er sich seit Beginn seines Studium unter Anleitung von JOHANN BERNOULLI mit Mathematik beschäftigt. 1726 beendete er sein Studium in Basel und bekam eine Stelle an der Petersburger Akademie der Wissenschaften, die er 1727 antrat. Auf Einladung FRIEDRICHS DES GROSSEN wechselte er 1741 an die preußische Akademie der Wissenschaften; nachdem sich das Verhältnis zwischen den beiden dramatisch verschlechtert hatte, kehrte er 1766 nach St. Petersburg zurück. Im gleichen Jahr erblindete er vollständig; trotzdem schrieb er rund die Hälfte seiner zahlreichen Arbeiten (73 Bände) danach. Sie enthalten bedeutende Beiträge zu vielen Gebieten der Mathematik, Physik, Astronomie und Kartographie.

Die wichtigste Eigenschaft der Γ-Funktion folgt durch partielle Integration:

$$\Gamma(x) = \int_0^\infty e^{-t} t^{x-1}\, dt = e^{-t}\frac{t^x}{x}\Big|_0^\infty + \frac{1}{x}\int_0^\infty e^{-t} t^x\, dt = \frac{\Gamma(x+1)}{x}$$

oder

$$\Gamma(x+1) = x\Gamma(x)\,.$$

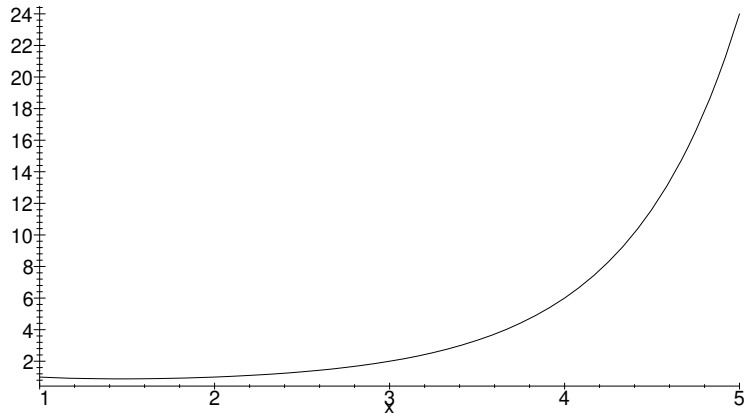


Abb. 66: Die Γ-Funktion

Aus dem elementar berechenbaren Wert

$$\Gamma(1) = \int_0^\infty e^{-t}\, dt = -e^{-t}\Big|_0^\infty = 1$$

ergibt sich somit, daß für alle natürliche Zahlen n gilt $\Gamma(n) = (n-1)!$; die Γ-Funktion ist also eine Art stetig gemachter Fakultätsfunktion. GAUSS definierte auf andere Weise eine stetige Funktion $\Pi(x)$, für die $\Pi(n) = n!$ ist, aber wie sich bald herausstellte, ist $\Pi(x) = \Gamma(x+1)$, so daß nur eine der beiden Funktionen wirklich gebraucht wird. Nach einigen Modewechseln im letzten Jahrhundert entscheidet man sich heute meist

für $\Gamma(x)$: Diese Funktionswerte sind in Tafelwerken tabelliert, und numerische Verfahren für ihre Berechnung stehen in den einschlägigen Unterprogrammbibliotheken und Computeralgebrasystemen zur Verfügung.

§4: Kurvenintegrale im $\mathbb{R}^n$

Bislang haben wir nur Funktionen einer Veränderlichen über ein Intervall integriert. Als ersten Schritt ins Mehrdimensionale wollen wir nun Funktionen mehrerer Veränderlicher betrachten und diese entlang einer Kurve im $\mathbb{R}^n$ integrieren. Entsprechende Integrale benötigt man einerseits, um Bogenlängen von Kurven zu berechnen, vor allem aber sind sie wichtig, um die aufzuwendende oder freiwerdende Energie bei der Bewegung eines Objekts in einem Kraftfeld (oder einem elektrisch geladenen Teilchens in einem elektrischen Feld *usw.*) zu berechnen. Wir beginnen mit der Definition von Kurven:

a) Kurven und Tangentenvektoren

Definition: *a)* Ein *Kurvenstück* γ ist eine stetig differenzierbare Abbildung

$$\gamma\colon [a,b] \to \mathbb{R}^n$$

eines Intervalls in den $\mathbb{R}^n$.
b) Eine *Kurve* γ ist eine endliche Folge von Kurvenstücken

$$\gamma_1, \ldots, \gamma_r\colon [a_i, b_i] \to \mathbb{R}^n$$

von Kurvenstücken mit der Eigenschaft, daß

$$\gamma_i(b_i) = \gamma_{i+1}(a_{i+1}) \quad \text{für } i = 1, \ldots, n-1.$$

c) Eine Kurve γ heißt geschlossen, wenn $\gamma_r(b_r) = \gamma_1(a_1)$ ist.

Die Bedingung im Teil *b)* der obigen Definition stellt sicher, daß Kurven, anschaulich betrachtet, zusammenhängend sind; es hat allein praktische Gründe, daß man nicht auch von den Intervallen verlangt, daß sie unmittelbar aneinander anschließen: Oft werden die Formeln für ein Kurvenstück einfacher, wenn man einen bestimmten Anfangswert wie etwa die Null für das Intervall nehmen kann.

Ein wesentlicher Unterschied zwischen einer Kurve und einem Kurvenstück liegt in der Differenzierbarkeit: Durch Verschiebung der Parameterintervalle könnte man jede Kurve durch eine stetige Abbildung $\gamma\colon [a,b] \to \mathbb{R}^n$ beschreiben, aber dort, wo zwei Kurvenstücke aneinanderstoßen, muß diese Abbildung nicht differenzierbar sein; anschaulich gesprochen kann die Kurve dort einen „Knick“ haben.

Die Ableitung $\dot\gamma(t)$ ist ein Vektor, dessen Komponenten die Ableitungen der Koordinatenfunktionen von $\gamma(t)$ sind; wir bezeichnen ihn als Tangentenvektor der Kurve im Punkt $\gamma(t)$. Gelegentlich wird es wichtig sein, daß dieser Vektor ungleich dem Nullvektor ist; wir definieren daher

Definition: *a)* Ein Kurvenstück $\gamma\colon [a,b] \to \mathbb{R}^n$ heißt *regulär,* wenn $\dot\gamma(t) \neq 0$ für alle $t \in (a,b)$.
b) Eine Kurve γ heißt *stückweise regulär,* wenn sie aus regulären Kurvenstücken zusammengesetzt werden kann.

Man beachte, daß für die Intervallendpunkte nichts gefordert wird: Selbst wenn die Ableitung dort existiert, muß sie nicht von Null verschieden kann. Auf diese Weise läßt sich ein Kurvenstück, bei dem $\dot\gamma(t)$ an endlich vielen Stellen gleich dem Nullvektor ist, immer noch als stückweise reguläre Kurve auffassen.

Anschaulich kann man sich ein Kurvenstück durch sein Bild im $\mathbb{R}^n$ vorstellen, allerdings muß man beachten, daß dasselbe Bild durch ganz verschiedene Funktionen parametrisiert werden kann. Als Beispiel für verschiedene Parametrisierungen einer und derselben Kurve betrachten wir den Einheitskreis $x^2 + y^2 = 1$. Seine bekannteste Darstellung als Kurvenstück ist die Parametrisierung

$$\gamma_1\colon [0, 2\pi] \to \mathbb{R}^2; \quad t \mapsto (\cos t, \sin t)\,,$$

aber natürlich wäre auch

$$\gamma_2\colon [0, 1] \to \mathbb{R}^2; \quad t \mapsto (\cos 2\pi t, \sin 2\pi t)$$

eine Möglichkeit. Es geht aber auch ganz anders, denn auch bei

$$\gamma_3\colon \mathbb{R} \to \mathbb{R}^2; \quad t \mapsto \left(\frac{t^2-1}{t^2+1}, \frac{2t}{t^2+1}\right)$$

liegen sämtliche Bildpunkte auf dem Einheitskreis, und zumindest für jedes endliche Teilintervall von $\mathbb{R}$ definiert auch γ_3 ein Kurvenstück. Wie eine kurze Kurvendiskussion (oder ein Blick auf Abbildung 67) zeigt, besteht das Bild von γ_3 aus allen Punkten des Einheitskreises außer $(1,0)$; letzterer Punkt ist der Grenzwert von $\gamma_3(t)$ für $t \to \pm\infty$.

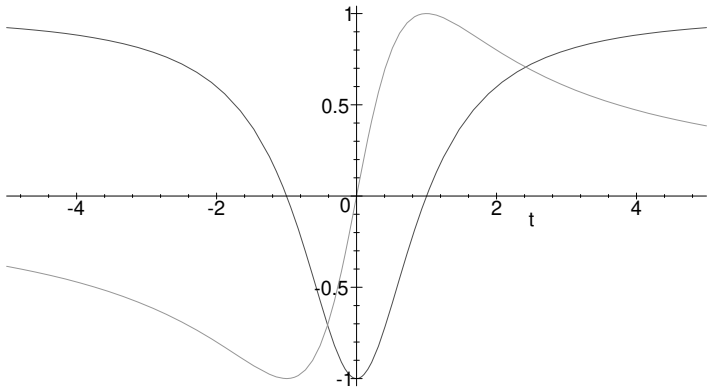


Abb. 67: Graph der Funktionen $x(t) = \frac{t^2-1}{t^2+1}$ und $y(t) = \frac{2t}{t^2+1}$

Jede dieser Parametrisierungen führt zu anderen Tangentenvektoren: Für einen festen Kurvenpunkt liegen zwar alle drei Vektoren auf ein und derselben Geraden, der Tangenten an den Kreis, aber sie haben verschiedene Länge: Für $(x,y) = \gamma_1(t)$ ist

$$\dot\gamma_1(t) = \begin{pmatrix} -\sin t \\ \cos t \end{pmatrix} = \begin{pmatrix} -y \\ x \end{pmatrix},$$

für $(x,y) = \gamma_2(t)$ ist

$$\dot\gamma_2(t) = \begin{pmatrix} -2\pi \sin 2\pi t \\ 2\pi \cos 2\pi t \end{pmatrix} = \begin{pmatrix} -2\pi y \\ 2\pi x \end{pmatrix},$$

und für $(x,y) = \gamma_3(t)$ schließlich zeigt eine kurze Rechnung, daß

$$\dot\gamma_3(t) = \frac{2}{(t^2+1)^2} \cdot \begin{pmatrix} 2t \\ 1-t^2 \end{pmatrix} = \frac{2}{t^2+1} \begin{pmatrix} y \\ -x \end{pmatrix}$$

ist. Bei der Parametrisierung mit γ_2 sind die Tangentenvektoren also jeweils 2π mal so lang wie bei der mit γ_1, was man anschaulich so interpretieren kann, daß der Kreis bei dieser Parametrisierung 2π mal so schnell durchlaufen wird wie bei der mit γ_1. Bei der Parametrisierung mit γ_3 ist die Länge der Vektoren variabel, und sie zeigen auch in die Gegenrichtung zu den Tangentenvektoren an γ_1 und γ_2; bei dieser Parametrisierung wird der Kreis also im Uhrzeigersinn durchlaufen.

b) Die Bogenlänge einer Kurve

In der Integralrechnung wird die Fläche unterhalb einer Kurve dadurch definiert, daß man sie durch Rechtecke angenähert; falls deren Gesamtfläche bei immer feiner werdenden Unterteilungen gegen einen festen Grenzwert konvergiert, bezeichnetet man diesen als Fläche unterhalb der Kurve oder auch als RIEMANN-Integral über die die Kurve beschreibende Funktion.

Nichts spricht dagegen, bei der Definition der Bogenlänge genauso vorzugehen: Die primitiven Bausteine, mit denen wir die Kurve annähern, sind nun natürlich keine Rechtecke mehr, sondern Strecken; am einfachsten nehmen wir dazu Tangentenvektoren. Dazu brauchen wir allerdings Differenzierbarkeit, wir müssen uns also zunächst auf ein Kurven*stück* beschränken.

Sei also $\gamma\colon [a,b] \to \mathbb{R}^n$ ein Kurvenstück; wir unterteilen das Intervall, wie wir es vom RIEMANN-Integral her gewohnt sind, durch Zwischenpunkte

$$a = t_0 < t_1 < \cdots < t_{N-1} < t_N = b$$

und wählen in jedem Teilintervall (t_i, t_{i+1}) einen Zwischenpunkt τ_i. Der Tangentenvektor im Punkt τ_i ist nach Definition $\dot\gamma(\tau_i)$, allerdings ist das nicht unbedingt der Vektor, den wir wollen: Wir wollen schließlich die Kurve durch ihre Tangente annähern, und dazu müssen wir die Länge an das Intervall anpassen, über dem wir die Kurve approximieren wollen. Indem wir den Vektor $\dot\gamma(\tau_i)$ der Differentialquotienten durch

$$\frac{1}{t_{i+1}-t_i} \cdot \overrightarrow{\gamma(t_i)\gamma(t_{i+1})} = \frac{1}{t_{i+1}-t_i} \begin{pmatrix} \gamma_1(t_{i+1}) - \gamma_1(t_i) \\ \vdots \\ \gamma_n(t_{i+1}) - \gamma_n(t_i) \end{pmatrix}$$

approximieren sehen wir, daß der Vektor $\overrightarrow{\gamma(t_i)\gamma(t_{i+1})}$ ungefähr gleich

$$\dot\gamma(\tau_i)\cdot(t_{i+1}-t_i)$$

ist, die Bogenlänge der Kurve kann also angenähert werden durch

$$\sum_{i=0}^{N-1}\left|\dot\gamma(\tau_i)\right|\cdot(t_{i+1}-t_i)\,.$$

Für die Kreislinie mit ihrer Parametrisierung $t\mapsto(\cos t,\sin t)$ ist dies in den Abbildungen 68 und 69 dargestellt für eine äquidistante Unterteilung des Intervalls $[0,2\pi]$ in zwanzig *bzw.* fünfzig Teilintervalle; im letzteren Fall ist in der Tat kaum mehr ein Unterschied zu sehen zwischen der Kreislinie und den fünfzig Vektoren, durch die sie approximiert wird.

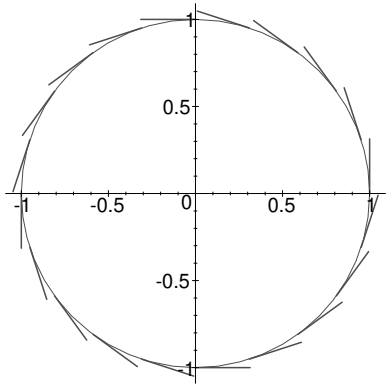


Abb. 68: Bogenlänge des Einheitskreises angenähert durch zwanzig Strecken

Da γ nach Definition eines Kurvenstücks stetig differenzierbar ist, ist auch $\left|\dot\gamma(t)\right|$ immerhin noch eine stetige und damit insbesondere RIEMANN-integrierbare Funktion; wir wissen also, daß der Grenzwert für immer enger werdende Verfeinerungen der Unterteilung existiert und gleich dem Integral über diese Funktion ist. Somit kommen wir auf ganz natürliche Weise auf die

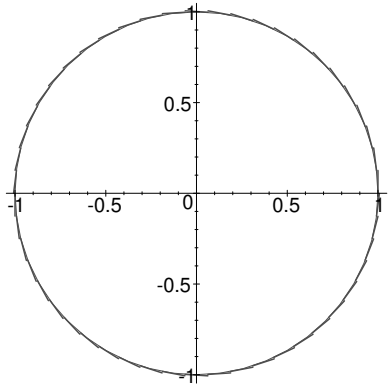


Abb. 69: Bogenlänge des Einheitskreises angenähert durch fünfzig Strecken

Definition: *a)* Die Bogenlänge eines Kurvenstücks $\gamma\colon[a,b]\to\mathbb{R}^n$ ist

$$\int_a^b\left|\dot\gamma(t)\right|\,dt\,.$$

b) Die Bogenlänge einer Kurve γ, bestehend aus den Kurvenstücken $\gamma_1,\dots,\gamma_r$ ist die Summe der Bogenlängen der Kurvenstücke γ_i.

Um zu sehen, ob das alles wirklich vernünftig ist, berechnen wir die Bogenlänge der Kreislinie

$$\gamma_1\colon[0,2\pi]\to\mathbb{R}^2;\quad t\mapsto(\cos t,\sin t)\,.$$

Hier ist

$$\dot\gamma_1(t)=\begin{pmatrix}-\sin t\\ \cos t\end{pmatrix},\quad\text{also}\quad\left|\dot\gamma_1(t)\right|=\sqrt{\sin^2 t+\cos^2 t}=1\,,$$

und damit ist die Bogenlänge

$$\int_0^{2\pi}\left|\dot\gamma_1(t)\right|\,dt=\int_0^{2\pi}dt=2\pi\,,$$

wie erwartet; zumindest für diese Parametrisierung ist also alles vernünftig.

Alternativ hatten wir die Kreislinie auch parametrisiert durch

$$\gamma_3\colon \mathbb{R} \to \mathbb{R}^2; \quad t \mapsto \left(\frac{t^2-1}{t^2+1}, \frac{2t}{t^2+1} \right) ;$$

hier ist die Bogenlänge im Sinne obiger Definition nicht erklärt, jedoch können wir natürlich γ_3 auf endliche Intervalle einschränken und diese immer größer werden lassen; falls ein Grenzwert existiert, bekommen wir also hier die Bogenlänge als uneigentliches Integral

$$\int_{-\infty}^{\infty} \left|\dot\gamma_3(t)\right| dt .$$

Berechnen wir zunächst die Ableitung von γ_3:

$$\frac{d}{dt}\left(\frac{t^2-1}{t^2+1} \right) = \frac{(t^2+1)\cdot 2t - (t^2-1)\cdot 2t}{(t^2+1)^2} = \frac{4t}{(t^2+1)^2}$$

und

$$\frac{d}{dt}\left(\frac{2t}{t^2+1} \right) = \frac{(t^2+1)\cdot 2 - 2t\cdot 2t}{(t^2+1)^2} = \frac{-2(t^2-1)}{(t^2+1)^2} ,$$

d.h.

$$\left|\dot\gamma_3(t)\right| = \frac{\sqrt{(4t)^2+4(t^2-1)^2}}{(t^2+1)^2} = \frac{2\sqrt{(t^2+1)^2}}{(t^2+1)^2} = \frac{2}{t^2+1} .$$

Wie Abbildung 70 zeigt, ist diese Annäherung der Kreislinie durch Strecken erheblich unregelmäßiger als die obige, aber beim Übergang zu immer feiner werdenden Unterteilungen gehen natürlich trotzdem alle Streckenlängen gegen null. (Eingezeichnet sind die Strecken zu einer Unterteilung des Intervalls $[-10, 10]$ in Teilintervalle der Länge $1/2$.)

Die Bogenlänge ist somit

$$\int_{-\infty}^{\infty} \frac{2\,dt}{t^2+1} = 2\arctan t \Big|_{-\infty}^{\infty} = 2\left(\frac{\pi}{2} - \frac{-\pi}{2} \right) = 2\pi ;$$

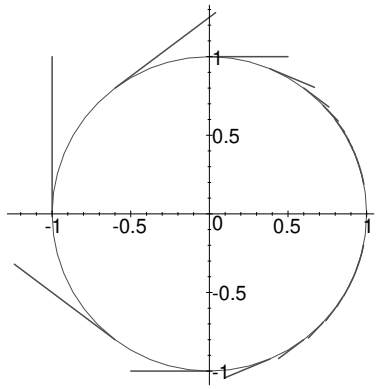


Abb. 70: Approximation durch Strecken in der rationalen Parametrisierung

auch bei dieser Parametrisierung erhalten wir also dasselbe Ergebnis, wie dies aus geometrischen Gründen auch sein muß: Die Bogenlänge ist schließlich eine Eigenschaft einer Kurve, nicht einer speziellen Parametrisierung der Kurve.

Ganz so einfach ist die Sache allerdings nun doch nicht: Schließlich parametrisiert auch

$$\gamma_4\colon [0, 20\pi] \to \mathbb{R}^2; \quad t \mapsto (\cos t, \sin t)$$

die Kreislinie, aber nun ist die Bogenlänge

$$\int_0^{20\pi} \left|\dot\gamma_4(t)\right| dt = \int_0^{20\pi} dt = 20\pi .$$

Auch das erscheint geometrisch durchaus sinnvoll: Wenn man nicht nur einmal, sondern zehnmal im Kreise herum geht, legt man schließlich einen zehnmal so langen Weg zurück.

Wir müssen also etwas sorgfältig sein, wenn wir präzisieren wollen, was die Unabhängigkeit der Bogenlänge von der Parametrisierung wirklich bedeuten soll; da das ganze dadurch auch etwas umfangreicher wird, sei das entsprechende Lemma in den nächsten Abschnitt verschoben, wo wir es gleich etwas allgemeiner beweisen werden.

c) Integration eines Vektorfelds längs einer Kurve

Die Hauptanwendung von Kurvenintegralen besteht nicht darin, die Längen aller möglicher Kurven zu berechnen; der Hauptgrund, warum wir solche Integrale betrachten, ist die Berechnung der aufzuwendenden oder freiwerdenden Energie bei der Bewegung eines Teilchens in einem Kraftfeld *bzw.* – im Fall eines elektrisch geladenen Teilchens – eines elektromagnetischen Felds.

Gehen wir der Einfachheit halber aus von einem Kraftfeld $\vec{F}(\boldsymbol{x})$ und einem Teilchen, das sich entlang eines Kurvenstücks $\gamma\colon [a,b] \to \mathbb{R}^n$ durch dieses Feld bewegt, d.h. der Vektor $\vec{F}\big(\gamma(t)\big)$ sei für alle t aus $[a,b]$ definiert. Im Punkt $\gamma(t)$ greift dann also der Kraftvektor $\vec{F}\big(\gamma(t)\big)$ an; die Arbeit, die das Teilchen verrichten muß oder gewinnt, ist das Skalarprodukt aus Kraftvektor und (Tangential-)Vektor des Wegs.

Um die Gesamtarbeit zunächst näherungsweise auszurechen, unterteilen wir wie gewohnt das Intervall $[a,b]$ durch Zwischenpunkte

$$a = t_0 < t_1 < \cdots < t_{N-1} < t_N = b$$

und wählen in jedem Teilintervall (t_i, t_{i+1}) einen Punkt τ_i. Zwischen den Punkten $\gamma(t_i)$ und $\gamma(t_{i+1})$ approximieren wir die Kurve wie im vorigen Abschnitt durch den Tangentenvektor $\dot\gamma(\tau_i)\cdot(t_{i+1}-t_i)$, die Arbeit kann dann angenähert werden durch

$$\sum_{i=0}^{N-1} \vec{F}\big(\gamma(\tau_i)\big)\cdot\dot\gamma(\tau_i)$$

Da γ nach Definition eines Kurvenstücks stetig differenzierbar ist, ist $\vec{F}\big(\gamma(t)\big)\cdot\dot\gamma(t)$ für ein stetiges Vektorfeld eine stetige und damit insbesondere RIEMANN-integrierbare Funktion; wir wissen also, daß der Grenzwert für immer enger werdende Verfeinerungen der Unterteilung existiert und gleich dem Integral über diese Funktion ist.

Definition: *a)* $\vec{V} \in \mathcal{C}^0(D,\mathbb{R}^n)$ sei ein stetiges Vektorfeld auf der offenen Teilmenge $D \subseteq \mathbb{R}^n$, und $\gamma\colon [a,b] \to D$ sei ein Kurvenstück. Das Integral

über $\vec{V}$ entlang γ ist

$$\int_\gamma \vec{V}(\boldsymbol{x})\,d\boldsymbol{x} \underset{\text{def}}{=} \int_a^b \vec{V}\big(\gamma(t)\big)\cdot\dot\gamma(t)\,dt\,.$$

b) Das Integral über $\vec{V}$ entlang einer Kurve γ ist die Summe der Integrale über $\vec{V}$ entlang der Kurvenstücke, aus denen γ zusammengesetzt ist.

Noch allgemeiner können wir ausgehen von einer beliebigen stetigen Funktion $f\colon C \to \mathbb{R}$ auf dem Bild $C = \gamma\big([a,b]\big)$ des Kurvenstücks $\gamma\colon [a,b] \to \mathbb{R}^n$ und diese entlang γ integrieren. Als Motivation könnte man sich etwa vorstellen, daß γ eine materielle Kurve ist und $f\big(\gamma(t)\big)$ die (lineare) Massendichte im Punkt $\gamma(t)$ angibt, oder γ könnte den Weg eines Teilchens durch eine Flüssigkeit mit räumlich variabler Viskosität beschreiben *usw.*

Wenn wir wieder von derselben Unterteilung wie oben ausgehen und das Kurvenstück durch eine Folge von Tangentenvektoren approximieren, müssen wir hier die Summen

$$\sum_{i=0}^{N-1} f\big(\gamma(\tau_i)\big)\,\big|\dot\gamma(\tau_i)\big|$$

betrachten, und diese konvergieren wegen der Stetigkeit von f gegen

$$\int_\gamma f(\boldsymbol{x})\,ds \underset{\text{def}}{=} \int_a^b f\big(\gamma(t)\big)\,\big|\dot\gamma(t)\big|\,dt\,.$$

Analog definieren wir auch das Integral über eine *Kurve* γ.

Definition: $\int_\gamma f(\boldsymbol{x})\,ds$ heißt RIEMANN-STIELTJES-Integral über f entlang γ.

(ds steht hier für das Bogenelement der Kurve, wobei der Buchstabe s wohl vom lateinischen *spatium* = Entfernung kommen dürfte.)

Der niederländische Mathematiker THOMAS JAN STIELTJES (1856–1894) studierte in Leiden, schwänzte dort aber viele Vorlesungen um stattdessen die Werke von GAUSS und JACOBI zu lesen. Obwohl er bei seinen Prüfungen dreimal durchfiel, wurde er Assistent am Observatorium von Leiden, dessen Direktor ein Freund seines Vaters war. 1883 heiratete er und wechselte unter anderem auf Betreiben seiner Frau von der Astronomie zur Mathematik. Nachdem eine Berufung nach Groningen an seinem fehlenden Hochschulabschluß gescheitert war, übersiedelte er 1885 nach Frankreich, wo er Professor an der Universität Toulouse wurde und bis an sein Lebensende blieb. Seine Arbeiten befassen sich vor allem mit Verbindungen zwischen der Zahlentheorie und der Analysis; das RIEMANN-STIELTJES-Integral wurde in einer Arbeit über Kettenbrüche eingeführt.

Dieses RIEMANN-STIELTJES-Integral wird im Spezialfall $f \equiv 1$ zur Bogenlänge, und für ein reguläres Kurvenstück γ ohne Selbstschnitte wird es für

$$f\big(\gamma(t)\big) = \frac{\vec{V}\big(\gamma(t)\big) \cdot \dot{\gamma}(t)}{\big|\dot{\gamma}(t)\big|}$$

zum Integral entlang γ über das Vektorfeld $\vec{V}$. (Falls sich die Kurve γ selbst schneidet, wie etwa eine 8, ist $\dot{\gamma}(t)$ im Schnittpunkt eventuell nicht nur vom Punkt $\gamma(t)$ abhängig, sondern vom Parameter t, so daß wir dann nicht in der Situation von RIEMANN-STIELTJES sind. Meist ist dies jedoch kein Problem, denn wenn so etwas nur endlich oft vorkommt, können wir die Kurve γ in endlich viele Kurvenstücke zerlegen, deren jedes die Voraussetzung erfüllt.)

Natürlich gelten auch für RIEMANN-STIELTJES-Integrale die üblichen, aus der klassischen Integralrechung bekannten Rechenregeln: Da wir Kurvenintegrale als spezielle RIEMANN-Integrale definiert haben, sind dies in der Tat einfach Spezialfälle der dortigen Regeln. Wir haben also, zum Beispiel, wieder die *Monotonieregel*

$$\int_\gamma f(\boldsymbol{x})\,ds \le \int_\gamma g(\boldsymbol{x})\,ds \quad \text{falls } f(\boldsymbol{x}) \le g(\boldsymbol{x}) \text{ für jeden Punkt auf } \gamma,$$

die *Linearität*seigenschaft

$$\int_\gamma (\alpha f + \beta g)\,ds = \alpha \int_\gamma f\,ds + \beta \int_\gamma g\,ds\,,$$

und für *Zusammensetzungen* gilt

$$\int_{\gamma+\delta} f\,ds = \int_\gamma f\,ds + \int_\delta f\,ds\,,$$

wobei $\gamma+\delta$ jene Kurve bezeichne, deren erste Kurvenstücke die Kurve γ definieren, während die restlichen die Bestandteile von δ sind.

Schließlich haben wir auch noch einen

Mittelwertsatz: Für jedes Kurvenstück $\gamma\colon [a,b] \to \mathbb{R}^n$ und jede stetige Funktion f auf $\gamma([a,b])$ gibt es einen Parameterwert $\tau \in [a,b]$, so daß gilt

$$\int_\gamma f\,ds = f(\tau) \cdot \textit{Bogenlänge von } \gamma\,.$$

Beweis: Dies ist einfach der gewöhnliche Mittelwertsatz der Integralrechnung, angewandt auf die Funktion

$$t \mapsto f\big(\gamma(t)\big) \cdot \dot{\gamma}(t)\,.$$

■

Es ist hier nicht wesentlich, daß γ als Kurven*stück* vorausgesetzt war; ein einfaches Argument mit dem Zwischenwertsatz zeigt, daß der Mittelwertsatz auch für Kurven gilt. Wesentlich ist dagegen die Stetigkeit von f, denn da der der Mittelwertsatz der Integralrechnung nur für stetige Funktionen gilt, können wir ihn nur anwenden, wenn $f\big(\gamma(t)\big)$ stetig ist. Ein einfaches Gegenbeispiel zum Mittelwertsatz wäre eine Kurve aus zwei gleichlangen Komponenten, auf deren einer f konstant gleich eins ist, während es auf der anderen verschwindet. Nach dem Mittelwertsatz, angewandt auf die beiden Komponenten, folgt, daß der Mittelwert über die gesamte Kurve gleich 1/2 ist, aber dieser Wert wird von f nirgends angenommen.

Als nächstes kommen wir zu der angekündigten Unabhängigkeit des Kurvenintegrals von der Parametrisierung der Kurve. Da diese, wie wir am Beispiel der mehrfach durchlaufenenen Kreislinie gesehen haben, nicht ganz uneingeschränkt gilt, müssen wir zunächst definieren, was wir meinen:

Definition: Zwei Kurvenstücke $\gamma\colon [a,b] \to \mathbb{R}^n$ und $\delta\colon [c,d] \to \mathbb{R}^n$ heißen äquivalent, wenn es eine bijektive stetig differenzierbare Abbildung $\varphi\colon [a,b] \to [c,d]$ gibt mit $\varphi(a) = c$ und $\varphi(b) = d$ derart, daß $\delta \circ \varphi = \gamma$ ist, d.h. für alle $x \in [a,b]$ ist $\delta\big(\varphi(x)\big) = \gamma(x)$.

Dese Definition stellt insbesondere sicher, daß $\gamma\big([a,b]\big)$ und $\delta\big([c,d]\big)$ dieselbe Teilmenge von $\mathbb{R}^n$ sind, so daß die beiden Kurven als Punktmengen übereinstimmen; durch die geforderte Bijektivität von φ ist aber auch sichergestellt, daß die Kurve bei beiden Parametrisierungen gleich oft durchlaufen wird. Daher erwarten wir

Lemma: $\gamma\colon [a,b] \to D \subseteq \mathbb{R}^n$ und $\delta\colon [c,d] \to D \subseteq \mathbb{R}^n$ seien äquivalente Kurvenstücke, und $f\colon D \to \mathbb{R}$ sei eine stetige Funktion. Dann ist

$$\int_\gamma f(\boldsymbol{x})\,ds = \int_\delta f(\boldsymbol{x})\,ds\,.$$

Der *Beweis* ist eine einfache Anwendung der Kettenregel und der Substitutionsregel: Nach Definition der Äquivalenz gibt es eine bijektive stetig differenzierbare Funktion φ mit $\varphi(a) = c$ und $\varphi(b) = d$, so daß $\gamma = \delta \circ \varphi$ und damit $\dot\gamma = (\dot\delta \circ \varphi) \cdot \dot\varphi$ ist. Offensichtlich muß φ monoton wachsen, also ist $\dot\varphi(t)$ nirgends negativ und somit gleich seinem Betrag. Nach der Substitutionsregel ist

$$\int_\gamma f(\boldsymbol{x})\,ds = \int_a^b f\big(\gamma(t)\big)\,|\dot\gamma(t)|\;dt = \int_a^b f\Big(\delta\big(\varphi(t)\big)\Big)\,\Big|\dot\delta\big(\varphi(t)\big)\Big|\,\dot\varphi(t)\,dt$$

$$= \int_c^d f\big(\delta(t)\big)\,\Big|\dot\delta(t)\Big|\;dt = \int_\delta f(\boldsymbol{x})\,ds\,.$$

■

d) Zirkulationsfreie und konservative Vektorfelder

Wir haben Kurvenintegrale über Vektorfelder eingeführt, um die Energie zu beschreiben, die zur Bewegung eines Teilchens durch ein Kraftfeld (oder ein elektromagnetisches Feld) aufgewendet werden muß oder dabei frei wird. Das gerade bewiesene Lemma zeigt, daß diese (Gesamt-)Energie nur vom Weg des Teilchens abhängt, nicht aber beispielsweise von seiner Geschwindigkeit. Wie aus der Physik bekannt ist, hängt aber beispielsweise bei reibungsfreier Bewegung eines Teilchens in einem mechanischen Kraftfeld der Weg nicht einmal von der Kurve ab, sondern nur von deren Anfangs- und Endpunkt oder, genauer gesagt, vom der potentiellen Energie des Anfangs- und des Endpunkts. Insbesondere verschwindet also das Integral längs einer jeden geschlossenen Kurve. In diesem Abschnitt wollen wir Vektorfelder mit dieser Eigenschaft genauer untersuchen.

Definition: *a)* Ein Vektorfeld $\vec V\colon D \to \mathbb{R}^n$ heißt *zirkulationsfrei,* wenn für jede geschlossene Kurve γ in D gilt:

$$\int_\gamma \vec V\,ds = 0\,.$$

b) Das Vektorfeld $\vec V$ heißt *konservativ,* wenn es eine differenzierbare Funktion $\varphi\colon D \to \mathbb{R}$ gibt, so daß $\vec V = \operatorname{grad}\varphi$ ist. φ heißt *Stammfunktion* von $\vec V$ und $-\varphi$ *Potentialfunktion.*

Zum Verständnis des Begriffs *konservativ* betrachten wir ein Beispiel aus der Physik: Ein Teilchen mit konstanter Masse m bewege sich durch das Gravitationsfeld

$$\vec F = -\operatorname{grad}\left(G\frac{Mm}{r}\right)$$

eines Himmelskörpers oder, allgemeiner, durch irgendein Potentialfeld $\vec F = -\operatorname{grad} U$. Nach dem zweiten NEWTONschen Gesetz gilt dann für die Bahn $\gamma\colon [a,b] \to \mathbb{R}^3$ des Teilchens mit der Zeit t als Parameter

$$\vec F\big(\gamma(t)\big) = m\ddot\gamma(t)\,,$$

d.h. die im Punkt $\gamma(t)$ wirkende Kraft ist gleich der Masse des Teilchens mal seiner Beschleunigung. Also ist in jedem Punkt $\gamma(t)$

$$m\ddot{\gamma}(t) + \operatorname{grad} U\big(\gamma(t)\big) = 0\,.$$

Um aus dieser Vektorgleichung eine skalare Beziehung abzuleiten, bilden wir das Skalarprodukt mit dem Geschwindigkeitsvektor $\dot{\gamma}(t)$:

$$\begin{aligned} 0 &= m\ddot{\gamma}(t)\cdot\dot{\gamma}(t) + \operatorname{grad} U\big(\gamma(t)\big)\cdot\dot{\gamma}(t) \\ &= \frac{d}{dt}\left(\frac{m}{2}\dot{\gamma}(t)\cdot\dot{\gamma}(t) + U\big(\gamma(t)\big)\right) \end{aligned}$$

nach Produkt- und Kettenregel. Also ist

$$E = \frac{m\big|\dot{\gamma}(t)\big|^2}{2} + U\big(\gamma(t)\big)$$

konstant, und das ist der klassische Energieerhaltungssatz: Die Summe aus kinetischer und potentieller Energie ist eine zeitlich unveränderliche Erhaltungsgröße.

Der Name „konservativ“ kommt vom lateinischen *conservare* = erhalten; der Grund für das negative Vorzeichen der Potentialfunktion liegt darin, daß der Gradient in Richtung des stärksten *Anstiegs* einer Funktion zeigt, wohingegen die Natur versucht, ein System zum Energie*minimum* zu bringen, so daß die Kräfte in Gegenrichtung zum Gradienten wirken.

Wie zu erwarten, sind die beiden Begriffe *konservativ* und *zirkulationsfrei* nicht unabhängig voneinander; tatsächlich werden wir gleich sehen, daß sie sogar äquivalent sind. Aus technischen Gründen wollen uns uns dabei, auch wenn es nicht unbedingt nötig wäre, auf sogenannte *zusammenhängende* Mengen beschränken:

Definition: Eine Teilmenge $D \subseteq \mathbb{R}^n$ heißt *zusammenhängend,* wenn es zu je zwei Punkten $\boldsymbol{x}, \boldsymbol{y} \in D$ eine Kurve γ gibt mit Anfangspunkt $\boldsymbol{x}$ und Endpunkt $\boldsymbol{y}$.

Diese Definition fordert zwar genau das, was wir gleich brauchen werden, sie ist aber nicht die übliche Definition einer zusammenhängenden Menge: In der Analysis wie auch in der Topologie bezeichnet man eine Menge D dann als zusammenhängend, wenn sie nicht als disjunkte Vereinigung zweier offener Teilmengen von D geschrieben werden kann; dies

ist im allgemeinen eine schwächere Bedingung als die hier geforderte. Für eine offene Teilmenge $D \subseteq \mathbb{R}^n$ sind die beiden Definitionen aber äquivalent:

Sei nämlich D_0 die Menge aller Punkte aus der offenen Menge D, die mit dem Punkt $\boldsymbol{x}_0 \in D$ durch einen Weg verbunden werden können. Dann ist D_0 offen, denn für $\boldsymbol{x}_1 \in D_0$ gibt es wegen der Offenheit von D eine ε-Umgebung von $\boldsymbol{x}_1$, die ganz in D liegt. Da jeder Punkt dieser ε-Umgebung durch eine Strecke mit dem Mittelpunkt $\boldsymbol{x}_1$ verbunden werden kann, liegt auch diese Umgebung in D_0, d.h. D_0 ist offen.

Aber auch das Komplement von D_0 in D ist offen, denn auch für einen Punkt $\boldsymbol{x}_2 \in D \setminus D_0$ enthält D eine ε-Umgebung. Läge diese nicht ganz in $D \setminus D_0$, gäbe es dort einen Punkt $\boldsymbol{x}_3$, der durch einen geeigneten Weg mit $\boldsymbol{x}_0$ verbunden werden könnte. Da aber $\boldsymbol{x}_3$ durch eine Strecke mit $\boldsymbol{x}_2$ verbunden werden kann, gäbe dann auch einen Weg von $\boldsymbol{x}_2$ nach $\boldsymbol{x}_0$, im Widerspruch zur Annahme. Also ist $D \setminus D_0$ offen.

Damit ist D also Vereinigung der offenen und disjunkten Teilmengen D_0 und $D \setminus D_0$; wenn D zusammenhängend ist, geht das nur, wenn eine der beiden Mengen leer ist. D_0 enthält den Punkt $\boldsymbol{x}_0$, also ist $D \setminus D_0 = \emptyset$ und somit $D = D_0$. Also kann jeder Punkt aus D durch einen Weg mit $\boldsymbol{x}_0$ verbunden werden, und damit ist D zusammenhängend im Sinne der obigen Definition.

Der folgdende Satz zeigt die Äquivalenz der Begriffe *konservativ* und *zirkulationsfrei;* wegen seiner zweiten Aussage kann er als das mehrdimensionale Analogon des Hauptsatzes der Differential- und Integralrechung aufgefaßt werden.

Satz: Für ein Vektorfeld $\vec{V}\colon D \to \mathbb{R}^n$ auf einer offenen zusammenhängenden Teilmenge $D \subseteq \mathbb{R}^n$ sind die folgenden Aussagen äquivalent:

1.) $\vec{V}$ ist konservativ.

2.) Es gibt eine Funktion $\varphi\colon D \to \mathbb{R}$, so daß für jede Kurve γ in D gilt:

$$\int_\gamma \vec{V}\cdot ds = \varphi\big(\gamma(b)\big) - \varphi\big(\gamma(a)\big)\,;$$

insbesondere hängt das Integral also nur von den Endpunkten von γ ab.

3.) $\vec{V}$ ist zirkulationsfrei.

Beweis: 1.) $\Rightarrow$ 2.): Falls $\vec{V}$ konservativ ist, gibt es eine Stammfunktion φ auf D, so daß $\vec{V} = \operatorname{grad}\varphi$ ist. Für eine Kurve $\gamma\colon [a,b] \to D$ ist daher

nach der Kettenregel

$$\int_\gamma \vec{V}\cdot ds = \int_a^b \vec{V}\big(\gamma(t)\big)\cdot\dot\gamma(t)\,dt = \int_a^b \operatorname{grad}\varphi\big(\gamma(t)\big)\cdot\dot\gamma(t)\,dt$$

$$= \int_a^b \frac{d}{dt}\varphi\big(\gamma(t)\big)\,dt = \varphi\big(\gamma(b)\big) - \varphi\big(\gamma(a)\big)\,.$$

Damit ist das Integral, genau wie wir es vom Eindimensionalen gewohnt sind, einfach gleich der Differenz der Werte der Stammfunktion am Endpunkt und am Anfangspunkt der Kurve.

2.) $\Rightarrow$ 3.): Diese Implikation ist klar, denn wenn der Wert des Integrals nicht von der Kurve abhängt, ist das Integral längs einer geschlossenen Kurve gleich dem Integral längs der zu einem Punkt degenerierten Kurve, und das ist natürlich gleich Null.

3.) $\Rightarrow$ 2.): Zumindest anschaulich ist auch hier klar, was wir machen: Wir durchlaufen zunächst die Kurve γ in der üblichen Weise und dann die Kurve δ „rückwärts". Dies ergibt eine geschlossene Kurve $\widetilde\gamma$, auf die wir *3.)* anwenden können.

Konkret sei für ein Kurvenstück $\delta_i\colon [c_i, d_i] \to \mathbb{R}^n$ von δ

$$\delta_i^*\colon [c_i, d_i] \to \mathbb{R}^n; \quad t \mapsto \delta(c_i + d_i - t)$$

das rückwärts durchlaufene Kurvenstück zu δ_i; ist r die Anzahl der δ_i, so bestehe die Kurve δ^* aus den Kurvenstücken $\delta_r^*, \ldots, \delta_1^*$. Da γ und δ dieselben Anfangs- und Endpunkte haben, ist der Endpunkt von δ^* gleich dem Anfangspunkt von γ und umgekehrt; die Folge von Kurvenstücken $\gamma_1, \ldots, \gamma_s, \delta_r^*, \ldots, \delta_1^*$ ist deshalb eine geschlossene Kurve $\widetilde\gamma$. Nach Voraussetzung verschwindet das Integral längs einer solchen Kurve, d.h.

$$0 = \int_{\widetilde\gamma} \vec{V}\,ds = \int_\gamma \vec{V}\,ds + \int_{\delta^*} \vec{V}\,ds = \int_\gamma \vec{V}\,ds - \int_\delta \vec{V}\,ds\,,$$

und damit sind die Integrale über γ und δ gleich.

2) $\Rightarrow$ 1): $\boldsymbol{x}_0$ sei irgendein beliebig ausgewählter Punkt von D. Da D zusammenhängend ist, gibt es dann für jeden Punkt $\boldsymbol{x} \in D$ eine Kurve $\gamma_{\boldsymbol{x}}$ mit Anfangspunkt $\boldsymbol{x}_0$ und Endpunkt $\boldsymbol{x}$.

Wir definieren nun einen Kandidaten für eine Stammfunktion durch

$$\varphi(\boldsymbol{x}) = \int_{\gamma_{\boldsymbol{x}}} \vec{V}\,ds\,;$$

wegen der Voraussetzung *2)* hängt $\varphi(\boldsymbol{x})$ in der Tat nur von $\boldsymbol{x}$ ab und nicht von der Wahl des Wegs $\gamma_{\boldsymbol{x}}$.

Für einen hinreichend kleinen Vektor $\vec{h}$ liegt auch die Verbindungsstrecke von $\boldsymbol{x}$ mit $\boldsymbol{x}+\vec{h}$ in D und kann durch eine Kurve η parametrisiert werden; dann ist

$$\varphi(\boldsymbol{x}+\vec{h}) - \varphi(\boldsymbol{x}) = \int_{\gamma_{\boldsymbol{x}+\vec{h}}} \vec{V}\,ds - \int_{\gamma_{\boldsymbol{x}}} \vec{V}\,ds = \int_\eta \vec{V}\,ds\,.$$

Speziell für einen Vektor $\vec{h} = h\vec{e}_i$ der Länge h in Richtung des Einheitsvektors $\vec{e}_i$ der Koordinatenachse x_i ist

$$\int_\eta \vec{V}\,ds = \int_\eta V_i\,dx_i = \int_0^h V_i(\boldsymbol{x}+t\vec{e}_i)\,dt\,,$$

und damit

$$\lim_{h\to 0}\frac{\varphi(\boldsymbol{x}+h\vec{e}_i)-\varphi(\boldsymbol{x})}{h} = \lim_{h\to 0}\frac{1}{h}\int_0^h V_i(\boldsymbol{x}+t\vec{e}_i)\,dt = V_i(\boldsymbol{x})$$

nach dem Hauptsatz der Differential- und Integralrechnung für Funktionen einer reellen Veränderlichen. Dies zeigt, daß $\operatorname{grad}\varphi = \vec{V}$ ist und das Vektorfeld somit eine Stammfunktion hat. ■

Im $\mathbb{R}^3$ sollten wir erwarten, daß es noch eine vierte äquivalente Charakterisierung konservativer Vektorfelder gibt: Da die Zirkulationsfreiheit mit Drehungen um eine Achse zusammenhängt, sollte für solche Felder auch die Rotation verschwinden. Diese Richtung ist trivial:

Lemma: Für ein zirkulationsfreies Vektorfeld $\vec{V}\colon D \to \mathbb{R}^3$ auf $D \subseteq \mathbb{R}^3$ ist $\operatorname{rot}\vec{V} \equiv 0$ auf D.

Beweis: Nach dem gerade bewiesenen Satz ist ein zirkulationsfreies Vektorfeld konservativ, und nach den Rechenregeln aus §2*f)* verschwindet die Rotation eines Gradienten. ∎

Die umgekehrte Richtung allerdings ist zumindest in dieser Allgemeinheit falsch. Als **Gegenbeispiel** betrachten wir das Magnetfeld eines geradlinigen stromdurchflossenen Leiters, d.h. also (abgesehen von konstanten Vorfaktoren) das Vektorfeld

$$\vec{V}(x,y,z) = \frac{1}{x^2+y^2}\begin{pmatrix} -y \\ x \\ 0 \end{pmatrix} .$$

Wie wir in §2*e2)* gesehen haben, verschwindet seine Rotation im gesamten Definitionsbereich des Vektorfelds, d.h. überall außerhalb der z-Achse.

Trotzdem ist das Vektorfeld nicht zirkulationsfrei, denn für den Einheitskreis

$$\gamma\colon [0,2\pi] \to \mathbb{R}^3;\, t \mapsto (\cos t, \sin t, 0)$$

ist

$$\int_\gamma \vec{V}\, ds = \int_0^{2\pi} \vec{V}\big(\gamma(t)\big)\dot{\gamma}(t)\, dt = \int_0^{2\pi} \begin{pmatrix} -\sin t \\ \cos t \\ 0 \end{pmatrix}\begin{pmatrix} -\sin t \\ \cos t \\ 0 \end{pmatrix} dt$$

$$= \int_0^{2\pi} dt = 2\pi .$$

Der Grund für dieses Verhalten liegt, wie wir bald sehen werden, darin, daß $\vec{V}$ auf der z-Achse nicht definiert ist: Obwohl die z-Achse im Vergleich zum gesamten $\mathbb{R}^3$ nur einen – sollte man meinen – vernachlässigbar geringen Teil ausmacht, genügt selbst diese minimale Definitionslücke, um die Umkehrung des Lemmas falsch zu machen.

§5: Mehrdimensionale Integrationstheorie

Die bisher betrachteten Kurvenintegrale waren alle zurückführbar auf gewöhnliche RIEMANN-Integrale über Funktionen einer reellen Veränderlichen. Mit ihrer Hilfe ist es möglich, die Länge von Kurven zu bestimmen, nicht aber Flächeninhalte oder Volumen von höherdimensionalen geometrischen Gebilden. Deren Berechnung, sowie die von naheliegenden Verallgemeinerungen wie dem Fluß durch die Oberfläche eines Bereichs sind Gegenstand dieses Paragraphen.

a) Flächeninhalte und Volumina

Beginnen wir mit Flächeninhalten und Volumina. Bereits das gewöhnliche RIEMANN-Integral kann als Flächeninhalt interpretiert werden, allerdings nur für die Fläche zwischen der x-Achse und einer Kurve zwischen zwei gegebenen x-Koordinaten. Hier soll es nun um beliebige Flächen im $\mathbb{R}^2$, Volumina im $\mathbb{R}^3$ und die entsprechenden höherdimensionalen Konzepte gehen.

Bei der Definition des RIEMANN-Integrals wird die Fläche unterhalb der Kurve durch Rechtecke angenähert, deren Kantenlänge in x-Richtung immer kleiner wird, während die Kantenlänge in y-Richtung durch die y-Koordinaten der Kurve $y = f(x)$ gegeben war. Im allgemeinen Fall, wo es keine ausgezeichnete Richtung mehr gibt, wird diese Unterscheidung zwischen x- und y-Richtung offensichtlich sinnlos; die einzig mögliche Verallgemeinerung des RIEMANN-Integrals besteht darin, daß man eine Fläche durch *beliebige* Rechtecke annähert und beim Grenzübergang *beide* Seiten gegen Null gehen läßt.

Ganz entsprechend müssen für dreidimensionale Bereiche Quader betrachtet werden, deren drei Seiten mit Verfeinerung der Überdeckung immer kleiner werden, *usw.*

Formal gehen wir dazu wie folgt vor: Wir wählen ein festes Koordinatensystem im $\mathbb{R}^n$ und betrachten Quader $Q_i \subset \mathbb{R}^n$, deren Kanten parallel zu denn Koordinatenachsen sind. (Für $n = 2$ sind diese „Quader“ natürlich Rechtecke, und für $n = 1$ Intervalle.) Das Volumen eines solchen Quaders soll gleich dem Produkt seiner Kantenlängen sein, genau wie wir es aus der Elementargeometrie gewohnt sind.

Definition: Eine *Elementarmenge* in $\mathbb{R}^n$ ist eine Teilmenge $E \subset \mathbb{R}^n$, die als Vereinigung endlich vieler Quader mit achsenparallelen Kanten geschrieben werden kann; dabei dürfen sich zwei Quader höchstens in gemeinsamen Randpunkten schneiden.

Im $\mathbb{R}^2$ wäre also beispielsweise jede Menge, deren Rand auf kariertem Papier so gezeichnet werden kann, daß alle Linien auf Karokanten liegen, eine Elementarmenge.

Das Volumen $\mu(E)$ einer Elementarmenge E definieren wir als die Summe der Volumina der endlich vielen Quader, aus denen die Menge besteht; man überlegt sich leicht, daß es unabhängig von der Art der Zerlegung der Menge in Quader ist.

Nun gehen wir im wesentlichen genauso vor, wie bei der Definition des RIEMANN-Integrals: Dort hatten wir die Fläche unterhalb einer Kurve $y = f(x)$ sowohl von oben als auch von unten durch Rechtecke angenähert; die Flächen der entsprechenden Elementarmengen hatten wir als RIEMANNsche Unter- *bzw.* Obersummen bezeichnet. Das Integral existierte nach Definition genau dann, wenn bei immer weiterer Verfeinerung der Überdeckung die Untersummen und die Obersummen gegen einen gemeinsamen Grenzwert konvergierten.

Entsprechend betrachten wir zur Definition des Volumens einer Teilmenge $B \subset \mathbb{R}^n$ Elementarmengen, die ganz in B enthalten sind und bezeichnen den Grenzwert bei immer feiner werdenden Quader als *unteres Volumen* $\underline{\mu}(B)$. Zur exakten Definition verwenden wir besser nicht einen Grenzwert, da es etwas umständlich wäre, hier zu definieren, über was genau wir den Grenzwert bilden, sondern wir definieren $\underline{\mu}(B)$ einfach als *Supremum* der Volumina aller Elementarmengen, die in B liegen:

$$\underline{\mu}(B) \underset{\text{def}}{=} \sup\left\{\mu(E) \;\middle|\; E \subseteq B \text{ Elementarmenge}\right\},$$

falls dieses Supremum existiert. (Es existiert offensichtlich genau dann, wenn die Menge B beschränkt ist; für unbeschränkte Mengen wie etwa den gesamten $\mathbb{R}^n$ ist die Menge aller $\mu(E)$ unbeschränkt, so daß kein Supremum existiert.)

Genauso definieren wir ein oberes Volumen $\overline{\mu}(B)$ als *Infimum* der Volumina aller Elementarmengen, die B enthalten:

$$\overline{\mu}(B) \underset{\text{def}}{=} \inf\left\{\mu(E) \;\middle|\; E \supseteq B \text{ Elementarmenge}\right\},$$

falls dieses Infimum existiert. (Auch hier ist die Existenz wieder an die Beschränktheit von B gekoppelt, denn für ein unbeschränkte Menge B gibt es keine Elementarmenge, die B enthält, so daß wir das Infimum über die leere Menge bilden müßten.)

Die Abbildungen 71 und 72 zeigen die Annäherung einer Kreisfläche durch Elementarmengen von innen und außen; Abbildung 73 zeigt entsprechende Elementarmengen für eine Halbkugel.

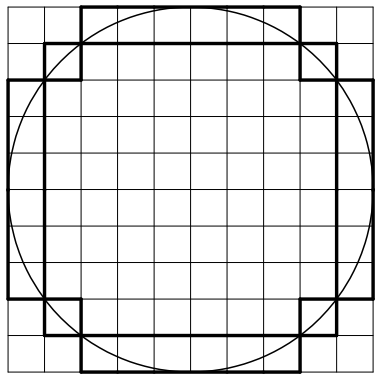


Abb. 71: Approximation einer Kreisfläche auf 10×10 Quadraten

Die Definition des Volumens ist nun fast selbstverständlich:

Definition: Wir sagen, die Menge $B \subset \mathbb{R}^n$ habe das Volumen $\mu(B)$, falls das untere Volumen $\underline{\mu}(B)$ und das obere Volumen $\overline{\mu}(B)$ beide existieren und gleich $\mu(B)$ sind.

Ganz entsprechend können wir auch Integrale über Funktionen definieren: $f\colon D \to \mathbb{R}$ sei eine Funktion auf der Teilmenge $D \subseteq \mathbb{R}^n$, und

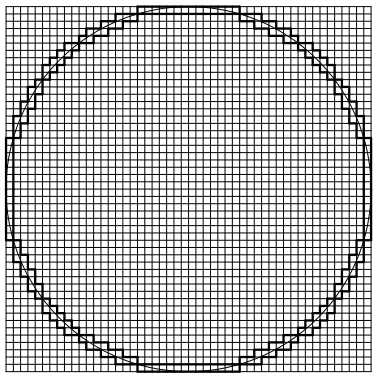


Abb. 72: Approximation einer Kreisfläche auf 50×50 Quadraten

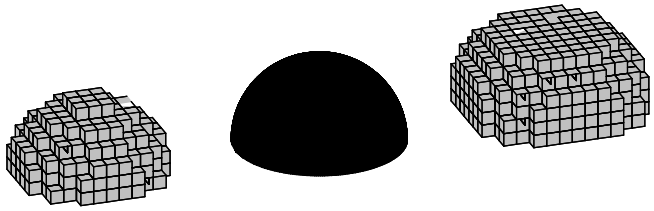


Abb. 73: Approximation einer Halbkugel durch Würfel

$B \subseteq D$ sei eine Teilmenge von D. Dann können wir wieder RIEMANNsche Untersummen definieren, indem wir eine Elementarmenge $E \subseteq D$ betrachten, bestehend etwa aus den Quadern $Q_1, \ldots, Q_N$, und dazu die RIEMANNsche Untersumme definieren als

$$\sum_{i=1}^{N} \mu(Q_i) \cdot \inf\{f(\boldsymbol{x}) \mid \boldsymbol{x} \in Q_i\} .$$

Die RIEMANNsche Obersumme für die Elementarmenge $E' \supseteq B$ beste-

hend aus den Quadern $Q'_1, \ldots, Q'_M$ ist entsprechend

$$\sum_{j=1}^{M} \mu(Q'_j) \cdot \sup\{f(\boldsymbol{x}) \mid \boldsymbol{x} \in Q_j\} ,$$

und wir definieren

Definition: $\int \cdots \int_B f(\boldsymbol{x})\, dx_1 \ldots dx_n$ existiert, wenn das Supremum der Menge aller RIEMANNscher Untersummen gleich dem Infimum der Menge aller RIEMANNscher Obersummen ist; dieser gemeinsame Wert ist der Wert des Integrals.

Die Schreibweise $\int \cdots \int_B f(\boldsymbol{x})\, dx_1 \ldots dx_n$ soll dabei bedeuten, daß wir n Integralzeichen schreiben, für $B \subset \mathbb{R}^2$ also

$$\iint_B f(x, y)\, dx\, dy$$

und für $B \subset \mathbb{R}^3$

$$\iiint_B f(x, y, z)\, dx\, dy\, dz .$$

(Nicht alle Lehrbücher verwenden bei der mehrdimensionalen Integration mehrere Integralzeichen; einige, wie etwa **[D]**, schreiben unabhängig von der Dimension immer nur ein Integralzeichen. In Physik und Technik scheint die Schreibweise mit mehreren Integralzeichen üblicher zu sein; deshalb soll auch hier diese Konvention verwendet werden.)

Integrale und Volumina sind natürlich eng miteinander verwandt: Einerseits läßt sich das Volumen eines Bereichs B auch als

$$\mu(B) = \int \cdots \int_B 1\, dx_1 \cdots dx_n$$

schreiben, andererseits ist, völlig analog zur Flächeninterpretation des RIEMANN-Integrals, für eine auf B nichtnegative Funktion f

$$\int \cdots \int_B f(\boldsymbol{x})\, dx_1 \cdots dx_n =$$

$$\mu\left(\left\{(x_1,\dots,x_n,y)\in\mathbb{R}^{n+1}\;\middle|\;\begin{array}{l}(x_1,\dots x_n)\in B \text{ und}\\ 0\le y\le f(x_1,\dots,x_n)\end{array}\right\}\right).$$

Auch für das so definierte mehrdimensionale Integral gelten aus offensichtlichen Gründen die Analoga der aus aus der eindimensionalen Integralrechnung bekannten Rechenregeln wie die *Monotonieregel*

$$\int\cdots\int_B f(\boldsymbol{x})\,dx_1\cdots dx_n \le \int\cdots\int_B g(\boldsymbol{x})\,dx_1\cdots dx_n\,,$$

falls $f(\boldsymbol{x})\le g(\boldsymbol{x})$ für alle $\boldsymbol{x}\in B$, und die *Linearität*seigenschaft

$$\int\cdots\int_B \big(\alpha f(\boldsymbol{x})+\beta g(\boldsymbol{x})\big)\,dx_1\cdots dx_n$$

$$=\alpha\int\cdots\int_B f(\boldsymbol{x})\,dx_1\cdots dx_n+\beta\int\cdots\int_B g(\boldsymbol{x})\,dx_1\cdots dx_n\,.$$

Außerdem haben wir wieder einen

Mittelwertsatz: Für einen zusammenhängenden abgeschlossenen und beschränkten Bereich $B\subset\mathbb{R}^n$ und eine stetige Funktion $f\colon B\to\mathbb{R}$ gibt es einen Punkt $\boldsymbol{x}_0\in B$, so daß gilt

$$\int\cdots\int_B f(\boldsymbol{x})\,dx_1\cdots dx_n = f(\boldsymbol{x}_0)\cdot\mu(B)\,.$$

Der *Beweis* geht ganz genauso wie in $\mathbb{R}$: Eine stetige Funktion nimmt auf einem abgeschlossenen Intervall sowohl ihr Maximum als auch ihr Minimum an; genauso zeigt man, daß eine stetige Funktion auf einer beschränkten abgeschlossen Teilmenge $B\subset\mathbb{R}^n$ ihr Maximum und ihr Minimum annimmt. Sei etwa $\boldsymbol{x}_1\in B$ ein Punkt, in dem f minimal wird, und $\boldsymbol{x}_2$ einer, in dem f maximal wird. Dann ist für jedes $\boldsymbol{x}\in B$

$$f(\boldsymbol{x}_1)\le f(\boldsymbol{x})\le f(\boldsymbol{x}_2)\,;$$

nach der Monotonieregel ist daher

$$\int\cdots\int_B f(\boldsymbol{x}_1)\,dx_1\cdots dx_n = f(\boldsymbol{x}_1)\cdot\mu(B)$$

kleiner oder gleich

$$\int\cdots\int_B f(\boldsymbol{x})\,dx_1\cdots dx_n\,,$$

was wiederum kleiner oder gleich

$$\int\cdots\int_B f(\boldsymbol{x}_2)\,dx_1\cdots dx_n = f(\boldsymbol{x}_2)\cdot\mu(B)$$

ist. Damit gibt es eine Zahl η zwischen $f(\boldsymbol{x}_1)$ und $f(\boldsymbol{x}_2)$, so daß

$$\int\cdots\int_B f(\boldsymbol{x})\,dx_1\cdots dx_n = \eta\cdot\mu(B)$$

ist. Wir müssen noch zeigen, daß dieses η ein Funktionswert von f auf B ist. Dazu verbinden wir $\boldsymbol{x}_1$ und $\boldsymbol{x}_2$ durch eine Kurve γ; indem wir nötigenfalls die Parameterintervalle der Kurvenstücke von γ verschieben, können wir annehmen, daß γ durch ein zusammenhängendes Parameterintervall $[a,\,b]$ parametrisiert wird. Wegen der Stetigkeit von f haben wir dann eine stetige Funktion

$$\varphi\colon[a,\,b]\to\mathbb{R};\quad t\mapsto f\big(\gamma(t)\big)$$

mit $\varphi(a)=f(\boldsymbol{x}_1)$ und $\varphi(b)=f(\boldsymbol{x}_2)$. Nach dem Zwischenwertsatz gibt es dazu ein $\tau\in[a,\,b]$ mit $\varphi(c)=\eta$; mit $\boldsymbol{x}_0=\gamma(\tau)$ ist also $f(\boldsymbol{x}_0)=\eta$. Damit ist der Satz vollständig bewiesen. ∎

Wie beim gewöhnlichen (eindimensionalen) RIEMANN-Integral ist die Existenz von mehrdimensionalen Integralen normalerweise kein Problem, allerdings kann man natürlich leicht Beispiele konstruieren, für die das Integral nicht existiert. Wir können etwa in Analogie zur DIRICHLETschen Sprungfunktion die Menge

$$B=\big\{(x,y)\in\mathbb{R}^2\;\big|\;x,y\in\mathbb{Q}\text{ und }0\le x,y\le 1\big\}$$

aller Punkte mit rationalen Koordianten im Einheitsquadrat betrachten. Da $\mathbb{Q}$ keine reellen Intervalle enthält, sind alle Rechtecke, die ganz in B enthalten sind, zu Punkten degeneriert. Eine in B enthaltene Elementarmenge besteht also aus endlich vielen Punkten und hat somit den

Flächeninhalt null. Eine Elementarmenge, die ganz B enthält, muß aber das gesamte Einheitsquadrat enthalten, da die rationalen Punkte dort dicht liegen; sie hat also mindestens die Fläche eins. Somit ist eins der obere und null der untere Flächeninhalt von B; der Flächeninhalt von B existiert also nicht.

Die folgende Überlegung liefert ein Kriterium für die Existenz des Volumens einer Teilmenge $B \subset \mathbb{R}^n$: Die Differenzmenge zwischen einer Elementarmenge, die B enthält, und einer Elementarmenge, die in B enthalten ist, kann offenbar wieder als Elementarmenge aufgefaßt werden und enthält den Rand von B. Je besser die äußere und die innere Elementarmenge B annähern, desto weniger unterscheidet sich diese Differenzmenge vom Rand; die Existenz eines Volumens von B ist daher äquivalent dazu, daß das Volumen des Rands von B existiert und verschwindet.

Für eine *unbeschränkte* Menge B können wir bislang weder Volumen noch Integrale definieren, denn unsere Konstruktion ist nur anwendbar, wenn B in einer Elementarmenge enthalten ist. Das ist allerdings nichts neues, denn beim eindimensionalen RIEMANN-Integral tritt genau das gleiche Problem auf und wird dadurch gelöst, daß man *uneigentliche* Integrale einführt: Beispielsweise ist

$$\int_a^\infty f(x)\,dx = \lim_{b\to\infty} \int_a^b f(x)\,dx\,,$$

falls dieser Grenzwert existiert. Genauso definieren wir jetzt

Definition: $B \subseteq \mathbb{R}^n$ sei eine unbeschränkte Menge und die Funktion $f\colon D \to \mathbb{R}$ sei in einer B umfassenden Menge $D \subseteq \mathbb{R}^n$ erklärt. Wir sagen, das *uneigentliche Integral*

$$\int \cdots \int_B f(\boldsymbol{x})\,dx_1 \cdots dx_n$$

existiere, wenn für jede Folge

$$B_1 \subset B_2 \subset \cdots \subset B_i \subset \cdots$$

von beschränkten Mengen B_i mit $\bigcup_{i\geq 1} B_i = B$ der Grenzwert

$$\lim_{i\to\infty} \int \cdots \int_{B_i} f(\boldsymbol{x})\,dx_1 \cdots dx_n$$

existiert, und wenn er für jede solche Folge denselben Wert hat. Diesen gemeinsamen Wert bezeichnen wir als den Wert des uneigentlichen Integrals.

Wie auch im Eindimensionalen genügt es hier nicht, nur eine einzige Folge von Mengen B_i zu betrachten; ansonsten könnte man beispielsweise dem uneigentlichen Integral

$$\iint_{\mathbb{R}^2} \sin x \sin y\,dx\,dy$$

je nachdem, ob man für die B_i Quadrate nimmt, in deren Eckpunkten der Cosinus für beide Koordinaten verschwindet, oder Quadrate, in deren Eckpunkten die beiden Cosinus einen konstanten anderen Wert haben, dem Integral die verschiedensten Werte zuordnen. Tatsächlich aber existiert dieses uneigentliche Integral natürlich genauso wenig wie das eindimensionale Integral $\int_{-\infty}^{\infty} \sin x\,dx$.

Für interessantere Beispiele sei auf den nächsten Abschnitt verwiesen, wo wir

$$\iint_{\mathbb{R}^2} e^{-(x^2+y^2)}\,dx\,dy$$

berechnen werden.

b) Integration über Normalbereiche

Der Grund dafür, daß wir bislang noch keine interessanten Beispiele für mehrdimensionale Integrale haben, liegt natürlich darin, daß man Werte von Integralen fast nie durch Anwendung der Definition bestimmt. Im Eindimensionalen wird stattdessen meist den Hauptsatz der Differential- und Integralrechnung benutzt, über den die Berechnung eines Integrals auf die (nicht immer explizit mögliche) Bestimmung einer Stammfunktion zurückgeführt wurde; hier, im Mehrdimensionalen, wollen wir die

Integration soweit wie möglich auf mehrfache eindimensionale Integration zurückführen. Zumindest in einem Fall wissen wir schon, wie das geht: Ist die Funktion f auf dem Intervall $[a,b]$ nichtnegativ, so hat

$$B = \left\{(x,y) \in \mathbb{R}^2 \;\middle|\; a \le x \le b \text{ und } 0 \le y \le f(x)\right\}$$

die Fläche

$$\iint\limits_B dx\,dy = \int\limits_a^b f(x)\,dx\,.$$

Auch den Flächeninhalt der etwas komplizierteren Menge

$$B = \left\{(x,y) \in \mathbb{R}^2 \;\middle|\; a \le x \le b \text{ und } g(x) \le y \le h(x)\right\}$$

läßt sich, wenn $g(x) \le h(x)$ ist für alle $x \in [a,b]$, durch das eindimensionale Integral

$$\iint\limits_B dx\,dy = \int\limits_a^b \big(h(x) - g(x)\big)\,dx$$

ausdrücken. Ähnliche Formeln gelten auch nach Vertauschung der Rollen von x und y.

Definition: Eine Teilmenge $B \subset \mathbb{R}^2$ heißt *Normalbereich vom Typ I,* wenn es reelle Zahlen $a \le b$ und stetig differenzierbare Funktionen $g,h\colon [a,b] \to \mathbb{R}$ gibt mit $g(x) \le h(x)$ für alle $x \in [a,b]$, so daß

$$B = \left\{(x,y) \in \mathbb{R}^2 \;\middle|\; a \le x \le b \text{ und } g(x) \le y \le h(x)\right\}$$

ist. B heißt *Normalbereich vom Typ II,* wenn es reelle Zahlen $c \le d$ und stetig differenzierbare Funktionen $g,h\colon [c,d] \to \mathbb{R}$ gibt mit $g(y) \le h(y)$ für alle $y \in [a,b]$, so daß

$$B = \left\{(x,y) \in \mathbb{R}^2 \;\middle|\; c \le y \le d \text{ und } g(y) \le x \le h(y)\right\}.$$

Für diese Normalbereiche können wir nicht nur die Fläche leicht ausrechnen, sondern auch beliebige Integrale über stetige Funktionen:

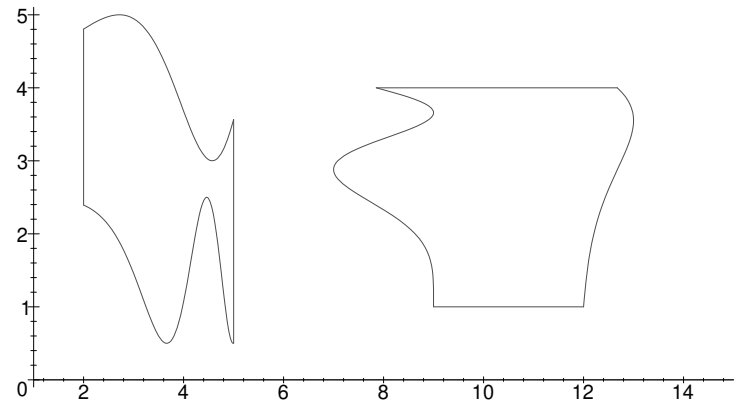


Abb. 74: Normalbereiche vom Typ I und II

Satz: $f\colon D \to \mathbb{R}$ sei eine stetige Funktion auf $D \subseteq \mathbb{R}^n$.
a) Für einen Normalbereich $B \subseteq D$ vom Typ I ist

$$\iint\limits_B f(x,y)\,dx\,dy = \int\limits_a^b \left(\int\limits_{g(x)}^{h(x)} f(x,y)\,dy \right) dx\,.$$

b) Für einen Normalbereich $B \subseteq D$ vom Typ II ist

$$\iint\limits_B f(x,y)\,dx\,dy = \int\limits_c^d \left(\int\limits_{g(y)}^{h(y)} f(x,y)\,dx \right) dy\,.$$

Beweis: Wir wissen bereits, daß der Flächeninhalt von B existiert; da f eine stetige Funktion ist, folgt dann auch ohne größere Schwierigkeiten die Existenz des Integrals von f über B. Wir wollen darauf nicht genauer eingehen, sondern nur benutzen, daß es dann ausreicht, spezielle Folgen von Elementarmengen zu betrachten. Eine solche spezielle Folge erhalten wir etwa dadurch, daß wir den ganzen $\mathbb{R}^2$ mit einem Quadrategitter überziehen; die Seitenlänge der Quadrate sei $k = \frac{b-a}{N}$ und ihr Flächeninhalt dementsprechend gleich k^2.

Da die Aussage *b)* durch Vertauschung der beiden Koordinaten in *a)* übergeht, reicht es, den Satz für Normalbereiche vom Typ I zu beweisen.

Als in B enthaltene Elementarmenge E_N wählen wir die Menge aller Quadrate, die ganz in B liegen; als B enthaltende Elementarmenge E'_N entsprechend die Menge aller Quadrate, die nichtleeren Durchschnitt mit B haben. Wir numerieren die Quadrate in E_N und E'_N mit zwei Indizes: Das Intervall $[a,b]$ wird durch das Quadrategitter in $N-1$ Teilintervalle der Länge k zerlegt; das Quadrat Q_{ij} liege über dem i-ten dieser Teilintervalle und sei, von unter her gesehen, das j-te Quadrat, das ganz in B liegt. Die Anzahl der ganz in B liegenden Quadrate im i-ten Streifen sei r_i; dann ist die RIEMANNsche Untersumme zu E_N gleich

$$\sum_{i=1}^{N-1}\sum_{j=1}^{r_i}\mu(Q_{ij})\cdot\inf_{(x,y)\in Q_{ij}} f(x,y) = \sum_{i=1}^{N-1}\left(\sum_{j=1}^{r_i}\inf_{(x,y)\in Q_{ij}} f(x,y)\cdot k\right)\cdot k\,.$$

Für die RIEMANNsche Obersumme müssen wir alle Quadrate betrachten, die mit B nichtleeren Durchschnitt haben; dabei treten im allgemeinen auch Quadrate Q_{ij} unterhalb von Q_{i1} auf, für deren Bezeichnung wir Indizes $j \le 0$ verwenden.

Im i-ten Streifen mögen die Q_{ij} mit $s_i \le j \le t_i$ auftreten; dann ist die RIEMANNsche Obersumme gleich

$$\sum_{i=1}^{N-1}\sum_{j=s_i}^{t_i}\mu(Q_{ij})\cdot\sup_{(x,y)\in Q_{ij}} f(x,y) = \sum_{i=1}^{N-1}\left(\sum_{j=s_i}^{t_i}\sup_{(x,y)\in Q_{ij}} f(x,y)\cdot k\right)\cdot k\,.$$

Da f als stetige Funktion RIEMANN-integrierbar ist, konvergieren für $N\to\infty$ und damit $k\to 0$

$$\sum_{j=1}^{r_i}\inf_{(x,y)\in Q_{ij}} f(x,y)\cdot k \quad\text{und}\quad \sum_{j=s_i}^{t_i}\sup_{(x,y)\in Q_{ij}} f(x,y)\cdot k$$

beide gegen

$$\int_{g(x)}^{h(x)} f(x,y)\,dy\,;$$

die RIEMANNschen Unter- und Obersummen daher entsprechend gegen

$$\int_a^b\left(\int_{g(x)}^{h(x)} f(x,y)\,dy\right)dx\,,$$

wie behauptet. ■

Als erstes Beispiel berechnen wir zur Kontrolle etwas Altbekanntes, die Fläche der Einheitskreisscheibe

$$B = \left\{(x,y)\in\mathbb{R}^2 \;\middle|\; x^2+y^2\le 1\right\},$$

die wir hier umständlich als

$$\iint_B dx\,dy$$

ausrechnen wollen.

B kann sowohl als Normalbereich vom Typ I wie auch als solcher vom Typ II geschrieben werden:

$$\begin{aligned} B &= \left\{(x,y)\in\mathbb{R}^2 \;\middle|\; -1\le x\le 1 \text{ und } -\sqrt{1-x^2}\le y\le\sqrt{1-x^2}\right\}\\ &= \left\{(x,y)\in\mathbb{R}^2 \;\middle|\; -1\le y\le 1 \text{ und } -\sqrt{1-y^2}\le x\le\sqrt{1-y^2}\right\}. \end{aligned}$$

In der ersten Darstellung ist nach Teil *a)* des gerade bewiesenen Satzes

$$\iint_B dx\,dy = \int_{-1}^{1}\left(\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} dy\right)dx = \int_{-1}^{1} 2\sqrt{1-x^2}\,dx\,.$$

Zur Berechnung dieses Integrals können wir die *Substitutionsregel* anwenden: Mit $x=\sin t$ ist

$$\int_{-1}^{1} 2\sqrt{1-x^2}\,dx = \int_{-\pi/2}^{\pi/2} 2\sqrt{1-\sin^2 t}\cos t\,dt\,,$$

und das rechtsstehende Integral läßt sich mit partieller Integration ausrechnen:

$$\begin{aligned} \int_{-\pi/2}^{\pi/2}\cos^2 t\,dt &= \sin t\cos t\,\Big|_{-\pi/2}^{\pi/2} - \int_{-\pi/2}^{\pi/2}\sin^2 t\,dt\\ &= 0 + \int_{-\pi/2}^{\pi/2}(1-\cos^2 t)\,dt = \pi - \int_{-\pi/2}^{\pi/2}\cos^2 t\,dt\,. \end{aligned}$$

Wenn wir das Integral ganz rechts auf die linke Seite bringen, erhalten wir den Flächeninhalt

$$\int_{-\pi/2}^{\pi/2} 2\sqrt{1-x^2}\,dx = \int_{-\pi/2}^{\pi/2} 2\cos^2 t\,dt = \pi\,,$$

wie erwartet.

Etwa interessanter ist die Berechnung des (hoffentlich auch bekannten) Volumens der Einheitskugel

$$K = \left\{(x,y,z) \in \mathbb{R}^3 \;\middle|\; x^2+y^2+z^2 \le 1\right\}.$$

Hierzu könnten wir eine dreidimensionale Integration durchführen, es geht aber zum Glück auch einfacher: Natürlich genügt es, das Volumen der Halbkugel $z \ge 0$ zu bestimmen, und dies ist die Menge aller Punkte zwischen der Einheitskreisscheibe B der (x,y)-Ebene und dem Graph der Funktion

$$f(x,y) = \sqrt{1-x^2-y^2}$$

über B. Somit ist das Volumen der Halbkugel gleich

$$\iint_B \sqrt{1-x^2-y^2}\,dx\,dy = \int_{-1}^{1}\left(\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \sqrt{1-x^2-y^2}\,dy\right)dx\,.$$

Schreibt man mit $a = \sqrt{1-x^2}$

$$\sqrt{1-x^2-y^2} = \sqrt{a^2-y^2} = a\sqrt{1-\left(\frac{y}{a}\right)^2}\,,$$

so ist das innere Integral

$$\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \sqrt{1-x^2-y^2}\,dy = a\int_{-a}^{a}\sqrt{1-\left(\frac{y}{a}\right)^2}\,dy\,,$$

was durch die Substitution $y = au$ zu

$$a\int_{-1}^{1}\sqrt{1-u^2}\,a\,du = a^2\int_{-1}^{1}\sqrt{1-u^2}\,du = \frac{\pi}{2}a^2$$

wird, wie wir oben gerade nachgerechnet haben. Somit ist

$$\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \sqrt{1-x^2-y^2}\,dy = \frac{\pi}{2}(1-x^2)$$

und

$$\iint_B \sqrt{1-x^2-y^2}\,dx\,dy = \int_{-1}^{1}\frac{\pi}{2}(1-x^2)\,dx$$

$$= \frac{\pi}{2}\left(x-\frac{x^3}{3}\right)\Bigg|_{-1}^{1} = \frac{\pi}{2}\cdot\frac{4}{3} = \frac{2\pi}{3}\,;$$

das Kugelvolumen als Volumen zweier Halbkugeln ist also in der Tat, wie es sein soll, gleich $4\pi/3$.

In diesen Beispielen hatten wir nur Kreisscheiben als Normalbereiche, aber der Leser kann sich leicht überzeugen selbst davon überzeugen, daß sich die Nützlichkeit dieser Mengen keineswegs auf solche trivialen Beispiele beschränkt: Man zeichne frei irgendeneinen beschränkten zweidimensionalen Bereich auf ein Blatt Papier und überzeuge sich davon, daß sich dieser durch Einfügen von (meist sehr wenigen) waagrechten und senkrechten Strecken als Vereinigung von Normalbereichen der Typen I und II darstellen läßt.

c) Die Transformationsformel

Von den vielen Regeln zur expliziten Bestimmung einer Stammfunktion ist sicherlich die Substitutionsregel die nützlichste; es lohnt sich daher, nach einer Verallgemeinerung dieser Regel für mehrdimensionale Integrale zu suchen.

Die Idee im Eindimensionalen ist bekanntlich, daß wir die Integrationsvariable x als Funktion $x = \varphi(t)$ einer neuen Variablen t schreiben: Mit $a = \varphi(t_0)$ und $b = \varphi(t_1)$ ist

$$\int_a^b f(x)\,dx = \int_{t_0}^{t_1} f\big(\varphi(t)\big)\dot\varphi(t)\,dt\,.$$

Genauso können wir auch bei einer Funktion mehrerer Veränderlicher diese als Funktionen neuer Variabler schreiben.

Beginnen wir der Anschaulichkeit halber mit einer Funktion $f(x, y)$ zweier Veränderlicher und schreiben wir diese als Funktionen

$$x = x(u, v) \quad \text{und} \quad y = y(u, v)$$

zweier neuer Variabler u und v. Ein wichtiges Beispiel, das man zur Veranschaulichung während der folgenden Rechnungen im Kopf behalten sollte, ist die Polarkoordinatensdarstellung

$$x = x(r, \varphi) = r \sin \varphi \quad \text{und} \quad y = y(r, \varphi) = r \cos \varphi \,.$$

Zur Definition des Integrals

$$\iint\limits_B f(x, y)\, dx\, dy$$

approximierten wir den Intgegrationsbereich B durch kleine achsenparallele Rechtecke. Wenn wir statt über x und y über u und v integrieren, müssen wir entsprechend den Bereich B', in dem sich diese neuen Variablen bewegen, in achsenparallele Rechtecke zerlegen; dabei fordern wir nun natürlich Parallelität zur u- und zur v-Achse. Wir betrachten ein festes dieser Rechtecke; es habe die Eckpunkte

$$(u_0, v_0), \quad (u_0 + h, v_0), \quad (u_0, v_0 + k) \quad \text{und} \quad (u_0 + h, v_0 + k)$$

und somit den Flächeninhalt hk.

Die Menge der Punkte (x, y), die zu den Punkten (u, v) aus diesem Rechteck gehören, d.h. also die Menge

$$\left\{ \big(x(u, v), y(u, v)\big) \;\middle|\; u_0 \le u \le u_0 + h \quad \text{und} \quad v_0 \le v \le v_0 + k \right\}$$

ist natürlich im allgemeinen kein Rechteck, sondern eine krummlinig begrenzte Figur; im Beispiel der Polarkoordinaten etwa wäre sie ein Winkelbereich zwischen zwei Kreisbögen. (An den bei Polarkoordinaten etwas problematischen Nullpunkt als Ecke denken wir in diesen Zusammenhang lieber nicht; es ist klar, daß sein Einfluß bei immer

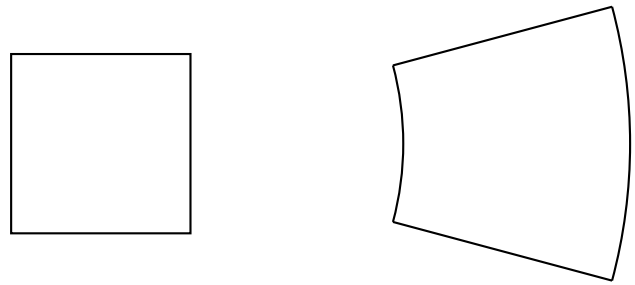


Abb. 75: Rechteck in (r, φ)-Ebene mit Bild in (x, y)-Ebene

kleiner werdenden Rechtecken für eine um den Nullpunkt beschränkte Funktion f immer kleiner wird.)

Trotzdem machen wir, wenn x und y *differenzierbare* Funktionen von u und v sind, bei kleinen Rechtecken keinen allzu großen Fehler, wenn wir die transformierte Menge als *Parallelogramm* betrachten, denn nach Definition der Differenzierbarkeit ist

$$\begin{aligned}
x(u_0 + h, v_0) &= x(u_0, v_0) + h \frac{\partial x}{\partial u}(u_0, v_0) + o(h) \\
y(u_0 + h, v_0) &= y(u_0, v_0) + h \frac{\partial y}{\partial u}(u_0, v_0) + o(h) \\
x(u_0, v_0 + k) &= x(u_0, v_0) + k \frac{\partial x}{\partial v}(u_0, v_0) + o(k) \\
y(u_0, v_0 + k) &= y(u_0, v_0) + k \frac{\partial y}{\partial v}(u_0, v_0) + o(k)
\end{aligned}$$

und

$$\begin{aligned}
x(u_0 + h, v_0 + k) &= x(u_0, v_0) + h \frac{\partial x}{\partial u}(u_0, v_0) + o(h) \\
&\quad + k \frac{\partial x}{\partial v}(u_0, v_0) + o(k) \\
y(u_0 + h, v_0 + k) &= y(u_0, v_0) + h \frac{\partial y}{\partial u}(u_0, v_0) + o(h) \\
&\quad + k \frac{\partial y}{\partial v}(u_0, v_0) + o(k)\,;
\end{aligned}$$

wenn wir Terme der Größenordnung $o(h)$ und $o(k)$ vernachlässigen, ist die transformierte Menge also ein Parallelogramm mit Kantenvektoren

$$h\begin{pmatrix}\frac{\partial x}{\partial u}(u_0,v_0)\\ \frac{\partial y}{\partial u}(u_0,v_0)\end{pmatrix} \quad \text{und} \quad k\begin{pmatrix}\frac{\partial x}{\partial v}(u_0,v_0)\\ \frac{\partial y}{\partial v}(u_0,v_0)\end{pmatrix}.$$

Der Flächeninhalt eines Parallelogramms ist bekanntlich gleich dem Produkt der Kantenlängen mal dem Sinus des eingeschlossenen Winkels; falls wir die beiden Vektoren in den $\mathbb{R}^3$ einbetten, indem wir ihnen eine Null als dritte Komponente geben, ist das gerade gleich dem Betrag des Vektorprodukts, das hier nur in der dritten Komponenten von null verschieden ist; die Fläche des Parallelogramms ist also

$$hk\cdot\left|\frac{\partial x}{\partial u}\frac{\partial y}{\partial v}-\frac{\partial x}{\partial v}\frac{\partial y}{\partial u}\right|.$$

Der gesamte Integrationsbereich wird approximiert durch die Vereinigung der Parallelogramme zu den sämtlichen Rechtecken, mit denen B' approximiert wurde; wenn wir also davon ausgehen, daß die Koordinatentransformation zwischen und B und B' bijektiv ist und wenn wir – was eigentlich noch durch genauere Abschätzungen zu rechtfertigen wäre – auch davon ausgehen, daß wir die oben erwähnten Fehler beim Grenzübergang $h\to 0$ und $k\to 0$ vernachlässigen können, erhalten wir die

Transformationsformel: $f\colon B\to\mathbb{R}$ sei eine integrierbare Funktion auf $B\subseteq\mathbb{R}^2$, und die Variablen x,y seien als differenzierbare Funktionen $x=x(u,v)$ und $y=y(u,v)$ neuer Variabler u,v dargestellt. Ist dann $B'\subseteq\mathbb{R}^2$ ein Integrationsbereich, für den die Abbildung

$$B'\to B;\quad (u,v)\mapsto\big(x(u,v),\ y(u,v)\big)$$

bijektiv ist, so gilt

$$\iint\limits_B f(x,y)\,dx\,dy=\iint\limits_{B'} f\big(x(u,v)\,y(u,v)\big)\left|\frac{\partial x}{\partial u}\frac{\partial y}{\partial v}-\frac{\partial x}{\partial v}\frac{\partial y}{\partial u}\right|\,du\,dv\,.$$

Für unbeschränkte Integrationsbereiche stehen hier natürlich uneigentliche Integrale; da diese – so sie existieren – Grenzwerte von üblichen Integralen sind, ist dies kein Problem.

Im Fall

$$x=r\cos\varphi \quad\text{und}\quad y=r\sin\varphi$$

der Polarkoordinaten ist

$$\frac{\partial x}{\partial r}\frac{\partial y}{\partial\varphi}-\frac{\partial x}{\partial\varphi}\frac{\partial y}{\partial r}=\cos\varphi\cdot(r\cos\varphi)-(-r\sin\varphi)\cdot\sin\varphi=r\,,$$

d.h.

$$\iint\limits_B f(x,y)\,dx\,dy=\iint\limits_{B'} f(r\cos\varphi,r\sin\varphi)\,r\,dr\,d\varphi.$$

Für diese Formel hätten wir eigentlich nicht den ganzen Apparat der Transformationsformel gebrauch; Abbildung 75 zeigt uns, wie wir die Fläche des Bilds eines Parallelogramms exakt ausrechnen können: Variiert r zwischen r und $r+h$ und φ zwischen φ und $\varphi+k$, so erhalten wir in der (x,y)-Ebene als Bild die Differenz zwischen zwei Kreissektoren mit Öffnungswinkel k und Radius $r+h$ beziehungsweise r; der Flächeninhalt ist also

$$\tfrac12 k(r+h)^2-\tfrac12 kr^2=r\cdot kh+\tfrac12 kh^2\,.$$

Wenn h und k simultan gegen null gehen, können wir kh^2 gegenüber kh vernachlässigen, der Flächeninhalt kh des Rechtecks aus der (r,φ)-Ebene wird also im wesentlichen nur mit r multipliziert – genau wie es die obige Rechnung auch zeigt.

Mit dieser Formel können wir beispielsweise noch einmal die Fläche eines Kreises ausrechnen: Die Punkte der Kreisscheibe B mit Radius R um den Nullpunkt haben Polarkoordinaten (r,φ) im Rechteck

$$B'=\big\{(r,\varphi)\;\big|\;0\le r\le R\quad\text{und}\quad 0\le\varphi<2\pi\big\}\,;$$

die Fläche von B ist also

$$\iint\limits_B dx\,dy=\iint\limits_{B'} r\,dr\,d\varphi=\int\limits_0^{2\pi}\left(\int\limits_0^R r\,dr\right)d\varphi=\int\limits_0^{2\pi}\frac{R^2}{2}\,d\varphi=\pi R^2\,.$$

Im Vergleich zum letzten Abschnitt, wo wir das Integral (für $R=1$) in kartesischen Koordinaten ausrechneten und dazu die Funktion $\sqrt{1-x^2}$ integrieren mußten, ist diese Rechnung erheblich einfacher; die Transformationsformel leistet also genau das, was wir von einer verallgemeinerten Substitutionsregel erwarten: Bei *geschickter,* an das Problem angepaßter Substitution kann sie die Berechnung eines Integrals erheblich vereinfachen.

Als nächstes Beispiel wollen wir endlich einmal ein Integral ausrechnen, bei dem wir das Ergebnis nicht besser und viel einfacher durch elementargeometrische Überlegungen bekommen können: das Integral

$$I \underset{\text{def}}{=} \int_{-\infty}^{\infty} e^{-x^2}\,dx\,.$$

Man kann zeigen, daß die Stammfunktion von e^{-x^2} nicht in geschlossener Form durch elementare Funktionen ausdrückbar ist; Integration mittels Stammfunktion ist also zwecklos. Stattdessen benutzen wir folgenden Trick:

Durch Grenzübergang können wir auch auch auf das unendliche „Rechteck“ $\mathbb{R}^2$ als Normalbereich auffassen; daher ist

$$\begin{aligned}\iint_{\mathbb{R}^2} e^{-(x^2+y^2)}\,dx\,dy &= \int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} e^{-(x^2+y^2)}\,dx\right)dy\\ &= \int_{-\infty}^{\infty} e^{-y^2}\left(\int_{-\infty}^{\infty} e^{-x^2}\,dx\right)dy = \int_{-\infty}^{\infty} e^{-y^2}\cdot I\,dy\\ &= I\cdot\int_{-\infty}^{\infty} e^{-y^2}\,dy = I^2\,.\end{aligned}$$

In Polarkoordinaten entspricht $\mathbb{R}^2$ dem Bereich $B' = \mathbb{R}_{\geq 0}\times[0,2\pi)$; also ist das betrachtete Integral nach der Transformationsformel auch gleich

$$\iint_{\mathbb{R}^2} e^{-(x^2+y^2)}\,dx\,dy = \iint_{B'} e^{-r^2}\,r\,dr\,d\varphi = \int_0^{2\pi}\left(\int_0^{\infty} re^{-r^2}\,dr\right)d\varphi\,.$$

Die Stammfunktion des neuen Integranden re^{-r^2} ist leicht zu finden: Aus

$$\frac{d}{dr}e^{-r^2} = -2re^{-r^2}$$

folgt sofort, daß

$$\int re^{-r^2}\,dr = -\tfrac{1}{2}e^{-r^2} + C\,.$$

Damit können wir weiterrechnen und erhalten das Ergebnis

$$\int_0^{2\pi}\left(\int_0^{\infty} re^{-r^2}\,dr\right)d\varphi = \int_0^{2\pi} -\tfrac{1}{2}e^{-r^2}\Big|_0^{\infty}\,d\varphi = \int_0^{2\pi}\tfrac{1}{2}\,d\varphi = \pi.$$

Also ist $I^2 = \pi$ und, da der Integrand von I überall positiv ist, folgt

$$I = \int_{-\infty}^{\infty} e^{-x^2}\,dx = \sqrt{\pi}\,.$$

Zur Verallgemeinerung der Transformationsformel auf höhere Dimensionen beachten wir zunächst, daß der Term

$$\frac{\partial x}{\partial u}\frac{\partial y}{\partial v} - \frac{\partial x}{\partial v}\frac{\partial y}{\partial u}$$

in der Formel

$$\iint_B f(x,y)\,dx\,dy = \iint_{B'} f\big(x(u,v)\,y(u,v)\big)\left|\frac{\partial x}{\partial u}\frac{\partial y}{\partial v} - \frac{\partial x}{\partial v}\frac{\partial y}{\partial u}\right|\,du\,dv$$

gerade gleich der Determinanten der JACOBI-Matrix

$$\begin{pmatrix}\dfrac{\partial x}{\partial u} & \dfrac{\partial x}{\partial v}\\[2mm] \dfrac{\partial y}{\partial u} & \dfrac{\partial y}{\partial v}\end{pmatrix}$$

des Koordinatenwechsels

$$(u,v)\mapsto\big(x(u,v),\ y(u,v)\big)$$

ist. Diese können wir uns auch geometrisch veranschaulichen, denn sie ist ja gerade der Flächeninhalt des Parallelogramms mit den Kantenvektoren

$$\begin{pmatrix}\dfrac{\partial x}{\partial u}(u_0,v_0)\\[2mm] \dfrac{\partial y}{\partial u}(u_0,v_0)\end{pmatrix}\quad\text{und}\quad\begin{pmatrix}\dfrac{\partial x}{\partial v}(u_0,v_0)\\[2mm] \dfrac{\partial y}{\partial v}(u_0,v_0)\end{pmatrix}.$$

Den Flächeninhalt eines Parallelogramms mit Kantenvektoren $\vec{a}$ und $\vec{b}$ können wir auch wie folgt ausrechnen:

Das (kartesische) Koordinatensystem in $\mathbb{R}^2$ sei so gewählt, daß der Vektor $\vec{a}$ ein Vielfaches $a\vec{e}_1$ des ersten Koordinateneinheitsvektors ist. Falls dann auch noch $\vec{b} = b\vec{e}_2$ ein Vielfaches des zweiten sein sollte, falls also das Parallelogramm sogar ein achsenparalleles Rechteck sein sollte, ist dessen Fläche gleich

$$ab = \det\begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} = \det(\vec{a}, \vec{b})\,.$$

Im allgemeinen wird das Parallelogramm natürlich kein Rechteck sein; da die Basisvektoren im Gegensatz zu $\vec{a}$ und $\vec{b}$ aufeinander senkrecht stehen, müssen wir $\vec{b}$ dann als Linearkombination $\vec{b} = c\vec{e}_1 + b\vec{e}_2$ schreiben. Dabei können wir $b\vec{e}_2$ als Projektion von $\vec{b}$ auf die von $\vec{e}_2$ aufgespannte Koordinatenachse auffassen. Da sich das Parallelogramm durch Scherung in das Rechteck mit Kantenvektoren $\vec{a}$ und $b\vec{e}_2$ überführen läßt und sich der Flächeninhalte bei Scherungen nicht ändern, hat das Parallelogramm immer noch den Flächeninhalt

$$ab = \det\begin{pmatrix} a & c \\ 0 & b \end{pmatrix} = \det(\vec{a}, \vec{b})\,.$$

Beim Übergang zu einem anderen orthonormalen Koordinatensystem werden die Koordinatenachsen gedreht, d.h. sie werden mit einer Matrix der Form

$$\begin{pmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{pmatrix}$$

multipliziert. Da deren Determinante gleich eins ist, ändert sich dabei nach dem Multiplikationssatz für Determinanten nichts am Wert von $\det(\vec{a}, \vec{b})$. Der Flächeninhalt eines von den Vektoren $\vec{a}$ und $\vec{b}$ aufgespannten Parallelogramms ist also stets gleich $\det(\vec{a}, \vec{b})$, egal bezüglich welcher Orthonormalbasis die Kantenvektoren dargestellt werden.

Genauso können wir auch in höheren Dimensionen argumentieren: Durch mehrfache Scherung läßt sich ein beliebiges Parallelepiped in einen Quader überführen; das Volumen des von n Kantenvektoren $\vec{a}_1, \ldots, \vec{a}_n$ aufgespannten Parallelepipeds ist also gleich

$$\det(\vec{a}_1, \ldots, \vec{a}_n)\,.$$

(Für $n = 3$ kennen wir diese Formel bereits, denn wir wissen, daß dort das Volumen gleich dem Spatprodukt $(\vec{a}_1 \times \vec{a}_2) \cdot \vec{a}_3$ der drei Vektoren ist, und dieses wiederum ist gleich der Determinanten.)

Damit gilt also im $\mathbb{R}^n$ die

Transformationsformel: $f\colon B \to \mathbb{R}$ sei eine integrierbare Funktion auf $B \subseteq \mathbb{R}^n$, und die Variablen $x_1, \ldots, x_n$ seien als differenzierbare Funktionen

$$\begin{aligned} x_1 &= x_1(u_1, \ldots, u_n) \\ &\vdots \\ x_n &= x_n(u_1, \ldots, u_n) \end{aligned}$$

neuer Variabler $u_1, \ldots, u_n$ dargestellt. Ist dann $B' \subseteq \mathbb{R}^2$ ein Integrationsbereich, für den die Abbildung

$$B' \to B; \quad \boldsymbol{u} \mapsto \boldsymbol{x}(\boldsymbol{u})$$

bijektiv ist, so gilt

$$\int \underset{B}{\cdots} \int f(\boldsymbol{x})\, dx_1 \cdots dx_n = \int \underset{B'}{\cdots} \int f\big(\boldsymbol{x}(\boldsymbol{u})\big)\, |\det J_{\boldsymbol{x}}(\boldsymbol{u})|\; du_1 \cdots du_n\,.$$

■

Ausgeschrieben wird die Formel für den $\mathbb{R}^3$, wenn wir kurz

$$F(u, v, w) \underset{\text{def}}{=} f\big(x(u, v, w),\ y(u, v, w),\ z(u, v, w)\big)$$

schreiben, zu

$$\iiint\limits_B f(x, y, z)\, dx\, dy\, dz =$$

$$\iiint\limits_{B'} F(u, v, w) \left| \det\begin{pmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} & \frac{\partial x}{\partial w} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} & \frac{\partial y}{\partial w} \\ \frac{\partial z}{\partial u} & \frac{\partial z}{\partial v} & \frac{\partial z}{\partial w} \end{pmatrix} \right| du\, dv\, dw\,.$$

Als Beispiel können wir wieder Koordinatensysteme betrachten: Für Kugelkoordinaten ist

$$\begin{aligned} x &= r\cos\varphi\sin\vartheta \\ y &= r\sin\varphi\sin\vartheta \quad , \\ z &= r\cos\vartheta \end{aligned}$$

die JACOBI-Matrix ist also

$$\begin{pmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \varphi} & \frac{\partial x}{\partial \vartheta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \varphi} & \frac{\partial y}{\partial \vartheta} \\ \frac{\partial z}{\partial r} & \frac{\partial z}{\partial \varphi} & \frac{\partial z}{\partial \vartheta} \end{pmatrix} = \begin{pmatrix} \cos\varphi\sin\vartheta & -r\sin\varphi\sin\vartheta & r\cos\varphi\cos\vartheta \\ \sin\varphi\sin\vartheta & r\cos\varphi\sin\vartheta & r\sin\varphi\cos\vartheta \\ \cos\vartheta & 0 & -r\sin\vartheta \end{pmatrix} .$$

Entwicklung nach der dritten Zeile ergibt für die Determinante den Wert

$$\begin{aligned} &\cos\vartheta \begin{vmatrix} -r\sin\varphi\sin\vartheta & r\cos\varphi\cos\vartheta \\ r\cos\varphi\sin\vartheta & r\sin\varphi\cos\vartheta \end{vmatrix} \\ &- r\sin\vartheta \begin{vmatrix} \cos\varphi\sin\vartheta & -r\sin\varphi\sin\vartheta \\ \sin\varphi\sin\vartheta & r\cos\varphi\sin\vartheta \end{vmatrix} \\ &= -r^2\cos\vartheta(\sin\vartheta\cos\vartheta) - r^2\sin\vartheta\cdot\sin^2\vartheta = -r^2\sin\vartheta . \end{aligned}$$

Der Betrag der Determinanten der JACOBI-Matrix ist also

$$r^2\,|\sin\vartheta| ,$$

und mit

$$F(r,\varphi,\vartheta) = f(r\cos\varphi\sin\vartheta,\ r\sin\varphi\sin\vartheta,\ r\cos\vartheta)$$

wird die Transformationsformel zu

$$\iiint\limits_B f(x,y,z)\,dx\,dy\,dz = \iiint\limits_{B'} F(r,\varphi,\vartheta)\cdot r^2\,|\sin\vartheta|\ dr\,d\varphi\,d\vartheta .$$

Berechnen wir zur Kontrolle schnell noch einmal das Volumen der Kugel B um den Nullpunkt mit Radius R:

$$\iiint\limits_B dx\,dy\,dz = \int_0^\pi \left(\int_0^{2\pi} \left(\int_0^R r^2\,|\sin\vartheta|\ dr \right) d\varphi \right) d\vartheta$$

$$= \int_0^\pi \left(\int_0^{2\pi} \frac{R^3}{3}\,|\sin\vartheta|\ d\varphi \right) d\vartheta = \frac{2\pi R^3}{3} \int_0^\pi \sin\vartheta\,d\vartheta = \frac{4\pi R^3}{3} .$$

Außer den Kugelkoordinaten hatten wir im $\mathbb{R}^3$ auch noch Zylinderkoordinaten betrachtet; für diese ist

$$\begin{aligned} x &= r\cos\varphi \\ y &= r\sin\varphi \qquad \text{und} \\ z &= z , \end{aligned}$$

also

$$\begin{pmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \varphi} & \frac{\partial x}{\partial z} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \varphi} & \frac{\partial y}{\partial z} \\ \frac{\partial z}{\partial r} & \frac{\partial z}{\partial \varphi} & \frac{\partial z}{\partial z} \end{pmatrix} = \begin{pmatrix} \cos\varphi & -r\sin\varphi & 0 \\ \sin\varphi & r\cos\varphi & 0 \\ 0 & 0 & 1 \end{pmatrix} ,$$

und die Determinante davon ist wie bei den ebenen Polarkoordinaten gleich r. Also haben wir hier im wesentlichen dieselbe Transformationsformel wie bei den ebenen Polarkoordinaten (Zylinderkoordinaten sind schließlich im wesentlichen auch nichts anderes als ebene Polarkoordinaten), nämlich

$$\iiint\limits_B f(x,y,z)\,dx\,dy\,dz = \iiint\limits_{B'} F(r,\varphi,z)\cdot r\,dr\,d\varphi\,dz ,$$

wobei wieder

$$F(r,\varphi,z) = f(r\cos\varphi, r\sin\varphi, z)$$

sein soll.

Genauso lassen sich auch beliebige andere Koordinatensysteme behandeln, beispielsweise könnte man zum Rechnen mit einem Ellipsoid auch Ellipsoidkoordinaten

$$\begin{aligned} x &= a\cos\varphi\sin\vartheta \\ y &= b\sin\varphi\sin\vartheta \\ z &= c\cos\vartheta \end{aligned}$$

einführen und damit das Volumen des Ellipsoids

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1$$

bestimmen. (Die Antwort ist natürlich *(?)* $\frac{4}{3}\pi abc$.)

In Physik und Technik sind noch zahlreiche weitere bewährte Koordinatensysteme im Einsatz, auf die wir hier nicht eingehen können; dank der obigen Transformationsformel kann man immer das benutzen, das der jeweiligen Situation am besten angepaßt ist und mithin am wenigsten Rechnung verlangt.

d) Der Satz von Green und der ebene Satz von Gauß

$B \subset \mathbb{R}^2$ sei ein Bereich, dessen Rand aus endlich vielen Kurvenstücken bestehe. Dann können wir sowohl über B als auch über dessen Randkurve ∂B integrieren; die beiden Sätze in diesem Abschnitt befassen sich mit der Beziehung zwischen diesen beiden Integrationen.

Im allgemeinen kann man natürlich nicht viel erwarten: Nimmt man aus B eine Kurve γ heraus, ist es bei der Integration einer stetigen Funktion gleichgültig, ob man über B oder über $B \setminus \gamma$ integriert, das Randintegral ändert sich aber um das im allgemeinen nicht verschwindende Kurvenintegral über γ.

Für eine geschlossene Kurve γ dagegen erwarten wir, wenn wir uns von der Anschauung leiten lassen und uns die Kurve als eine Art deformierten Kreis vorstellen, daß sie die Ebene in zwei Bereiche zerlegt: einen beschränkten Bereich B und einen unbeschränkten Bereich B'. Hier sollten sich γ und B gegenseitig so stark beeinflussen, daß eigentlich auch die Integration über B etwas mit der Integration über γ zu tun haben müßte.

Nun ist es natürlich etwas gefährlich, sich *nur* von der Anschauung leiten zu lassen, denn es gibt sicherlich erheblich mehr Kurven, als man sich gemeinhin vorstellt. Beispielsweise gibt es auch Kurven, die wie eine Ziffer 8 aussehen, etwa die Lemniskate, und diese zerlegen die Ebene in *drei* Bereiche. Wir müssen also noch zusätzlich fordern, daß sich die Kurve nicht selbst überkreuzt, aber dann gilt in der Tat

Jordanscher Kurvensatz: γ sei eine geschlossene ebene Kurve ohne Überkreuzungen. Dann hat $\mathbb{R}^2 \setminus \gamma$ zwei Zusammenhangskomponenten, von denen die eine beschränkt und die andere unbeschränkt ist.

Der Beweis dieses anschaulich fast selbstverständlichen Satzes erfordert einen erstaunlich großen Aufwand: Man muß die Ebene in kleine Quadrate (oder Dreiecke oder etwas ähnliches) unterteilen, γ durch Kantenzüge aus deren Seiten annähern, den Satz durch aufwendige Rechnungen mit formalen Summen von Quadraten und Quadratseiten für die angenäherten Kurven beweisen und dann schließlich noch durch ein Limesargument zeigen, daß die Aussage auch für γ selbst gilt. Die genaue Durchführung dieses Beweises würde etwa drei Vorlesungstermine erfordern – ein Aufwand, der für uns in keinem vernünftigen Zusammenhang mit seinem Nutzen steht: Bei den meisten in der Praxis vorkommenden Kurven wird die Aussage des JORDANschen Kurvensatzes zumindest für diese speziellen Kurven ohnehin klar sein.

MARIE ENNEMOND CAMILLE JORDAN (1838–1922) leistete wesentliche Beiträge zur Entwicklung der Topologie und der Gruppentheorie. Er beschäftigte sich auch mit der Lösbarkeit nichtlinearer Gleichungen und stellte in einem vielbeachteten Buch die GALOISsche Theorie über die Nichtlösbarkeit allgemeiner Gleichungen vom Grad größer vier dar. Seine Untersuchungen über endliche Körper führten ihn zu der heute als JORDAN-Zerlegung bekannten Darstellung von Matrizen, mit der wir uns im nächsten Semester beschäftigen werden; außerdem bewies er Satze über die Konvergenz von FOURIER-Reihen.

Um den JORDANschen Kurvensatz zu umgehen, starten wir einfach mit einem Bereich $B \subset \mathbb{R}^2$, von dem wir all das verlangen, was uns nachher das Leben einfach macht; in der Praxis sollte es (hoffentlich) selten schwierig sein, diese Forderungen für die Bereiche aus konkreten Anwendungen zu verifizieren.

Satz von Green: $B \subset \mathbb{R}^2$ sei eine abgeschlossene Menge, die sowohl als Vereinigung endlich vieler Normalbereiche vom Typ I wie auch als Vereinigung endlich vieler Normalbereiche vom Typ II geschrieben werden kann; ihre Randkurve γ sei so orientiert, daß B im Gegenuhrzeigersinn umlaufen wird. Weiter sei $\vec{V} \in \mathcal{C}^1(D, \mathbb{R}^2)$ ein differenzierbares

Vektorfeld auf einer offenen Teilmenge $D \supset B$ von $\mathbb{R}^2$. Dann ist

$$\int_\gamma \vec{V}\, ds = \iint_B \left(\frac{\partial V_2}{\partial x} - \frac{\partial V_1}{\partial y} \right) dx\, dy\,.$$

Beweis: Wir betrachten zunächst nur ein Vektorfeld $\vec{V} = \binom{V_1}{0}$ mit verschwindender y-Komponente und setzen B als Normalbereich

$$N = \left\{ (x,y) \in \mathbb{R}^2 \;\middle|\; a \le x \le b \quad \text{und} \quad f(x) \le y \le g(x) \right\}$$

voraus. Das Integral über die Randkurve γ berechnen wir am besten komponentenweise:

$$\gamma_1 \colon [a,b] \to \mathbb{R}^2; \quad t \mapsto \big(t, f(t)\big)$$

sei die untere Begrenzungskurve,

$$\gamma_2 \colon \big[f(b), g(b)\big] \to \mathbb{R}^2; \quad t \mapsto (b,t)$$

die rechte Randstrecke,

$$\gamma_3 \colon [a,b] \to \mathbb{R}^2; \quad t \mapsto \big(t, g(t)\big)$$

die obere Begrenzungskurve, und

$$\gamma_4 \colon \big[f(a), g(a)\big] \to \mathbb{R}^2; \quad t \mapsto \big(a, t\big)$$

schließlich die linke Seitenstrecke.

Dann ist $\dot{\gamma}_1(t) = \binom{1}{f'(t)}$, also

$$\int_{\gamma_1} \vec{V}\, ds = \int_a^b \begin{pmatrix} V_1\big(t, f(t)\big) \\ 0 \end{pmatrix} \begin{pmatrix} 1 \\ f'(t) \end{pmatrix} dt = \int_a^b V_1\big(t, f(t)\big)\, dt$$

und entsprechend

$$\int_{\gamma_3} \vec{V}\, ds = \int_a^b \begin{pmatrix} V_1\big(t, g(t)\big) \\ 0 \end{pmatrix} \begin{pmatrix} 1 \\ g'(t) \end{pmatrix} dt = \int_a^b V_1\big(t, g(t)\big)\, dt\,.$$

Für die beiden Seitenstrecken ist $\dot{\gamma}_2(t) = \binom{0}{1} = \dot{\gamma}_4(t) = \binom{0}{1}$, also ist

$$\int_{\gamma_2} \vec{V}\, ds = \int_{f(a)}^{g(a)} \begin{pmatrix} V_1(a,t) \\ 0 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} dt = 0\,,$$

und genauso folgt auch das Verschwinden des Integrals über γ_4.

Wenn wir im Gegenuhrzeigersinn um N herum integrieren, werden die Kurvenstücke γ_3 und γ_4 rückwärts durchlaufen, also ist

$$\begin{aligned}
\int_\gamma \vec{V}\, ds &= \int_{\gamma_1} \vec{V}\, ds + \int_{\gamma_2} \vec{V}\, ds - \int_{\gamma_3} \vec{V}\, ds - \int_{\gamma_4} \vec{V}\, ds \\
&= \int_a^b V_1\big(t, f(t)\big)\, dt - \int_a^b V_1\big(t, g(t)\big)\, dt \\
&= \int_a^b \Big(V_1\big(t, f(t)\big) - V_1\big(t, g(t)\big) \Big)\, dt\,.
\end{aligned}$$

Da N Normalbereich vom Typ I ist, können wir auch das Flächenintegral über N leicht ausrechnen:

$$\begin{aligned}
\iint_N \left(\frac{\partial V_2}{\partial x} - \frac{\partial V_1}{\partial y} \right) dx\, dy &= \iint_N -\frac{\partial V_1}{\partial y}\, dx\, dy \\
= -\int_a^b \left(\int_{f(x)}^{g(x)} \frac{\partial V_1}{\partial y}\, dy \right) dx &= \int_a^b \Big(V_1\big(x, g(x)\big) - V_1\big(x, f(x)\big) \Big)\, dx\,.
\end{aligned}$$

Zumindest für das hier betrachtete spezielle Vektorfeld V ohne y-Komponente und einen Normalbereich vom Typ I ist der Satz also richtig.

Damit gilt er aber zumindest für dieses spezielle Vektorfeld auch für jeden Bereich B, der sich in Normalbereiche vom Typ I zerlegen läßt: Das Flächenintegral über B ist gleich der Summe der Flächenintegrale über die Normalbereiche vom Typ I, in die wir B zerlegt haben und somit

problemlos. Die Randkurve von B besteht einerseits aus Kurvenstücken, die auch zum Rand von B gehören, andererseits aus solchen, die durch die Zerlegung eingeführt wurden und zwei Normalbereiche voneinander trennen. Letztere werden aber als Randkurven der beiden angrenzenden Normalkurven jeweils in entgegengesetzter Richtung durchlaufen, so daß sich die Integrale längs solcher Kurvenstücke gegenseitig wegheben. Summiert man also die Integrale über die Ränder aller Normalbereiche auf, bleiben am Ende nur die Integrale längs jener Kurvenstücke übrig, die auf der Randkurve von B liegen, die Summe ist somit gleich dem Integral über die Randkurve.

Damit ist der Satz bewiesen für alle Vektorfelder $\vec{V}$ mit verschwindender y-Komponente.

Als nächstes beweisen wir ihn für Vektorfelder $\vec{V}$ mit verschwindender x-Komponente; wenn man mit Normbereichen vom Typ II argumentiert statt mit solchen vom Typ I, kann man dazu die obigen Argumente fast wörtlich wiederholen.

Der Rest des Beweises ist nun einfach: Ein beliebiges Vektorfeld $\vec{V}$ läßt sich als Summe

$$\vec{V}(x,y) = \begin{pmatrix} V_1(x,y) \\ V_2(x,y) \end{pmatrix} = \begin{pmatrix} V_1(x,y) \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ V_2(x,y) \end{pmatrix}$$

schreiben, und der Satz gilt für beide Summanden. Da beide Seiten der Behauptung linear in $\vec{V}$ sind, folgt der Satz auch für $\vec{V}$ selbst. ■

GEORGE GREEN (1793–1841) war der Sohn eines Bäckers aus Nottingham. Er besuchte nur von 1801 bis 1802 eine Schule, danach arbeitete er in der Bäckerei und später in der dazugekauften Mühle. Es ist nicht bekannt, wann und wie er Mathematik lernte. 1827 veröffentlichte er sein Buch *An Essay on the Application of Mathematical Analysis to the Theory of Electricity and Magnetism,* von dem 51 Exemplare verkauft wurden, größtenteils in Nottingham selbst. Einer der Leser stellte Kontakte zur Cambridge Philosophical Society und zur Royal Academy in Edinburgh her, so daß GREENs weitere Arbeiten (über Elektrizität und über Hydrodynamik) dort veröffentlicht wurden. 1833 begann er mit dem Studium der Mathematik an der Universität Cambridge; nach dessen Abschluß blieb er in Cambridge, bis er 1840 wegen gesundheitlicher Probleme nach Nottingham zurückkehrte. Weder GREEN selbst noch seine Zeitgenossen erkannten die Bedeutung seiner Arbeiten, die, in heutiger Sprechweise, Potentialfunktionen einführten und für die

Physik nutzbar machten. Auf dieser Grundlage bauten später JAMES CLERK MAXWELL (1831–1879) und andere die Elektrodynamik auf.

Wenn wir für $\vec{V}$ speziell ein Vektorfeld der Form

$$\vec{V}(x,y) = \begin{pmatrix} -\Phi_y(x,y) \\ \Phi_x(x,y) \end{pmatrix} \quad \text{mit} \quad \Phi \in \mathcal{C}^2(D,\mathbb{R})$$

einsetzen, erhalten wir für das Flächenintegral

$$\iint\limits_B \left(\frac{\partial V_2}{\partial x} - \frac{\partial V_1}{\partial y} \right) dx\,dy = \iint\limits_B (\Phi_{xx} + \Phi_{yy})\,dx\,dy = \iint\limits_B \Delta\Phi\,dx\,dy\,,$$

wobei Δ wie üblich den LAPLACE-Operator

$$\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$$

bezeichnet.

Das Integral über die Randkurve $\gamma = (\gamma_1, \gamma_2)$ wird zu

$$\begin{aligned} &\int\limits_\gamma \begin{pmatrix} -\Phi_y\big(\gamma(t)\big) \\ \Phi_x\big(\gamma(t)\big) \end{pmatrix} \begin{pmatrix} \dot\gamma_1(t) \\ \dot\gamma_2(t) \end{pmatrix} dt \\ &= \int\limits_\gamma \Big(-\Phi_y\big(\gamma(t)\big) \cdot \dot\gamma_1(t) + \Phi_x\big(\gamma(t)\big) \cdot \dot\gamma_2(t) \Big)\,dt \\ &= \int\limits_\gamma \begin{pmatrix} \Phi_x\big(\gamma(t)\big) \\ \Phi_y\big(\gamma(t)\big) \end{pmatrix} \begin{pmatrix} \dot\gamma_2(t) \\ -\dot\gamma_1(t) \end{pmatrix} dt\,. \end{aligned}$$

Der erste Vektor im Integranden der dritten Zeile ist einfach der Gradient von Φ. Der zweite Vektor hat Skalarprodukt null mit dem Tangentenvektor $\dot\gamma(t) = \binom{\dot\gamma_1(t)}{\dot\gamma_2(t)}$, steht also auf diesem senkrecht und ist somit ein *Normalenvektor.* Das Skalarprodukt des Gradienten einer Funktion mit einem Vektor ist bekanntlich die *Richtungsableitung* der Funktion in Richtung dieses Vektors; mit

$$\vec{n} = \begin{pmatrix} \dot\gamma_2(t) \\ -\dot\gamma_1(t) \end{pmatrix} \quad \text{und} \quad \partial_{\vec{n}} = \vec{n} \cdot \operatorname{grad}$$

erhalten wir somit den

Satz von Gauß (ebener Fall)**:** Für B, D und γ wie im Satz von GREEN und $\Phi \in \mathcal{C}^2(D, \mathbb{R}^2)$ gilt

$$\iint\limits_B \Delta\Phi \; dx\, dy = \int\limits_\gamma \partial_{\vec{n}}\Phi \; ds \, .$$

■

Mit einer Verallgemeinerung sowie auch mit der anschaulichen Interpretation dieses Satzes werden wir uns in Abschnitt *f)* noch genauer befassen.

e) Oberflächenintegrale

Im dreidimensionalen Fall wird der Satz von GAUSS, wie die meisten wohl bereits aus der Physik wissen, ein Volumenintegral mit dem Fluß durch eine Oberfläche in Verbindung bringen. Was Volumenintegrale sind, haben wir inzwischen auch hier definiert; in diesem Abschnitt geht es um den Begriff des Oberflächenintegrals, um den Fluß eines Vektorfelds durch eine Oberfläche und um verwandte Themen.

Die Vorgehensweise bei der Definition eines Oberflächenintegrals entspricht genau der bei der Definition des Kurvenintegrals: Dort hatten wir Kurvenstücke als Funktionen $\gamma\colon [a,b] \to \mathbb{R}^n$ definiert; hier definieren wir entsprechend *Flächenstücke* durch Funktionen zweier Parameter. Wir müssen dabei allerdings etwas sorgfältiger vorgehen, denn im Vergleich zum Kurvenfall ist die Situation zumindest in zweierlei Hinsicht komplexer: Erstens gibt es auf der reellen Geraden im wesentlichen nur eine Art von zusammenhängenden Teilmengen, nämlich die Intervalle. In der Ebenen gibt es erheblich mehr Möglichkeiten – aber damit hatten wir uns ja schon bei der Definition von Flächen- und Volumenintegralen beschäftigt. Zweitens mußten wir bei Kurven nur fordern, daß $\dot\gamma(t)$ nirgends verschwindet, um die Existenz eines Tangentenvektors zu garantieren. Hier im Zweidimensionalen recht das nicht mehr: Ein Flächenstück hat in jedem Punkt Anspruch nicht nur auf eine Tangente, sondern auf eine ganze *Tangentialebene;* wir müssen also sicherstellen, daß es in jedem Punkt mindestens zwei linear unabhängige Tangentenvektoren gibt.

Die dazu notwendigen Definitionen kann man problemlos für Flächenstücke in jedem $\mathbb{R}^n$ hinschreiben; wir wollen uns aber, da dies für alle hier betrachteten Anwendungen ausreicht, auf den etwas anschaulicheren Fall von Flächenstücken im $\mathbb{R}^3$ beschränken.

Ein Flächenstück soll demnach gegeben sein durch eine Funktion

$$f\colon B \to \mathbb{R}^3; \quad (u,v) \mapsto f(u,v) = \begin{pmatrix} x(u,v) \\ y(u,v) \\ z(u,v) \end{pmatrix}$$

auf einer Teilmenge $B \subset \mathbb{R}^2$; mit Blick auf unsere Erfahrungen bei Kurvenstücken wollen wir von vornherein fordern, daß f differenzierbar sein soll in einer offenen Umgebung von B. Für kleine reelle Zahlen h, k ist dann für einen Punkt $(u_0, v_0) \in B$

$$f(u_0+h, v_0+k) = f(u_0,v_0) + f_u(u_0,v_0)\cdot h + f_v(u_0,v_0)\cdot k + o\left(\sqrt{h^2+k^2}\right) \, .$$

Da hier nur h und k variabel sind, bedeutet die Forderung nach zwei linear unabhängigen Tangentialvektoren, daß die beiden Vektoren $f_u(u_0,v_0)$ und $f_v(u_0,v_0)$ in allen Punkten $(u_0,v_0) \in B$ linear unabhängig sein müssen.

An dieser Stelle führt die Beschränkung auf den $\mathbb{R}^3$ zu einer kleinen Vereinfachung: Im $\mathbb{R}^3$, und nur dort, existiert ein Vektorprodukt; zwei Vektoren sind genau dann linear unabhängig, wenn dieses Vektorprodukt ungleich dem Nullvektor ist. Sowohl der Betrag als auch die Richtung dieses Vektors werden uns schon bald in mehrfacher Hinsicht nützlich sein; wir definieren daher

Definition: Ein reguläres Flächenstück im $\mathbb{R}^3$ ist gegeben durch einen Bereich $B \subset \mathbb{R}^2$ und eine in einer offenen Menge $D \supseteq B$ differenzierbaren Funktion $f\colon D \to \mathbb{R}^3; \quad (u,v) \mapsto f(u,v)$, für die gilt:
1.) Die Einschränkung $f_{|B}$ von f auf B ist injektiv.
2.) In jedem Punkt $(u_0,v_0) \in B$ ist $f_u(u_0,v_0) \times f_v(u_0,v_0) \neq \vec{0}$.

Die wichtigste Oberfläche, mit der wir es im folgenden zu tun haben werden, ist die der Kugel. Die Oberfläche einer Kugel vom Radius R hat in Kugelkoordinaten die Gleichung $r = R$ mit beliebigen Winkelkoordinaten φ und ϑ, sie ist also (in kartesischen Kordinaten) gegeben

durch die Funktion

$$f\colon [0,2\pi)\times[0,\pi]\to\mathbb{R}^3;\quad (\varphi,\vartheta)\mapsto\begin{pmatrix} R\cos\varphi\sin\vartheta\\ R\sin\varphi\sin\vartheta\\ R\cos\vartheta\end{pmatrix}.$$

Die Differenzierbarkeit ist hier überhaupt kein Problem: Die angegebene Funktion ist sogar auf ganz $\mathbb{R}^2$ beliebig oft stetig differenzierbar.

Bei der Injektivität gibt es allerdings Probleme: Für $\vartheta = 0$ und $\vartheta = \pi$ ist $\sin\vartheta = 0$; daher werden alle Parameterpaare $(\varphi, 0)$ auf $(0,0,R)$ und aller Parameterpaare (φ,π) auf $(0,0,-R)$ abgebildet.

Das ist aber glücklicherweise auch bereits alles was passieren kann, denn falls $\sin\vartheta \neq 0$ ist, können wir aus der dritten Komponenten $R\cos\vartheta$ den Winkel $\vartheta\in[0,\pi]$ eindeutig bestimmen. Damit kennen wir auch $R\sin\vartheta$ und bekommen aus den ersten beiden Komponenten von f den Cosinus und den Sinus von φ, wodurch $\varphi\in[0,2\pi)$ eindeutig festgelegt ist.

Die partiellen Ableitungen von f sind

$$f_\varphi = \begin{pmatrix} -R\sin\varphi\sin\vartheta\\ R\cos\varphi\sin\vartheta\\ 0\end{pmatrix}\quad\text{und}\quad f_\vartheta = \begin{pmatrix} R\cos\varphi\cos\vartheta\\ R\sin\varphi\cos\vartheta\\ -R\sin\vartheta\end{pmatrix},$$

also ist

$$f_\varphi\times f_\vartheta = \begin{pmatrix} R^2\cos\varphi\sin^2\vartheta\\ R^2\sin\varphi\sin^2\vartheta\\ R^2\sin\vartheta\cos\vartheta\end{pmatrix} = R\sin\vartheta\cdot f(\varphi,\vartheta).$$

Da für eine echte Kugel $R>0$ sein muß und $f(\varphi,\vartheta)$ die Länge R hat, ist dieser Vektor genau dann gleich dem Nullvektor, wenn $\sin\vartheta = 0$ ist; wie bei der Injektivität gibt es also wieder Probleme an den Polen, aber auch nur dort.

Somit können wir die Kugel nur dann als reguläres Flächenstück auffassen, wenn wir die beiden Pole herausnehmen, d.h. wenn wir uns auf den Parameterbereich $[0,2\pi)\times(0,\pi)$ beschränken.

Für die meisten praktischen Zwecke ist das natürlich völlig unproblematisch: Bei der Flächenbestimmung oder allgemeiner der Integration einer beschränkte Funktion sind einzelne, isoliert liegende Punkte irrelevant. Lediglich im Falle von uneigentlichen Integranden muß man hier sehr vorsichtig sein; hier in dieser Vorlesung sollen entsprechende Integrale deshalb vorsichtshalber gleich gar nicht erst definiert werden.

Nachdem wir nun wissen, was ein reguläres Flächenstück ist, müssen wir als nächstes lernen, damit zu rechnen. Bevor wir uns an die Bestimmung von Flächeninhalten machen, ist es vielleicht ganz nützlich (wenn auch nicht unbedingt notwendig), daß wir uns ein paar Gedanken über die *Längenmessung* machen.

Längen sind Eigenschaften von Kurven; wir müssen also eine Kurve oder – da wir Längen auch stückweise aneinandersetzen können – einfacher ein Kurvenstück auf einem Flächenstück betrachten. Genau wie man etwa bei der Navigation auf der Erde nicht von einem dreidimensionalen kartesischen Koordinatensystem ausgeht, sondern von der geographischen Länge und Breite, empfiehlt es sich auch hier, Kurven über die Parameter des Flächenstücks zu definieren – und sei es auch nur, um sicher zu sein, daß die Kurve auch wirklich auf dem Flächenstück liegt.

Wir beschreiben eine Kurve auf einem Flächenstück $f\colon B\to\mathbb{R}^3$ daher durch eine Funktion

$$\delta\colon [a,b]\to B;\quad t\mapsto\big(u(t),v(t)\big);$$

die eigentliche Kurve im $\mathbb{R}^3$ wird dann beschrieben durch die zusammengesetzte Abbildung

$$\gamma = f\circ\delta\colon [a,b]\to\mathbb{R}^3;\quad t\mapsto f\big(u(t),v(t)\big).$$

Wir wollen natürlich, daß diese Kurve immer dieselbe Länge hat, egal ob wir sie einfach so im $\mathbb{R}^3$ oder auf einem Flächenstück betrachten. Die Länge des Kurvenstücks γ auf dem Flächenstück f ist daher gleich der bekannten Länge des Kurvenstücks γ im $\mathbb{R}^3$, also

$$\int_a^b |\dot\gamma(t)|\,dt = \int_a^b \left| f_u\big(u(t),v(t)\big)\frac{du}{dt} + f_v\big(u(t),v(t)\big)\frac{dv}{dt}\right| dt.$$

Der Betrag des Vektors $f_u\big(u(t),v(t)\big)\frac{du}{dt} + f_v\big(u(t),v(t)\big)\frac{dv}{dt}$ ist gleich der Wurzel aus dem Skalarprodukt des Vektors mit sich selbst, also der

Wurzel aus

$$
\begin{aligned}
& f_u\big(u(t),v(t)\big)^2 \left(\frac{du}{dt}(t)\right)^2 \\
& + 2 f_u\big(u(t),v(t)\big)\cdot f_v\big(u(t),v(t)\big)\ \left(\frac{du}{dt}(t)\right)\left(\frac{dv}{dt}(t)\right) \\
& + f_v\big(u(t),v(t)\big)^2 \left(\frac{dv}{dt}\right)^2 .
\end{aligned}
$$

In dieser Formel können wir zwei Arten von Termen unterscheiden: Rechts stehen jeweils die Ableitungen von u und v nach t; diese hängen nur von der jeweiligen Kurve ab, die ja gerade durch die Funktionen $u(t)$ und $v(t)$ definiert ist. Links dagegen stehen Ausdrücke in den partiellen Ableitungen von f nach u und v; diese sind Funktionen auf dem Flächenstück und sind insbesondere unabhängig von jeder Kurve. Wir bezeichnen daher diese drei letzteren Größen

$$
\begin{aligned}
E(u,v) &= f_u(u,v)\cdot f_u(u,v)\,, \\
F(u,v) &= f_u(u,v)\cdot f_v(u,v) \quad \text{und} \\
G(u,v) &= f_v(u,v)\cdot f_v(u,v)
\end{aligned}
$$

als *Fundamentalgrößen* des Flächenstücks f.

Um etwas Übung im Umgang mit diesen Größen zu bekommen, wollen wir sie für die Kugeloberfläche in ihrer oben angegebenen Parametrisierung berechnen. Hier ist, wie wir schon wissen,

$$
f_\varphi = \begin{pmatrix} -R\sin\varphi\sin\vartheta \\ R\cos\varphi\sin\vartheta \\ 0 \end{pmatrix} \quad \text{und} \quad f_\vartheta = \begin{pmatrix} R\cos\varphi\cos\vartheta \\ R\sin\varphi\cos\vartheta \\ -R\sin\vartheta \end{pmatrix} ;
$$

also ist

$$
\begin{aligned}
E(\varphi,\vartheta) &= R^2(\sin^2\varphi\sin^2\vartheta + \cos^2\varphi\sin^2\vartheta) = R^2\sin^2\vartheta \\
F(\varphi,\vartheta) &= R^2(-\sin\varphi\sin\vartheta\cos\varphi\cos\vartheta + \cos\varphi\sin\vartheta\sin\varphi\cos\vartheta) = 0 \\
G(\varphi,\vartheta) &= R^2(\cos^2\varphi\cos^2\vartheta + \sin^2\varphi\cos^2\vartheta + \sin^2\vartheta) = R^2 .
\end{aligned}
$$

Dieses Ergebnis sollte man auch geometrisch interpretieren: Das Verschwinden von F, die Orthogonalität von f_φ und f_ϑ also, bedeutet,

daß die Längenkreise und die Breitenkreise in jedem Punkt der Kugel (außer den beiden Polen) aufeinander senkrecht stehen; die Konstans von G kommt daher, daß alle Längenkreise gleich lang sind, und die ϑ-Abhängigkeit von E schließlich reflektiert die Tatsache, daß die Breitenkreise verschiedene Radien haben.

Allgemein kann man die Länge einer Kurve auf einem Flächenstück somit als Integral

$$
\int_a^b \sqrt{E\,\dot u^2 + 2F\,\dot u\dot v + G\,\dot v^2}\,dt
$$

berechnen; dem interessierten Leser sei empfohlen, damit beispielsweise überprüfen, daß Breitenkreise i.a. nicht die kürzeste Verbindung zwischen zweien ihrer Punkte sind, sondern daß der Kreis um den Kugelmittelpunkt durch diese beiden Punkte (der sogenannte Großkreis) eine kürzere Verbindungskurve liefert.

Unser Hauptziel hier sind allerdings nicht Längenberechnungen, sondern die Berechnung von Oberflächen *bzw.* von Flüssen von Vektorfeldern durch diese Oberflächen. Dabei gehen wir genauso vor, wie bei der Transformationsformel im vorigen Abschnitt: Wir unterteilen den Parameterbereich B durch Rechtecke, und betrachten deren Bilder auf dem Flächenstück. Diese sind, da wir f als differenzierbar vorausgesetzt haben, in erster Näherung Parallelogramme, wobei die Näherung bei zunehmender Verfeinerung des Rechteckgitters auf B immer besser wird. Der Flächeninhalt eines solchen Parallelogramms mit Kantenvektoren $f_u\cdot h$ und $f_v\cdot k$ ist

$$
hk\cdot|f_u\times f_v|\ ,
$$

also sollte die Fläche des Flächenstücks gleich

$$
\iint_B |f_u\times f_v|\ du\,dv
$$

sein, und genau das definieren wir auch als den Flächeninhalt eines Flächenstücks.

Definition: Der Flächeninhalt eines Flächenstücks $f\colon B \to \mathbb{R}^3$ ist

$$\iint\limits_f dO \underset{\text{def}}{=} \iint\limits_B |f_u \times f_v| \, du \, dv \, .$$

(Das O in dO soll dabei an Oberfläche erinnern.)

Zur Kontrolle, ob diese Definition sinnvoll sein kann, berechnen wir die Oberfläche der Kugel: Hier ist, wie wir bereits wissen,

$$f_\varphi \times f_\vartheta = \begin{pmatrix} R^2 \cos\varphi \sin^2\vartheta \\ R^2 \sin\varphi \sin^2\vartheta \\ R^2 \sin\vartheta \cos\vartheta \end{pmatrix} = R \sin\vartheta \cdot f(\varphi, \vartheta) \, ;$$

die Länge dieses Vektors ist also, da $f(\varphi, \vartheta)$ als Radiusvektor natürlich die Länge R haben muß, gleich $R^2 \sin\vartheta$. Die Oberfläche der Kugel berechnet sich daher zu

$$\iint\limits_B R^2 \sin\vartheta = \int\limits_0^{2\pi} \left(\int\limits_0^{\pi} R^2 \sin\vartheta \, d\vartheta \right) d\varphi$$
$$= 2\pi \cdot R^2 \cdot (-\cos\pi + \cos 0) = 4\pi R^2 \, ,$$

womit wir uns wieder einmal in hundertprozentiger Übereinstimmung mit der Schulmathematik befinden.

Genau wie Kurvenintegrale sind auch Oberflächenintegrale unabhängig von der Parametrisierung des Flächenstücks; im wesentlichen geht alles genau wie bei den Kurvenintegralen, ist aber etwas aufwendiger. Der Vollständigkeit halber sei die entsprechende Aussagen samt Beweis im Kleindruck abgedruckt:

Lemma: $f\colon B \to \mathbb{R}^3$ und $g\colon B' \to \mathbb{R}^3$ seien zwei Flächenstücke, für die es eine in einer Umgebung von B stetig differenzierbare Funktion φ gebe, die B bijektiv auf B' abbildet derart, daß $g = f \circ \varphi$ ist. Dann ist

$$\iint\limits_f du \, dv = \iint\limits_g du \, dv \, .$$

Beweis: Nach Definition ist

$$\iint\limits_f dO = \iint\limits_B |f_u \times f_v| \, du \, dv \quad \text{und} \quad \iint\limits_g dO = \iint\limits_B |g_u \times g_v| \, du \, dv \, .$$

Wenn wir die drei Komponenten von f *bzw.* g mit $f^{(x)}, f^{(y)}, f^{(z)}$ *bzw.* $g^{(x)}, g^{(y)}, g^{(z)}$ bezeichnen, ist

$$f_u \times f_v = \begin{pmatrix} f_u^{(y)} f_v^{(z)} - f_u^{(z)} f_v^{(y)} \\ f_u^{(z)} f_v^{(x)} - f_u^{(x)} f_v^{(z)} \\ f_u^{(x)} f_v^{(y)} - f_u^{(y)} f_v^{(x)} \end{pmatrix} = \begin{pmatrix} \det \begin{pmatrix} f_u^{(y)} & f_v^{(y)} \\ f_u^{(z)} & f_v^{(z)} \end{pmatrix} \\ \det \begin{pmatrix} f_u^{(z)} & f_v^{(z)} \\ f_u^{(x)} & f_v^{(x)} \end{pmatrix} \\ \det \begin{pmatrix} f_u^{(x)} & f_v^{(x)} \\ f_u^{(y)} & f_v^{(y)} \end{pmatrix} \end{pmatrix} ,$$

und genau entsprechend natürlich auch für g. Die links stehenden Determinanten sind offensichtlich gerade die Determinaten der JACOBI-Matrizen der verschiedenen zweidimensionalen Projektionen von f; an der ersten Stelle etwa die der Funktion

$$f^{(yz)} \underset{\text{def}}{=} \begin{pmatrix} f^{(y)} \\ f^{(z)} \end{pmatrix} : B \to \mathbb{R}^2 \, .$$

Analog dazu definieren wir auch Funktionen $f^{(zx)}, f^{(xy)}$ und die entsprechenden Funktionen für g.

Aus $g = f \circ \varphi$ folgt, daß auch $g^{(x)} = f^{(x)} \circ \varphi$ und entsprechend für die anderen Komponenten. Damit ist auch $f^{(yz)} = g^{(yz)} \circ \varphi$ *usw.*

Nach der zweidimensionalen Kettenregel ist dann

$$J_{g^{(yz)}} = \left(J_{f^{(yz)}} \circ \varphi \right) \cdot J_\varphi \, ,$$

also ausgeschrieben

$$\begin{pmatrix} g_u^{(y)} & g_v^{(y)} \\ g_u^{(z)} & g_v^{(z)} \end{pmatrix} = \begin{pmatrix} f_u^{(y)} \circ \varphi & f_v^{(y)} \circ \varphi \\ f_u^{(z)} \circ \varphi & f_v^{(z)} \circ \varphi \end{pmatrix} \cdot J_\varphi \, .$$

Entsprechendes gilt auch für die anderen Indexkombinationen, und nach dem Multiplikationssatz für Determinanten gilt eine entsprechende Produktbeziehung auch für die Determinaten der hier stehenden Matrizen; insgesamt erhalten wir also, daß

$$g_u \times g_v = \Big((f_u \circ \varphi) \times (f_v \circ \varphi) \Big) \cdot \det J_\varphi$$

ist und dementsprechend

$$|g_u \times g_v| = \Big((f_u \circ \varphi) \times (f_v \circ \varphi) \Big) \cdot |\det J_\varphi| \, .$$

Damit folgt die Behauptung aus der Transformationsformel. ■

Die Berechnung von Vektorprodukten ist etwas umständlich und sehr anfällig für Vorzeichenfehler; wir wollen daher sehen, daß die Fundamentalgrößen einer Fläche ihrem Namen gerecht werden und uns auch diese Berechnung abnehmen können: Der Flächeninhalt eines Parallelogramms mit Kantenvektoren $\vec{a}$ und $\vec{b}$ ist gleich dem Produkt der Längen

der beiden Vektoren mal dem Sinus des eingeschlossenen Winkels α. Das Quadrat des Flächeninhalts ist also

$$(\vec{a}\cdot\vec{a})(\vec{b}\cdot\vec{b})\sin^2\alpha = (\vec{a}\cdot\vec{a})(\vec{b}\cdot\vec{b})(1-\cos^2\alpha)$$
$$= (\vec{a}\cdot\vec{a})(\vec{b}\cdot\vec{b}) - (\vec{a}\cdot\vec{a})(\vec{b}\cdot\vec{b})\cos^2\alpha = (\vec{a}\cdot\vec{a})(\vec{b}\cdot\vec{b}) - (\vec{a}\cdot\vec{b})^2\,,$$

und der Flächeninhalt selbst somit

$$\sqrt{(\vec{a}\cdot\vec{a})(\vec{b}\cdot\vec{b}) - (\vec{a}\cdot\vec{b})^2}\,.$$

Wir interessieren uns für das Parallelogramm mit Kantenvektoren $h\cdot f_u$ und $k\cdot f_v$; hier wird diese Formel zu

$$hk\cdot\sqrt{(f_u\cdot f_u)(f_v\cdot f_v) - (f_u\cdot f_v)^2} = hk\sqrt{EG-F^2}\,.$$

Die Fläche eines Flächenstücks kann also auch berechnet werden als

$$\iint\limits_B \sqrt{EG-F^2}\,du\,dv\,.$$

Im konkreten Fall steht hier natürlich genau dasselbe Integral wie bei der Formel mit dem Betrag des Vektorprodukts, allerdings erfordert diese Formel, sofern man die Fundamentalgrößen einer Fläche bereits kennt, deutlich weniger Rechenaufwand. Dem Leser sei empfohlen, sich am Beispiel der Kugeloberfläche hiervon zu überzeugen!

Obwohl wir uns auf Flächen im $\mathbb{R}^3$ beschränken wollen, sei hier zumindest kurz darauf hingewiesen, daß wir bei der Herleitung dieser neuen Darstellung des Flächeninhalts nirgends benutzen mußten, daß wir im $\mathbb{R}^3$ sind; diese Formel gilt also auch für Flächenstücke im $\mathbb{R}^n$ mit $n>3$.

Eine für uns wichtige Besonderheit, die *nur* im $\mathbb{R}^3$ gilt, ist dagegen die Tatsache, daß wir jedem Flächenstück eine eindeutig bestimmte Normalenrichtung zuordnen können: Im Dreidimensionalen gibt es nur eine Richtung, die auf der Tangentialebenen eines Flächenstücks senkrecht steht; in Dimension $n>3$ gibt es dagegen einen ganzen $(n-2)$-dimensionalen Raum mit dieser Eigenschaft.

Wenn wir von der Parameterdarstellung $f\colon D\to B$ eines regulären Flächenstücks im $\mathbb{R}^3$ und einer festen Reihenfolge der Parameter ausgehen, können wir sogar unterscheiden, nach welcher Seite diese Richtung

sich von der Fläche entfernt: Die Tangentialebene wird aufgespannt von den partiellen Ableitungen f_u und f_v von f nach den beiden Parametern u,v auf $D\subseteq\mathbb{R}^2$, und das Kreuzprodukt $f_u\times f_v$ liefert einen Vektor, der auf diesen beiden Tangentialvektoren und somit der gesamten Tangentialebenen senkrecht steht.

Damit können wir jedem Punkt eines *regulären* Flächenstücks eine eindeutig bestimmte Normalenrichtung zuordnen. Man beachte, daß dies nur möglich ist aufgrund der Forderungen, die wir an ein reguläres Flächenstück gestellt haben: Die Abbildung f muß injektiv sein, so daß die Parameterwerte (u,v) zu jedem Punkt des Flächenstücks eindeutig bestimmt sind, und $f_u\times f_v$ darf nirgends verschwinden, da der Nullvektor keine wohlbestimmte Richtung hat.

Typisches Beispiel eines nichtregulären Flächenstücks ist das MÖBIUS-Band

$$f\colon D=[0,1]\times[0,2\pi)\to\mathbb{R}^3;\quad (u,v)\mapsto\begin{pmatrix}(2+u\cos v)\cos 2v\\(2+u\cos v)\sin 2v\\u\sin v\end{pmatrix}\,,$$

das durch Zusammenkleben der Enden eines einmal verdrillten rechteckigen Streifens entsteht. Hier ist f für $u=0$ nicht injektiv, denn $f(0,v)=f(0,v+\pi)$; wir haben also kein reguläres Flächenstück. In der Tat wir für die Parameterwerte $(0,v)$ der Mittelkreis des Bandes zweimal durchlaufen. Dort ist

$$f_u(0,v)=\begin{pmatrix}\cos v\cos 2v\\\cos v\sin 2v\\\sin v\end{pmatrix}\quad\text{und}\quad f_v(0,v)=\begin{pmatrix}-4\sin 2v\\4\cos 2v\\0\end{pmatrix}\,,$$

also

$$f_u(0,v)\times f_v(0,v)=\begin{pmatrix}-4\sin v\cos 2v\\-4\sin v\sin 2v\\4\cos v\end{pmatrix}$$

und damit

$$f_u(0,v)\times f_v(0,v)=-f_u(0,v+\pi)\times f_v(0,v+\pi)\,.$$

es gibt also auf dem Mittelkreis keine eindeutig bestimmte Normalenrichtung.

Für Kurven hatten wir einerseits RIEMANN-STIELTJES-Integrale definiert, die für eine beliebige Funktion auf der Kurve erklärt sind und in deren Definition die Länge des Tangentenvektors eingeht; andererseits hatten wir Integrale über Vektorfelder, in deren Definition der Tangentenvektor selbst einging. Genauso gehen wir auch hier bei den Oberflächenintegralen vor, nur daß wir jetzt den Normalenvektor anstelle des Tangentenvektors benutzen:

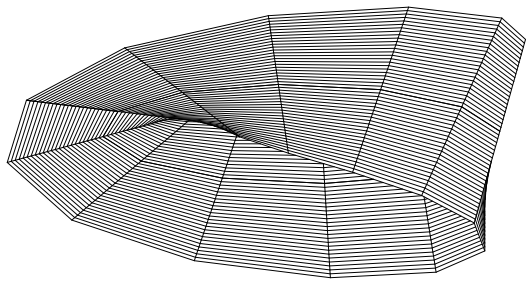


Abb. 76: Das Möbiusband

Definition: $f\colon D \to \mathbb{R}^3$ sei ein reguläres Flächenstück.
a) Ist g eine auf $f(D)$ definierte stetige Funktion, so bezeichnen wir

$$\iint\limits_f g\,dO \underset{\text{def}}{=} \iint\limits_D g\big(f(u,v)\big)\ \big|f_u(u,v) \times f_v(u,v)\big|\ du\,dv$$

als *Oberflächenintegral* von g über f.
b) Ist $\vec{V}$ ein in einer Umgebung von $f(D)$ definiertes Vektorfeld, so bezeichnen wir

$$\iint\limits_f \vec{V} \cdot d\vec{O} \underset{\text{def}}{=} \iint\limits_D \vec{V}\big(f(u,v)\big) \cdot \big(f_u(u,v) \times f_v(u,v)\big)\,du\,dv$$

als den *Fluß* des Vektorfelds $\vec{V}$ durch die Oberfläche f.

Eine unproblematische Verallgemeinerung des obigen Lemmas zeigt, daß auch diese Integrale im dort definierten Sinne unabhängig von der Parametrisierung sind.

Der Name *Fluß* für das unter *b)* definierte Integral wird klar, wenn man sich f etwa als eine Kugeloberfläche vorstellt. In jedem Punkt $\boldsymbol{x}$ auf dieser Oberfläche kann man ein kartesisches Koordinatensystem aus einem Normaleneinheitsvektor $\vec{n}_{\boldsymbol{x}}$ und zwei Tangenteneinheitsvektoren

verankern. Drückt man nun den Vektor $\vec{V}(\boldsymbol{x})$ in diesem Koordinatensystem aus, so beschreibt die $\vec{n}_{\boldsymbol{x}}$-Komponente jenen Teil des Vektors, der durch die Kugeloberfläche hindurch nach innen oder außen geht (welches von beiden hängt ab vom Vorzeichen der Komponente und von der Orientierung des Normaleneinheitsvektors), während die beiden anderen Komponenten den Teil beschreiben, der auf der Oberfläche bleibt, also sozusagen den Fluß *auf* der Oberfläche im Gegensatz zum Fluß *durch* die Oberfläche. Letzterer läßt sich berechnen als das Skalarprodukt $\vec{V}(\boldsymbol{x}) \cdot \vec{n}_x$ mit dem Normaleneinheitsvektor, und das Integral über diese Funktion ist

$$\iint\limits_f \big(\vec{V} \cdot \vec{n}_{\boldsymbol{x}}\big)\,dO = \iint\limits_f \vec{V} \cdot d\vec{O}\,,$$

denn

$$f_u \times f_v = |f_u \times f_v|\ \vec{n}_{\boldsymbol{x}}\,.$$

f) Die Sätze von Stokes und Gauß

In diesem Abschnitt geht es um drei der zentralsten Sätze der mehrdimensionalen Analysis: Außer den beiden im Titel erwähnten Sätzen soll noch das mehrdimensionale Analogon des Hauptsatzes der Differential- und Integralrechnung bewiesen werden.

Wir beginnen mit letzterem sowie dem Satz von STOKES: Bei beiden geht es darum, die Zirkulation eines Vektorfelds zu bestimmen. Beide Sätze gelten in beliebiger Dimension, jedoch wollen wir sie der Einfachheit halber nur für den $\mathbb{R}^3$ beweisen.

Ausgangspunkt ist ein reguläres Flächenstück $f\colon D \to \mathbb{R}^3$; dessen Bild $B = f(D)$, das „eigentliche“ geometrische Flächenstück also, sei beschränkt und habe eine reguläre Kurve $\gamma\colon [a,b] \to \mathbb{R}^3$ als Rand. Wie schon mehrfach erwähnt, wollen wir Kurvenintegrale längs γ mit Oberflächenintegralen über f oder – wie wir wegen der Parameterunabhängigkeit auch einigermaßen korrekt sagen können – B in Verbindung bringen.

Wir gehen also aus von einem Vektorfeld $\vec{V}\colon G \to \mathbb{R}^3$, dessen (offener) Definitionsbereich G sowohl B als auch γ enthält, und wollen Informa-

tionen über das Kurvenintegral

$$\int_\gamma \vec{V}\, ds\,.$$

Dazu approximieren wir B durch ein Flächenstück B^*, das wir aus endlich vielen ungefähr rechteckförmigen Teilmengen R_i zusammensetzen können; ein grobes Bild einer solchen Unterteilung ist in Abbildung 77 zu sehen.

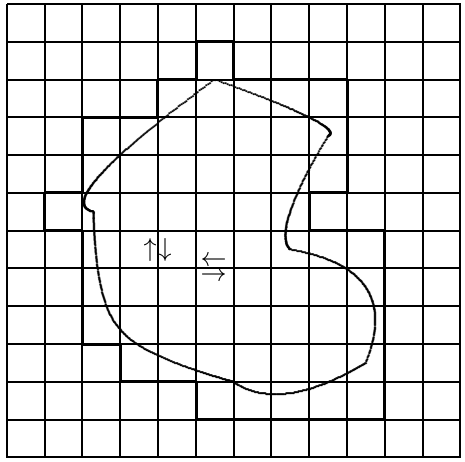


Abb. 77: Unterteilung eines Bereichs

Da f als reguläres Flächenstück vorausgesetzt war, ist überall in B ein wohldefinierter Normalenvektor bestimmt; damit können wir auch für jedes der Rechtecke einen Umlaufsinn festlegen: Dieser soll, wenn man von der Spitze des Normalenvektors aus auf des Rechteck schaut, der Gegenuhrzeigersinn sein. Da der Normalenvektor eines regulären Flächenstücks stetig von den Parametern der Fläche abhängt und nirgends verschwindet, ist es gleichgültig, über welchem Punkt eines (hinreichend kleinen) Rechtecks wir ihn betrachten.

In Abbildung 77 wurde davon ausgegangen, daß alle Normalenvektoren nach oben zeigen; der entsprechende Umlaufsinn ist für einige der

Rechteckkanten eingezeichnet. Wie man sieht, wird jede gemeinsame Kante zweier benachbarter Rechtecke von diesen beiden Rechtecken in verschiedener Weise orientiert. Falls wir daher für jedes Rechteck R_i das Wegintegral entlang seines Rands berechnen und alle diese Integrale aufaddieren, bleiben nur die Integrale längs der in Abbildung 77 fett ausgezogenen Kanten erhalten, d.h.

$$\sum_{i=1}^{M} \int_{\partial R_i} \vec{V}\, ds = \int_{\partial B^*} \vec{V}\, ds\,,$$

wobei M die Anzahl der „Rechtecke“ R_i bezeichnet. Natürlich ist erst recht für jede Funktion h auf B^*

$$\sum_{i=1}^{M} \iint_{R_i} h\, dO = \iint_{B^*} h\, dO\,,$$

wenn wir eine Formel für die einzelnen Rechtecke R_i haben, gilt diese also automatisch auch für B^* und somit, nach einem Grenzübergang, auch für B.

Betrachten wir also ein festes Rechteck R_i und das Kurvenintegral entlang seines Umfangs. Da wir die Rechtecke als klein voraussetzen, können wir $\vec{V}(\boldsymbol{x})$ auf R_i ohne großen Fehler linearisieren.

Dazu sei $\boldsymbol{x}_i \in R_i$ ein beliebiger Punkt des Rechtecks. Ein beliebiger Punkt $\boldsymbol{x} \in R_i$ hat dann die Form $\boldsymbol{x} = \boldsymbol{x}_i + \vec{h}$ mit einem nicht allzu großen Vektor $\vec{h}$. Nach Definition der Differenzierbarkeit ist dann

$$\begin{aligned}\vec{V}(\boldsymbol{x}) &= \vec{V}(\boldsymbol{x}_0 + \vec{h}) = \vec{V}(x_0) + J_{\vec{V}}(\boldsymbol{x}_0)\vec{h} + o(\|\vec{h}\|)\\ &= \vec{V}(\boldsymbol{x}_0) + S\cdot\vec{h} + \tfrac{1}{2}\operatorname{rot}\vec{V}(\boldsymbol{x}_0)\times h + o(\|\vec{h}\|)\,,\end{aligned}$$

wobei S für den symmetrischen Anteil der JACOBI-Matrix $J_{\vec{V}}(\boldsymbol{x}_0)$ steht.

Wenn wir den Term $o(\|\vec{h}\|)$ vernachlässigen, ist also

$$\int\limits_{\partial R_i} \vec{V}(\boldsymbol{x})\,ds = \int\limits_{\partial R} \vec{V}(\boldsymbol{x}_0 + \vec{x})\,ds$$

$$\approx \int\limits_{\partial R} \vec{V}(\boldsymbol{x}_0)\,ds + \int\limits_{\partial R} S\vec{x}\,ds + \tfrac{1}{2}\int\limits_{\partial R} \operatorname{rot}\vec{V}(\boldsymbol{x}_0) \times \vec{x}\,ds\,, \qquad (*)$$

wobei R das um den negativ genommenen Ortsvektor von $\boldsymbol{x}_0$ verschobene Rechteck R_i sei, d.h., wenn $\boldsymbol{x}$ den Rand von R durchläuft, durchläuft $\boldsymbol{x}_0 + \boldsymbol{x}$ den Rand von R_i.

Das erste Integral in dieser Summe ist ein Kurvenintegral über das konstante Vektorfeld, das jedem Punkt den Vektor $\vec{V}(\boldsymbol{x}_0)$ zuordnet. Dieses Vektorfeld ist offensichtlich ein Potentialfeld, denn sind V_1, V_2 und V_3 die Komponenten dieses Vektors, so ist

$$\operatorname{grad}\big(\vec{V}(\boldsymbol{x}_0)\cdot\boldsymbol{x}\big) = \operatorname{grad}(V_1x + V_2y + V_3z) = \vec{V}(x_0)\,.$$

Genauso ist auch $S\vec{h}$ ein Potentialfeld, denn der Gradient der quadratischen Form

$$\begin{aligned} {}^t\vec{x}S\vec{x} &= S_{11}x^2 + S_{22}y^2 + S_{33}z^2 \\ &\quad + (S_{12}+S_{21})xy + (S_{13}+S_{31})xz + (S_{23}+S_{32})yz \end{aligned}$$

ist wegen $S_{k\ell} = S_{\ell k}$ gleich dem Zweifachen dieses Vektorfelds.

Damit sind also die beiden Vektorfelder $\vec{V}(x_0)$ und $S\vec{x}$ zirkulationsfrei, und die Integrale über den (als Kurve geschlossenen) Rand von R_i verschwinden.

Damit haben wir (bis auf die hier unterdrückten, für einen richtigen Beweis aber unbedingt notwendige Abschätzung der Fehlerterme) den folgenden Satz bewiesen:

Satz: Das Vektorfeld $\vec{V}\colon G \to \mathbb{R}^3$ habe eine symmetrische JACOBI-Matrix, d.h. $\operatorname{rot}\vec{V}$ sei identisch null. Dann gilt für jedes reguläre

Flächenstück $f\colon D \to G$, dessen Rand eine Kurve γ ist,

$$\int\limits_{\gamma} \vec{V}\,ds = 0\,.$$

Denn nach obiger Rechnung verschwindet bei symmetrischer JACOBI-Matrix das Kurvenintegral entlang eines jeden Rechtecks, und die Summe all dieser Kurvenintegrale konvergiert bei immer feinerer Rechteckunterteilung des Flächenstücks gegen das Kurvenintegral längs des Rands γ von $B = f(D)$.

Dies reicht schon für den Hauptsatz der Differential- und Integralrechnung im $\mathbb{R}^3$: Wir wissen bereits, daß ein Vektorfeld genau dann eine Stammfunktion hat, wenn es zirkulationsfrei ist – die eine Richtung dieser Aussage haben wir gerade wieder angewandt. Außerdem wissen wir, daß für ein Vektorfeld mit Stammfunktion die Rotation verschwindet, denn für eine zweimal stetig differenzierbare Funktion φ ist $\operatorname{rot}\operatorname{grad}\varphi = 0$. Für ein leicht nachprüfbares Kriterium zur Existenz einer Stammfunktion fehlt also nur noch die Aussage, daß aus dem Verschwinden der Rotation die Zirkulationsfreiheit folgt.

Diese Aussage ist aber leider falsch: Das Beispiel des Magnetfelds eines stromdurchflossenen Leiters zeigte, daß die Rotation sehr wohl identisch verschwinden kann, ohne daß das Vektorfeld zirkulationsfrei ist. Dieses Magnetfeld $\vec{V}(x,y,z) = \frac{c}{x^2+y^2}\binom{-y}{x}$ ist allerdings auf der z-Achse nicht definiert, und wir hatten im Beispiel Kreise um die z-Achse betrachtet.

Bei einem überall definiertes Vektorfeld hätten wir für einen solchen Kreis einfach die geschlossene Kreisscheibe B betrachten können, deren Rand der betrachtete Integrationsweg ist, und nach obigem Satz wäre das Integral längs des Randes verschwunden.

Das Problem bei diesem Gegenbeispiel liegt also offensichtlich darin, daß der betrachtete Integrationsweg nicht als Rand eines regulären Flächenstücks im Definitionsbereich des Vektorfelds geschrieben werden kann: Da das Vektorfeld auf der z-Achse nicht definiert ist, muß der Schnittpunkt mit der z-Achse aus der Kreisscheibe herausgenommen

werden, wir haben also nur noch einen punktierten Kreis, und dessen Rand besteht aus der Kreislinie *plus* dem herausgenommenen Punkt.

Um solche Fälle auszuschließen definieren wir:

Definition: Eine offene zusammenhängende Teilmenge $G \subseteq \mathbb{R}^3$ heißt *einfach zusammenhängend,* wenn jede geschlossene Kurve γ in G Rand eines regulären Flächenstücks ist. Anschaulich kann man dies auch so interpretieren, daß die Kurve innerhalb von G auf einen Punkt zusammengezogen werden kann: Man denke sich die Kurve als einen stark angespannten Gummiring; wenn man diesen auf ein Flächenstück legt, zieht er sich automatisch zusammen. Umgekehrt überstreicht der Ring beim Zusammenziehen auf einen Punkt ein Flächenstück, dessen Rand die Ausgangsposition des Gummis ist.

Mit dieser Definition gilt dann offensichtlich die folgende Verallgemeinerung des Hauptsatzes der Differential- und Integralrechnung auf den $\mathbb{R}^3$:

Satz: Ein differenzierbares Vektorfeld $\vec{V}: G \to \mathbb{R}^3$ auf einer einfach zusammenhängenden Teilmenge $G \subseteq \mathbb{R}^3$ hat genau dann eine Stammfunktion, wenn $\operatorname{rot} \vec{V}$ dort identisch verschwindet, d.h. also, wenn die JACOBI-Matrix symmetrisch ist. ∎

Für ein ebenes Vektorfeld gilt entsprechend

Satz: Ein differenzierbares Vektorfeld $\vec{V}: G \to \mathbb{R}^2$ auf einer einfach zusammenhängenden Teilmenge $G \subseteq \mathbb{R}^2$ hat genau dann eine Stammfunktion, wenn

$$\frac{\partial V_1}{\partial y} - \frac{\partial V_2}{\partial x}$$

dort identisch verschwindet, d.h. also, wenn die JACOBI-Matrix symmetrisch ist.

Beweis: Dies folgt sofort aus dem Satz von GREEN. ∎

Auf höhere Dimension soll nicht genauer eingegangen werden; nur soviel sei erwähnt: Die eindeutige Bestimmtheit des Normalenvektors

eines Flächenstücks im $\mathbb{R}^3$ hat in höheren Dimensionen keine Entsprechung mehr; dort muß man von einem Flächenstück explizit *fordern,* daß es orientierbar ist im folgenden Sinne: Man kann es durch kleine rechteckförmige Flächen überdecken und jedem dieser „Rechtecke“ einen Umlaufsinn zuordnen derart, daß die gemeinsame Kante zweier Nachbarrechtecke von diesen beiden Rechtecken entgegengesetzt orientiert wird.

Wenn man dann den einfachen Zusammenhang einer offenen zusammenhängenden Teilmenge $G \subseteq \mathbb{R}^n$ dadurch definiert, daß jede geschlossene Kurve in G Rand eines so orientierbaren Flächenstücks sein soll, was auch wieder im wesentlichen äquivalent ist zur Zusammenziehbarkeit der Kurve, gilt auch hier der

Satz: Ein differenzierbares Vektorfeld $\vec{V}: G \to \mathbb{R}^n$ auf einer einfach zusammenhängenden Teilmenge $G \subseteq \mathbb{R}^n$ hat genau dann eine Stammfunktion, wenn die JACOBI-Matrix von $\vec{V}$ symmetrisch ist. ∎

Damit genug zum ersten der drei Hauptsätze dieses Paragraphen; wir machen weiter, wo wir vor der Spezialisierung auf symmetrische Vektorfelder aufgehört haben und folgern, daß nach $(*)$ für ein beliebiges differenzierbares Vektorfeld gilt

$$\int\limits_{\partial R_i} \vec{V}(\boldsymbol{x})\, ds \approx \tfrac{1}{2} \int\limits_{\partial R} \operatorname{rot} \vec{V}(\boldsymbol{x}_0) \times \vec{x}\, ds\,;$$

wir müssen uns also dieses Integral genauer ansehen. Dabei setzen wir zur Abkürzung

$$\vec{a} = \operatorname{rot} \vec{V}(x_0)\,;$$

dann ist der Integrand

$$\tfrac{1}{2}(\vec{a} \times \vec{x}) \cdot ds\,.$$

Nach Definition eines Kurvenintegrals ist das Integral hierüber Grenzwert einer Summe von Termen der Art

$$\tfrac{1}{2}(\vec{a} \times \vec{x}_j) \cdot \vec{t}_j\,,$$

wobei $\boldsymbol{x}_j$ Punkte auf der Kurve sind und t_j die Tangentenvektoren in diesen Punkten. Dieses Skalarprodukt eines Vektorprodukts mit einem

Vektor ist bekanntlich das *Spatprodukt,* und es ist gleich der Determinante mit den Spaltenvektoren $\vec{a}, \vec{x}_j$ und t_j. Diese Determinante ändert ihr Vorzeichen, wenn man zwei Spalten vertauscht; tut man dies zweimal, kehrt sie wieder zu ihrem alten Wert zurück. Daher ist

$$(\vec{a} \times \vec{x}_j) \cdot \vec{t}_j = (\vec{x}_j \times \vec{t}_j) \cdot \vec{a} \,.$$

Der Vektor $\vec{x}_j \times \vec{t}_j$ steht senkrecht auf $\vec{x}_j$ und auf $\vec{t}_j$; falls wir also von einem *flachen* Rechteck ausgehen, steht er auch senkrecht auf diesem und hat somit die Richtung des Normalenvektors. Sein Betrag ist gleich der Fläche des von $\vec{x}_j$ und $\vec{t}_j$ aufgespannten Parallelogramms, die Hälfte davon also gleich der Fläche des von diesen beiden Vektoren aufgespannten Dreiecks. Die Summe aller dieser Dreiecksflächen ist im Limes gleich der Fläche des Rechtecks, über dessen Rand wir integrieren, also ist das Integral über den Rand des Rechtecks für kleine Rechtecke näherungsweise gleich einem Skalarprodukt $\vec{n} \cdot \vec{a}$, wobei $\vec{n}$ als Länge den Flächeninhalt des Rechtecks hat und als Richtung die des Normalenvektors. Als Grenzwert einer Summe solcher Skalarprodukte ist aber gerade das Integral

$$\iint\limits_f \operatorname{rot} \vec{V} \, d\vec{O}$$

erklärt; wenn wir immer feinere Rechteckunterteilungen betrachten, konvergiert die Summe der Kurvenintegrale über die Ränder dieser Rechtecke also einerseits gegen das Kurvenintegral des Vektorfelds über den Rand γ des betrachteten Flächenstücks, andererseits aber auch gegen das Integral der Rotation von $\vec{V}$ über das Flächenstück selbst. Dies ist der

Satz von Stokes: $\vec{V}: G \to \mathbb{R}^3$ sei ein differenzierbares Vektorfeld auf der offenen Teilmenge $G \subseteq \mathbb{R}^3$. Weiter sei $f: D \to G$ ein reguläres Flächenstück mit einer stückweise regulären Kurve γ als Rand. Dann ist

$$\int\limits_\gamma \vec{V} \, ds = \iint\limits_f \operatorname{rot} \vec{V} \, d\vec{O} \,.$$

GEORGE GABRIEL STOKES (1819–1903) wurde in Irland geboren als jüngster von sechs Söhnen eines protestantischen Pfarrers. Nach dem Tod seines Vaters kam er im Alter von 16 Jahren an eine Schule nach Bristol in England und begann zwei Jahre später sein Studium an der Universität Cambridge. Er wurde vor allem bekannt durch seine Arbeiten zur mathematischen Physik, für die er viele mathematische Techniken entwickelte. Er gilt als der Begründer sowohl der Strömungslehre als auch der Geodäsie und hatte unter anderem großen Einfluß auf die Entwicklung von MAXWELL. Ab 1849 lehrte er als Professor an der Universität Cambridge.

Aus dem Satz von STOKES können wir eine neue Charakterisierung der Rotation herleiten:

Satz: $\vec{V}: G \to \mathbb{R}^3$ sei ein differenzierbares Vektorfeld, $\vec{e}$ ein Einheitsvektor, und $\gamma_{\vec{e},r}$ ein Kreis mit Radius r um $\boldsymbol{x}_0$ ist, der in einer Ebenen senkrecht zu $\vec{e}$ liege und von $\vec{e}$ aus gesehen im Gegenuhrzeigersinn durchlaufen werde; die gesamte von $\gamma_{\vec{e},r}$ berandete Kreisscheibe liege in G. Dann ist

$$\operatorname{rot} \vec{V}(\boldsymbol{x}_0) = \lim_{r \to 0} \frac{1}{\pi r^2} \int\limits_{\gamma_{\vec{e},r}} \vec{V} \, ds \,.$$

Beweis: $\gamma_{\vec{e},r}$ ist Rand einer Kreisscheibe D vom Radius r mit Flächeninhalt πr^2; nach dem Satz von STOKES ist

$$\int\limits_{\gamma_{\vec{e},r}} \vec{V} \, ds = \iint\limits_D \operatorname{rot} \vec{V} \, d\vec{O} = \iint\limits_D (\operatorname{rot} \vec{V}) \cdot \vec{e} \, dO \,,$$

und für immer kleiner werdende Werte von r stimmt dies immer besser überein mit $\operatorname{rot} \vec{V}(\boldsymbol{x}_0) \cdot \vec{e}$ mal dem Flächeninhalt der Kreisscheibe. ■

Falls die Voraussetzungen dieses Satzes erfüllt sind, falls also insbesondere die Kreislinie Rand einer im Definitionsbereich von $\vec{V}$ enthaltenen Fläche ist, läßt sich die Rotation somit als Limes einer Art „normierter" Zirkulation auffassen, d.h. als Integral längs einer geschlossenen Kurve dividiert durch die Länge dieser Kurve. Die Rotation gibt also auch ein Maß für die Abweichung eines Vektorfelds von der Zirkulationsfreiheit.

Damit wäre auch der zweite Hauptsatz dieses Abschnitts bewiesen; bleibt noch der

Satz von Gauß: $\vec{W}: G \to \mathbb{R}^3$ sei ein differenzierbares Vektorfeld und $f: D \to G$ sei ein reguläres Flächenstück derart, daß $B = f(D)$ Rand eines beschränkten dreidimensionalen Bereichs V sei. Dann ist

$$\iint\limits_f \vec{W}\, d\vec{O} = \iiint\limits_V \operatorname{div} \vec{W}\, dx\, dy\, dz\,.$$

Bemerkung: Der Satz gilt auch, mit den offensichtlichen Definitionen, falls der Rand von V kein Flächen*stück* ist, sondern eine Fläche, die aus endlich vielen regulären Stücken zusammengesetzt ist. Am Beweis ändert sich dabei abgesehen vom größeren Schreibaufwand praktisch nichts.

Der *Beweis* des Satzes von GAUSS beruht auf derselben Idee wie der des Satzes von STOKES: Dort hatten wir ein Flächenstück durch Rechtecke angenähert, um das Kurvenintegral längs seines Randes zu einem Integral über das Flächenstück in Beziehung zu setzen; hier unterteilen wir entsprechend das Volumen V in kleine, der Einfachheit halber als achsenparallel vorausgesetzte Quader, um das Oberflächenintegral über den Rand von V mit einem Integral über V in Beziehung zu setzen.

Ein solcher Quader sei Q_i, und $\vec{h}, \vec{k}, \vec{\ell}$ seien die drei Kantenvektoren von Q_i in Richtung der x-, y- und z-Achse. Dann ist das Integral von $\vec{W}$ über die Oberfläche von Q_i gleich

$$\iint\limits_{\partial Q_i} \vec{W}\, d\vec{O} = \iint\limits_{\substack{\text{vordere und}\\ \text{hintere Seitenfläche}}} \vec{W}\, d\vec{O} + \iint\limits_{\substack{\text{linke und}\\ \text{rechte Seitenfläche}}} \vec{W}\, d\vec{O} + \iint\limits_{\substack{\text{obere und}\\ \text{untere Seitenfläche}}} \vec{W}\, d\vec{O}\,,$$

und weiter ist beispielsweise

$$\begin{aligned}\iint\limits_{\substack{\text{linke und}\\ \text{rechte Seitenfläche}}} \vec{W}\, d\vec{O} &= \iint\limits_{\substack{\text{linke}\\ \text{Seitenfläche}}} \vec{W}\, d\vec{O} + \iint\limits_{\substack{\text{rechte}\\ \text{Seitenfläche}}} \vec{W}\, d\vec{O}\\ &= \iint\limits_{\substack{\text{linke}\\ \text{Seitenfläche}}} \vec{W}(\boldsymbol{x})\, d\vec{O} - \iint\limits_{\substack{\text{linke}\\ \text{Seitenfläche}}} \vec{W}(\boldsymbol{x}+\vec{h})\, d\vec{O}\\ &= -\iint\limits_{\substack{\text{linke}\\ \text{Seitenfläche}}} \big(\vec{W}(\boldsymbol{x}+\vec{h}) - \vec{W}(\boldsymbol{x})\big)\, d\vec{O}\,,\end{aligned}$$

denn da alle Normalenvektoren nach außen zeigen, sind die linke und die rechte Seite des Quaders verschieden orientiert.

Nun kommt die Differenzierbarkeit von $\vec{W}$ ins Spiel und sagt uns, daß

$$\vec{W}(\boldsymbol{x}+\vec{h}) = \vec{W}(\boldsymbol{x}) + J_{\vec{W}}(\boldsymbol{x}) \cdot \vec{h} + o(\|\vec{h}\|)$$

ist. Da der Vektor $\vec{h}$ in Richtung der x-Achse zeigt, können wir das (mit $h = \left|\vec{h}\right|$) auch einfacher schreiben als

$$\vec{W}(x+h, y, z) = \vec{W}(x, y, z) + \left(\frac{\partial}{\partial x}\vec{W}\right) \cdot h + o(h)\,,$$

wobei die partielle Ableitung von $\vec{W}$ nach x für jenes Vektorfeld stehen soll, dessen Komponenten die partiellen Ableitungen der entsprechenden Komponenten von $\vec{W}$ sind. Da wir in diesem Semester noch viel zu tun haben, vernachlässigen wir den Term $o(h)$ in der Hoffnung, daß er für immer feiner werdende Unterteilungen trotz der dann ansteigenden *Anzahl* dieser Terme keine Rolle mehr spielen wird.

In der Tat sollte ein mit den Abschätzungstechniken aus der Analysis I vertrauter Leser keine Schwierigkeiten haben, dies mathematisch streng zu zeigen. Die Heuristik, nach der man dabei vorgeht, ist (hoffentlich) klar: Die Anzahl aufeinanderliegender Quader in Richtung der x-Achse ist proportional zu $1/\,|\vec{h}|$, und $1/\,|\vec{h}|$ Terme der Größenordnung $o(|\vec{h}|$ addieren sich zu einem Term der Größenordnung $o(1)$, d.h. zu einer Funktion, die gegen Null geht. (Die Aussage „geht schneller als 1 gegen Null“ ist natürlich gleichbedeutend damit, daß die Funktion überhaupt gegen Null geht.)

Der Einheitsnormalenvektor der linken Seite des Quaders zeigt in Richtung der negativen x-Achse, denn die linke Seite ist parallel zur (y,z)-Ebene und zeigt nach außen. Wenn wir den Einheitsvektor der x-Achse mit $\vec{e}_x$ bezeichnen, ist daher

$$d\vec{O} = -\vec{e}_x \, dy \, dz$$

und damit

$$\iint\limits_{\substack{\text{linke und}\\ \text{rechte Seitenfläche}}} \vec{W} \, d\vec{O} \approx -\iint\limits_{\substack{\text{linke}\\ \text{Seitenfläche}}} \left(\vec{W}(\boldsymbol{x}+\vec{h}) - \vec{W}(\boldsymbol{x})\right) d\vec{O} \,,$$

$$= \iint\limits_{\substack{\text{linke}\\ \text{Seitenfläche}}} h \cdot \frac{\partial}{\partial x}\vec{W} \cdot \vec{e}_x = h \cdot \iint\limits_{\substack{\text{linke}\\ \text{Seitenfläche}}} \begin{pmatrix} \dfrac{\partial V_1}{\partial x} \\ \dfrac{\partial V_1}{\partial y} \\ \dfrac{\partial V_1}{\partial z} \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} dx \, dy$$

$$= h \iint\limits_{\substack{\text{linke}\\ \text{Seitenfläche}}} \frac{\partial V_1}{\partial x} \, dy \, dz \,.$$

Ganz entsprechend ist

$$\iint\limits_{\substack{\text{vordere und}\\ \text{hintere Seitenfläche}}} \vec{W} \, d\vec{O} \approx k \iint\limits_{\substack{\text{vordere}\\ \text{Seitenfläche}}} \frac{\partial V_2}{\partial y} \, dx \, dz$$

und

$$\iint\limits_{\substack{\text{obere und}\\ \text{untere Seitenfläche}}} \vec{W} \, d\vec{O} \approx \ell \iint\limits_{\substack{\text{obere}\\ \text{Seitenfläche}}} \frac{\partial V_3}{\partial z} \, dx \, dy \,.$$

Nach dem Mittelwertsatz für Flächenintegrale können wir diese Integrale weiter abschätzen: Es gibt Punkte ξ auf der linken, η auf der vorderen und ζ auf der unteren Seitenfläche, so daß

$$\iint\limits_{\substack{\text{linke und}\\ \text{rechte Seitenfläche}}} \vec{W} \, d\vec{O} \approx h \iint\limits_{\substack{\text{linke}\\ \text{Seitenfläche}}} \frac{\partial V_1}{\partial x} \, dy \, dz$$

$$= h \frac{\partial V_1}{\partial x}(\xi) \iint\limits_{\substack{\text{linke}\\ \text{Seitenfläche}}} dy \, dy = hkl \frac{\partial V_1}{\partial x}(\xi) \,.$$

Genauso ist

$$\iint\limits_{\substack{\text{vordere und}\\ \text{hintere Seitenfläche}}} \vec{W} \, d\vec{O} \approx hkl \frac{\partial V_2}{\partial y}(\eta)$$

und

$$\iint\limits_{\substack{\text{obere und}\\ \text{untere Seitenfläche}}} \vec{W} \, d\vec{O} \approx hkl \frac{\partial V_3}{\partial z}(\zeta) \,.$$

Nun ist es wieder an der Zeit, eine etwas komplexere Abschätzung unter den Tisch fallen zu lassen: Die Punkte ξ, η und ζ liegen auf den Seitenflächen eines Quaders, der immer kleiner werden soll. Damit rücken auch diese Punkte immer weiter zusammen und wir sollten wohl keinen allzu großen Fehler machen, wenn wir einfach *irgendeinen* Punkt $\boldsymbol{x}_i$ im Quader Q_i auswählen und sowohl ξ, η als auch ζ durch diesen Punkt ersetzen; die formale Rechtfertigung hierfür geht wieder aus von der Differenzierbarkeit des Vektorfelds V und beruht im übrigen auf Abschätzungen.

Wenn wir das alles glauben, ist also

$$\iint\limits_{\substack{\text{linke und}\\ \text{rechte Seitenfläche}}} \vec{W} \, d\vec{O} \approx hkl \frac{\partial V_1}{\partial x}(\boldsymbol{x}_i) \,,$$

$$\iint\limits_{\substack{\text{vordere und}\\ \text{hintere Seitenfläche}}} \vec{W} \, d\vec{O} \approx hkl \frac{\partial V_2}{\partial y}(\boldsymbol{x}_i) \quad \text{und}$$

$$\iint\limits_{\substack{\text{obere und}\\ \text{untere Seitenfläche}}} \vec{W} \, d\vec{O} \approx hkl \frac{\partial V_3}{\partial z}(\boldsymbol{x}_i) \,.$$

Das Oberflächenintegral über die gesamte Quaderoberfläche ist die Summe dieser drei Teilintegrale, und da bekanntlich

$$\frac{\partial V_1}{\partial x}(\boldsymbol{x}_i) + \frac{\partial V_2}{\partial y}(\boldsymbol{x}_i) + \frac{\partial V_3}{\partial z}(\boldsymbol{x}_i) = \operatorname{div} \vec{W}(\boldsymbol{x}_i)$$

ist, haben wir somit gezeigt, daß

$$\iint\limits_{\partial Q_i} \vec{W} \, d\vec{O} \approx hkl \, \operatorname{div} \vec{W}(\boldsymbol{x}_i) = \operatorname{div} \vec{W}(\boldsymbol{x}_i) \cdot \operatorname{Vol}(Q_i) \qquad (*)$$

ist.

Damit haben wir eine lokale Version des Satzes von GAUSS gezeigt; zum Beweis des Satzes selbst nähern wir das Volumen V an durch die Quader Q_i. Dann ist

$$\iiint\limits_V \operatorname{div} \vec{W}\, dx\, dy\, dz \approx \sum_i \operatorname{div} \vec{W}(\boldsymbol{x}_i) \cdot \operatorname{Vol}(Q_i)\,,$$

und wenn wir die Quader immer weiter verkleinern, wird im Limes aus dem Ungefährzeichen ein Gleichheitszeichen – genau so hatten wir schließlich Volumenintegrale definiert.

Im Falle des Integrals über die Randfläche $B = f(D)$ von V ist die Situation etwas komplizierter: Wie beim Beweis des Satzes von STOKES überlegt man sich zunächst leicht, daß eine gemeinsame Seitenfläche zweier benachbarter Quader von den beiden Quadern verschieden orientiert wird, so daß sich die beiden Integrale über diese Fläche gegenseitig aufheben; die Summe über alle Oberflächenintegrale über die Q_i ist also gleich der Summe über alle Oberflächenintegrale über jene Seitenflächen von Quadern Q_i, die nur Seitenfläche eines einzigen Quaders sind, d.h. also über die Randfläche der Vereinigung aller Quader Q_i.

Die Quaderflächen, die diesen Rand ausmachen, sind allesamt parallel zu Koordinatenebenen; ihre Normalenvektoren sind also parallel zu Koordinaten*achsen,* wohingegen die Normalenvektoren von B natürlich beliebige Richtungen haben können. Wir müssen uns überlegen, warum die Integrale im Limes trotzdem gleich sein können.

Der Integrand bei einem Oberflächenintegral über ein Vektorfeld ist das Skalarprodukt aus dem Normalenvektor von B im jeweils betrachteten Punkt und dem Wert $\vec{W}(\boldsymbol{x})$ des Vektorfelds in diesem Punkt. Konkret sei etwa $\vec{n}$ der Normalenvektor im Punkt $\boldsymbol{x} \in B$, und in Koordinaten sei

$$\vec{n} = \begin{pmatrix} n_1 \\ n_2 \\ n_3 \end{pmatrix} \quad \text{und} \quad \vec{W}(\boldsymbol{x}) = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}.$$

Dann ist

$$\begin{aligned} \vec{W}(\boldsymbol{x}) \cdot \vec{n} &= v_1 n_1 + v_2 n_2 + v_3 n_3 \\ &= \vec{W}(\boldsymbol{x}) \cdot (n_1 \vec{e}_1) + \vec{W}(\boldsymbol{x}) \cdot (n_2 \vec{e}_2) + \vec{W}(\boldsymbol{x}) \cdot (n_3 \vec{e}_3)\,; \end{aligned}$$

wir erhalten also dasselbe Skalarprodukt auch, indem wir $\vec{W}(\boldsymbol{x})$ nacheinander mit geeigneten Normalenvektoren von Quaderflächen parallel zu den Koordinatenebenen multiplizieren und die Ergebnisse aufaddieren.

Falls $\vec{n}$ drei nichtverschwindende Komponenten hat, ist klar, daß die Fläche B in der Nähe von $\boldsymbol{x}$ nur so durch Quader angenähert werden kann, daß es dort freie Randflächen parallel zu allen drei Koordinatenebenen gibt; ein nicht allzu schwieriges Argument über den Satz von PYTHAGORAS zusammen mit Limesbetrachtungen zeigt, daß auch mit den Längen alles gut geht, so daß auch das Oberflächenintegral über B gleich der Summe der Integrale über die Quaderoberflächen ist.

Wie wir oben gesehen haben, ist der Satz für einen einzelnen Quader richtig; da auf beiden Seiten das Integral durch Summen entsprechender Integrale für Quader angehähert werden kann, ist also der Satz von GAUSS (modulo zahlreicher Auslassungen) bewiesen. ■

Genau wie der Satz von STOKES zu einer alternativen Definition der Rotation führte, kann mit dem Satz von GAUSS die *Divergenz* auf andere Weise ausgedrückt werden; zumindest für Quader haben wir die entsprechende Formel im gerade beendeten Beweis als Formel $(*)$ bereits hergeleitet:

Satz: Für ein differenzierbares Vektorfeld $\vec{W}\colon G \to \mathbb{R}^3$ ist

$$\operatorname{div} \vec{W}(\boldsymbol{x}_0) = \lim_{r \to 0} \frac{3}{4\pi r^3} \iint\limits_{\partial B_r} \vec{W}\, d\vec{O}\,,$$

wobei B_r eine Kugel mit Radius r um $\boldsymbol{x}_0$ ist, deren Normalenvektoren nach außen zeigen.

Beweis: Nach dem Mittelwertsatz ist das Volumenintegral über B_r gleich dem Produkt des Volumens $\frac{4}{3}\pi r^3$ von B_r mit dem Wert des Integranden $\operatorname{div} \vec{W}$ an einem geeigneten Punkt der Kugel. Falls r gegen Null geht, muß dieser Punkt immer näher an den Mittelpunkt der Kugel rücken, bis er im Limes mit diesem zusammenfällt. ■

Auch der Satz von GAUSS läßt sich wieder anschaulich interpretieren: Bei der Definition der Divergenz haben wir uns bereits überlegt, daß diese mißt, inwieweit ein Punkt eher eine Quelle oder eine Senke für ein Vektorfeld ist. Da das Oberflächenintegral gerade gleich dem Fluß des Vektorfelds durch die betrachtete Oberfläche ist und wir die Oberfläche so orientiert haben, daß der Normalenvektor nach außen zeigt, sagt der Satz von GAUSS also einfach, daß der gesamte Fluß einer Vektorfelds durch die Oberfläche eines Volumens V genau das ist, was im Innern von V erzeugt oder (bei negativen Vorzeichen auf beiden Seiten) vernichtet wird.

Der obige Satz zur Charakterisierung der Divergenz zeigt dementsprechend noch einmal, warum die Divergenz auch als *Quellendichte* bezeichnet wird: Nach dem Satz von GAUSS ist der Fluß durch die Oberfläche einer Kugel gleich der Summe des im Innern erzeugten oder vernichteten; dividiert man dies durch das Volumen der Kugel, entsteht die Quellen*dichte,* deren Grenzwert für immer kleiner werdende Kugeln gleich der Divergenz ist.

E N D E

S C H Ö N E F E R I E N !