

Kapitel 6

Mehrdimensionale Integrale

Auch im Mehrdimensionalen ist die Integralrechnung nach der Differentialrechnung das zweite Standbein der Analysis. Während allerdings für Funktionen einer Veränderlichen Differentiation und Integration einander entgegengesetzte Operationen sind, ist die Situation hier komplizierter: Die Ableitung einer reellwertigen Funktion mehrerer Veränderlicher ist schließlich keine reellwertige Funktion mehr, sondern eine vektorwertige; so etwas wie eine Stammfunktion kann daher nur für Funktionen in n Variablen mit Werten in \mathbb{R}^n existieren, die sogenannten *Vektorfeldern*. Mit Vektorfeldern und Integralen darüber beschäftigt sich die Analysis III; hier in der Analysis II geht es um einen Integralbegriff, der die Flächeninterpretation des eindimensionalen *bestimmten* Integrals verallgemeinert. Anstelle eines Integrationsintervalls $[a, b]$ können (und müssen) wir im \mathbb{R}^n natürlich sehr viel allgemeinere Mengen betrachten; um möglichst schnell zu Beispielen zu kommen, für die wir die bekannte eindimensionale Integrationstheorie verwenden können, beginnen wir aber mit Integralen über Quader.

§1: Integration stetiger Funktionen über Quader

Unter einen Quader im \mathbb{R}^n wollen wir im folgenden, wenn nichts anderes gesagt wird, stets einen *abgeschlossenen achsenparallelen Quader* verstehen, also eine Menge der Form

$$Q = \prod_{i=1}^n [a_i, b_i] = \{(x_1, \dots, x_n) \mid a_i \leq x_i \leq b_i \text{ für alle } i\},$$

wobei jeweils $a_i \leq b_i$ sein soll. Als zugehörigen *offenen* Quader be-

zeichnen wir das entsprechende Produkt der offenen Intervalle, also

$$\overset{\circ}{Q} \stackrel{\text{def}}{=} \prod_{i=1}^n (a_i, b_i) = \{(x_1, \dots, x_n) \mid a_i < x_i < b_i \text{ für alle } i\}.$$

Im Eindimensionalen sind Quader natürlich einfach Intervalle; im Zweidimensionalen sind sie achsenparallele Rechtecke.

Das mehrdimensionale Integral, das wir in diesem Kapitel betrachten wollen, soll insbesondere auch zur Volumenbestimmung oder – besser – zur *Volumendefinition* dienen: Für eine von krummen Flächen begrenzten Teilmenge wissen wir schließlich selbst im Dreidimensionalen nicht wirklich, wie wir ihr Volumen definieren sollen.

Dieses Problem stellte sich bereits den klassischen griechischen Mathematikern; auf sie geht auch der Ansatz zurück, auf dem jeder heute gebräuchliche Integralbegriff beruht: Wir können mit elementaren Methoden die Volumina vieler geradlinig begrenzter Flächen und Körper berechnen; durch *Exhaustion* oder *Ausschöpfung* konnten EUDOXOS (erste Hälfte des vierten vorchristlichen Jahrhunderts), ARCHIMEDES (287–212) und andere viele kompliziertere Flächen und Volumina zurückführen auf die bekannter Polygone und Körper.

Am einfachsten lassen sich die Flächen von Rechtecken und Volumina von Quadern berechnen; wir definieren allgemein das Volumen $\mu(Q)$ eines Quaders im \mathbb{R}^n als Produkt seiner n Kantenlängen, d.h.

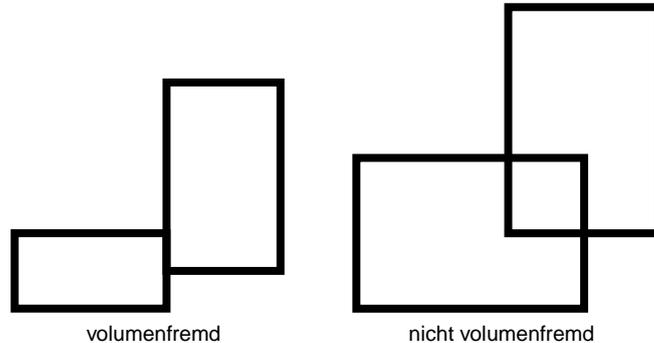
$$\mu(Q) = \mu(\overset{\circ}{Q}) = \prod_{i=1}^n (b_i - a_i) \quad \text{für} \quad Q = \prod_{i=1}^n [a_i, b_i].$$

Falls irgendein a_i gleich dem zugehörigen b_i ist, verschwindet dieses Produkt; solche *entarteten* Quader haben also das Volumen Null. Das entspricht insofern unserem Alltagsgebrauch des Wortes Volumen, als wir auch da einem Rechteck zwar eine Fläche zuordnen; wenn wir aber vom (dreidimensionalen) Volumen reden, betrachten wir es als Quader der Dicke Null und damit auch mit Volumen Null.

In der Tradition von EUDOXOS, HIPPOKRATES und ARCHIMEDES wollen wir relativ komplizierte Volumina „ausschöpfen“ durch Vereinigungen von Quadern; dabei müssen wir natürlich darauf achten, daß sich diese

nicht überschneiden, zumindest nicht so, daß der Durchschnitt ein positives Volumen hat.

Definition: Zwei Quader Q und Q' heißen *volumenfremd*, wenn die zugehörigen offenen Quader leeren Durchschnitt haben.



Als ersten Schritt in Richtung auf die Definition eines n -dimensionalen Integrals definieren wir nun Integrale stetiger Funktionen über Quader in einer zwar sehr willkürlichen, dafür aber unmittelbar zum Rechnen geeigneten Weise:

Definition: $f: D \rightarrow \mathbb{R}^n$ sei eine stetige Funktion und $Q = \prod_{i=1}^n [a_i, b_i]$ sei ein in D enthaltener Quader. Dann ist

$$\int_Q f = \int_{a_n}^{b_n} \left(\cdots \left(\int_{a_2}^{b_2} \left(\int_{a_1}^{b_1} f(x_1, \dots, x_n) dx_1 \right) dx_2 \right) \cdots \right) dx_n$$

für $Q \neq \emptyset$ und $\int_{\emptyset} f = 0$.

Bei dieser Definition bezieht sich jede einzelne Integration nur auf *eine* Variable; wenn wir über dx_i integrieren, betrachten wir die übrigen Variablen x_j als festgehaltene Parameter. Beim innersten Integral über dx_1 halten wir also x_2 bis x_n fest und integrieren nur über x_1 ; das Ergebnis ist für jedes feste $(n-1)$ -Tupel (x_2, \dots, x_n) eine reelle Zahl und insgesamt betrachtet eine Funktion, die nun nur noch von den Variablen x_2 bis x_n abhängt. Die zweite Integration eliminiert auch noch die Variable x_2

und so weiter, bis wir zuletzt nur noch eine Funktion von x_n integrieren und als Ergebnis eine reelle Zahl erhalten. Somit lassen sich bei allen n Integralen die in Kapitel 4 entwickelten Techniken anwenden.

Für konkret gegebene Funktionen f lassen sich die so definierten Integrale, wie wir bald anhand von Beispielen sehen werden, meist problemlos ausrechnen – zumindest, wenn wir die Stammfunktionen der jeweiligen Integranden angeben können. Ansonsten haben wir mit dieser Definition allerdings mindestens zwei Probleme:

Erstens können wir nicht sicher sein, daß das Integral überhaupt existiert: Wie wir in Kapitel 4 gesehen haben, gibt es durchaus Funktionen einer Veränderlichen, die *nicht* integrierbar sind. Zwar wissen wir, daß jede stückweise stetige Funktion einer Veränderlichen integrierbar ist, und wir haben f sogar als stetige Funktion vorausgesetzt, aber wir wissen nicht, ob nach der ersten Integration die entstehende Funktion von x_2 bis x_n noch stetig zumindest in x_2 ist.

Zweitens haben wir uns willkürlich darauf festgelegt, zuerst über x_1 , dann über x_2 und so weiter zu integrieren. Wenn wir bei einer anderen Integrationsreihenfolge ein anderes Ergebnis bekommen, widerspricht dies allem, was wir von einem sinnvoll definierten Integral erwarten. Außerdem kann es auch aus praktischen Gründen gelegentlich sehr viel einfacher sein, die Integrationen in einer anderen Reihenfolge auszuführen.

Das erste Problem läßt sich glücklicherweise recht einfach lösen: Wie das folgende Lemma induktiv zeigt, sind für eine stetige Funktion f auch alle Zwischenintegranden stetig:

Lemma: $[a, b] \subset \mathbb{R}$ sei ein abgeschlossenes Intervall und $D \subseteq \mathbb{R}^{n-1}$. Für eine stetige Funktion $f: [a, b] \times D \rightarrow \mathbb{R}$ definiert die Zuordnung

$$F(y) \stackrel{\text{def}}{=} \int_a^b f(x, y) dx$$

eine stetige Funktion $F: D \rightarrow \mathbb{R}$.

Beweis: Wegen der Stetigkeit von f insbesondere auch in x integrieren wir für jedes feste $y \in D$ über eine stetige Funktion; das Integral

existiert also. Um die Stetigkeit von F zu beweisen, benutzen wir die Charakterisierung durch Folgen: F ist genau dann stetig in einem Punkt $y_0 \in D$, wenn für jede gegen y_0 konvergierende Folge $(y_k)_{k \in \mathbb{N}}$ auch die Folge der Funktionswerte $F(y_k)$ gegen $F(y_0)$ konvergiert, wenn es also zu jedem $\varepsilon > 0$ ein $N \in \mathbb{N}$ gibt, so daß $|F(y_k) - F(y_0)| < \varepsilon$ ist für alle $k \geq N$.

Wegen der Linearität der (eindimensionalen) Integration und der Monotonieregel ist

$$\begin{aligned} |F(y_k) - F(y_0)| &= \left| \int_a^b f(x, y_k) dx - \int_a^b f(x, y_0) dx \right| \\ &= \left| \int_a^b (f(x, y_k) - f(x, y_0)) dx \right| \\ &\leq \int_a^b |f(x, y_k) - f(x, y_0)| dx, \end{aligned}$$

und das können wir weiter abschätzen, falls wir eine handhabbare obere Schranke für den Integranden finden. Dazu verhilft uns ein Kompaktheitsargument:

Wir beginnen mit der Menge $Y \subset \mathbb{R}^{n-1}$ bestehend aus y_0 und den y_k mit $k \in \mathbb{N}$. Diese ist kompakt, denn jede offene Überdeckung \mathcal{U} von Y enthält insbesondere eine offene Menge U , die y_0 enthält. Damit enthält U auch für irgendein $\varepsilon > 0$ alle $y \in \mathbb{R}^{n-1}$ mit $\|y - y_0\|_\infty < \varepsilon$, also gibt es wegen der Konvergenz der Folge $(y_k)_{k \in \mathbb{N}}$ gegen y_0 ein $N \in \mathbb{N}$, so daß $y_k \in U$ für alle $k \geq N$. Damit enthält U alle y_k mit höchstens endlich vielen Ausnahmen. Für jede dieser Ausnahmen gibt es (mindestens) eine offene Menge $V \in \mathcal{U}$, in der sie enthalten ist; also gibt es eine endliche Teilüberdeckung von \mathcal{U} .

Somit ist Y nach dem Satz von HEINE-BOREL abgeschlossen und beschränkt, und damit gilt natürlich dasselbe auch für $[a, b] \times Y$, so daß, wieder nach dem Satz von HEINE-BOREL, auch diese Menge kompakt ist. Nach dem letzten Lemma aus Kapitel 5, §4a) ist die stetige

Funktion f daher auf der Teilmenge $[a, b] \times Y$ gleichmäßig stetig. Es gibt somit zu jedem $\varepsilon > 0$ ein $\delta > 0$, so daß $|f(x, y) - f(x^*, y^*)| < \varepsilon$ ist für alle $(x, y), (x^*, y^*) \in [a, b] \times Y$ mit

$$\|(x, y) - (x^*, y^*)\|_\infty = \max\{|x - x^*|, \|y - y^*\|_\infty\} < \delta.$$

Dies gilt insbesondere für alle Punkte (x, y_0) und (x, y_k) aus $[a, b] \times Y$ mit $\|y_0 - y_k\|_\infty < \delta$, und wegen der Konvergenz der Folge $(y_k)_{k \in \mathbb{N}}$ gegen y_0 gibt es ein $N \in \mathbb{N}$, so daß dies für alle $k \geq N$ erfüllt ist. Für diese k ist daher auch $|f(x, y_0) - f(x, y_k)| < \varepsilon$ für alle $x \in [a, b]$.

Um nun die Konvergenz der Folge der $F(y_k)$ gegen $F(y_0)$ zu beweisen, betrachten wir irgendein $\varepsilon > 0$. Wie wir gerade gesehen haben, gibt es dazu ein $N \in \mathbb{N}$, so daß

$$|f(x, y_0) - f(x, y_k)| < \frac{\varepsilon}{b-a} \quad \text{für alle } k \geq N \text{ und alle } x \in [a, b].$$

Damit ist nach der ersten Abschätzung dieses Beweises für $k \geq N$ auch

$$|F(y_k) - F(y_0)| \leq \int_a^b \frac{\varepsilon}{b-a} dx = \varepsilon,$$

die Folge konvergiert also gegen $F(y_0)$ und F ist stetig in y_0 . Da $y_0 \in D$ beliebig war, ist F stetig auf D . ■

Auch für unserer zweites Problem brauchen wir als ersten Schritt zu einer Lösung ein Lemma nach Art des gerade bewiesenen, jetzt allerdings für differenzierbare Funktionen, wobei wir uns auf zunächst auf \mathbb{R}^2 beschränken wollen:

Lemma: $I \subseteq \mathbb{R}$ sei ein nicht notwendigerweise endliches Intervall, und die Funktion $f: [a, b] \times I \rightarrow \mathbb{R}$ erfülle die beiden Bedingungen

- 1.) Für jedes $y \in I$ ist die Funktion $x \mapsto f(x, y)$ stetig auf $[a, b]$
- 2.) Die partielle Ableitung f_y existiert in jedem Punkt (x, y) aus $[a, b] \times I$ und ist dort stetig. Dann definiert die Vorschrift

$$F(y) \stackrel{\text{def}}{=} \int_a^b f(x, y) dx$$

eine in jedem Punkt $y \in I$ stetig differenzierbare Funktion F und

$$F'(y) = \int_a^b f_y(x, y) dx.$$

Beweis: Wir beschränken uns zunächst auf den Fall, daß $I = [c, d]$ ein kompaktes Intervall ist und zeigen, daß der Differenzenquotient für $h \rightarrow 0$ gegen das rechtsstehende Integral konvergiert, d.h.

$$\lim_{h \rightarrow 0} \frac{F(y_0 + h) - F(y_0)}{h} = \int_a^b f_y(x, y_0) dx.$$

für alle $y_0 \in [c, d]$. Liegt auch $y_0 + h$ in $[c, d]$, können wir den Zähler schreiben als

$$\begin{aligned} F(y_0 + h) - F(y_0) &= \int_a^b f(x, y_0 + h) dx - \int_a^b f(x, y_0) dx \\ &= \int_a^b (f(x, y_0 + h) - f(x, y_0)) dx. \end{aligned}$$

Da f bezüglich y eine Stammfunktion von f_y ist, können wir den Integranden weiter ausrechnen als

$$f(x, y_0 + h) - f(x, y_0) = \int_{y_0}^{y_0+h} f_y(x, y) dy,$$

und nach dem Mittelwertsatz der Integralrechnung aus Kapitel 4, §a3) gibt es ein $\theta \in [0, 1]$, so daß dieses Integral gleich $h f_y(x, y_0 + \theta h)$ ist. Somit ist

$$f(x, y_0 + h) - f(x, y_0) = h f_y(x, y_0 + \theta h),$$

also

$$\frac{F(y_0 + h) - F(y_0)}{h} = \int_a^b f(x, y_0 + \theta h) dx.$$

Wir müssen zeigen, daß wir beim Grenzübergang $h \rightarrow 0$ auch rechts einfach $h = 0$ setzen können. Dazu verwenden wir die Kompaktheit

von $[a, b] \times [c, d]$: Nach Voraussetzung ist f_y dort stetig, also wegen der Kompaktheit nach dem letzten Lemma aus Kapitel 5, §4a) sogar gleichmäßig stetig. Daher gibt es zu jedem $\eta > 0$ ein $\delta > 0$, so daß

$$|f_y(x, y) - f_y(x^*, y^*)| < \eta \quad \text{falls } \|(x, y) - (x^*, y^*)\|_\infty < \delta.$$

Insbesondere ist also $|f(x, y_0 + \theta h) - f(x, y_0)| < \eta$, falls $|\theta h| < \delta$, also erst recht, wenn $|h| < \delta$ ist. Somit ist

$$\begin{aligned} &\left| \frac{F(y_0 + h) - F(y_0)}{h} - \int_a^b f_y(x, y_0) dx \right| \\ &= \left| \int_a^b f_y(x, y_0 + \theta h) dx - \int_a^b f_y(x, y_0) dx \right| \\ &\leq \int_a^b |f_y(x, y_0 + \theta h) - f_y(x, y_0)| dx < \eta \cdot (b - a) \end{aligned}$$

für alle h mit $|h| < \delta$.

Zum Abschluß des Beweises für kompakte Intervalle betrachten wir irgendein $\varepsilon > 0$ und setzen $\eta = \varepsilon / (b - a)$. Für $|h| < \delta$ ist dann

$$\left| \frac{F(y_0 + h) - F(y_0)}{h} - \int_a^b f_y(x, y_0) dx \right| < \eta \cdot (b - a) = \varepsilon,$$

das heißt

$$\lim_{h \rightarrow 0} \frac{F(y_0 + h) - F(y_0)}{h} = \int_a^b f_y(x, y_0) dx.$$

Damit ist die Differenzierbarkeit von F und auch die behauptete Formel für F' gezeigt.

Bleibt noch der Fall, daß I kein kompaktes Intervall ist. Dann liegt trotzdem jedes $y_0 \in I$ in einem kompakten Teilintervall; da die Behauptung für alle Punkte aus diesem Teilintervall richtig ist, gilt sie insbesondere für y_0 , also, da y_0 beliebig war, für alle Punkte aus I . ■

Damit können wir nun zeigen, daß das oben definierte Integral einer stetigen Funktion über einen Quader unabhängig ist von der Integrationsreihenfolge. Wir betrachten zunächst den Fall einer stetigen Funktion zweier Variablen; hier ist die Behauptung die folgende schwache Form des Satzes von FUBINI:

Satz: Für jede stetige Funktion $f: [a, b] \times [c, d] \rightarrow \mathbb{R}$ ist

$$\int_a^b \left(\int_c^d f(x, y) dy \right) dx = \int_c^d \left(\int_a^b f(x, y) dx \right) dy.$$

Beweis: Wir betrachten die beiden Funktionen

$$F(x) = \int_a^x \left(\int_c^d f(\xi, y) dy \right) d\xi \quad \text{und} \quad G(x) = \int_c^d \left(\int_a^x f(\xi, y) d\xi \right) dy;$$

die Behauptung ist dann äquivalent zur Gleichung $F(b) = G(b)$. Nach dem Hauptsatz der Differential- und Integralrechnung und den beiden vorigen Lemmata ist

$$F'(x) = \int_c^d f(x, y) dx = G'(x),$$

F und G unterscheiden sich also höchstens um eine additive Konstante. Da $F(a) = G(a) = 0$ ist, muß diese aber verschwinden, d.h. $F(x) = G(x)$ für alle x und damit insbesondere für $x = b$. ■

Wie einige Beispiele aus den Übungen zeigen, kann sich der Aufwand für die Berechnung der beiden Seiten der Gleichung im Satz von FUBINI deutlich unterscheiden; der Satz hat also durchaus auch praktische Bedeutung. Um ihn auf Funktionen von $n > 2$ Variablen zu verallgemeinern, beachten wir, daß sich jede Permutation π der n Variablen x_1, \dots, x_n als Produkt von Transpositionen schreiben läßt, d.h. also als Produkt von Vertauschungen zweier Variablen x_i und x_j . Tatsächlich können wir π sogar als Produkt von Vertauschungen benachbarter Variablen schreiben: Bezeichnet nämlich $(i j)$ für $i < j$ die Vertauschung von x_i und x_j , so ist die Hintereinanderausführung $(j j+1)(i j)(j j+1)$

gleich $(i j+1)$. Somit folgt induktiv, daß sich die Transposition $(i j)$ mit $i < j$ schreiben läßt als

$$(j-1 j)(j-2 j) \cdots (i+1 i+2)(i i+1)(i+1 i+2) \cdots (j-2 j-1)(j-1 j).$$

Auf die Vertauschung zweier benachbarter Variablen können wir aber den gerade bewiesenen Satz anwenden.

Daher können wir in der Formel

$$\int_Q f = \int_{a_n}^{b_n} \left(\cdots \left(\int_{a_2}^{b_2} \left(\int_{a_1}^{b_1} f(x_1, \dots, x_n) dx_1 \right) dx_2 \right) \cdots \right) dx_n$$

die Reihenfolge der n Integrationen beliebig permutieren.



Der italienische Mathematiker GUIDO FUBINI (1879–1943) arbeitete zunächst auf dem Gebiet der Differentialgeometrie, interessierte sich dann aber immer mehr für analytische Themen wie Differentialgleichungen und Funktionen mehrerer komplexer Veränderlicher. 1901 wurde er Professor in Catania auf Sizilien, später in Genua und ab 1908 in Turin, wo er blieb, bis er 1939 trotz seiner angegriffenen Gesundheit wegen des italienischen Faschismus nach USA emigrierte und ans Institute for Advanced Study in Princeton wechselte. Der hier zitierte Satz ist zwar sein bekanntestes, aber ganz sicher nicht sein bedeutendstes Ergebnis.

Damit sind beide der gleich nach der Definition erwähnten Probleme gelöst, und wir können unbesorgt mit den so definierten Integral rechnen. Wir wollen als nächstes einige zwar ziemlich offensichtliche, aber dennoch zu beweisende weitere Eigenschaften zusammenstellen:

Zerlegungsregel: Läßt sich der Quader Q als Vereinigung der paarweise volumenfremden Quader Q_1, \dots, Q_r schreiben, so ist

$$\int_Q f = \int_{Q_1} f + \cdots + \int_{Q_r} f.$$

Zum *Beweis* schreiben wir

$$Q = \prod_{i=1}^n [a_i, b_i] \quad \text{und} \quad Q_j = \prod_{i=1}^n [a_i^{(j)}, b_i^{(j)}].$$

Im Falle $n = 1$ haben wir dann einfach eine Zerlegung des Intervalls $[a_1, b_1]$ in Teilintervalle; durch Umordnung der Q_j können wir erreichen, daß

$$a_1 = a_1^{(1)} \leq b_1^{(1)} = a_1^{(2)} \leq b_1^{(2)} = a_1^{(3)} \leq \dots \leq b_1^{(r-1)} = a_1^{(r)} \leq b_1^{(r)} = b_1$$

ist, und wie wir aus Kapitel 4 wissen, ist

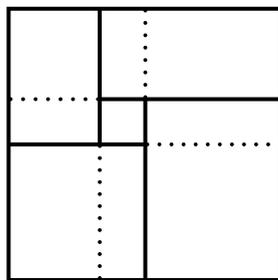
$$\int_{a_1}^{b_1} f(x_1) dx_1 = \sum_{j=1}^r \int_{a_1^{(j)}}^{b_1^{(j)}} f(x_1) dx_1.$$

Leider können wir diese Zerlegung für $n > 1$ nicht einfach durch die Definition durchziehen, denn unsere Quaderzerlegung muß nicht von festen Zerlegungen der Intervalle $[a_i, b_i]$ induziert sein. Wir können sie allerdings so verfeinern, daß dies der Fall ist, indem wir die Quader Q_j noch weiter zerlegen: Für jedes i betrachten wir die Menge der sämtlichen a_i^j und b_i^j und ordnen sie der Größe nach zu einer Folge

$$a_i = c_{i1} < c_{i2} < \dots < c_{ir_i} = b_i,$$

wobei wir mehrfach vorkommende Zahlen nur einfach berücksichtigen. Dann betrachten wir die (in der Zeichnung gestrichelten) $\prod_{i=1}^n r_i$ Quader der Form

$$\prod_{i=1}^n [c_{ij_i}, c_{i,j_i+1}].$$



Auch sie bilden eine Zerlegung des Quaders Q durch paarweise volumenfremde Teilquader, und für jeden der Quader Q_j bildet die Teilmenge jener Quader, die in Q_j liegen, eine Zerlegung von Q_j , die

ebenfalls von dieser speziellen Form ist. Es genügt daher, wenn wir zeigen, daß die Behauptung für diese spezielle Zerlegung von Q gilt, und da folgt sie induktiv aus der Definition von $\int_Q f$ und der obigen Formel für den Fall $n = 1$. ■

Monotonieregel: Ist $f(x) \leq g(x)$ für alle $x \in Q$, so ist $\int_Q f \leq \int_Q g$.

Insbesondere ist $|\int_Q f| \leq \int_Q |f|$.

Beweis: Für den eindimensionalen Fall kennen wir die Monotonieregel aus Kapitel 4; die Verallgemeinerung auf n -dimensionale Quader folgt aus der Definition von $\int_Q f$ sofort durch vollständige Induktion. Die Aussage über den Betrag wiederum folgt daraus wegen der Ungleichung $-|x| \leq x \leq |x|$. ■

Aus den beiden vorigen Regeln folgt sofort

Korollar: $f: D \rightarrow \mathbb{R}$ sei stetig und nichtnegativ auf $D \subseteq \mathbb{R}^n$, und die Quader Q, Q_1, \dots, Q_r seien Teilmengen von D .

a) Ist $Q \subseteq Q_1 \cup \dots \cup Q_r$, so ist

$$\int_Q f \leq \int_{Q_1} f + \dots + \int_{Q_r} f.$$

b) Ist $Q_1 \cup \dots \cup Q_r \subseteq Q$ und sind die Q_i paarweise volumenfremd, so ist

$$\int_{Q_1} f + \dots + \int_{Q_r} f \leq \int_Q f.$$

Beweis: a) Durch Unterteilung können wir annehmen, daß Q die Vereinigung gewisser volumenfremder Q_i ist; dann ist $\int_Q f$ die Summe der entsprechenden $\int_{Q_i} f$. Die restlichen $\int_{Q_j} f$ sind wegen der Nichtnegativität von f allesamt größer oder gleich null.

b) Hier können wir durch Unterteilung annehmen, daß Q die volumenfremde Vereinigung der Q_i zusammen mit einigen weiteren Quadern $Q_j \subset Q$ ist. Da $\int_{Q_j} f \geq 0$ für alle j , folgt auch hier die Behauptung. ■

Linearität: Für zwei stetige Funktionen $f, g: D \rightarrow \mathbb{R}$ und $a, b \in \mathbb{R}$ gilt für einen Quader $Q \subset D \subseteq \mathbb{R}^n$

$$\int_Q (af + bg) = a \int_Q f + b \int_Q g.$$

Der *Beweis* folgt induktiv aus der entsprechenden Aussage im Eindimensionalen. ■

Lemma: Für jedes $c \in \mathbb{R}$ ist $\int_Q c = c \cdot \mu(Q)$.

Beweis: Auch dies folgt induktiv aus dem entsprechenden Lemma für den eindimensionalen Fall

$$\int_a^b c \, dx = c(b - a).$$

Damit haben wir Integrale über Quader recht gut verstanden; oft müssen wir aber über allgemeinere Teilmengen von \mathbb{R}^n integrieren. Als ersten Zwischenschritt in Richtung auf entsprechende Integrale betrachten wir stetige Funktionen, die außerhalb eines Quaders verschwinden:

Definition: a) Für eine Teilmenge $X \subseteq \mathbb{R}^n$ ist der *Abschluß* \overline{X} von X die kleinste abgeschlossene Teilmenge $Y \subseteq \mathbb{R}^n$, die X enthält, d.h. der Durchschnitt aller abgeschlossener Teilmengen $Z \subseteq \mathbb{R}^n$, die X enthalten.

b) $f: D \rightarrow \mathbb{R}$ sei eine stetige Funktion auf $D \subseteq \mathbb{R}^n$. Der *Träger* von f ist der Abschluß der Menge aller $x \in D$, für die $f(x)$ nicht verschwindet, d.h. die Menge $\text{Tr}(f) = \overline{\{x \in D \mid f(x) \neq 0\}}$.

c) Eine stetige Funktion $f: D \rightarrow \mathbb{R}$ heißt *Funktion mit kompaktem Träger*, falls $\text{Tr}(f)$ eine kompakte Teilmenge von \mathbb{R}^n ist.

d) $C^0(D, \mathbb{R})$ ist die Menge aller stetiger Funktionen $D \rightarrow \mathbb{R}$ und $K^0(D, \mathbb{R})$ die Teilmenge der Funktionen mit kompaktem Träger.

$C^0(D, \mathbb{R})$ ist ein Vektorraum, da jede Linearkombination stetiger Funktionen wieder stetig ist, und $K^0(D, \mathbb{R})$ ist ein Untervektorraum, denn der Träger einer Linearkombination zweier Funktionen f, g ist enthalten in der Vereinigung der Träger von f und von g . Bei der Definition eines

möglichst allgemeinen n -dimensionalen Integrals werden die Funktionen mit kompaktem Träger im folgenden die gleiche Rolle spielen wie die Treppenfunktionen bei der Definition des RIEMANN-Integrals für Funktionen einer Veränderlichen.

Da jede kompakte Teilmenge von \mathbb{R}^n beschränkt ist, liegt sie insbesondere in einem Quader; für eine Funktion $f: \mathbb{R}^n \rightarrow \mathbb{R}$ mit kompaktem Träger gibt es also einen Quader $Q \subseteq \mathbb{R}^n$, so daß $f(x) = 0$ für alle $x \notin Q$. Da f nach Definition stetig auf ganz \mathbb{R}^n und damit insbesondere auch auf Q ist, existiert das Integral über Q und wir *definieren*

$$\int_{\mathbb{R}^n} f = \int_Q f.$$

Diese Definition ist nur sinnvoll, wenn wir zeigen können, daß sie nicht vom Quader Q abhängt; das ist aber einfach, denn sind Q, \tilde{Q} zwei Quader, die den Träger von f enthalten, so liegt der Träger auch im Quader $Q \cap \tilde{Q}$. Wie im Beweis der Zerlegungsregel können wir Quader Q_1, \dots, Q_r und $\tilde{Q}_1, \dots, \tilde{Q}_s$ finden derart, daß

$Q = (Q \cap \tilde{Q}) \cup Q_1 \cup \dots \cup Q_r$ und $\tilde{Q} = (Q \cap \tilde{Q}) \cup \tilde{Q}_1 \cup \dots \cup \tilde{Q}_s$ ist. Da die Q_i und die \tilde{Q}_j disjunkt zum Träger von f sind, verschwindet das Integral von f über diese Quader; nach der Zerlegungsregel ist also

$$\int_Q f = \int_{Q \cap \tilde{Q}} f = \int_{\tilde{Q}} f.$$

Für die Konstruktion allgemeinerer Integrale wollen wir als erstes Integrale allgemeinerer als der bisher betrachteten Funktionen über \mathbb{R}^n definieren; danach wollen wir auch über allgemeinere Teilmengen von \mathbb{R}^n integrieren.

Für beides ist es wichtig, $K^0(\mathbb{R}^n, \mathbb{R})$ zu einem *normierten* Vektorraum zu machen. Dazu definieren wir gleich zwei Normen, von denen wir gleich sehen werden, daß die nicht äquivalent sind.

Die erste Norm, die L^1 -Norm ist definiert als

$$\|f\|_1 = \int_{\mathbb{R}^n} |f|;$$

nach Definition einer Funktion mit kompaktem Träger existiert dieses Integral für jede Funktion mit kompaktem Träger.

Als zweite Norm auf $K^0(\mathbb{R}^n, \mathbb{R})$ betrachten wir die Supremumsnorm

$$\|f\|_\infty = \sup\{|f(x)| \mid x \in \mathbb{R}^n\}.$$

Für Funktionen mit kompaktem Träger ist das Supremum des Betrags über \mathbb{R}^n gleich dem Supremum von $|f(x)|$ über den kompakten Träger, denn auch $|f(x)|$ ist eine stetige Funktion. Da diese auf jedem Kompaktum ihr Maximum annimmt, ist dieses gleich dem Supremum; insbesondere existiert also ein Supremum.

Wir bezeichnen den normierten Vektorraum $K^0(\mathbb{R}^n, \mathbb{R})$ mit der L^1 -Norm als $K_1^0(\mathbb{R}^n, \mathbb{R})$; mit der Supremumsnorm bezeichnen wir ihn als $K_\infty^0(\mathbb{R}^n, \mathbb{R})$. Bezüglich keiner der beiden Normen ist der Vektorraum vollständig:

Betrachten wir etwa die Folge der Funktionen $f_k: \mathbb{R} \rightarrow \mathbb{R}$ mit

$$f_k(x) = \begin{cases} 0 & \text{falls } x \notin (0, 2) \\ kx & \text{falls } 0 \leq x \leq \frac{1}{k} \\ 1 & \text{falls } \frac{1}{k} \leq x \leq 2 - \frac{1}{k} \\ k(2-x) & \text{falls } 2 - \frac{1}{k} \leq x \leq 2 \end{cases}.$$



Sie konvergiert punktweise gegen die Funktion $f: \mathbb{R} \rightarrow \mathbb{R}$ mit

$$f(x) = \begin{cases} 0 & \text{falls } x \notin (0, 2) \\ 1 & \text{falls } x \in (0, 2) \end{cases},$$

denn für jedes $x \in (0, 2)$ gibt es ein $N \in \mathbb{N}$, so daß $\frac{1}{k} \leq x \leq 2 - \frac{1}{k}$ für alle $k \geq N$. Diese Funktion ist unstetig an den Sprungstellen bei $x = 0$ und $x = 2$, liegt also insbesondere nicht in $K^0(\mathbb{R}, \mathbb{R})$. Trotzdem ist die Folge der f_k eine CAUCHY-Folge, denn für $\ell \geq k$ stimmen die Funktionen f_k und f_ℓ überein außer in den beiden Intervallen $(0, \frac{1}{k})$ und $(2 - \frac{1}{k}, 2)$. Diese haben jeweils die Länge $\frac{1}{k}$, und da alle Funktionswerte

in $[0, 1]$ liegen, ist dort $|f_\ell(x) - f_k(x)| \leq 1$. Daher ist

$$\begin{aligned} \|f_\ell - f_k\|_1 &= \int_{\mathbb{R}} |f_\ell - f_k| \\ &= \int_0^{1/k} |f_\ell(x) - f_k(x)| dx + \int_{2-1/k}^2 |f_\ell(x) - f_k(x)| dx \leq \frac{2}{k}. \end{aligned}$$

Wählen wir also zu vorgegebenem $\varepsilon > 0$ ein $N \in \mathbb{N}$ mit $2/N < \varepsilon$, so ist $\|f_\ell - f_k\| < \varepsilon$ für alle $k, \ell \geq N$, die Folge $(f_k)_{k \in \mathbb{N}}$ ist somit eine CAUCHY-Folge bezüglich der L^1 -Norm, konvergiert aber nicht gegen eine Funktion aus $K^0(\mathbb{R}, \mathbb{R})$.

Bezüglich der Supremumsnorm ist $(f_k)_{k \in \mathbb{N}}$ keine CAUCHY-Folge, denn für $k \leq \ell$ ist $f_\ell(\frac{1}{\ell}) - f_k(\frac{1}{\ell}) = 1 - \frac{k}{\ell}$, und das ist für $\ell \geq 2k$ stets größer als $\frac{1}{2}$. Damit ist insbesondere klar, daß unsere Normen nicht äquivalent sind, denn äquivalente Normen führen zu denselben CAUCHY-Folgen.

Es gibt allerdings auch CAUCHY-Folgen bezüglich der Supremumsnorm, die nicht gegen ein Element von $K^0(\mathbb{R}, \mathbb{R})$ konvergieren. Als Beispiel betrachten wir die Funktionen $g(x) = e^{-x^2}$ und, für jedes $k \in \mathbb{N}$,

$$g_k(x) = \begin{cases} e^{-x^2} & \text{falls } |x| \leq k \\ e^{-k^2} (|k+1| - |x|) & \text{falls } k \leq |x| \leq k+1 \\ 0 & \text{sonst} \end{cases}.$$

Offensichtlich liegen alle g_k in $K^0(\mathbb{R}, \mathbb{R})$, denn sie sind stetig und der Träger $[-k-1, k+1]$ ist kompakt. Für $\ell \geq k$ ist $|g_\ell(x) - g_k(x)| \leq e^{-k^2}$ für alle $x \in \mathbb{R}$, also ist $\|g_k - g_\ell\| < e^{-N^2}$ für alle $k, \ell \geq N$ und wie haben eine CAUCHY-Folge bezüglich der Supremumsnorm. Da die Grenzfunktion g für kein $x \in \mathbb{R}$ verschwindet, hat g aber natürlich keinen kompakten Träger.

Man könnte die Unvollständigkeit von $K^0(\mathbb{R}, \mathbb{R})$ bezüglich beider Normen als Nachteil sehen, tatsächlich aber das genaue Gegenteil der Fall: Die Funktionen, die uns wirklich interessieren, haben schließlich nur in den seltensten Fällen kompakte Träger, und zumindest gelegentlich müssen wir auch Funktionen mit Sprungstellen betrachten, beispielsweise bei Kostenfunktionen mit Rabattstufen. Wie wir sehen werden,

lassen sich fast alle solche Funktionen als Grenzwerte von CAUCHY-Folgen von Funktionen mit kompaktem Träger bezüglich der L^1 -Norm realisieren; die zur Vollständigkeit von $K_1^0(\mathbb{R}, \mathbb{R})$ „fehlenden“ Funktionen sind also genau die Funktionen, die uns interessieren.

§2: Das Lebesgue-Integral auf \mathbb{R}^n

Bei der Definition eines mehrdimensionalen Integrals wollen wir gleichzeitig auch das eindimensionale Integral noch etwas verallgemeinern. Beim RIEMANN-Integral lernten wir die DIRICHLETSche Sprungfunktion

$$f: \begin{cases} \mathbb{R} \rightarrow \mathbb{R} \\ x \mapsto \begin{cases} 1 & \text{falls } x \in \mathbb{Q} \\ 0 & \text{sonst} \end{cases} \end{cases}$$

kennen als Beispiel einer nicht integrierbaren Funktion: Wenn wir sie über irgendeinem Intervall $[a, b]$ durch Treppenfunktionen approximieren, müssen wir für die RIEMANNschen Obersummen stets Stufen der Höhe eins wählen und für die Untersummen Stufen der Höhe null, denn in jedem noch so kleinen Intervall positiver Länge gibt es sowohl rationale als auch irrationale Zahlen.

Nun gibt es aber nur abzählbar viele rationale Zahlen, während die irrationalen Zahlen nicht abzählbar sind. Wenn wir ein Integral wollen, das mit Grenzwerten vertauschbar ist, müssen wir daher auch der DIRICHLETSchen Sprungfunktion ein Integral zuordnen: Ist nämlich $\varphi: \mathbb{N} \rightarrow \mathbb{Q}$ eine Abzählung der rationalen Zahlen, eine bijektive Abbildung also, so können wir für jedes $n \in \mathbb{N}$ eine Funktion $f_n: \mathbb{R} \rightarrow \mathbb{R}$ definieren durch

$$f_n(x) = \begin{cases} 1 & \text{falls es ein } k \leq n \text{ gibt mit } \varphi(k) = x \\ 0 & \text{sonst} \end{cases}$$

Da f_n somit nur an den endlich vielen Stellen $\varphi(1), \dots, \varphi(n)$ von Null verschieden ist, ist f_n RIEMANN-integrierbar, und das Integral über jedes (endliche oder unendliche) Intervall verschwindet. Für $n \rightarrow \infty$ konvergiert die Folge der Funktionen f_n gegen die DIRICHLETSche Sprungfunktion, denn für die bezüglich der Abzählung durch φ n -te rationale Zahl $x = \varphi(n)$ ist $f_k(x) = 1$ für alle $k \geq n$, und für irrationale Zahlen verschwinden alle f_k . Wenn wir also ein Integral suchen, für das gilt

$$\lim_{n \rightarrow \infty} \int_a^b f_n(x) dx = \int_a^b \lim_{n \rightarrow \infty} f_n(x) dx,$$

dann müssen wir auch Integranden wie die DIRICHLETSche Sprungfunktion zulassen und ihr für jedes Intervall das Integral Null zuordnen.

Dazu definieren wir nun die sogenannten Nullmengen; das sollen solche Teilmengen von \mathbb{R}^n sein, auf die es bei der Integration nicht ankommt.

Definition: Eine Teilmenge $Z \subset \mathbb{R}^n$ heißt *Nullmenge*, wenn es zu jedem $\varepsilon > 0$ Quader Q_1, Q_2, \dots gibt derart, daß

$$Z \subseteq \bigcup_{k \geq 1} Q_k \quad \text{und} \quad \sum_{k \geq 1} \mu(Q_k) < \varepsilon.$$

Man beachte, daß die Folge Q_1, Q_2, \dots nicht endlich sein muß; wir können auch abzählbar unendlich viele Quader nehmen. Deshalb ist beispielsweise $\mathbb{Q} \subset \mathbb{R}$ eine Nullmenge, denn wir können für eine bijektive Funktion $\varphi: \mathbb{N} \rightarrow \mathbb{Q}$ als „Quader“ Q_k einfach das einpunktige Intervall $[\varphi(k), \varphi(k)]$ nehmen. Genauso folgt auch allgemeiner, daß *jede* endliche oder abzählbare Teilmenge von \mathbb{R}^n eine Nullmenge ist.

Umgekehrt muß nicht jede Nullmenge abzählbar sein: Beispielsweise ist auch $Z = \mathbb{R}^{n-1} \times \{0\}$, also die Menge aller Punkte aus \mathbb{R}^n mit letzter Koordinate Null eine Nullmenge, denn sie wird überdeckt durch die Quader

$$Q_k = \underbrace{[-k, k] \times \dots \times [-k, k]}_{n-1 \text{ Faktoren}} \times [0, 0],$$

die allesamt das Volumen Null haben. Damit ist auch die Summe ihrer Volumina Null und somit kleiner als jedes vorgegebene $\varepsilon > 0$.

Ein Quader mit von Null verschiedenem Volumen kann dagegen nie eine Nullmenge sein: Nehmen wir für ε beispielsweise das halbe Volumen des Quaders, so kann es unmöglich eine Überdeckung durch Quader geben, deren Volumina sich zu einer Summe von kleiner ε addieren.

Gelegentlich wird es einfacher sein, statt mit kompakten Quadern mit Würfeln oder mit offenen Quadern zu arbeiten; das ist auch problemlos möglich nach dem folgenden

Lemma: Für eine Teilmenge $A \subseteq \mathbb{R}^n$ sind folgende Aussagen äquivalent:

- a) A ist eine Nullmenge.

b) Für jedes $\varepsilon > 0$ gibt es offene Quader Q_1, Q_2, \dots derart, daß A in der Vereinigung aller Q_i enthalten ist und die Summe der Volumina der sämtlichen Q_i kleiner ist als ε .

c) Für jedes $\varepsilon > 0$ gibt es Würfel W_1, W_2, \dots derart, daß A in der Vereinigung aller W_i enthalten ist und die Summe der Volumina der sämtlichen W_i kleiner ist als ε .

Beweis: Natürlich folgt a) aus c), denn jeder Würfel ist insbesondere ein Quader. Genauso schnell folgt a) aus b), denn hier können wir einfach die offenen Quader Q_i durch ihre Abschlüsse ersetzen.

Für die Umkehrungen gehen wir aus von einer Nullmenge A und einem $\varepsilon > 0$. Dazu gibt es Quader Q_1, Q_2, \dots derart, daß A in der Vereinigung aller Q_i enthalten ist und die Summe der Volumina der sämtlichen Q_i kleiner ist als $\varepsilon/2$.

Zum Beweis von b) aus a) betrachten wir jeden Quader

$$Q_i = [a_1, b_1] \times \dots \times [a_n, b_n] \quad \text{auch} \quad \tilde{Q}_i = (\tilde{a}_1, \tilde{b}_1) \times \dots \times (\tilde{a}_n, \tilde{b}_n),$$

wobei \tilde{a}_k und \tilde{b}_k jeweils so gewählt sind, daß $\tilde{a}_k < a_k \leq b_k < \tilde{b}_k$ ist und $\tilde{b}_k - \tilde{a}_k < \sqrt[n]{2}(b_k - a_k)$. Dann ist \tilde{Q}_i ein offener Quader, der Q_i enthält und höchstens das doppelte Volumen von Q_i hat. Somit ist die Summe der Volumina der \tilde{Q}_i kleiner als ε , und b) ist bewiesen.

Zum Beweis von c) (aus a) können wir fast genauso vorgehen; jetzt wählen wir aber die \tilde{a}_k, \tilde{b}_k so, daß sie rationale Zahlen sind und im übrigen dieselben Ungleichungen erfüllen wir oben. Dann ist

$$\tilde{Q}_i = [\tilde{a}_1, \tilde{b}_1] \times \dots \times [\tilde{a}_n, \tilde{b}_n]$$

zwar kein Würfel, aber eine Vereinigung volumenfremder Würfel: Bezeichnet N den Hauptnenner aller $\tilde{b}_k - \tilde{a}_k$, so können wir jeden Quader \tilde{Q}_i unterteilen in Würfel der Kantenlänge $1/N$, und die Summe der Volumina der sämtlichen so erhaltenen Würfel ist gleich der Summe der Volumina aller \tilde{Q}_i , also kleiner als ε . ■

Einige weitere elementare Eigenschaften von Nullmengen sind im folgenden Lemma zusammengefaßt:

Lemma: a) Jede Teilmenge einer Nullmenge ist eine Nullmenge.

b) Die Vereinigung endlich oder abzählbar vieler Nullmengen ist wieder eine Nullmenge.

Beweis: a) Ist $Y \subseteq Z$ und $Z \subseteq \mathbb{R}^n$ eine Nullmenge, so gibt es für jedes $\varepsilon > 0$ eine Folge von Quadern Q_k , so daß Z in der Vereinigung aller Q_k liegt und die Summe der Volumina $\mu(Q_k)$ kleiner ist als ε . Da Y in Z enthalten ist, überdecken die Q_k auch Y .

b) Z_1, Z_2, \dots seien endlich oder abzählbar viele Nullmengen. Zu jedem $\varepsilon > 0$ gibt es dann für jedes Z_i Quader Q_{i1}, Q_{i2}, \dots derart, daß Z_i in der Vereinigung dieser Quader enthalten ist, und die Summe der Volumina $\mu(Q_{ij})$ ist kleiner als $\varepsilon/2^i$. Nehmen wir nun alle Q_{ij} zusammen, so ist die Vereinigung der Z_i darin enthalten und die Summe der Volumina ist kleiner als $\sum_{i=1}^{\infty} \frac{\varepsilon}{2^i} = \varepsilon$. Somit ist auch die Vereinigung der Z_i eine Nullmenge. ■

Lemma: Eine Nullmenge hat keine innere Punkte. Insbesondere ist keine nichtleere offene Menge Nullmenge.

Beweis: Ein Punkt $x \in Z \subseteq \mathbb{R}^n$ ist bekanntlich genau dann ein innerer Punkt, wenn es ein $\varepsilon > 0$ gibt, so daß auch alle Punkte $y \in \mathbb{R}^n$ mit $\|x - y\| < \varepsilon$ in Z liegen. Da alle Normen auf \mathbb{R}^n äquivalent sind, gibt es dann auch ein $\varepsilon > 0$, so daß dies für die Maximumnorm gilt; Z enthält also für ein gewisses $\varepsilon > 0$ den offenen Würfel aus allen $y \in \mathbb{R}^n$ mit $\|x - y\|_{\infty} < \varepsilon$, insbesondere also auch den abgeschlossenen Würfel aus allen $y \in \mathbb{R}^n$ mit $\|x - y\|_{\infty} \leq \varepsilon/2$. Dieser ist aber, wie wir uns oben überlegt haben, keine Nullmenge. Damit kann auch Z keine Nullmenge sein, denn nach dem vorigen Lemma ist jede Teilmenge einer Nullmenge selbst eine Nullmenge. Da jeder Punkt einer offenen Menge innerer Punkt ist, kann damit auch keine nichtleere offene Menge Nullmenge sein. ■

Wie wir oben gesehen haben, ist die durch $x_n = 0$ definierte Hyperebene in \mathbb{R}^n eine Nullmenge. Entsprechend sollte man erwarten, daß jede Teilmenge einer Dimension echt kleiner n Nullmenge ist. Bevor wir so etwas beweisen können, müßten wir aber zuallererst einmal wissen, was die Dimensionen einer beliebigen Teilmenge von \mathbb{R}^n ist. Wir können

nicht einfach sagen, eine Teilmenge $A \subseteq \mathbb{R}^n$ sei m -dimensional, wenn sie sich als Bild einer Teilmenge $A = \varphi(B)$ einer injektiven Abbildung einer Teilmenge $B \subseteq \mathbb{R}^m$ darstellen läßt, denn durch eine Variante des aus Kapitel 1 bekannten ersten CANTORSchen Diagonalverfahrens läßt sich zeigen, daß \mathbb{R} und \mathbb{R}^2 gleichmächtig sind, daß es also eine bijektive Abbildung $\mathbb{R} \rightarrow \mathbb{R}^2$ gibt. Mit vollständiger Induktion folgt dann leicht, daß es für beliebige $n, m \in \mathbb{N}$ auch eine bijektive Abbildung $\mathbb{R}^m \rightarrow \mathbb{R}^n$ gibt.

Als nächsten Versuch könnten wir verlangen, daß die Abbildung φ stetig sein soll. Leider bekommen wir auch damit Probleme, zumindest wenn wir auch surjektive Abbildungen an Stelle von bijektiven zulassen: 1890 konstruierte GUISEPPE PEANO eine stetige surjektive Abbildung $\varphi: [0, 1] \rightarrow [0, 1] \times [0, 1]$. Die Idee dazu, in einer 1891 von DAVID HILBERT angegebenen Vereinfachung, ist folgende: Wir teilen im ersten Schritt sowohl das Intervall $I = [0, 1]$ als auch das Quadrat $Q = [0, 1] \times [0, 1]$ in vier gleich große Teilintervalle bzw. -quadrate. Bei I soll das i -te Teilintervall einfach das von $(i - 1)/4$ bis $i/4$ sein, bei Q wird die Sache etwas komplizierter, denn wir wollen eine Unterteilung, bei der das i -te Quadrat stets benachbart zum $(i + 1)$ -ten ist. Das können wir zum Beispiel durch das Schema $\begin{matrix} Q_2 & Q_3 \\ Q_1 & Q_4 \end{matrix}$ erreichen.

Für $k > 1$ gehen wir von der vorhandenen Unterteilung des k -ten Schritts aus und unterteilen wieder jedes einzelne Intervall oder Quadrat in vier Teile, so daß Quadrate mit aufeinanderfolgender Nummer benachbart sind. Im zweiten Schritt könnten wir dazu nach dem Schema

$$\begin{matrix} Q_6 & Q_7 & Q_{10} & Q_{11} \\ Q_5 & Q_8 & Q_9 & Q_{12} \\ Q_4 & Q_3 & Q_{14} & Q_{13} \\ Q_1 & Q_2 & Q_{15} & Q_{16} \end{matrix}$$

vorgehen: Die ersten vier Quadrate der k -ten Unterteilung sollen im ersten der $(k - 1)$ -ten liegen, die nächsten vier im zweiten, und so weiter. Damit auch noch nach der Unterteilung aufeinanderfolgende Quadrate benachbart sind, müssen wir bei der Vierteilung darauf achten, daß das jeweils vierte Teilquadrat an das nächste Quadrat der vorigen Stufe angrenzt; die Numerierung der neuen Quadrate kann also nicht immer nach dem Schema $\begin{matrix} Q_2 & Q_3 \\ Q_1 & Q_4 \end{matrix}$ erfolgen, sondern wir müssen beispielsweise

auch die relative Reihenfolge $\begin{matrix} Q_4 & Q_3 \\ Q_1 & Q_2 \end{matrix}$ verwenden.

Jeder Punkt $x \in I$ kann nun definiert werden durch eine Intervallschachtelung aus Intervallen J_k , wobei J_k eines der Teilintervalle der k -ten Stufe ist. (Bei Punkten der Form $i/4^k$ haben wir hier mehrere Möglichkeiten, was jedoch für das Ergebnis irrelevant ist.) Die Abbildung wird nun so konstruiert, daß das i -te Intervall der k -ten Unterteilung auf das i -te Quadrat der k -ten Unterteilung von Q abgebildet werden soll. Man überlegt sich leicht, daß wir dadurch eine Folge von ineinander enthaltenen Quadraten bekommen, deren waagrechte und senkrechte Kanten jeweils Intervallschachtelungen für reelle Zahlen y und z sind. Mit der Definition $\varphi(x) = (y, z)$ ist die HILBERTSche Kurve erklärt, und man kann ohne größere Schwierigkeiten zeigen, daß sie alle behaupteten Eigenschaften hat. φ ist übrigens auch ein Beispiel einer stetigen, aber nirgends differenzierbaren Funktion.



GUISEPPE PEANO (1858–1932) war Sohn eines Landarbeiters und wuchs auf einem Bauernhof nahe Cuneo im Piemont auf. 1870 brachte ihn ein Bruder seiner Mutter nach Turin, wo er weiterführende Schulen und schließlich die Universität besuchte. Dort wurde er 1880 Assistent und 1890 Professor. Bekannt ist er unter durch einen Existenzsatz für Lösungen von Differentialgleichungen und ein Beispiel für nichteindeutige Lösungen. Die berühmten PEANO-Axiome für die natürlichen Zahlen veröffentlichte er 1889, und zwar aus unerfindlichen Gründen in lateinischer Sprache. Später beschäftigte er sich vor allem mit Logik.



DAVID HILBERT (1862–1943) wurde in Königsberg geboren, wo er auch zur Schule und zur Universität ging. Er promovierte dort 1885 mit einem Thema aus der Invariantentheorie, habilitierte sich 1886 und bekam 1893 einen Lehrstuhl. 1895 wechselte er an das damalige Zentrum der deutschen wie auch internationalen Mathematik, die Universität Göttingen, wo er bis zu seiner Emeritierung im Jahre 1930 lehrte. Seine Arbeiten umfassen ein riesiges Spektrum aus unter anderem Invariantentheorie, Zahlentheorie, Geometrie, Funktionalanalysis, Logik und Grundlagen der Mathematik sowie auch zur Relativitätstheorie. Er gilt als einer der Väter der modernen Algebra.

Das Beispiel der von PEANO und HILBERT konstruierten Kurven zeigt insbesondere, daß das Bild einer Nullmenge unter einer stetigen Abbildung keine Nullmenge mehr sein muß: Identifizieren wir das Einheitsintervall mit der Menge aller Punkte $(x, 0) \in \mathbb{R}^2$ mit $0 \leq x \leq 1$, so ist dies eine Nullmenge; das Bild dieser Menge unter der oben konstruierten stetigen Abbildung $(x, 0) \mapsto \varphi(x)$ aber ist ein Quadrat mit Seitenlänge eins, also definitiv keine Nullmenge.

Solche Beispiele sind nicht mehr möglich, wenn wir an Stelle der bloßen Stetigkeit eine sogenannte LIPSCHITZ-Bedingung fordern:

Definition: Die Funktion $f: Z \rightarrow \mathbb{R}^n$ auf der Teilmenge $Z \subseteq \mathbb{R}^n$ erfüllt eine LIPSCHITZ-Bedingung mit LIPSCHITZ-Konstante $\lambda > 0$, wenn für alle $x, y \in Z$ gilt:

$$\|f(x) - f(y)\|_\infty \leq \lambda \|x - y\|_\infty .$$

Ist Z eine offene Menge, so ist eine solche Funktion nicht nur stetig, sondern sogar gleichmäßig stetig, denn für jedes $\varepsilon > 0$ gilt für $\delta = \varepsilon/\lambda$, daß $\|f(x) - f(y)\|_\infty < \varepsilon$ ist wann immer $\|x - y\|_\infty < \delta$ ist.



RUDOLF OTTO SIGISMUND LIPSCHITZ (1832–1903) wurde in Königsberg geboren und starb in Bonn; er lehrte an den Universitäten Berlin, Breslau und Bonn. Am besten bekannt ist er durch die gerade definierte LIPSCHITZ-Bedingung die für die Existenz und Eindeutigkeit der Lösungen von Differentialgleichungen eine große Rolle spielt. Sein Hauptarbeitsgebiet waren Differentialgleichungen und andere Hilfsmittel der mathematischen Physik, beispielsweise Matrizengruppen. Seine Arbeiten über dynamische Systeme haben wichtige Anwendungen in der Himmelsmechanik.

Lemma: Erfüllt die Abbildung $f: Z \rightarrow \mathbb{R}^n$ eine LIPSCHITZ-Bedingung und ist $Z \subset \mathbb{R}^n$ eine Nullmenge, so ist auch $f(Z)$ eine Nullmenge.

Beweis: λ sei eine LIPSCHITZ-Konstante für f und $Q \subset \mathbb{R}^n$ sei ein kompakter Würfel. Wir wollen uns als erster überlegen, daß das Bild $f(Q \cap Z)$ in einem Würfel \tilde{Q} liegt mit $\mu(\tilde{Q}) \leq (2\lambda)^n \mu(Q)$. Falls $Z \cap Q$ leer ist, gibt es nichts zu beweisen; andernfalls sei x ein Punkt aus dem

Durchschnitt. Ist a die Kantenlänge von Q , so erfüllt jeder Punkt $y \in Q$ die Ungleichung $\|x - y\|_\infty \leq a$, denn in keiner der n Koordinaten können sich zwei Punkte eines Würfels um mehr als a unterscheiden. Wegen der LIPSCHITZ-Bedingung ist daher für alle $y \in Q \cap Z$

$$\|f(x) - f(y)\|_\infty \leq \lambda \|x - y\|_\infty = \lambda a .$$

Somit liegen alle diese Punkte im Würfel

$$\tilde{Q} = \{z \in \mathbb{R}^n \mid \|f(x) - z\|_\infty \leq \lambda a\}$$

mit Mittelpunkt $f(x)$ und Kantenlänge $2\lambda a$. Sein Volumen ist

$$\mu(\tilde{Q}) = (2\lambda a)^n = (2\lambda)^n a^n = (2\lambda)^n \mu(Q) ,$$

wie behauptet.

Um nun zu zeigen, daß $f(Z)$ eine Nullmenge ist, geben wir uns ein $\varepsilon > 0$ vor. Da Z eine Nullmenge ist, können wir Würfel Q_1, Q_2, \dots finden, so daß Z in der Vereinigung dieser Würfel liegt und die Summe von deren Volumina kleiner ist als $\varepsilon/(2\lambda)^n$. Für jeden dieser Würfel liegt $f(Z \cap Q_i)$ in einer Würfel mit Volumen kleiner $(2\lambda)^n \mu(Q_i)$, also wird $f(Z)$ von diesen Würfeln überdeckt und ihr Gesamtvolumen ist kleiner als ε . Damit ist $f(Z)$ als Nullmenge erkannt. ■

LIPSCHITZ-Konstanten sind oft nicht leicht direkt zu finden; darin unterscheiden sie sich nicht von den Konstanten aus dem BANACHSchen Fixpunktsatz. Dort haben wir in den Übungen gesehen, daß für differenzierbare Funktionen der Mittelwertsatz der Differentialrechnung hier oft nützlich ist. Genauso wird das auch hier sein; bevor wir das zeigen können, müssen wir uns zunächst überlegen, daß sich beliebige offene Mengen durch Quader „ausschöpfen“ lassen:

Lemma: Jede offene Teilmenge $U \subseteq \mathbb{R}^n$ läßt sich als Vereinigung höchstens abzählbar vieler volumenfremder (kompakter) Quader schreiben.

Beweis: Wir bezeichnen einen Würfel der Form

$$\left[\frac{k_1}{2^g}, \frac{k_1 + 1}{2^g} \right] \times \dots \times \left[\frac{k_n}{2^g}, \frac{k_n + 1}{2^g} \right] \subset \mathbb{R}^n$$

als *Baustein* der Generation $g \in \mathbb{N}$. Offensichtlich ist \mathbb{R}^n für jedes g die volumenfremde Vereinigung aller Bausteine der Generation g , und

es gibt abzählbar viele solcher Bausteine. E_g sei die Vereinigung aller Bausteine der Generation g , die vollständig in U enthalten sind; das ist offensichtlich für jedes g eine Teilmenge von U , und auch die Vereinigung aller E_g liegt in U .

Tatsächlich ist die Vereinigung aller E_g sogar gleich U , denn jeder Punkt $x \in U$ ist innerer Punkt, es gibt also ein $\varepsilon > 0$, so daß jeder Punkt $y \in \mathbb{R}^n$ mit $\|y - x\|_\infty < \varepsilon$ in U liegt. Wir wählen ein $g \in \mathbb{N}$ so, daß $2^g > 1/\varepsilon$ ist und setzen $k_i = \lceil 2^g x_i \rceil$, wobei x_i die i -te Koordinate von x bezeichnet. Dann liegt x im Baustein $[\frac{k_1}{2^g}, \frac{k_1+1}{2^g}] \times \dots \times [\frac{k_n}{2^g}, \frac{k_n+1}{2^g}]$, und da dessen Kantenlänge 2^{-g} kleiner ist als ε ist $\|y - x\|_\infty < \varepsilon$ für alle y aus diesem Baustein. Somit liegt dieser ganz in U , und $x \in E_g$.

Um U als Vereinigung volumenfremder Quader darzustellen, beginnen wir mit der Menge aller Bausteinen der ersten Generation, die ganz in U enthalten sind, und nehmen dann im g -ten Schritt, $g \geq 2$, alle Bausteine der Generation g dazu, die erstens ganz in U liegen und zweitens nicht in einem der bereits vorhandenen Quader enthalten sind. Nach dem g -ten Schritt ist die Vereinigung aller bis dahin betrachteter Quader gleich E_g ; insgesamt erhalten wir also die Vereinigung aller E_g , das heißt die Menge U . ■

Mit diesem Lemma können wir nun das vorige in eine Form bringen, in der es deutlich einfacher angewendet werden kann:

Lemma: $f: D \rightarrow \mathbb{R}$ sei eine stetig differenzierbare Abbildung auf der offenen Menge $D \subseteq \mathbb{R}^n$, und $Z \subset D$ sei eine Nullmenge. Dann ist auch $f(Z)$ eine Nullmenge.

Beweis: Nach dem gerade bewiesenen Lemma läßt sich D als abzählbare Vereinigung kompakter Quader Q_k schreiben. Auf jedem dieser Quader erfüllt f eine LIPSCHITZ-Bedingung, denn für $x, y \in Q_k$ können wir die Funktion $g(t) = f(x + t(y - x))$ betrachten, und nach dem Mittelwertsatz der Differentialrechnung ist

$$f(y) - f(x) = g(1) - g(0) = \frac{g(1) - g(0)}{1 - 0} = g'(\tau)$$

für ein $\tau \in [0, 1]$. Nach der Kettenregel ist

$$g'(\tau) = \nabla f(x + \tau(y - x)) \cdot (y - x),$$

also insgesamt

$$\begin{aligned} \|f(y) - f(x)\|_\infty &= \|\nabla f(x + \tau(y - x)) \cdot (y - x)\|_\infty \\ &\leq \max\{\|\nabla f(z)\|_\infty \mid z \in Q\} \cdot \|y - x\|_\infty, \end{aligned}$$

wobei das Maximum existiert, da wir f als *stetig* differenzierbar vorausgesetzt haben und den Gradienten auf einer kompakten Menge betrachten. Somit genügt die Einschränkung von f auf Q_k einer LIPSCHITZ-Bedingung mit diesem Maximum als Konstante, und damit ist $f(Q_k \cap Z)$ nach dem vorletzten Lemma eine Nullmenge. $f(Z)$ ist die Vereinigung dieser abzählbar vielen Nullmengen $f(Q_k \cap Z)$ und daher ebenfalls eine Nullmenge. ■

Korollar: $D \subset \mathbb{R}^m$ sei eine offene Menge und $f: D \rightarrow \mathbb{R}^n$ eine stetig differenzierbare Abbildung. Ist $m < n$, so ist $f(Z)$ für jede Teilmenge $Z \subseteq D$ eine Nullmenge in \mathbb{R}^n .

Beweis: Wir betten \mathbb{R}^m ein in \mathbb{R}^n , indem wir einfach die letzten $n - m$ Koordinaten auf Null setzen. Das Bild von \mathbb{R}^m liegt dann insbesondere in der Hyperebene $x_n = 0$ von \mathbb{R}^n , ist also eine Nullmenge; erst recht ist die Einbettung \tilde{A} von A eine Nullmenge in \mathbb{R}^n . Wir definieren eine neue Abbildung $g: D \times \mathbb{R}^{n-m} \rightarrow \mathbb{R}^n$, die jedem Punkt $x = (x_1, \dots, x_m)$ den Punkt $f(x_1, \dots, x_m)$ zuordnet; die letzten $n - m$ Koordinaten werden also einfach ignoriert. Dann ist $f(A) = g(\tilde{A})$ nach dem vorigen Lemma eine Nullmenge, wie behauptet. ■

Damit sind also Kurven in \mathbb{R}^2 sowie Kurven und Flächen in \mathbb{R}^3 Nullmengen, falls sie eine stetig differenzierbare Parametrisierung zulassen, genauso alle Vereinigungen endlich oder abzählbar vieler solcher Mengen. Entsprechendes gilt natürlich auch in höheren Dimensionen.

Für die Zwecke des noch zu definierenden LEBESGUE-Integrals wollen wir Nullmengen weitgehend ignorieren; um dies sprachlich kurz und einfach ausdrücken zu können, führen wir eine neue Sprechweise ein:

Definition: $a)$ Zwei Funktionen $f, g: D \rightarrow \mathbb{R}$ auf $D \subseteq \mathbb{R}^n$ heißen *fast überall gleich*, wenn es eine Nullmenge $Z \subseteq D$ gibt, so daß $f(x) = g(x)$ für alle $x \in D \setminus Z$ ist.

b) Eine Folge $(f_k)_{k \in \mathbb{N}}$ von Funktionen $f_k: D \rightarrow \mathbb{R}$ konvergiert fast überall punktweise gegen die Funktion $f: D \rightarrow \mathbb{R}$, wenn es eine Nullmenge $Z \subset D$ gibt, so daß

$$\lim_{k \rightarrow \infty} f_k(x) = f(x) \quad \text{für alle } x \in D \setminus Z.$$

c) Die Folge $(f_k)_{k \in \mathbb{N}}$ von Funktionen $f_k: D \rightarrow \mathbb{R}$ konvergiert fast überall punktweise, wenn es eine Funktion $f: D \rightarrow \mathbb{R}$ gibt, gegen die sie fast überall punktweise konvergiert.

Die Strategie zur Definition des LEBESGUE-Integrals ist nun einfach zu formulieren:

Definition: a) Eine Folge $(f_k)_{k \in \mathbb{N}}$ von Funktionen $f_k \in K_1^0(\mathbb{R}^n, \mathbb{R})$ heißt *approximierend* für die Funktion $f: \mathbb{R}^n \rightarrow \mathbb{R}$, wenn sie erstens eine CAUCHY-Folge ist und zweitens fast überall punktweise gegen f konvergiert.

b) Eine Funktion $f: \mathbb{R}^n \rightarrow \mathbb{R}$ heißt *integrierbar*, wenn es dazu eine approximierende Folge $(f_k)_{k \in \mathbb{N}}$ von Funktionen aus $K_1^0(\mathbb{R}^n, \mathbb{R})$ gibt; in diesem Fall bezeichnen wir

$$\int_{\mathbb{R}^n} f \stackrel{\text{def}}{=} \lim_{k \rightarrow \infty} \int_{\mathbb{R}^n} f_k$$

als das LEBESGUE-Integral über f .



HENRI LÉON LEBESGUE (1875–1941) studierte ab 1894 an der Ecole Normale Supérieure. Nach seinem ersten Abschluß 1897 studierte er noch zwei Jahre auf eigene Faust weiter, hauptsächlich in der Bibliothek, bis er 1899 eine Stelle als Gymnasiallehrer in Nancy erhielt. Dort schrieb er 1901 seine Arbeit über die Definition des LEBESGUE-Integrals. 1902 promovierte er mit einer weiteren Arbeit in Paris und bekam daraufhin eine Stelle an der Universität von Rennes. 1906 wechselte er nach Poitiers, 1910 an die Sorbonne in Paris. Ab 1921 war er Professor am Collège de France. Seine Arbeiten befassen sich mit Integrationstheorie, Potentialtheorie, Variationsrechnung, Topologie und anderen Gebieten.

Auch wenn es auf den ersten Blick so aussieht, haben wir damit noch nicht wirklich das LEBESGUE-Integral definiert: *Erstens* wissen wir noch nicht, ob der Grenzwert der Folge von Integralen existiert, und *zweitens*

wissen wir noch nicht, daß alle approximierenden Folgen für f zum selben Grenzwert führen.

Das erste Problem ist leicht zu lösen: Wegen der Linearität und der Monotonieeigenschaft des Integrals für Funktionen mit kompaktem Träger gilt für jede approximierende Folge $(f_k)_{k \in \mathbb{N}}$

$$\left| \int_{\mathbb{R}^n} f_\ell - \int_{\mathbb{R}^n} f_k \right| = \left| \int_{\mathbb{R}^n} (f_\ell - f_k) \right| \leq \int_{\mathbb{R}^n} |f_\ell - f_k| = \|f_\ell - f_k\|_1.$$

Da eine approximierende Folge nach Definition eine CAUCHY-Folge bezüglich der L^1 -Norm ist, gibt es zu jedem $\varepsilon > 0$ ein $N \in \mathbb{N}$, so daß die rechte Seite kleiner als ε ist für $k, \ell \geq N$. Damit gilt dasselbe erst recht für die linke Seite, d.h. die Folge der Integrale über die f_k ist eine CAUCHY-Folge reeller Zahlen und konvergiert somit nach dem CAUCHYSchen Konvergenzkriterium.

Das zweite Problem erfordert mehr Arbeit. Wir betrachten zwei approximierende Folgen $(f_k)_{k \in \mathbb{N}}$ und $(g_k)_{k \in \mathbb{N}}$ für dieselbe Funktion $f: \mathbb{R}^n \rightarrow \mathbb{R}$ und überlegen uns als erstes, daß $(f_k - g_k)_{k \in \mathbb{N}}$ eine approximierende Folge für die Nullfunktion ist.

Dazu müssen wir zunächst zeigen, daß $(f_k - g_k)_{k \in \mathbb{N}}$ eine CAUCHY-Folge ist: Zu jedem $\varepsilon > 0$ gibt es natürliche Zahlen N_1, N_2 , so daß

$$\|f_\ell - f_k\|_1 < \frac{\varepsilon}{2} \quad \text{für alle } k, \ell \geq N_1 \quad \text{und}$$

$$\|g_\ell - g_k\|_1 < \frac{\varepsilon}{2} \quad \text{für alle } k, \ell \geq N_2.$$

Für $k, \ell \geq \max\{N_1, N_2\}$ ist daher nach der Dreiecksungleichung

$$\begin{aligned} \|(f_\ell - g_\ell) - (f_k - g_k)\|_1 &= \|(f_\ell - f_k) - (g_\ell - g_k)\|_1 \\ &\leq \|f_\ell - f_k\|_1 + \|g_\ell - g_k\|_1 < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon. \end{aligned}$$

Weiter müssen wir noch zeigen, daß es eine Nullmenge Z gibt, außerhalb derer die Folge der $f_k(x) - g_k(x)$ gegen Null konvergiert. Wir wissen, daß die Folge der $f_k(x)$ für x außerhalb einer Nullmenge Z_1 gegen $f(x)$ konvergiert; die der $g_k(x)$ konvergiert außerhalb einer Nullmenge Z_2 gegen diesen Wert. Auch die Vereinigung $Z = Z_1 \cup Z_2$ ist eine Nullmenge, und für $x \notin Z$ konvergiert die Folge der $f_k(x) - g_k(x)$ nach unseren Rechenregeln für Grenzwerte aus dem letzten Semester

gegen $f(x) - f(x) = 0$. Damit ist gezeigt, daß $(f_k - g_k)_{k \in \mathbb{N}}$ eine approximierende Folge für die Nullfunktion ist.

Wegen der Linearität der Integration ist

$$\begin{aligned} \lim_{k \rightarrow \infty} \int_{\mathbb{R}^n} (f_k - g_k) &= \lim_{k \rightarrow \infty} \left(\int_{\mathbb{R}^n} f_k - \int_{\mathbb{R}^n} g_k \right) \\ &= \lim_{k \rightarrow \infty} \int_{\mathbb{R}^n} f_k - \lim_{k \rightarrow \infty} \int_{\mathbb{R}^n} g_k ; \end{aligned}$$

wenn wir zeigen können, daß der Grenzwert links verschwindet, ist daher

$$\lim_{k \rightarrow \infty} \int_{\mathbb{R}^n} f_k = \lim_{k \rightarrow \infty} \int_{\mathbb{R}^n} g_k .$$

Die Folge der $(f_k - g_k)$ ist eine approximierende Folge für die Nullfunktion; wir sind also fertig, wenn wir zeigen können, daß allgemein für jede approximierende Folge $(f_k)_{k \in \mathbb{N}}$ zur Nullfunktion die Folge der $\int_{\mathbb{R}^n} f_k$ eine Nullfolge ist. Das Problem dabei ist, daß die Folge der Funktionswerte $f_k(x)$ nicht für jeden Punkt $x \in \mathbb{R}^n$ gegen Null konvergieren muß; wir müssen uns überlegen, daß diese Abweichungen für das Integral irrelevant sind.

Dazu verallgemeinern wir zunächst den Begriff der Nullmenge:

Definition: A sei eine Teilmenge von \mathbb{R}^n . Wir betrachten die Menge M_A aller reeller Zahlen s mit der Eigenschaft, daß A durch eine abzählbare Menge von Quadern Q_k überdeckt werden kann derart, daß $\sum_{k \geq 1} \mu(Q_k) < s$ ist. Falls diese Menge leer ist, definieren wir das äußere Maß $\mu^*(A)$ als ∞ ; andernfalls ist $\mu^*(A)$ das Infimum der Menge M_A .

Offensichtlich ist $\mu^*(A)$ stets größer oder gleich Null, denn M_A kann keine negativen Zahlen enthalten. Ist $\mu^*(A) = 0$, so gibt es zu jedem $\varepsilon > 0$ eine Überdeckung durch Quader Q_k mit $\sum_{k \geq 1} \mu(Q_k) < \varepsilon$, die Menge A ist also eine Nullmenge; umgekehrt ist auch das äußere Maß einer jeden Nullmenge gleich Null. Das äußere Maß einer Teilmenge kann selbstverständlich nicht größer sein als das der größeren Menge, denn jede Überdeckung dieser Menge ist erst recht eine Überdeckung der Teilmenge. Außerdem ist klar, daß für zwei Teilmengen $A, B \subseteq \mathbb{R}^n$

stets $\mu^*(A \cup B) \leq \mu^*(A) + \mu^*(B)$ sein muß, wobei wir eine Summe, in der (mindestens) einer der Summanden ∞ ist, als ∞ interpretieren.

Nicht ganz so offensichtlich ist die Tatsache, daß dies auch für abzählbare Vereinigungen gilt:

$$\mu^* \left(\bigcup_{k=1}^{\infty} A_k \right) \leq \sum_{k=1}^{\infty} \mu^*(A_k),$$

wobei die Summe gleich ∞ sein soll, wenn sie divergiert oder wenn mindestens einer der Summanden gleich ∞ ist. In diesen beiden Fällen ist die Ungleichung trivialerweise erfüllt; andernfalls beachten wir, daß wir für jede der Mengen A_k und jedes $\varepsilon > 0$ eine Überdeckung von A_k durch Quader finden können, so daß die Summe von deren Volumina kleiner ist als $\mu^*(A_k) + \varepsilon/2^k$. Nehmen wir alle diese Quader zusammen, so überdecken sie die Vereinigung aller A_k , und die Summe ihrer Volumina ist kleiner als

$$\sum_{k=1}^{\infty} \left(\mu^*(A_k) + \frac{\varepsilon}{2^k} \right) = \sum_{k=1}^{\infty} \mu^*(A_k) + \sum_{k=1}^{\infty} \frac{\varepsilon}{2^k} = \sum_{k=1}^{\infty} \mu^*(A_k) + \varepsilon .$$

Somit kann das Infimum über die Menge aller $s \in \mathbb{R}$ derart, daß die Vereinigung der A_k eine Überdeckung durch Quader mit Gesamtvolumen höchstens s hat, nicht kleiner sein als die Summe der $\mu^*(A_k)$, und genau das war zu beweisen. ■

Als nächstes wollen wir uns überlegen, daß die L^1 -Norm einer Funktion mit kompaktem Träger zwar keine obere Schranke für die Funktionswerte liefern kann, wohl aber eine obere Schranke für das äußere Maß jener Teilmenge von \mathbb{R}^n , auf der die Funktion Werte oberhalb einer gegebenen Schranke annimmt. Die entsprechende Ungleichung von TSCHEBYSCHEFF gilt, wie wir bald sehen werden, auch noch unter deutlich schwächeren Voraussetzungen als im folgenden Lemma; sie ist auch ein wichtiges Hilfsmittel der Statistik.

Lemma: Die Funktion $f \in K_1^0(\mathbb{R}^n, \mathbb{R})$ nehme keine negativen Werte an. Dann ist für jede reelle Zahl $c > 0$

$$\mu^* \left(\{x \in \mathbb{R}^n \mid f(x) > c\} \right) \leq \frac{\|f\|_1}{c} .$$

Beweis: Als Urbild der offenen Menge $\{x \in \mathbb{R} \mid x > c\}$ unter der stetigen Abbildung f ist $U = \{x \in \mathbb{R}^n \mid f(x) > c\}$ eine offene Teilmenge von \mathbb{R}^n ; da sie im (kompakten) Träger von f liegt ist sie außerdem beschränkt; es gibt also einen Quader Q mit $U \subset Q$. Wie wir vor kurzem gesehen haben, läßt sich U als abzählbare Vereinigung von Quadern Q_k schreiben; nach den bekannten Eigenschaften des Quaderintegrals ist dann für jedes $m \in \mathbb{N}$

$$\|f\|_1 = \int_{\mathbb{R}^n} f = \int_Q f \geq \sum_{k=1}^m \int_{Q_k} f \geq \sum_{k=1}^m c \mu(Q_k) = c \sum_{k=1}^m \mu(Q_k).$$

Für $m \rightarrow \infty$ folgt $\|f\|_1 \geq c \sum_{k=1}^{\infty} \mu(Q_k)$ und damit $\sum_{k=1}^{\infty} \mu(Q_k) \leq \frac{\|f\|_1}{c}$.

Die Menge U kann somit überdeckt werden durch Quader deren Gesamtvolumen höchstens gleich der rechten Seite ist; damit muß auch das äußere Maß von U kleiner oder gleich dieser Schranke sein. ■



PAFNUTI LWOWITSCH TSCHEBYSCHJEFF (1821–1894) ist die in Deutschland übliche Transkription von Пафну́ты Львович Чебышев; im Englischen schreibt man heute meist PAFNUTY LVOVICH CHEBYSHEV. Er studierte Mathematik in Moskau und publizierte bereits während seines Studiums in deutschen und französischen Fachzeitschriften. 1847 bekam er eine Stelle an der Universität von St. Petersburg, wo er bis 1882 lehrte, ab 1860 als Professor. Seine Ergebnisse spielen noch heute eine große Rolle in der Wahrscheinlichkeitstheorie, der Zahlentheorie (insbesondere Primzahlverteilung), der Approximationstheorie und in anderen Gebieten der Mathematik.

Als erste Anwendung der TSCHEBYSCHJEFFSchen Ungleichung wollen wir zeigen, daß eine CAUCHY-Folge bezüglich der L^1 -Norm zumindest eine Teilfolge hat, die abgesehen von einer kleinen Teilmenge auch punktweise und sogar gleichmäßig konvergiert, genauer:

Lemma: Jede CAUCHY-Folge $(f_k)_{k \in \mathbb{N}}$ aus $K_1^0(\mathbb{R}^n, \mathbb{R})$ hat eine Teilfolge, die fast überall punktweise konvergiert. Außerdem gibt es für jedes $\varepsilon > 0$ eine Teilmenge $Z \subset \mathbb{R}^n$ mit $\mu^*(Z) < \varepsilon$, so daß die Teilfolge auf $\mathbb{R}^n \setminus Z$ sogar gleichmäßig konvergiert.

Beweis: Wir konstruieren die Folge der Indizes ν_k für die Teilfolge rekursiv: ν_1 sei so gewählt, daß $\|f_q - f_p\|_1 < \frac{1}{4}$ ist für alle $p, q \geq \nu_1$. Die folgenden ν_k werden so gewählt, daß

$$\|f_q - f_p\|_1 < 4^{-k} \text{ für alle } p, q \geq \nu_k \text{ und } \nu_k > \nu_{k-1}$$

ist. Mit $g_k \stackrel{\text{def}}{=} f_{\nu_k}$ ist dann eine Teilfolge der gegebenen Folge definiert, für die gilt:

$$\|g_\ell - g_k\|_1 < 4^{-k} \text{ für alle } \ell \geq k.$$

Als Kandidat für eine Grenzfunktion betrachten wir die Funktion $g: \mathbb{R}^n \rightarrow \mathbb{R}$ mit

$$g(x) = g_1(x) + \sum_{k=1}^{\infty} (g_{k+1}(x) - g_k(x)).$$

Ihre $(r-1)$ -te Teilsumme ist

$$g_1(x) + \sum_{k=1}^{r-1} (g_{k+1}(x) - g_k(x)) = g_r(x),$$

denn jedes g_k mit $k < r$ steht in dieser endlichen Summe einmal mit positivem und einmal mit negativem Vorzeichen. Falls die Reihe konvergiert, ist sie also in der Tat ein Kandidat für eine Grenzfunktion.

Hier kommt nun die Ungleichung von TSCHEBYSCHJEFF ins Spiel: Danach ist das äußere Maß der Menge

$$Y_k = \{x \in \mathbb{R}^n \mid |g_{k+1}(x) - g_k(x)| > 2^{-k}\}$$

höchstens gleich

$$\frac{\|g_{k+1} - g_k\|_1}{2^{-k}} < \frac{4^{-k}}{2^{-k}} = 2^{-k}.$$

Das äußere Maß der Vereinigung Z_k aller Y_ℓ mit $\ell \geq k$ ist daher kleiner als $2^{-k} + 2^{-k-1} + \dots = 2^{1-k}$. Für $x \notin Z_k$ ist nach Definition dieser Menge $|g_{\ell+1}(x) - g_\ell(x)| < 2^{-\ell}$, also ist die geometrische Reihe $\sum 2^{-\ell}$ eine konvergente Majorante der definierenden Reihe von $g(x)$. Somit konvergiert diese Reihe auf $\mathbb{R}^n \setminus Z_k$ gleichmäßig. Da wir zu jedem $\varepsilon > 0$ ein $k \in \mathbb{N}$ finden können mit $2^{1-k} < \varepsilon$, ist damit die Behauptung über die gleichmäßige Konvergenz bewiesen.

Bleibt noch die punktweise Konvergenz außerhalb einer Nullmenge. Der Durchschnitt Z aller Z_k ist in jeder der Mengen Z_k enthalten, also ist $\mu^*(Z) < 2^{1-k}$ für alle $k \in \mathbb{N}$ und Z somit eine Nullmenge. Für $x \notin Z$ gibt es Mengen Z_k , in denen x nicht enthalten ist, und wie wir gerade gesehen haben, konvergiert die Reihe in jeder der Mengen $\mathbb{R}^n \setminus Z_k$. Somit konvergiert sie für alle $x \in \mathbb{R}^n \setminus Z$, also fast überall. ■

Damit haben wir alles zusammen, was wir zum Nachweis der Wohldefiniertheit des LEBESGUE-Integrals brauchen; wie wir uns bereits überlegt haben, folgt diese aus dem folgenden

Lemma: $(f_k)_{k \in \mathbb{N}}$ sei eine CAUCHY-Folge aus $K_1^0(\mathbb{R}^n, \mathbb{R})$, und die Folge $(f_k(x))_{k \in \mathbb{N}}$ konvergiere für fast alle $x \in \mathbb{R}^n$ gegen Null. Dann konvergiert auch die Folge der L^1 -Normen

$$\|f_k\|_1 = \int_{\mathbb{R}^n} f_k$$

gegen Null.

Beweis: Wir betrachten die Teilfolge $(g_k)_{k \in \mathbb{N}}$ aus dem vorigen Lemma. Wenn wir zeigen können, daß die Teilfolge der Zahlen $\|g_k\|_1$ gegen Null konvergiert, folgt dasselbe auch für die Folge der $\|f_k\|_1$, denn zu jedem $\varepsilon > 1$ gibt es, da $(f_k)_{k \in \mathbb{N}}$ eine CAUCHY-Folge ist, ein $N \in \mathbb{N}$, so daß $\|f_\ell - f_k\|_1 < \frac{1}{2}\varepsilon$ ist für alle $k, \ell > N$. Dies gilt insbesondere auch für solche Indizes $\ell = \nu_p$, für die $f_\ell = g_p$ in der Teilfolge liegt. Für die Teilfolge wiederum gibt es ein $M \in \mathbb{N}$, so daß $\|g_p\|_1 < \frac{1}{2}\varepsilon$ ist für alle $p \geq M$. Wählen wir N so, daß $\nu_N \geq M$ ist, folgt

$$\|f_k\|_1 = \|f_k + (f_\ell - f_k)\|_1 \leq \|f_k\|_1 + \|f_\ell - f_k\|_1 < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Somit konvergiert auch die Folge der Normen der f_k gegen Null.

Um zu zeigen, daß die Folge der $\|g_k\|_1$ eine Nullfolge ist, betrachten wir ein $\varepsilon > 0$ und einen Quader Q , der den Träger von g_k enthält. Durch eine langwierige Abschätzung werden wir gleich zeigen, daß dann gilt

$$\|g_k\|_1 \leq \frac{1}{3 \cdot 4^{k-1}} + \varepsilon(\mu(Q) + \|g_k\|_\infty).$$

Wenn wir dies für alle $\varepsilon > 0$ gezeigt haben, können wir ε gegen Null gehen lassen und erhalten die Abschätzung

$$\|g_k\|_1 \leq \frac{1}{3 \cdot 4^{k-1}},$$

aus der natürlich sofort folgt, daß die Normen eine Nullfolge bilden.

Bleibt also noch die obige Ungleichung zu zeigen.

Wir definieren für alle Paare $(k, \ell) \in \mathbb{N} \times \mathbb{N}$ von natürlichen Zahlen $\ell \geq k$ Hilfsfunktionen $\gamma_{k\ell}: \mathbb{R}^n \rightarrow \mathbb{R}$ als

$$\gamma_{k\ell}(x) \stackrel{\text{def}}{=} \sum_{i=k}^{\ell} |g_{i+1}(x) - g_i(x)|.$$

Da die g_i als Elemente von $K_1^0(\mathbb{R}^n, \mathbb{R})$ stetig sind, sind auch diese Funktionen stetig.

In einem ersten Schritt wollen wir $|g_k(x)|$ durch die gerade definierten Funktionen abschätzen. Dabei unterscheiden wir zwei Fälle:

1. Fall: x liegt nicht in der Teilmenge Z aus dem vorigen Beweis, und $\lim_{k \rightarrow \infty} g_k(x) = 0$.

Dann ist

$$\begin{aligned} |g_k(x)| &= \left| \lim_{\ell \rightarrow \infty} (g_{\ell+1}(x) - g_k(x)) \right| = \left| \lim_{\ell \rightarrow \infty} \sum_{i=k}^{\ell} (g_{i+1}(x) - g_i(x)) \right| \\ &\leq \lim_{\ell \rightarrow \infty} \sum_{i=k}^{\ell} |g_{i+1}(x) - g_i(x)| \leq \lim_{\ell \rightarrow \infty} \gamma_{k\ell}(x); \end{aligned}$$

der letzte Grenzwert existiert, wie wir im Beweis des vorigen Lemmas gesehen haben, da er durch eine geometrische Reihe majorisiert werden kann.

Zu jedem $\varepsilon > 0$ gibt es ein $N \in \mathbb{N}$, so daß sich $\gamma_{k\ell}(x)$ für $\ell \geq N$ um weniger als ε von diesem Grenzwert unterscheidet; für solche ℓ ist dann auch $|g_k(x)| < \gamma_{k\ell}(x) + \varepsilon$. Da beide Seiten dieser Ungleichung stetige Funktionen von x sind, gilt diese Ungleichung auch noch in einer gewissen Umgebung von x ; es gibt also insbesondere einen offenen Quader Q_x , der x enthält, so daß $|g_k(y)| < \gamma_{k\ell}(y) + \varepsilon$ für alle $y \in Q_x$.

2. Fall: $x \in Z$ oder $\lim_{k \rightarrow \infty} g_k(x)$ existiert nicht oder der Limes existiert und ist von Null verschieden.

Wie wir wissen, ist Z eine Nullmenge. Außerdem ist mit der Folge der f_k auch die Teilfolge der g_k eine approximierende Folge für die Nullfunktion; daher ist auch die Menge aller $x \in \mathbb{R}^n$, für die $\lim_{k \rightarrow \infty} g_k(x)$ nicht existiert oder aber von Null verschieden ist, eine Nullmenge. Damit ist auch die Vereinigung Z' dieser beiden Mengen eine Nullmenge. Es gibt somit für jedes $\varepsilon > 0$ eine abzählbare Menge von offenen Quadern Q_j derart, daß Z' in der Vereinigung der Q_j liegt und die Summe von deren Volumina kleiner ist als ε .

Wir betrachten nun die Menge \mathfrak{U} , die aus den sämtlichen im ersten Fall betrachteten Quadern Q_x sowie den sämtlichen im zweiten Fall betrachteten Quadern Q_j besteht. Sie ist eine offene Überdeckung des kompakten Quaders Q , hat also eine endliche Teilüberdeckung. Diese besteht aus endlich vielen der Q_x – dies seien die Quader P_1, \dots, P_r –, sowie aus endlich vielen der Q_j ; dies seien die Quader R_1, \dots, R_s .

Für jeden der Quader P_i gibt es ein $M_i \in \mathbb{N}$, so daß $|g_k(x)| < \gamma_{k\ell}(x)$ ist für alle $\ell \geq M_i$; bezeichnet M das Maximum dieser M_i und P die Vereinigung der P_i , ist also $|g_k(x)| < \gamma_{k\ell}(x)$ für alle $x \in P$ und alle $\ell \geq M$. Indem wir die P_i nötigenfalls noch unterteilen, mit Q schneiden und abschließen, können wir annehmen, daß $P \cap Q$ die volumenfremde Vereinigung von abgeschlossenen Quadern P_i ist.

Für jedes $\ell \geq k$ ist

$$\begin{aligned} \int_Q \gamma_{k\ell} &= \int_Q \sum_{i=k}^{\ell} \int_Q |g_{i+1} - g_i| = \sum_{i=k}^{\ell} \int_Q |g_{i+1} - g_i| \\ &= \sum_{i=k}^{\ell} \int_{R^n} |g_{i+1} - g_i| = \sum_{i=k}^{\ell} \|g_{i+1} - g_i\| < \sum_{i=k}^{\ell} 4^{-i} \\ &< \sum_{i=k}^{\infty} 4^{-i} = \frac{1}{3 \cdot 4^{k-1}}; \end{aligned}$$

für $\ell \geq M$ erhalten wir daher, wenn wir nun auch noch die R_j durch

ihre Abschlüsse ersetzen, die Abschätzung

$$\begin{aligned} \|g_k\|_1 &= \int_{\mathbb{R}^n} |g_k| = \int_Q |g_k| \leq \sum_{i=1}^r \int_{P_i} |g_k| + \sum_{j=1}^s \int_{R_j} |g_k| \\ &\leq \sum_{i=1}^r \int_{P_i} (\gamma_{k\ell} + \varepsilon) + \sum_{j=1}^s \|g_k\|_{\infty} \mu(R_j) \\ &\leq \sum_{i=1}^r \int_{P_i} (\gamma_{k\ell} + \varepsilon) + \|g_k\|_{\infty} \sum_{j=1}^s \mu(R_j) \\ &< \int_Q \gamma_{k\ell} + \varepsilon (\mu(Q) + \|g_k\|_{\infty}) \\ &\leq \lim_{\ell \rightarrow \infty} \int_Q \gamma_{k\ell} + \varepsilon (\mu(Q) + \|g_k\|_{\infty}) \\ &\leq \frac{1}{3 \cdot 4^{k-1}} + \varepsilon (\mu(Q) + \|g_k\|_{\infty}), \end{aligned}$$

wie behauptet. Damit ist das Lemma und die Wohldefiniertheit des LEBESGUE-Integrals bewiesen. \blacksquare

Wir bezeichnen die Menge aller LEBESGUE-integrierbarer Funktionen auf \mathbb{R}^n mit $\text{Leb}_1(\mathbb{R}^n, \mathbb{R})$, wobei der Index 1 bedeuten soll, daß wir auch für $f \in \text{Leb}_1(\mathbb{R}^n, \mathbb{R})$ die Schreibweise

$$\|f\|_1 \stackrel{\text{def}}{=} \int_{\mathbb{R}^n} f$$

verwenden wollen. Es ist klar, daß $\text{Leb}_1(\mathbb{R}^n, \mathbb{R})$ ein Vektorraum ist, aber $\|\cdot\|_1$ definiert keine Norm auf $\text{Leb}_1(\mathbb{R}^n, \mathbb{R})$: Schließlich ist $\|f\|_1 = 0$ nicht nur für die Nullfunktion, sondern auch für jede Funktion, die nur *fast überall* verschwindet.

Es gibt zwei Ansätze, um mit diesem Problem fertig zu werden: Zum ersten könnten wir den Faktorraum $L^1(\mathbb{R}^n, \mathbb{R})$ des Vektorraums $\text{Leb}_1(\mathbb{R}^n, \mathbb{R})$ nach dem Untervektorraum der fast überall verschwindenden Funktionen betrachten; nach dem gerade bewiesenen Lemma ist das ein normierter Vektorraum. Elemente von $L^1(\mathbb{R}^n, \mathbb{R})$ sind dann allerdings nicht mehr Funktionen, sondern Äquivalenzklassen von Funktio-

nen, wobei zwei Funktionen genau dann zur gleichen Äquivalenzklasse gehören, wenn sie fast überall gleich sind. Dieser Ansatz wird oft in der Funktionalanalysis verwendet.

Alternativ können wir auch den Begriff der Norm etwas abschwächen:

Definition: Eine *Halbnorm* auf einem \mathbb{R} - oder \mathbb{C} -Vektorraum V ist eine Abbildung $\|\cdot\| : V \rightarrow \mathbb{R}$ mit den Eigenschaften

- a) $\|\lambda v\| = |\lambda| \|v\|$ für alle $\lambda \in \mathbb{R}$ bzw. \mathbb{C} und $v \in V$
- b) $\|v\| \geq 0$ für alle $v \in V$
- c) $\|v + w\| \leq \|v\| + \|w\|$ (*Dreiecksungleichung*)

Wir verlangen also alle Eigenschaften einer Norm außer der Forderung, daß sie nur für den Nullvektor verschwinden darf. Offensichtlich ist $\|\cdot\|_1$ eine Halbnorm auf $\text{Leb}_1(\mathbb{R}^n, \mathbb{R})$, und mit dieser wollen wir im folgenden rechnen.

Da jede Funktion aus $\text{Leb}_1(\mathbb{R}^n, \mathbb{R})$ Grenzwert einer approximierenden Folge von Funktionen mit kompaktem Träger ist, können wir die bereits für das Quaderintegral und damit auch für Integrale über Funktionen mit kompaktem Träger bewiesenen Eigenschaften des Integrals ohne großen Aufwand auf das LEBESGUE-Integral übertragen: Insbesondere ist das definierte Integral *linear*, d.h.

$$\int_{\mathbb{R}^n} (f + g) = \int_{\mathbb{R}^n} f + \int_{\mathbb{R}^n} g,$$

wir haben die Monotonieeigenschaft

$$\int_{\mathbb{R}^n} f \leq \int_{\mathbb{R}^n} g \quad \text{falls } f(x) \leq g(x) \text{ für alle } x \in \mathbb{R}^n,$$

und nach Definition ist auch

$$\int_{\mathbb{R}^n} f = \int_{\mathbb{R}^n} g \quad \text{falls } f \text{ und } g \text{ fast überall gleich sind,}$$

(Damit folgt natürlich auch, daß wir bei der Monotonieregel nur fordern müssen, daß $f(x) \leq g(x)$ fast überall gilt, d.h. die eventuellen Ausnahmepunkte bilden eine Nullmenge.)

Um zu sehen, daß das mit so großem Aufwand definierte Integral auch zu etwas nütze ist, wollen wir als nächstes einige Beispiele von LEBESGUE-integrierbaren Funktionen betrachten.

Selbstverständlich existiert das LEBESGUE-Integral für die DIRICHLETSche Sprungfunktion

$$f: \begin{cases} \mathbb{R} \rightarrow \mathbb{R} \\ x \mapsto \begin{cases} 1 & \text{falls } x \in \mathbb{Q} \\ 0 & \text{sonst} \end{cases} \end{cases},$$

denn wegen der Abzählbarkeit von \mathbb{Q} ist diese Funktion fast überall gleich der Nullfunktion, also ist $\int_{\mathbb{R}^n} f = 0$, wie erwartet.

Allgemeiner können wir für eine beliebige Teilmenge $A \subseteq \mathbb{R}^n$ die sogenannte *charakteristische Funktion*

$$\chi_A: \begin{cases} \mathbb{R}^n \rightarrow \mathbb{R} \\ x \mapsto \begin{cases} 1 & \text{falls } x \in A \\ 0 & \text{sonst} \end{cases} \end{cases}$$

definieren und fragen, wann diese integrierbar ist.

Beginnen wir mit einem Intervall $A = (a, b) \subset \mathbb{R}$. Dann ist

$$\chi_A(x) = \begin{cases} 1 & \text{falls } a < x < b \\ 0 & \text{sonst} \end{cases}.$$

Völlig analog zum Fall $A = (0, 2)$, anhand dessen wir uns überlegt hatten, daß $K_1^0(\mathbb{R}, \mathbb{R})$ nicht vollständig ist, zeigt man auch hier, daß $\chi_A(x)$ Grenzwert der Folge der Funktionen

$$f_k(x) = \begin{cases} 0 & \text{falls } x \notin (a, b) \\ kx & \text{falls } a \leq x \leq a + \frac{1}{k} \\ 1 & \text{falls } a + \frac{1}{k} \leq x \leq b - \frac{1}{k} \\ k(b-x) & \text{falls } b - \frac{1}{k} \leq x \leq b \end{cases}$$

aus $K_1^0(\mathbb{R}, \mathbb{R})$ ist, wobei wir allerdings nur Werte von k betrachten dürfen, für die $b - a > 2/k$ ist. Diese Folge konvergiert auch punktweise gegen χ_A , ist also eine approximierende Folge.

Nach Definition des Quaderintegrals, das hier einfach das RIEMANN-Integral aus Kapitel 4 ist, haben wir

$$\begin{aligned} \int_{\mathbb{R}} f_k &= \int_a^{a+1/k} kx \, dx + \int_{a+1/k}^{b-1/k} dx + \int_{b-1/k}^b k(b-x) \, dx \\ &= \frac{1}{2k} + \left(b - a - \frac{2}{k}\right) + \frac{1}{2k} = b - a - \frac{1}{k}, \end{aligned}$$

was für $k \rightarrow \infty$ gegen $b - a$ konvergiert. Somit ist

$$\int_{\mathbb{R}} \chi_A = (b - a),$$

was wohl niemanden überraschen dürfte.

Ersetzen wir A durch das abgeschlossene Intervall $[a, b]$ oder eines der beiden halboffenen Intervalle $(a, b]$ oder $[a, b)$, ändert sich daran nichts, denn die charakteristischen Funktionen sind fast überall gleich: Höchstens an den Stellen $x = a$ und $x = b$ kann es Abweichungen geben.

Ist $Q = [a_1, b_1] \times \dots \times [a_n, b_n]$ ein Quader in \mathbb{R}^n , so ist entsprechend $\chi_Q(x) = \chi_{[a_1, b_1]}(x_1) \cdot \dots \cdot \chi_{[a_n, b_n]}(x_n)$ in jedem Punkt $x = (x_1, \dots, x_n)$, und auch diese Funktion können wir approximieren durch stetige Funktionen mit kompaktem Träger, indem wir einfach als k -te Funktion das Produkt der $f_k(x_i)$ nehmen, wobei wir natürlich in der i -ten Komponente die Funktion f_k für das Intervall $[a_i, b_i]$ nehmen müssen. Nach der rekursiven Definition des Quaderintegrals folgt dann schnell, daß auch χ_Q LEBESGUE-integrierbar ist und daß wir auch hier das zu erwartete Ergebnis

$$\int_Q \chi_Q = \prod_{i=1}^n (b_i - a_i) = \mu(Q)$$

bekommen. Wegen der Linearität der Integration gilt etwas allgemeiner

$$\int_{\mathbb{R}^n} c \chi_Q = c \cdot \mu(Q).$$

Zumindest einfache unstetige Funktionen lassen sich also integrieren.

Die Ungleichung von TSCHEBYSCHEFF kennen wir bislang nur für stetige Funktionen mit kompaktem Träger; wir wollen sie verallgemeinern auf beliebige LEBESGUE-integrierbare Funktionen:

Lemma: Die Funktion $f \in \text{Leb}_1(\mathbb{R}^n, \mathbb{R})$ nehme keine negativen Werte an. Dann ist für jede reelle Zahl $c > 0$

$$\mu^* \left(\{x \in \mathbb{R}^n \mid f(x) > c\} \right) \leq \frac{\|f\|_1}{c}.$$

Beweis: Da f LEBESGUE-integrierbar ist, gibt es dazu eine approximierende Folge von Funktionen mit kompaktem Träger. Zu dieser wiederum

können wir nach dem vorletzten Lemma für jedes $\varepsilon > 0$ eine Teilfolge finden, die außerhalb einer Menge Z mit $\mu(Z) < \varepsilon$ gleichmäßig gegen f konvergiert. Diese Teilfolge, für ein fest gewähltes $\varepsilon \in (0, c)$, sei $(f_k)_{k \in \mathbb{N}}$. Wegen der gleichmäßigen Konvergenz außerhalb von Z gibt es ein $N \in \mathbb{N}$, so daß

$$|f(x) - f_k(x)| < \varepsilon \quad \text{für alle } k \geq N \text{ und alle } x \notin Z.$$

Ist $f(x) > c$, ist für $x \notin Z$ und $k \geq N$ insbesondere $f_k(x) > c - \varepsilon$, d.h.

$$\{x \in \mathbb{R}^n \mid f(x) > c\} \subseteq \{x \in \mathbb{R}^n \mid f_k(x) > c - \varepsilon\} \cup Z.$$

Nach der bereits bewiesenen TSCHEBYSCHEFF-Ungleichung für Funktionen mit kompaktem Träger ist das Maß der ersten Menge rechts höchstens gleich $\|f\|_1 / (c - \varepsilon)$, das von Z ist höchstens gleich ε . Somit ist

$$\mu^* \left(\{x \in \mathbb{R}^n \mid f(x) > c\} \right) \leq \frac{\|f\|_1}{c - \varepsilon} + \varepsilon.$$

Dies gilt für jedes $\varepsilon \in (0, c)$, und da die rechte Seite eine stetige Funktion in ε ist, bleibt es auch gültig, wenn wir ε gegen Null gehen lassen. Dann erhalten genau die zu beweisende Ungleichung. ■

Korollar: Für $f \in \text{Leb}_1(\mathbb{R}^n, \mathbb{R})$ ist $\|f\|_1 = 0$ genau dann, wenn f fast überall verschwindet.

Beweis: Falls f fast überall verschwindet, also fast überall gleich der Nullfunktion ist, sind die Integrale von f und der Nullfunktion über \mathbb{R}^n gleich, also beide Null. Wenn umgekehrt $\|f\|_1$ verschwindet, so ist nach der Ungleichung von TSCHEBYSCHEFF für jedes $c > 0$

$$\mu^* \left(\{x \in \mathbb{R}^n \mid |f(x)| > c\} \right) \leq \frac{\|f\|_1}{c} = 0,$$

also ist $\{x \in \mathbb{R}^n \mid |f(x)| > c\}$ für jedes $c > 0$ eine Nullmenge. Da die Vereinigung abzählbar vieler Nullmengen wieder eine Nullmenge ist, gilt dann das gleiche auch für

$$\{x \in \mathbb{R}^n \mid f(x) \neq 0\} = \bigcup_{k \geq 0} \{x \in \mathbb{R}^n \mid |f(x)| > \frac{1}{k}\},$$

das heißt, f verschwindet fast überall. ■

Als unmittelbare Folgerung erhalten wir das weitere

Korollar: Für zwei Funktionen $f, g \in \text{Leb}_1(\mathbb{R}^n, \mathbb{R})$ ist genau dann $\|f - g\|_1 = 0$, wenn f und g fast überall gleich sind. ■

Eine Motivation für die Einführung des LEBESGUE-Integrals an Stelle des einfacheren RIEMANN-Integrals war die Beobachtung, daß der Vektorraum aller RIEMANN-integrierbarer Funktionen von \mathbb{R} nach \mathbb{R} nicht vollständig ist. Als letztes Resultat dieses Paragraphen wollen wir uns überlegen, daß wir dieses Problem bei LEBESGUE-integrierbaren Funktionen nicht haben. Da $\text{Leb}_1(\mathbb{R}^n, \mathbb{R})$ kein normierter Vektorraum ist, können wir hier zwar nicht von Vollständigkeit reden, aber wir können doch zeigen

Satz: Jede CAUCHY-Folge aus $\text{Leb}_1(\mathbb{R}^n, \mathbb{R})$ konvergiert gegen eine Funktion $f \in \text{Leb}_1(\mathbb{R}^n, \mathbb{R})$.

(Man beachte, daß die Funktion f hier nicht eindeutig bestimmt ist: Nach obigem Korollar konvergiert die Folge dann auch gegen jede Funktion $g \in \text{Leb}_1(\mathbb{R}^n, \mathbb{R})$ mit $\|f - g\|_1 = 0$.)

Beweis: $(f_k)_{k \in \mathbb{N}}$ sei eine CAUCHY-Folge aus $\text{Leb}_1(\mathbb{R}^n, \mathbb{R})$. Wir müssen zeigen, daß es eine Funktion $f \in \text{Leb}_1(\mathbb{R}^n, \mathbb{R})$ gibt, gegen die diese Folge konvergiert.

Zu jeder der Funktionen f_k gibt es nach Definition der LEBESGUE-Integrierbarkeit eine approximierende Folge; die bezüglich der L^1 -Norm gegen f_k konvergiert; wir wählen daraus jeweils ein Folgenglied g_k mit der Eigenschaft, daß $\|f_k - g_k\|_1 < 1/k$ ist. Die Funktionen g_k sind Funktionen mit kompaktem Träger, und auch sie bilden eine CAUCHY-Folge, denn für jedes $\varepsilon > 0$ können wir zunächst ein $N_1 \in \mathbb{N}$ finden, so daß $\|f_\ell - f_k\|_1 < \frac{1}{3}\varepsilon$ ist für alle $k, \ell \geq N_1$. Ist $N \geq N_1$ eine natürliche Zahl, die gleichzeitig größer ist als $3/\varepsilon$, gilt dann für alle Indizes $k, \ell \geq N$

$$\begin{aligned} \|g_\ell - g_k\|_1 &= \|(g_\ell - f_\ell) + (f_\ell - f_k) + (f_k - g_k)\|_1 \\ &\leq \|f_\ell - g_\ell\|_1 + \|f_\ell - f_k\|_1 + \|f_k - g_k\|_1 \\ &< \frac{1}{\ell} + \frac{\varepsilon}{3} + \frac{1}{k} \leq \frac{2}{N} + \frac{\varepsilon}{3} \leq 2 \times \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon. \end{aligned}$$

Somit ist $(g_k)_{k \in \mathbb{N}}$ eine CAUCHY-Folge von Funktionen mit kompaktem Träger; wie wir weiter oben gesehen haben, gibt es dazu eine Teilfolge

$(g_{k_\nu})_{\nu \in \mathbb{N}}$, so daß für fast alle $x \in \mathbb{R}^n$ die reelle Zahlenfolge $(g_{k_\nu}(x))_{\nu \in \mathbb{N}}$ konvergiert. Für alle x , für die der Fall ist, bezeichnen wir den Grenzwert der Folge mit $f(x)$, ansonsten setzen wir $f(x) = 0$. Damit ist eine Funktion $f: \mathbb{R}^n \rightarrow \mathbb{R}$ definiert, die offensichtlich LEBESGUE-integrierbar ist, denn $(g_{k_\nu})_{\nu \in \mathbb{N}}$ ist eine approximierende Folge für f . Mit der Folge der g_{k_ν} konvergiert auch die der f_{k_ν} bezüglich der L^1 -Norm gegen f , denn

$$\|f_{k_\nu} - f\|_1 \leq \|f_{k_\nu} - g_{k_\nu}\|_1 + \|g_{k_\nu} - f\|_1,$$

wobei der erste Summand kleiner ist als $1/k_\nu$ und der zweite nach Definition von f gegen Null geht. Da die Folge der f_k eine CAUCHY-Folge ist, konvergiert dann aber auch die ganze Folge gegen f . ■

Aus dem Beweis folgt auch die zu erwartende Gleichung

$$\begin{aligned} \lim_{k \rightarrow \infty} \int_{\mathbb{R}^n} f_k &= \int_{\mathbb{R}^n} \lim_{k \rightarrow \infty} f_k, \text{ denn} \\ \left| \int_{\mathbb{R}^n} f - \int_{\mathbb{R}^n} f_k \right| &= \left| \int_{\mathbb{R}^n} (f - f_k) \right| \leq \int_{\mathbb{R}^n} |f - f_k| = \|f - f_k\|_1. \end{aligned}$$

§3: Das Lebesgue-Integral auf Teilmengen von \mathbb{R}^n

Nicht alle Funktionen sind auf ganz \mathbb{R}^n definiert, und selbst wenn sie es sind, interessiert oft nur ein Teilbereich. Das ist nicht neues gegenüber der Situation im Eindimensionalen, wo wir ja auch meist nicht über ganz \mathbb{R} , sondern nur über ein Intervall integrieren.

Wenn wir speziell die Konstante 1 über ein Intervall $[a, b]$ integrieren, erhalten wir als Ergebnis die Länge $b - a$ des Intervalls; entsprechend sollten bei Integration der Eins über eine Teilmenge $B \subseteq \mathbb{R}^n$ deren Fläche bzw. Volumen erhalten. Was ist aber das Volumen einer beliebigen Teilmenge eines \mathbb{R}^n ? Bislang kennen wir nur das Volumen achsenparalleler Quader, das wir entweder einfach als Produkt der Kantenlängen berechnen können oder aber kompliziert, indem wir die charakteristische Funktion des Quaders über den \mathbb{R}^n integrieren. Da wir für jede Teilmenge des \mathbb{R}^n eine charakteristische Funktion haben, können wir diese zweite Möglichkeit verallgemeinern:

Definition: Eine Teilmenge $B \subseteq \mathbb{R}^n$ heißt *meßbar*, wenn ihre charakteristische Funktion χ_B in $\text{Leb}_1(\mathbb{R}^n, \mathbb{R})$ liegt; als *Volumen* von B bezeichnen wir dann den Wert des Integrals $\int_{\mathbb{R}^n} \chi_B$.

Beispiele nicht meßbarer Mengen lassen sich leicht angeben, beispielsweise der gesamte \mathbb{R}^n . Hier liegt der Grund allerdings darin, daß das Volumen eben nicht endlich ist. Interessanter ist die Frage, ob jede *beschränkte* Teilmenge von \mathbb{R}^n meßbar ist. Bislang konnte noch niemand ein explizites Beispiel einer nicht meßbaren beschränkten Teilmenge angeben, und vielleicht ist das sogar unmöglich, die *Existenz* solcher Mengen läßt sich aber beweisen. Das zentrale Ergebnis hierzu ist der folgende

Satz von Banach-Tarski: Für $n \geq 3$ gibt es zu je zwei beschränkten Teilmengen $A, B \subset \mathbb{R}^n$ eine natürliche Zahl m sowie disjunkte Zerlegungen $A = \bigcup_{k=1}^m A_k$ und $B = \bigcup_{k=1}^m B_k$ derart, daß für jedes k die Mengen A_k und B_k durch eine Kongruenzabbildung ineinander übergeführt werden können.

Kongruenzabbildungen sind zusammengesetzt aus Drehungen und Verschiebungen; jeder vernünftige Volumenbegriff sollte also den beiden Mengen A_k und B_k dasselbe Volumen zuordnen. (Nach der weiter unten betrachteten Transformationsformel hat das oben definierte Integral diese Eigenschaft.) Genauso sollte das Volumen einer disjunkten Vereinigung einfach die Summe der Volumina der Teilmengen sein; im obigen Satz müßten also, falls alle A_k und B_k meßbar wären, A und B dasselbe Volumen haben. Da wir für A und B beispielsweise zwei Würfel mit unterschiedlichen Kantenlängen nehmen können, ist das natürlich absurd. Somit können zumindest nicht alle der Mengen A_k und B_k meßbar sein; es muß auch nicht meßbare Mengen geben. Da der Beweis von BANACH und TARSKI nicht konstruktiv ist (er beruht wesentlich auf dem sogenannten Auswahlaxiom), liefert er allerdings keine expliziten Beispiele für solche Mengen, und zumindest bislang konnte auch noch niemand so ein Beispiel angeben.

Obwohl der Satz von BANACH-TARSKI unserer Anschauung völlig widerspricht, ist sein Beweis nicht sonderlich schwer; man findet ihn natürlich in der Originalarbeit

S. BANACH, A. TARSKI: Sur la décomposition des ensembles de points en parties respectivement congruentes, *Fund. Math.* **6** (1924), 244–277

oder, vereinfacht, beispielsweise bei

KARL STROMBERG: The Banach-Tarski Paradox, *American Mathematical Monthly* **86** (1979), 151–161;

eine sehr ausführliche populärwissenschaftliche Darstellung bietet das Buch

LEONARD M. WAPNER: The Pea and the Sun – A Mathematical Paradox, *A.K. Peters*, 2005



ALFRED TARSKI (1902–1983) wurde als ALFRED TEITELBAUM im damals russischen Warschau geboren. Er studierte zunächst Biologie, wechselte dann aber, nachdem er eine Logik-Vorlesung gehört hatte, zur Mathematik. Seine erste Arbeit, über Mengenlehre, erschien 1921, der Satz von BANACH-TARSKI 1924, nachdem er 1923 seinen Namen geändert hatte. Nach seinem Studium unterrichtete er zunächst an verschiedenen Warschauer Institutionen. 1930 traf er in Wien KURT GÖDEL (1908–1978) und publizierte, wohl dadurch be-

einflußt, 1933 seine sicherlich bekannteste Arbeit, in der er die *Wahrheit* einer Aussage in einem formalen System dadurch definierte, daß die Aussage in allen Interpretationen dieses Systems korrekt ist. Zwei Wochen vor dem deutschen Einmarsch in Polen war er 1939 zu einer Konferenz an die Harvard University nach USA gereist; er konnte dort bleiben und arbeitete an verschiedenen amerikanischen Universitäten, bis er 1942 nach Berkeley ging, wo er trotz seiner Emeritierung 1968 bis 1973 lehrte.

Nachdem wir über das LEBESGUE-Integral ein Volumen definiert haben, ist nun schon fast klar, wie wir ein Integrale über Teilmengen des \mathbb{R}^n mit beliebigen Integranden definieren:

Definition: $f: B \rightarrow \mathbb{R}$ sei eine Funktion auf einer Teilmenge $B \subseteq \mathbb{R}^n$. Falls die Funktion $\tilde{f}: \mathbb{R}^n \rightarrow \mathbb{R}$ mit $\tilde{f}(x) = f(x)$ für $x \in B$ und $\tilde{f}(x) = 0$ sonst LEBESGUE integrierbar ist, definieren wir

$$\int_B f(x) \stackrel{\text{def}}{=} \int_{\mathbb{R}^n} \tilde{f}(x);$$

andernfalls existiert das linksstehende Integral nicht.

Kurz könnte man auch schreiben $\int_B f(x) = \int_{\mathbb{R}^n} f(x)\chi_B(x)$, allerdings ist das insofern nicht ganz exakt, als $f(x)$ nicht für alle $x \in \mathbb{R}^n$ definiert

sein muß. Interpretiert man das Produkt $f(x)\chi_B(x)$ allerdings so, daß es für alle $x \notin B$ verschwinden soll, unabhängig davon ob $f(x)$ definiert ist oder nicht, ist dies genau die obige Definition.

Im letzten Semester hatten wir Integrale in der Form $\int_a^b f(x) dx$ geschrieben; das neu eingeführte LEBESGUE-Integral dagegen schreiben wir zumindest bislang immer nur als $\int_B f$. Dies geschah vor allem, um das neue Integral vom alten zu unterscheiden; in der Literatur sind auch für das LEBESGUE-Integral eher Schreibweisen wie

$$\int_B f(x) dx, \quad \int_B f(x) d\mu \quad \text{oder} \quad \int_B f(x_1, \dots, x_n) dx_1 dx_2 \cdots dx_n$$

üblich. Für allem in Naturwissenschaft und Technik ist auch heute noch üblich, bei einem Integral über einen n -dimensionalen Bereich auch n Integralzeichen zu schreiben, also beispielsweise

$$\iint_B f(x, y) dx dy \quad \text{und} \quad \iiint_C f(x, y, z) dx dy dz$$

für $B \subseteq \mathbb{R}^2$ und $C \subseteq \mathbb{R}^3$. Für einen Integrationsbereich im allgemeinen n -dimensionalen Raum ergibt sich entsprechend ein Monstrum wie

$$\underbrace{\int \cdots \int}_n f(x_1, \dots, x_n) dx_1 \cdots dx_n.$$

Da das oben definierte Integral ein Spezialfall des Integrals über den \mathbb{R}^n ist und dieses wiederum auf Quaderintegrale zurückgeführt werden kann, ist klar, daß auch für dieses Integral die „üblichen“ Rechenregeln gelten; etwas genauer ansehen wollen wir uns nur eine, die nicht in offensichtlicher Weise auf Regeln für eindimensionale Integrale zurückgeführt werden kann: die Transformationsformel.

Von den vielen Regeln zur expliziten Bestimmung einer Stammfunktion ist sicherlich die Substitutionsregel die nützlichste; die Transformationsformel verallgemeinert sie auf mehrdimensionale Integrale.

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

$$\int_a^b f(x) dx = \int_{t_0}^{t_1} f(\varphi(t)) \dot{\varphi}(t) dt.$$

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 Polarkoordinatendarstellung

$$x = x(r, \varphi) = r \sin \varphi \quad \text{und} \quad y = y(r, \varphi) = r \cos \varphi.$$

Wir betrachten zunächst ein Integral

$$\int_R f(x, y) dx dy$$

über ein achsenparalleles Rechteck R mit Eckpunkten

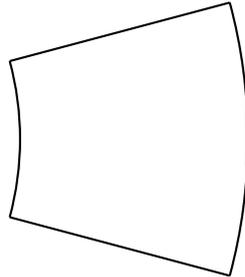
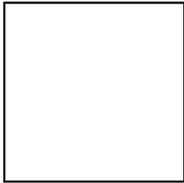
$$(u_0, v_0), \quad (u_0 + h, v_0), \quad (u_0, v_0 + k) \quad \text{und} \quad (u_0 + h, v_0 + k).$$

Wenn wir statt über x und y über u und v integrieren, müssen wir über jenen Bereich B' integrieren, in dem sich diese neuen Variablen bewegen, also die Menge

$$\{(x(u, v), y(u, v)) \mid u_0 \leq u \leq u_0 + h \quad \text{und} \quad v_0 \leq v \leq v_0 + k\}.$$

Diese 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 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

$$x(u_0 + h, v_0) = x(u_0, v_0) + h \frac{\partial x}{\partial u}(u_0, v_0) + o(h)$$

$$y(u_0 + h, v_0) = y(u_0, v_0) + h \frac{\partial y}{\partial u}(u_0, v_0) + o(h)$$

$$x(u_0, v_0 + k) = x(u_0, v_0) + k \frac{\partial x}{\partial v}(u_0, v_0) + o(k)$$

$$y(u_0, v_0 + k) = y(u_0, v_0) + k \frac{\partial y}{\partial v}(u_0, v_0) + o(k)$$

und

$$x(u_0 + h, v_0 + k) = x(u_0, v_0) + h \frac{\partial x}{\partial u}(u_0, v_0) + o(h) + k \frac{\partial x}{\partial v}(u_0, v_0) + o(k)$$

$$y(u_0 + h, v_0 + k) = y(u_0, v_0) + h \frac{\partial y}{\partial u}(u_0, v_0) + o(h) + k \frac{\partial y}{\partial v}(u_0, v_0) + o(k);$$

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

$$hk \cdot \left| \frac{\partial x}{\partial u} \frac{\partial y}{\partial v} - \frac{\partial x}{\partial v} \frac{\partial y}{\partial u} \right|.$$

Integrieren wir nicht über ein Rechteck, sondern über einen komplizierteren Integrationsbereich $B \subset \mathbb{R}^2$, so können wir diesen durch immer kleinere Rechtecke approximieren; führt man dies genauer aus, erhält man die folgende

Transformationsformel: $f: B \rightarrow \mathbb{R}$ 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

$$B' \rightarrow B; \quad (u, v) \mapsto (x(u, v), y(u, v))$$

bijektiv ist, so gilt

$$\int_B f(x, y) dx dy = \int_{B'} f(x(u, v), y(u, v)) \left| \frac{\partial x}{\partial u} \frac{\partial y}{\partial v} - \frac{\partial x}{\partial v} \frac{\partial y}{\partial u} \right| du dv.$$

(Hier ist es nützlich, die Integrationsvariablen explizit durch dx, dy und so weiter anzugeben, da wir links und rechts in verschiedenen Koordinatensystemen arbeiten.)

Im Fall

$$x = r \cos \varphi \quad \text{und} \quad y = r \sin \varphi$$

der Polarkoordinaten ist

$$\frac{\partial x}{\partial r} \frac{\partial y}{\partial \varphi} - \frac{\partial x}{\partial \varphi} \frac{\partial y}{\partial r} = \cos \varphi \cdot (r \sin \varphi) - (-r \cos \varphi) \cdot \sin \varphi = r,$$

d.h.

$$\int_B f(x, y) dx dy = \int_{B'} f(r \cos \varphi, r \sin \varphi) r dr d\varphi.$$

Für diese Formel hätten wir eigentlich nicht den ganzen Apparat der Transformationsformel gebraucht; die obige Abbildung 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

$$\frac{1}{2}k(r+h)^2 - \frac{1}{2}kr^2 = r \cdot kh + \frac{1}{2}kh^2.$$

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

$$B' = \{(r, \varphi) \mid 0 \leq r \leq R \text{ und } 0 \leq \varphi < 2\pi\};$$

die Fläche von B ist also

$$\int_B dx dy = \int_{B'} r dr d\varphi = \int_0^{2\pi} \left(\int_0^R r dr \right) d\varphi = \int_0^{2\pi} \frac{R^2}{2} d\varphi = \pi R^2.$$

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 letztes 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

$$I \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} e^{-x^2} dx.$$

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 den \mathbb{R}^2 als ein unendliches „Rechteck“ auffassen; daher ist

$$\begin{aligned} \int_{\mathbb{R}^2} e^{-(x^2+y^2)} dx dy &= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} e^{-(x^2+y^2)} dx \right) dy \\ &= \int_{-\infty}^{\infty} e^{-y^2} \left(\int_{-\infty}^{\infty} e^{-x^2} dx \right) dy = \int_{-\infty}^{\infty} e^{-y^2} \cdot I dy \\ &= I \cdot \int_{-\infty}^{\infty} e^{-y^2} dy = I^2. \end{aligned}$$

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

$$\int_{\mathbb{R}^2} e^{-(x^2+y^2)} dx dy = \int_{B'} e^{-r^2} r dr d\varphi = \int_0^{2\pi} \left(\int_0^{\infty} r e^{-r^2} dr \right) d\varphi.$$

Die Stammfunktion des neuen Integranden $r e^{-r^2}$ ist leicht zu finden: Aus

$$\frac{d}{dr} e^{-r^2} = -2r e^{-r^2}$$

folgt sofort, daß

$$\int r e^{-r^2} dr = -\frac{1}{2} e^{-r^2} + C.$$

Damit können wir weiterrechnen und erhalten das Ergebnis

$$\int_0^{2\pi} \left(\int_0^{\infty} r e^{-r^2} dr \right) d\varphi = \int_0^{2\pi} -\frac{1}{2} e^{-r^2} \Big|_0^{\infty} d\varphi = \int_0^{2\pi} \frac{1}{2} d\varphi = \pi.$$

Also ist $I^2 = \pi$ und, da der Integrand von I überall positiv ist, folgt

$$I = \int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi}.$$