

Reelle Analysis mehrerer Veränderlicher

Vorlesungsskriptum Sommersemester 1995

J.-H. Eschenburg
Institut für Mathematik
Universität Augsburg

Dieses Skriptum ist die Ausarbeitung meiner Vorlesung Analysis II im Sommersemester 1995. Es besteht aus drei Abschnitten:

- I Konvergenz und Stetigkeit
- II Differentiation
- III Integration

Der Abschnitt I hat die Geometrie und die topologischen Begriffe im euklidischen Raum sowie die Definition und die wichtigsten Sätze über stetige Funktionen zum Inhalt. Abschnitt II behandelt die Differentialrechnung mehrerer Veränderlicher: Begriffe von partieller und totaler Ableitung, Kettenregel, Satz von Taylor, lokale Extrema, Umkehrsatz und impliziter Funktionensatz, Mannigfaltigkeiten, Extrema mit Nebenbedingungen. In Abschnitt III wird der Riemannsche Volumen- und Integralbegriff eingeführt und der Satz von Fubini, der Transformationsatz und der Divergenzatz bewiesen. Das Lebesgue-Integral gehört schon zum Stoff des nachfolgenden Semesters und ist in einem Anhang dargestellt. Die eingestreuten “Bemerkungen” führen über den eigentlichen Vorlesungsstoff hinaus und können beim ersten Lesen übergangen werden. Ich danke meinen Hörern für ihre Geduld und hoffe auf Reaktionen und Anregungen zu diesem Skriptum.

März 1996 (korrigiert Mai 2000, März 2008, Juli 2008)
Jost Eschenburg

Was ist Mathematik? Gegenstand dieser Wissenschaft ist der Teil unserer Sprache und unseres Denkens, der sich mit Größen und Formen beschäftigt. Dazu gehören nicht nur die Zahlwörter, sondern auch viele Gegensatzpaare wie groß und klein, nah und fern, glatt und rau, rund und eckig, gerade und krumm. Diese Begriffe beschreiben mathematische *Ideen*, und sie ähneln einem Fraktal: Wir können sie wie das Mandelbrot'sche "Apfelmännchen" einfach erzeugen und mit einem Blick erfassen; wenn wir sie aber näher untersuchen, so entfalten sie sich und zeigen immer neue Details, und je genauer wir hinschauen und je feinere Untersuchungsmethoden wir entwickeln, desto überraschendere Perspektiven tun sich auf. Ein vier- bis sechsjähriges Kind versteht bereits die Zahlen, und doch beschäftigt sich damit auch aktuellste mathematische Forschung. Man sollte daher mathematische Gegenstände auf jedem Wissensstand *von außen nach innen* zu verstehen versuchen: von einem anfänglichen groben zu einem immer tieferen, reicheren und feineren Verständnis hin.

Was ist Analysis? Nach H.Weyl gibt es zwei Grundbewegungen des mathematischen Denkens: Das Zusammensetzen (Synthesis) und das Teilen (Analysis). Beim Zusammensetzen wird aus der Einheit die Vielheit aufgebaut; durch vielfaches (aber immer endliches) Zusammenfügen derselben Einheit entstehen die natürlichen Zahlen. Die umgekehrte Bewegung des Teilens setzt ein bereits gegebenes Ganzes ("*Kontinuum*") voraus, das nun in Abschnitte eingeteilt wird; eine Position innerhalb dieses Ganzen kann nur durch immer feinere Stückelung festgelegt werden; so entstehen die reellen Zahlen. Es gibt einen grundlegenden Unterschied zwischen diesen beiden Bewegungen: Die Fixierung einer einzigen Position mit absoluter Genauigkeit erfordert im Prinzip unendlich viele solcher Teilungsschritte, während das Zusammensetzen immer nur endlich oft durchgeführt werden muß; die Analysis hat es daher in starker Weise mit dem *Unendlichen* zu tun. Diese mathematische Idee "Kontinuum" wird durch die topologischen Begriffe wie Konvergenz, Vollständigkeit und Stetigkeit entfaltet. Auf diesem Hintergrund ist die Aufgabe der Analysis die Beschreibung der *Größe* (Integration) und der *Veränderung* (Differentiation). Die eine Aufgabe scheint statisch (Messung einer Größe), die andere dynamisch (Abhängigkeit einer Größe von einer anderen), und doch hängen beide Aufgaben sehr eng miteinander zusammen: Die Größe (Flächeninhalt) einer ebenen Figur

etwa kann ich dadurch messen, dass ich ein Lineal darüber hinweggleiten lasse und in jeder Position des Lineals die Grenzpunkte der Figur und ihre Veränderung notiere; aus der zweidimensionalen statischen (Flächeninhalt) mache ich damit eine eindimensionale dynamische Aufgabe (Position auf dem eindimensionalen Lineal in Abhängigkeit von seiner Lage). Die Veränderungen (*“Funktionen”* genannt), die die Analysis beschreibt, sind allerdings nicht völlig willkürlich, sondern relativ gutartig (*“natura non facit saltus”* - die Natur macht keine Sprünge - was natürlich nur bedingt und auf einer nicht zu feinen Maßstabsskala stimmt) und können daher durch einfache Modelle - *lineare Approximation* - beschrieben werden. Will man auch willkürlichere Veränderungen zulassen, von denen es eine schier unübersehbare Fülle gibt, so tritt man gleichsam einen Schritt zurück und betrachtet statt individueller Funktionen ganze Klassen von Funktionen in ihrer Gesamtheit (Funktionsräume).

Inhaltsverzeichnis

I. Konvergenz und Stetigkeit

1. Der euklidische Raum 6
2. Folgen und Konvergenz 10
3. Offene, abgeschlossene und kompakte Teilmengen 11
4. Stetige Abbildungen 14

II. Differentiation

5. Kurven 19
6. Partielle Ableitungen 24
7. Differenzierbarkeit 27
8. Beispiele und Rechenregeln 32
9. Exkurs: Die Matrixnorm 34
10. Richtungsableitungen und Gradient 36
11. Der Schrankensatz 39
12. Exkurs: Der Banachsche Fixpunktsatz 40
13. Der Umkehrsatz 43
14. Der Satz über implizite Abbildungen 47
15. Mannigfaltigkeiten und reguläre Niveaumengen 52
16. Lokale Extrema (mit Nebenbedingungen) 58
17. Exkurs über Polynome 61
18. Höhere Ableitungen und Satz von Taylor 62
19. Lokale Extrema: Bedingungen höherer Ordnung 44
20. Koordinaten 67

III. Integration

21. Flächeninhalt, Volumen, Maß 73
22. Integrierbarkeit 76
23. Klassen integrierbarer Funktionen 80
24. Nullmengen 84
25. Der Satz von Fubini 87
26. Der Transformationssatz 96
27. Lineare Invarianz des Maßes 102
28. Die Integralsätze von Gauß und Stokes 104
29. Integration über Mannigfaltigkeiten und Divergenzsatz 109

Anhang: Das Lebesgue-Integral

- A1. Mängel des Riemannsches Integralbegriffs 117
- A2. Vervollständigung metrischer Räume 119
- A3. Limesfunktionen von L^1 -Cauchyfolgen 122
- A4. Riemann- und Lebesgue-Maß 131
- A5. Konvergenzsätze 136

A6. Integration über Teilmengen, Transformationsatz, Fubini	139
Literatur	140
Symbolverzeichnis	141
Index	143

I. Konvergenz und Stetigkeit

1. Der euklidische Raum

In der Ebene und im Raum unserer Anschauung können wir zwei bzw. drei Koordinatenachsen festlegen und damit jeden Punkt durch zwei bzw. drei Zahlen festlegen, den *kartesischen Koordinaten* dieses Punktes. Abstrakt definieren wir deshalb die Ebene als Menge aller Paare und den Raum als Menge aller Tripel reeller Zahlen:

$$\mathbb{R}^2 = \{(x, y); x, y \in \mathbb{R}\}, \quad \mathbb{R}^3 = \{(x, y, z); x, y, z \in \mathbb{R}\}.$$

Mathematisch hindert uns nichts daran, die Dimensionszahlen 2 und 3 durch eine beliebige natürliche Zahl n zu ersetzen. Dann gehen uns allerdings die Buchstaben aus, weshalb wir statt immer neuer Buchstaben x, y, z, \dots (?) lieber Buchstaben mit einer Zahl als Index x_1, x_2, x_3, \dots verwenden:

$$\mathbb{R}^n = \{(x_1, \dots, x_n); x_1, \dots, x_n \in \mathbb{R}\}.$$

Einem solchen n -Tupel (auch *Vektor* genannt) (x_1, \dots, x_n) geben wir wieder den Namen x , was manchmal zu Mißverständnissen führt, da in \mathbb{R}^2 und \mathbb{R}^3 mit x häufig die erste Koordinate bezeichnet wird. Die Menge \mathbb{R}^n zusammen mit den weiter unten definierten Strukturen nennt man den *n -dimensionalen euklidischen Raum*. Diese Verallgemeinerung auf beliebige Dimensionszahlen ist keine Spielerei der Mathematiker; erst dieser Schritt macht Probleme mit vielen Variablen einer geometrischen Sprache ähnlich wie in Ebene und Raum zugänglich, indem man nämlich die vielen Variablen zu einem neuen Objekt (einem Vektor in \mathbb{R}^n) zusammenfaßt. Alle nun folgenden Begriffe und Sätze haben in der Ebene und im Raum eine anschauliche Interpretation, die wir in vielen Fällen durch Figuren illustrieren werden. Durch richtig gelesene Figuren lassen sich Ideen häufig schneller klarmachen (und auch besser behalten) als durch Formeln und Begriffe; man sollte auf das mächtige Hilfsmittel unserer Anschauung auf keinen Fall ohne Not verzichten. Allerdings ist eine Rückübersetzung in die strenge Formelsprache der Mathematik unbedingt nötig, da die Anschauung auch trügerisch ist und viele mathematische Begriffe weit über sie hinausreichen.

Bemerkung. Bei der Identifizierung von Ebene und Raum mit \mathbb{R}^2 und \mathbb{R}^3 stoßen wir bereits auf ein Problem, das die ganze Mathematik durchzieht: \mathbb{R}^2 und \mathbb{R}^3 modellieren nämlich Ebene und Raum *mit gegebenen Koordinatenachsen*. In vielen geometrischen Problemen (denken wir z.B. an die Geometrie der Schule oder auch an physikalische Probleme) sind solche Achsen gar nicht gegeben, sondern sie müssen von

uns erst *gewählt* werden, wenn wir das Problem “analytisch” behandeln wollen. Da wir von unserem eigenen Zutun unabhängige Aussagen anstreben, interessieren wir uns dafür, wie sich die mathematische Beschreibung verändert, wenn wir andere Achsen wählen (*Basis-* oder *Koordinatentransformation*; vgl. Lineare Algebra). Wir werden auf dieses Problem in Kap. 20 in allgemeinerem Zusammenhang zurückkommen.

Komponentenweise Addition und Multiplikation mit reellen Zahlen definiert die Vektorraum-Struktur des \mathbb{R}^n : Sind $x = (x_1, \dots, x_n)$ und $y = (y_1, \dots, y_n)$ Elemente des \mathbb{R}^n und ist $\alpha \in \mathbb{R}$, so definiert man $x + y$, $\alpha x \in \mathbb{R}^n$ als

$$x + y = (x_1 + y_1, \dots, x_n + y_n), \quad \alpha x = (\alpha x_1, \dots, \alpha x_n).$$

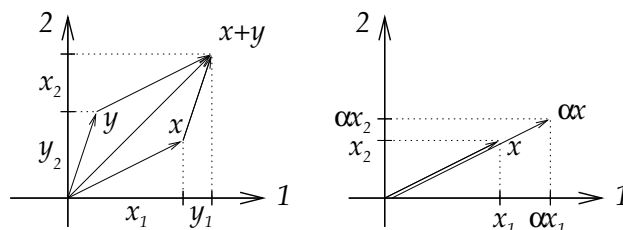


ABBILDUNG 1

Die speziellen Vektoren $e_1 = (1, 0, \dots, 0)$, $e_2 = (0, 1, 0, \dots, 0)$, ..., $e_n = (0, \dots, 0, 1)$ bilden eine Basis dieses Vektorraums, die wir *Standardbasis* nennen wollen. Vektoren in \mathbb{R}^2 und \mathbb{R}^3 veranschaulichen wir uns als Pfeile vom Ursprung zu dem Punkt mit den entsprechenden Koordinaten. Abb. 1 zeigt die anschauliche Bedeutung der Vektorraum-Struktur.

Der für uns wichtigste Begriff ist die *Länge* oder der *Betrag* eines Vektors $x \in \mathbb{R}^n$, nämlich

$$|x| := \sqrt{x_1^2 + \dots + x_n^2}.$$

Für zwei Punkte $x, y \in \mathbb{R}^n$ nennen wir die Zahl $|x - y|$ auch den *Abstand* zwischen x und y . Abb. 2 zeigt mit Hilfe des Satzes von Pythagoras, dass die obige Definition dem anschaulichen Begriff der Länge oder des Abstandes in Ebene und Raum entspricht. Aus der Definition entnehmen wir sofort die folgenden Ungleichungen zwischen den Beträgen des Vektors x und seinen Komponenten x_i für $i = 1, \dots, n$:

$$(*) \quad \max_i |x_i| \leq |x| \leq \sqrt{n} \cdot \max_i |x_i|.$$

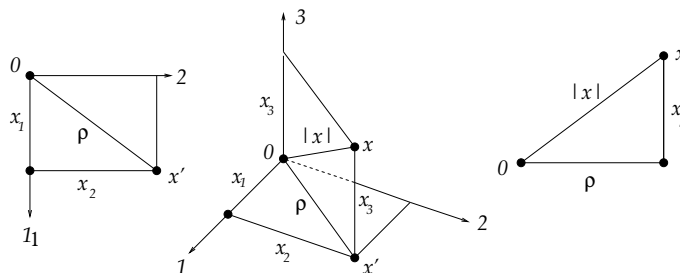


ABBILDUNG 2

Der Betrag eines Vektors hängt eng mit dem *Skalarprodukt* zusammen, das für zwei Vektoren $x, y \in \mathbb{R}^n$ folgendermaßen definiert ist:

$$x \cdot y = \langle x, y \rangle = \sum_{i=1}^n x_i y_i.$$

Damit ist $|x| = \sqrt{\langle x, x \rangle}$. Zwei Vektoren $x, y \in \mathbb{R}^n$ heißen *senkrecht*, wenn $\langle x, y \rangle = 0$.

1. Satz 1. Für alle $x, y \in \mathbb{R}^n$ gelten die *Cauchy-Schwarz-Ungleichung* und die *Dreiecksungleichung*:

$$\begin{aligned} \langle x, y \rangle^2 &\leq |x|^2 |y|^2, \\ |x + y| &\leq |x| + |y|. \end{aligned}$$

Beweis. Ohne Einschränkung sei $x \neq 0$. Für alle $\alpha \in \mathbb{R}$ ist dann

$$\begin{aligned} 0 &\leq \langle \alpha x + y, \alpha x + y \rangle \\ &= \alpha^2 \langle x, x \rangle + 2\alpha \langle x, y \rangle + \langle y, y \rangle \\ &= \langle x, x \rangle \left(\alpha^2 + 2\alpha \frac{\langle x, y \rangle}{\langle x, x \rangle} \right) + \langle y, y \rangle \\ &= \langle x, x \rangle \left(\alpha + \frac{\langle x, y \rangle}{\langle x, x \rangle} \right)^2 + \langle y, y \rangle - \frac{\langle x, y \rangle^2}{\langle x, x \rangle} \end{aligned}$$

(quadratische Ergänzung!). Wählen wir $\alpha = -\frac{\langle x, y \rangle}{\langle x, x \rangle}$, so verschwindet der erste Summand rechts, also folgt

$$0 \leq \langle y, y \rangle - \frac{\langle x, y \rangle^2}{\langle x, x \rangle},$$

was die erste Ungleichung (Cauchy-Schwarz-Ungleichung) beweist. Die zweite Ungleichung (Dreiecksungleichung) ist eine direkte Folgerung davon, denn

$$\begin{aligned} |x + y|^2 &= \langle x + y, x + y \rangle \\ &= \langle x, x \rangle + \langle y, y \rangle + 2\langle x, y \rangle, \\ (|x| + |y|)^2 &= \langle x, x \rangle + \langle y, y \rangle + 2|x||y|. \end{aligned}$$

□

Zur Notation: Ist ein Dreieck mit den Eckpunkten $x, y, z \in \mathbb{R}^n$ gegeben, so sagt die zweite Ungleichung in Satz 1:

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

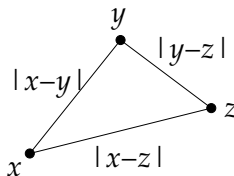


ABBILDUNG 3

d.h. die Summe von zwei Seitenlängen des Dreiecks ist stets größer als die dritte Seitenlänge. Daher stammt der Name “Dreiecksungleichung”. (Abb. 3)

Bemerkung. In den folgenden Abschnitten wird die Betragsfunktion $x \mapsto |x|$ eigentlich nicht explizit benutzt, sondern nur die folgenden drei Eigenschaften:

- (N1) $|x| > 0$ für alle $x \in \mathbb{R}^n \setminus \{0\}$,
- (N2) $|\alpha x| = |\alpha||x|$ für alle $x \in \mathbb{R}^n$ und $\alpha \in \mathbb{R}$,
- (N3) $|x + y| \leq |x| + |y|$ für alle $x, y \in \mathbb{R}^n$.

Allgemein nennt man eine Funktion $|\cdot| : \mathbb{R}^n \rightarrow \mathbb{R}$ mit diesen drei Eigenschaften eine *Norm*. Andere Normen sind z.B. $|x|_0 := \max_i |x_i|$ oder $|x|_1 := \sum_i |x_i|$. Man kann sich überlegen (vgl. z.B. [Barner-Flohr], S. 13), dass es für jede Norm $|\cdot|'$ Konstanten $A, B > 0$ gibt, so dass

$$\forall_{x \in \mathbb{R}^n} A \cdot |x|' \leq |x| \leq B \cdot |x|'.$$

Man sagt dann, dass die Normen $|\cdot|$ und $|\cdot|'$ *äquivalent* sind. Für $|\cdot|_0$ z.B. sind $A = 1$ und $B = \sqrt{n}$ (vgl. (*)). Den Begriff der Norm kann man nicht nur für den \mathbb{R}^n , sondern ebenso für jeden reellen Vektorraum V erklären. Wenn V unendliche Dimension hat, gibt es nicht-äquivalente Normen auf V .

2. Folgen und Konvergenz

Eine *Folge* im \mathbb{R}^n ist eine Abbildung von den natürlichen Zahlen in den \mathbb{R}^n ; wir bezeichnen sie mit $(x_k)_{k \in \mathbb{N}}$ oder (x_k) oder $(x(k))$. Jedes Folgenglied x_k ist selbst ein Vektor (x_k^1, \dots, x_k^n) ; die Folge (x_k) in \mathbb{R}^n definiert also n reellwertige Folgen

$$(x_k^1)_{k \in \mathbb{N}}, \dots, (x_k^n)_{k \in \mathbb{N}}.$$

(Um Verwechslungen mit dem Folgenindex zu vermeiden, schreiben wir den Komponentenindex vorübergehend als Hochzahl; bitte nicht als Potenz lesen!) Eine Folge (x_k) in \mathbb{R}^n *konvergiert gegen* $x \in \mathbb{R}^n$ (kurz: $x_k \rightarrow x$) oder x ist *Grenzwert* oder *Limes* der Folge (x_k) (kurz: $x = \lim x_k$) wenn die reellwertige Folge $(|x - x_k|)$ eine Nullfolge ist, wenn also gilt:

$$\forall \epsilon > 0 \exists N \forall k \geq N |x_k - x| < \epsilon.$$

Eine Folge (x_k) in \mathbb{R}^n heißt *Cauchyfolge*, wenn gilt

$$\forall \epsilon > 0 \exists N \forall k, m \geq N |x_k - x_m| < \epsilon.$$

Wegen der Dreiecksungleichung sind konvergente Folgen auch Cauchyfolgen, denn aus $|x_k - x| < \epsilon$ und $|x_m - x| < \epsilon$ für $k, m > N$ folgt $|x_k - x_m| < 2\epsilon$. Wegen der Ungleichungen

$$|x_k^i - x^i| \leq |x_k - x|$$

für alle $i = 1, \dots, n$ sowie

$$|x_k - x| \leq \sqrt{n} \cdot \max_i |x_k^i - x^i|$$

erhalten wir (nachprüfen!):

2. Satz 2.1 *Eine Folge (x_k) in \mathbb{R}^n ist genau dann konvergent gegen $x = (x^1, \dots, x^n) \in \mathbb{R}^n$, wenn ihre Komponentenfolgen (x_k^i) für alle $i = 1, \dots, n$ gegen x^i konvergieren.*

Da eine entsprechende Aussage (aus demselben Grund) auch für Cauchyfolgen gilt und Cauchyfolgen in \mathbb{R} konvergieren (Vollständigkeitsaxiom für \mathbb{R}), erhalten wir als Folgerung:

3. Satz 2.2 *Der \mathbb{R}^n ist vollständig, d.h. jede Cauchyfolge in \mathbb{R}^n konvergiert.*

Bemerkung. Man kann Satz 2.1 und die analoge Aussage für Cauchyfolgen auch so deuten, dass Konvergenz und Cauchyfolgen-Eigenschaft erhalten bleiben, wenn man von der euklidischen Betragsnorm $|\cdot|$ zur Maximumsnorm $|x|_0 = \max_i |x_i|$ übergeht. Allgemein sieht man sofort,

dass diese Eigenschaften nicht von der Wahl der Norm abhängen, da alle Normen im \mathbb{R}^n äquivalent sind (vgl. vorige Bemerkung).

Eine Folge (x_k) in \mathbb{R}^n heißt *beschränkt*, wenn es eine Konstante $C \in \mathbb{R}$ (unabhängig von k) gibt, so dass $|x_k| \leq C$ für alle $k \in \mathbb{N}$. Eine *Teilfolge* (x_{k_j}) einer Folge (x_k) in \mathbb{R}^n ist wie üblich die Verkettung (Hintereinanderschaltung, Komposition) der Folge (x_k) mit einer streng monotonen Folge (k_j) mit Werten in den natürlichen Zahlen \mathbb{N} .

4. Satz 2.3 (*Bolzano-Weierstraß-Eigenschaft*) *Jede beschränkte Folge in \mathbb{R}^n besitzt eine konvergente Teilfolge.*

Beweis. Wir beweisen den Satz durch Induktion über die Dimension n . Für $n = 1$ ist die Aussage bekannt. Nun sei $n \geq 2$ und (z_k) eine beschränkte Folge in \mathbb{R}^n ; es gelte also $|z_k| < C$ für alle $k \in \mathbb{N}$. Wir spalten den \mathbb{R}^n auf als $\mathbb{R}^n = \mathbb{R}^{n-1} \times \mathbb{R}$. Jedes Folgenglied $z(k) = z_k$ hat also die Darstellung $z(k) = (x(k), y(k))$ mit $x(k) \in \mathbb{R}^{n-1}$ und $y(k) \in \mathbb{R}$. Da $|x(k)|, |y(k)| \leq |z(k)| < C$ für alle k , sind die Folgen $(x(k))$ in \mathbb{R}^{n-1} und $(y(k))$ in \mathbb{R} beschränkt. Nach Induktionsvoraussetzung besitzt $(x(k))$ eine konvergente Teilfolge $x(k_j) \rightarrow x$. Da die Teilfolge $y(k_j)$ ebenfalls beschränkt ist, besitzt sie (nach der Aussage des Satzes für $n = 1$) wiederum eine konvergente Teilfolge $y(k_{j_m}) \rightarrow y$. Teilfolgen konvergenter Folgen konvergieren gegen denselben Limes (überlegen!), also gilt auch $x(k_{j_m}) \rightarrow x$ und damit $z(k_{j_m}) \rightarrow z = (x, y)$. Da $(z(k_{j_m}))$ eine Teilfolge der gegebenen Folge $(z(k))$ ist, ist der Satz bewiesen. \square

Den Limes einer Teilfolge nennt man auch einen *Häufungspunkt* der Folge, so dass man den obigen Satz auch so formulieren kann: Jede beschränkte Folge in \mathbb{R}^n besitzt einen Häufungspunkt.

3. Offene, abgeschlossene und kompakte Teilmengen

In der Analysis einer Veränderlichen spielen offene und abgeschlossene Intervalle eine Rolle: Abgeschlossene Intervalle enthalten ihre Randpunkte, offene dagegen nicht. Um die Begriffe “offen” und “abgeschlossen” auf den \mathbb{R}^n zu verallgemeinern, müssen wir zunächst sagen, was wir unter “Randpunkten” verstehen wollen. Für jedes $x \in \mathbb{R}^n$ und jede Zahl $r > 0$ definieren wir zunächst den *Ball um x mit Radius r* ,

$$B_r(x) = \{x' \in \mathbb{R}^n; |x' - x| < r\}.$$

Ist nun $T \subset \mathbb{R}^n$, so heißt $x \in \mathbb{R}^n$ ein *Randpunkt* von T , wenn jeder Ball um x (mit beliebig kleinem Radius) sowohl Punkte von T als auch von $\mathbb{R}^n \setminus T$ enthält (Abb. 4). Die Menge aller Randpunkte von T nennen wir ∂T (*Rand von T*). Einen Randpunkt können wir auch so kennzeichnen:



ABBILDUNG 4

$x \in \partial T$ genau dann, wenn es Folgen (t_k) in T und (s_k) in $\mathbb{R}^n \setminus T$ gibt, die beide gegen x konvergieren. Ist nämlich $x \in \partial T$, so können wir Punkte $t_k \in B_{1/k}(x) \cap T$ und $s_k \in B_{1/k}(x) \cap (\mathbb{R}^n \setminus T)$ für jedes $k \in \mathbb{N}$ finden und haben damit die gewünschten Folgen. Ist x dagegen *kein* Randpunkt von T , so gibt es einen Ball $B_r(x)$, der entweder ganz in T oder ganz in $\mathbb{R}^n \setminus T$ liegt, und entweder kann keine Folge in T oder keine Folge in $\mathbb{R}^n \setminus T$ gegen x konvergieren, denn sie kann nicht näher als Abstand r an x herankommen.

Eine Teilmenge $T \subset \mathbb{R}^n$ heißt nun *abgeschlossen*, wenn $\partial T \subset T$, und *offen*, wenn $\partial T \subset \mathbb{R}^n \setminus T$. Offene Mengen sind also durch die Eigenschaft gekennzeichnet, dass sie mit jedem Punkt auch noch einen kleinen Ball um diesen Punkt ganz enthalten (andernfalls wäre dieser Punkt ja Randpunkt). Eine abgeschlossene Menge T kann man so kennzeichnen: Jede Folge in T , die in \mathbb{R}^n konvergiert, hat Grenzwert in T . Ist diese Eigenschaft nämlich erfüllt, so liegt insbesondere jeder Randpunkt in T , also ist T abgeschlossen. Ist sie nicht erfüllt, gibt es eine Folge (t_k) in T mit $t_k \rightarrow x$ und $x \notin T$. Dann ist $x \in \partial T \setminus T$, also ist T auch nicht abgeschlossen.

Beispiele offener Mengen sind die Bälle $B_r(x)$, denn für alle $x' \in B_r(x)$ gilt $B_{r'}(x') \subset B_r(x)$ für $r' = r - |x' - x|$. Dies gilt nach Dreiecksungleichung (Abb. 5): Ist $x'' \in B_{r'}(x')$, so ist

$$|x'' - x| \leq |x'' - x'| + |x' - x| < r' + |x - x'| = r,$$

also ist $x'' \in B_r(x)$.

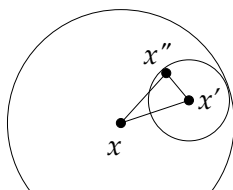


ABBILDUNG 5

Nach Definition ist das Komplement einer offenen Menge abgeschlossen und das Komplement einer abgeschlossenen Menge offen. Für jede Teilmenge $T \subset \mathbb{R}^n$ sind ∂T und $T \cup \partial T$ abgeschlossen (überlegen!), die letztere Menge $T \cup \partial T$ nennt man auch den *Abschluß* von T und

bezeichnet ihn mit \bar{T} oder $\text{Clos}(T)$. Natürlich sind die meisten Mengen $T \subset \mathbb{R}^n$ weder offen noch abgeschlossen, d.h. ∂T schneidet sowohl T als auch $\mathbb{R}^n \setminus T$. Die halboffenen Intervalle in \mathbb{R} sind die einfachsten Beispiele solcher Mengen. Gleichzeitig offen und abgeschlossen kann eine Menge $T \subset \mathbb{R}^n$ nur dann sein, wenn ∂T leer ist; dies ist nur dann der Fall, wenn entweder T oder $\mathbb{R}^n \setminus T$ leer sind, also genau für $T = \emptyset$ oder $T = \mathbb{R}^n$. Ist nämlich $x \in T$ und $x' \in \mathbb{R}^n \setminus T$, so gibt es auf der Strecke von x nach x' , parametrisiert durch $x(t) = tx' + (1-t)x$, $t \in [0, 1]$, einen letzten Punkt, der noch in \bar{T} liegt, nämlich $x(t_0)$ mit $t_0 = \sup\{t; x(t) \in \bar{T}\}$, und es folgt $x(t_0) \in \partial T$ (überlegen!).

Bemerkung. Man kann die Begriffe offen und abgeschlossen von Teilmengen von \mathbb{R}^n auf Teilmengen von $M \subset \mathbb{R}^n$ erweitern: Ist $T \subset M$, so nennt man $x \in M$ Randpunkt von T *relativ zu M* , wenn jeder Ball um x Punkte sowohl von T als auch von $M \setminus T$ enthält; die Menge dieser relativen Randpunkte heie $\partial_M T$ (*relativer Rand*) (Abb. 6). Offene und

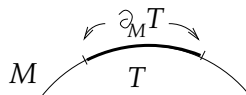


ABBILDUNG 6

abgeschlossene Mengen *relativ zu M* werden wie oben definiert, wobei ∂T durch $\partial_M T$ zu ersetzen ist. Man kann relativ offene und abgeschlossene Mengen in M auch kennzeichnen als Durchschnitte gewöhnlicher offener und abgeschlossener Mengen mit M .

Der Begriff “abgeschlossen” gestattet noch eine wichtige Verfeinerung. Wenn wir noch einmal an den eindimensionalen Fall denken, so gibt es zwei Sorten abgeschlossener Intervalle, nämlich beschränkte $([a, b])$ und unbeschränkte $((-\infty, b], [a, \infty), (-\infty, \infty))$. Gute Sätze galten meist nur in beschränkten abgeschlossenen Intervallen. Die richtige Verallgemeinerung davon ist der Begriff der Kompaktheit. Eine Teilmenge $K \subset \mathbb{R}^n$ heißt *kompakt*, wenn jede Folge in K einen Häufungspunkt in K (also eine in K konvergente Teilfolge) besitzt. (Vgl. S. 86 für eine alternative Definition.)

5. Satz 3. *Eine Teilmenge $K \subset \mathbb{R}^n$ ist kompakt genau dann, wenn sie beschränkt (also in einem Ball enthalten) und abgeschlossen ist.*

Beweis. “ \Rightarrow ”: Ist $K \subset \mathbb{R}^n$ kompakt, so ist K beschränkt, denn andernfalls gäbe es eine Folge (x_k) in K mit $|x_k| \geq k$ für alle $k \in \mathbb{N}$; jede Teilfolge einer solchen Folge muß aber divergieren, da die Beträge gegen unendlich streben. Außerdem ist K abgeschlossen, denn ist $x \in \partial K$

und (x_k) eine Folge in K mit $x_k \rightarrow x$, so gibt es eine in K konvergente Teilfolge $x_{k_j} \rightarrow x' \in K$; da aber die ganze Folge (x_k) gegen x konvergiert, muß $x = x' \in K$ und damit $\partial K \subset K$ gelten.

“ \Leftarrow ”: Ist $K \subset \mathbb{R}^n$ beschränkt und abgeschlossen und (x_k) eine Folge in K , so ist (x_k) beschränkt und besitzt daher nach Satz 2.3 eine konvergente Teilfolge. Da K abgeschlossen ist, liegt der Limes dieser Teilfolge wieder in K . Dies zeigt die Kompaktheit von K . \square

4. Stetige Abbildungen

Von jetzt an wollen wir den Hauptgegenstand der Analysis betrachten, nämlich Funktionen und Abbildungen; Abbildungen mit Werten in \mathbb{R} (oder \mathbb{C}) heißen *Funktionen*. Es sei also eine Abbildung $f : D \rightarrow \mathbb{R}^p$ mit $D \subset \mathbb{R}^n$ gegeben. Wir geben drei äquivalente Definitionen der Stetigkeit in einem Punkt: Die Abbildung f heißt *stetig* im Punkt $x \in D$, wenn

- (1) für jede Folge $x_k \rightarrow x$ in D gilt $f(x_k) \rightarrow f(x)$,
- (2) $\forall \epsilon > 0 \exists \delta > 0 \forall x' \in D [|x' - x| < \delta \Rightarrow |f(x') - f(x)| < \epsilon]$,
- (3) $\forall \epsilon > 0 \exists \delta > 0 f(B_\delta(x) \cap D) \subset B_\epsilon(f(x))$.

Die Abbildung f heißt *stetig*, wenn f stetig in x ist für alle $x \in D$.

(2) und (3) sind äquivalent, denn die Aussage $\forall x' \in D [|x' - x| < \delta \Rightarrow |f(x') - f(x)| < \epsilon]$ ist äquivalent zu $\forall x' [x' \in B_\delta(x) \cap D \Rightarrow f(x) \in B_\epsilon(f(x))]$ und damit zu $f(B_\delta(x) \cap D) \subset B_\epsilon(f(x))$. dass (1) und (2) äquivalent sind, sieht man so: Ist (2) erfüllt und $x_k \rightarrow x$ eine in D konvergente Folge, und ist $\epsilon > 0$ vorgegeben und δ wie in (2), so gilt $|x_k - x| < \delta$ für *genügend großes* k (d.h. $\exists N \forall k \geq N \dots$), somit $|f(x_k) - f(x)| < \epsilon$ und also $f(x_k) \rightarrow f(x)$, d.h. (1) gilt. Ist dagegen (2) nicht erfüllt, so gilt durch Negation von (2) und Spezialisierung $\delta = 1/k$ für $k \in \mathbb{N}$:

$$\exists \epsilon > 0 \forall k \in \mathbb{N} \exists x_k \in D [|x_k - x| < \frac{1}{k} \wedge |f(x_k) - f(x)| \geq \epsilon]$$

(\wedge ist die Abkürzung für “und”). So entsteht eine Folge (x_k) in D mit $x_k \rightarrow x$, aber $f(x_k) \not\rightarrow f(x)$, also ist auch (1) nicht erfüllt. Somit sind (1) und (2) äquivalent.

Bemerkung. (3) kann auch folgendermaßen geschrieben werden:

$$(4) \quad \forall \epsilon > 0 \exists \delta > 0 B_\delta(x) \cap D \subset f^{-1}B_\epsilon(f(x)),$$

wobei $f^{-1}(A)$ das *Urbild* einer Menge $A \subset \mathbb{R}^p$ bezeichnet, also

$$f^{-1}A = \{x \in D; f(x) \in A\}.$$

Dies führt zu einer weiteren Kennzeichnung der Stetigkeit: *Eine Abbildung $f : D \rightarrow \mathbb{R}^p$ ist stetig genau dann, wenn Urbilder offener Mengen offen sind.* Ist f nämlich stetig, $U \subset \mathbb{R}^p$ offen und $x \in f^{-1}U$, so ist $B_\epsilon(f(x)) \subset U$ für ein genügend kleines $\epsilon > 0$, und nach (4) gehört noch $B_\delta(x) \cap D$ für ein $\delta > 0$ zu $f^{-1}U$, so dass $x \notin \partial_D(f^{-1}U)$ und daher $f^{-1}U$ offen ist. Umgekehrt: Sind Urbilder offener Mengen stets offen, so ist für jedes $x \in D$ auch $f^{-1}B_\epsilon(f(x))$ offen in D und enthält x . Also gibt es ein $\delta > 0$, so dass $B_\delta(x) \cap D \subset f^{-1}B_\epsilon(f(x))$, und (4) ist erfüllt.

Vier Sätze über stetige Abbildungen sind es, die wir immer wieder verwenden werden: Zwischenwertsatz, Existenz von Extrema und gleichmäßige Stetigkeit bei kompaktem Definitionsbereich, Erhaltung der Stetigkeit bei gleichmäßiger Konvergenz.

Eine Menge $D \subset \mathbb{R}^n$ heißt *wegweise zusammenhängend*, wenn es für alle $x, x' \in D$ eine stetige Abbildung $w : [0, 1] \rightarrow D$ gibt mit $w(0) = x$ und $w(1) = x'$. Solch eine Abbildung heißt *Weg von x nach x' in D* .

6. Satz 4.1 (Zwischenwertsatz) *Ist $D \subset \mathbb{R}^n$ wegweise zusammenhängend, $f : D \rightarrow \mathbb{R}$ eine stetige Funktion und $x_0, x_1 \in D$ mit $f(x_0) \leq f(x_1)$, dann gibt es für jedes $y \in [f(x_0), f(x_1)]$ ein $x \in D$ mit $f(x) = y$.*

Beweis. Wir wenden den Zwischenwertsatz der eindimensionalen Analysis (Intervallhalbierungsverfahren!) auf $f \circ w : [0, 1] \rightarrow \mathbb{R}$ an, wobei w ein Weg von x_0 nach x_1 in D ist. \square

Ein (*globales*) *Maximum* und ein (*globales*) *Minimum* einer Funktion $f : D \rightarrow \mathbb{R}$ sind Punkte $x_+, x_- \in D$ mit $f(x_-) \leq f(x) \leq f(x_+)$ für alle $x \in D$. (Manchmal nennt man allerdings auch die Werte $f(x_+), f(x_-)$ Maximum und Minimum.)

7. Satz 4.2 *Ist $D \subset \mathbb{R}^n$ kompakt und $f : D \rightarrow \mathbb{R}$ stetig, so besitzt f (mindestens) ein Maximum und ein Minimum.*

Beweis. Es sei $s = \sup\{f(x); x \in D\} \leq \infty$. Dann ist $s \geq f(x)$ für alle $x \in D$, und es gibt eine Folge (x_k) in D mit $f(x_k) \rightarrow s$. Da D kompakt ist, besitzt (x_k) einen Häufungspunkt $x_+ \in D$, d.h. es gibt eine Teilfolge $x_{k_j} \rightarrow x_+$. Da f stetig ist (Def.(1)), folgt $f(x_{k_j}) \rightarrow f(x_+)$, andererseits aber $f(x_{k_j}) \rightarrow s$, also gilt $f(x_+) = s$, also ist x_+ Maximum, und insbesondere ist $s < \infty$. Ebenso gewinnt man ein Minimum, oder man wendet den vorstehenden Beweis auf $-f$ an. \square

8. Korollar *Das Bild einer kompakten, wegweise zusammenhängenden Menge $D \subset \mathbb{R}$ unter einer stetigen Funktion $f : D \rightarrow \mathbb{R}$ ist ein kompaktes Intervall.*

Beweis. Wir setzen $a = f(x_-)$ und $b = f(x_+)$ (vgl. Satz 4.2). Nach Satz 4.1 ist $f(D) = [a, b]$. \square

In vielen Fällen, besonders in der Integrationstheorie, ist eine leichte Verschärfung des Stetigkeitsbegriffs nützlich: Eine Abbildung $f : D \rightarrow \mathbb{R}^p$ (mit $D \subset \mathbb{R}^n$) heißt *gleichmäßig stetig*, wenn gilt:

$$\forall \epsilon > 0 \exists \delta > 0 \forall x, x' \in D [|x' - x| < \delta \Rightarrow |f(x') - f(x)| < \epsilon].$$

Im Unterschied zur gewöhnlichen Stetigkeit ist das δ jetzt auch unabhängig von x (und nicht nur von x'). Natürlich folgt die Stetigkeit aus der gleichmäßigen Stetigkeit, aber bei kompaktem Definitionsbereich gilt eben auch die Umkehrung:

9. Satz 4.3 *Ist $D \subset \mathbb{R}^n$ kompakt und $f : D \rightarrow \mathbb{R}^p$ eine stetige Abbildung, so ist f gleichmäßig stetig.*

Beweis. Durch Widerspruch: Wir nehmen an, f sei *nicht* gleichmäßig stetig. Dann folgt wie üblich durch Spezialisieren von δ zu $1/k$:

$$(*) \quad \exists \epsilon > 0 \forall k \in \mathbb{N} \exists x_k, x'_k \in D \left[|x'_k - x_k| < \frac{1}{k} \wedge |f(x'_k) - f(x_k)| \geq \epsilon \right].$$

Da D kompakt ist, besitzt (x_k) einen Häufungspunkt x , es gibt also eine Teilfolge $x_{k_j} \rightarrow x$. Da $|x'_{k_j} - x_{k_j}| < 1/k_j \rightarrow 0$, folgt auch $x'_{k_j} \rightarrow x$. Wegen der Stetigkeit in x konvergieren die Folgen $f(x_{k_j})$ und $f(x'_{k_j})$ beide gegen $f(x)$, insbesondere folgt $|f(x_{k_j}) - f(x'_{k_j})| \rightarrow 0$, was im Widerspruch zu $(*)$ steht. \square

Beispiele stetiger Abbildungen $f = (f_1, \dots, f_p) : D \rightarrow \mathbb{R}^p$ sind zunächst solche, deren Komponenten f_i mit nur mit Hilfe der vier Grundrechenarten aus den Koordinatenfunktionen $x_1, \dots, x_n : \mathbb{R}^n \rightarrow \mathbb{R}$ gebildet werden, d.h. *rationale Funktionen* sind; dies folgt sofort aus Def. (1) und den Rechenregeln für konvergente Folgen. Weiterhin folgt aus Def. (1) unmittelbar, dass Verkettungen stetiger Funktionen wieder stetig sind; deshalb ist z.B. die euklidische Norm $x \mapsto |x|$ stetig (als Verkettung der Quadratwurzel mit der rationalen Funktion $\sum_j x_j^2$). Weitere Funktionen (die allermeisten) entstehen als Grenzwerte konvergenter Funktionenfolgen $f_k : D \rightarrow \mathbb{R}^p$, $k \in \mathbb{N}$. Eine solche Funktionenfolge heißt *punktweise konvergent*, wenn die Folge $(f_k(x))$ in \mathbb{R}^p für alle $x \in D$ konvergiert; der Grenzwert $f(x) := \lim f_k(x)$ definiert eine neue Funktion $f : D \rightarrow \mathbb{R}^p$. Wir schreiben dafür $f \xrightarrow{pw} f$. Leider muß f nicht stetig sein, auch wenn alle f_k stetig sind, wie das Beispiel $f_k : [0, 1] \rightarrow \mathbb{R}$, $f_k(x) = x^k$ zeigt: Die Limesfunktion f ist 0 auf $[0, 1)$

und springt bei 1 auf den Wert 1. Um die Stetigkeit der Limesabbildung zu sichern, benötigt man einen stärkeren Konvergenzbegriff für Folgen von Abbildungen: Für eine Abbildung $f : D \rightarrow \mathbb{R}^p$ definieren wir die *Supremumsnorm*

$$\|f\| := \sup_{x \in D} |f(x)| \leq \infty.$$

f heißt *beschränkt*, wenn dieses Supremum endlich ist. Eine Folge von Abbildungen $f_k : D \rightarrow \mathbb{R}^p$ heißt *gleichmäßig konvergent* gegen eine Abbildung $f : D \rightarrow \mathbb{R}^p$, symbolisch $f_k \xrightarrow{glm} f$, wenn $\|f_k - f\| \rightarrow 0$, oder mit anderen Worten, wenn

$$\forall \epsilon > 0 \exists N \forall k \geq N \forall x \in D |f_k(x) - f(x)| < \epsilon.$$

Der Unterschied zur punktweisen Konvergenz ist, dass die Konstante N hier von x unabhängig ist. (Abb. 7)

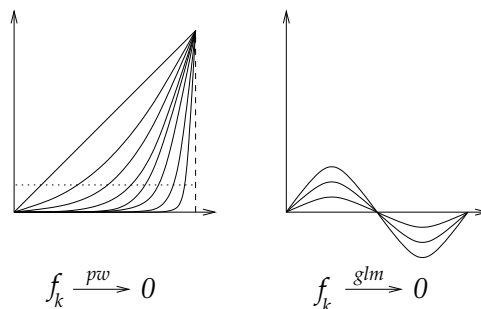


ABBILDUNG 7

10. Satz 4.4 *Ist $f_k : D \rightarrow \mathbb{R}^p$ eine Folge stetiger Funktionen mit $f_k \xrightarrow{glm} f$, so ist auch $f : D \rightarrow \mathbb{R}^p$ stetig.*

Beweis. Es sei $x \in D$ fest. Für alle $x' \in D$ gilt nach Dreiecksungleichung:

$$|f(x') - f(x)| \leq |f(x') - f_k(x')| + |f_k(x') - f_k(x)| + |f_k(x) - f(x)|$$

für jedes feste $k \in \mathbb{N}$. Wir wählen k so groß, dass $\|f_k - f\| < \epsilon$. Damit sind der erste und der dritte Term auf der rechten Seite durch ϵ beschränkt. Wegen der Stetigkeit von f_k in x können wir ein $\delta > 0$ finden, so dass $|f_k(x') - f_k(x)| < \epsilon$ für alle $x' \in D$ mit $|x' - x| < \delta$. Für solche x' ist also auch der mittlere Term durch ϵ beschränkt und wir erhalten $|f(x') - f(x)| < 3\epsilon$ für solche x' , also die Stetigkeit von f in x . \square

Bemerkung. Die beschränkten stetigen Abbildungen $f : D \rightarrow \mathbb{R}^p$ bilden einen (unendlich-dimensionalen) Vektorraum V , und die Supremumsnorm ist eine Norm auf V im Sinne der Definition in §1. Eine Konsequenz des Satzes 4.4 ist, dass dieser Vektorraum mit Norm *vollständig* ist, d.h. jede Supremumsnorm-Cauchyfolge konvergiert bezüglich der Supremumsnorm. Ist nämlich (f_k) eine solche Cauchyfolge in V , so ist $(f_k(x))$ für jedes $x \in D$ eine Cauchyfolge in \mathbb{R}^p , konvergiert also gegen einen Wert $f(x) \in \mathbb{R}^p$, und aus der Cauchyfolgen-Eigenschaft $\|f_k - f_m\| < \epsilon$ für alle genügend großen $k, m \in \mathbb{N}$ folgt sofort $\|f_k - f\| \leq \epsilon$, also $f_k \xrightarrow{glm} f$. Insbesondere ist f beschränkt und nach Satz 4.4 stetig, also ist $f \in V$, und (f_k) konvergiert gegen f bezüglich der Supremumsnorm, was zu zeigen war. Vektorräume mit Norm, die bezüglich dieser Norm vollständig sind, nennt man auch *Banachräume*; der Vektorraum der beschränkten stetigen Funktionen auf D mit der Supremumsnorm ist also ein Banachraum. Der Vorteil dieser Begriffsbildungen ist, dass man geometrische Ideen auf sehr abstrakte Gebilde wie Räume von Abbildungen anwenden kann. Wir werden in einer späteren Bemerkung eine solche Anwendung kennen lernen (§12, S. 42).

II. Differentiation

5. Kurven

Unter einer *Kurve* wollen wir eine differenzierbare Abbildung $c = (c_1, \dots, c_n) : I \rightarrow \mathbb{R}^n$ verstehen, wobei $I \subset \mathbb{R}$ ein Intervall ist. “*Differenzierbar*” soll heißen, dass alle Komponentenfunktionen $c_j : I \rightarrow \mathbb{R}$ für $j = 1, \dots, n$ differenzierbar sind. Wollen wir eine solche Abbildung anschaulich darstellen, so zeichnen wir nicht den Graphen (wie bei Funktionen von I nach \mathbb{R}), sondern die Bildmenge $c(I)$ (Abb. 8); dieser gilt unser Hauptinteresse. Die Variable $t \in I$ wird auch *Parameter der*

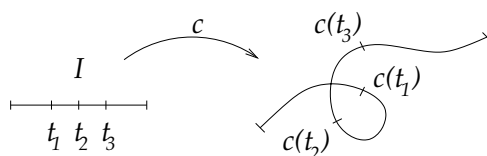


ABBILDUNG 8

Kurve genannt. Physikalisch wird der Parameter oft als *Zeit* gedeutet, daher der Name t (“time”). Die *Ableitung der Kurve c im Punkte $t \in I$* ist der Vektor

$$c'(t) = (c'_1(t), \dots, c'_n(t)) = \lim_{h \rightarrow 0} \frac{c(t+h) - c(t)}{h}.$$

Da dieser Vektor den Grenzwert des mit $1/h$ gestreckten Sekantenvektors $c(t+h) - c(t)$ darstellt, wird er auch als *Tangentenvektor* von c in t bezeichnet (Abb. 9), oder als *Geschwindigkeitsvektor*, wenn der Pa-

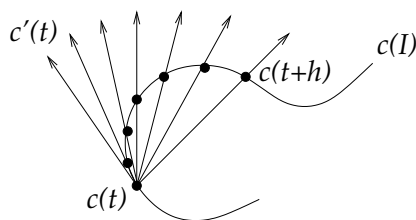


ABBILDUNG 9

rameter als “Zeit” interpretiert wird (Geschwindigkeit = Weg / Zeit). Die Kurve c heißt *regulär*, wenn $c'(t) \neq 0$ für alle $t \in I$.

In der Schule bezeichnet man als Kurve gewöhnlich den Graphen einer differenzierbaren Funktion $y : I \rightarrow \mathbb{R}$, also die Menge $\text{Graph}(y) = \{(x, y(x)) \in \mathbb{R}^2; x \in I\}$. Dies ist ein Spezialfall unserer Definition, denn $\text{Graph}(y)$ ist das Bild der Kurve $c : I \rightarrow \mathbb{R}^2$, $c(x) = (x, y(x))$. Diese

Graphen-Kurven im \mathbb{R}^2 haben die Eigenschaft, dass sie *schlicht* über der x -Achse liegen, d.h. dass über jedem Punkt der x -Achse höchstens ein Punkt der Kurve liegt. Für eine beliebige Kurve im \mathbb{R}^2 braucht dies nicht zu gelten (vgl. Abb. 8). Manche Ideen lassen sich aber von den Graphen-Kurven auf beliebige Kurven in \mathbb{R}^2 verallgemeinern, z.B. dass jede Sekante parallel zu einer Tangente ist (Abb. 10):

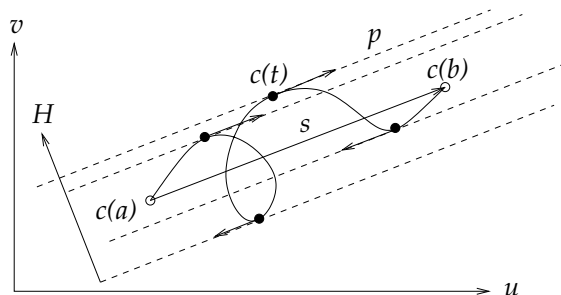


ABBILDUNG 10

11. Satz 5.1 (*Erweiterter Mittelwertsatz*) Ist $c = (u, v) : [a, b] \rightarrow \mathbb{R}^2$ stetig und auf dem offenen Intervall (a, b) differenzierbar, und gilt $c(a) \neq c(b)$, so gibt es (mindestens) ein $t \in (a, b)$, so dass der Tangentenvektor $c'(t)$ parallelgerichtet zum Sekantenvektor $c(b) - c(a)$ ist, d.h. $c'(t) \in \mathbb{R} \cdot (c(b) - c(a))$, oder mit anderen Worten:

$$(*) \quad (u(b) - u(a))v'(t) = (v(b) - v(a))u'(t).$$

Beweis. Wir wenden den Satz von Rolle (“Zwischen zwei Stellen gleichen Wertes einer Funktion liegt eine Nullstelle der Ableitung”; genaue Voraussetzungen?) an auf die Funktion $h : [a, b] \rightarrow \mathbb{R}$,

$$h(t) = (u(b) - u(a))v(t) - (v(b) - v(a))u(t)$$

Die Werte $h(a)$ und $h(b)$ sind gleich, nämlich gleich $u(b)v(a) - v(b)u(a)$, und für jede Nullstelle t von h' gilt (*). Also folgt die Behauptung aus dem Satz von Rolle. \square

Bemerkung. Wie leider üblich fällt hier die im Beweis benötigte Funktion h “vom Himmel”. Abb. 10 suggeriert aber eine geometrische Beweisidee, die sofort auf h führt: Die Gerade durch die Punkte $c(a)$ und $c(b)$ heiße s . Man schiebe eine Parallele zu s von “weit draußen” an das Bild von c heran; diejenige Parallele p , die zum ersten Mal das Bild berührt, muß eine Tangente parallel zu s sein. Der obige Beweis übersetzt dieses geometrische Argument in die Sprache der Analysis: Die zu s parallelen Geraden sind die Konstanzlinien der Koordinate H senkrecht zu s , der “Höhenfunktion” $H(x) = \langle n, x \rangle$, wobei n ein Vektor

ist, der senkrecht auf s steht. Da s in Richtung des Vektors $c(b) - c(a) = (u(b) - u(a), v(b) - v(a))$ verläuft, ist $n = -(v(b) - v(a), u(b) - u(a))$ senkrecht zu s . Die im Beweis von Satz 5.1 verwendete Funktion h ist gleich $H \circ c$, denn

$$H(c(t)) = \langle n, c(t) \rangle = -(v(b) - v(a))u(t) + (u(b) - u(a))v(t) = h(t).$$

Die Parallele p zu s berührt das Bild von c genau dort, wo h ein Extremum hat; man erinnere sich, dass die Nullstelle von h' im Beweis des Satzes von Rolle eine Extremalstelle von h ist.

12. Korollar (Satz von de l'Hôpital)¹ Es seien $u, v : (0, \infty) \rightarrow \mathbb{R}$ differenzierbare Funktionen mit $u(x), v(x) \rightarrow \infty$ für $x \rightarrow \infty$. Dann existiert der Limes des Quotienten $u(x)/v(x)$ für $x \rightarrow \infty$, wenn $v'(x) \neq 0$ für hinreichend große x und der Quotient der Ableitungen $u'(x)/v'(x)$ für $x \rightarrow \infty$ einen Limes besitzt, und die beiden Limiten sind gleich:

$$\lim_{x \rightarrow \infty} \frac{u(x)}{v(x)} = \lim_{x \rightarrow \infty} \frac{u'(x)}{v'(x)} =: L.$$

Beweis. Wir wenden Satz 5.1 an mit $a = x$ und $b = \tilde{x} > x$:

$$\frac{u(\tilde{x}) - u(x)}{v(\tilde{x}) - v(x)} = \frac{u'(\xi)}{v'(\xi)} \quad (*)$$

für ein $\xi \in (x, \tilde{x})$. Für $\tilde{x} \rightarrow \infty$ gilt $u(\tilde{x}), v(\tilde{x}) \rightarrow \infty$, daher $\frac{u(\tilde{x})}{u(x)} \rightarrow 0$ und $\frac{v(\tilde{x})}{v(x)} \rightarrow 0$ für festes x , somit $\frac{u(\tilde{x}) - u(x)}{u(x)} \rightarrow 1$ und $\frac{v(\tilde{x}) - v(x)}{v(x)} \rightarrow 1$. Deshalb können wir für festes x die linke Seite von (*) durch $\frac{u(\tilde{x})}{v(\tilde{x})}$ ersetzen, wenn \tilde{x} genügend groß ist. Die rechte Seite von (*) aber geht gegen L , wenn \tilde{x} (und damit $\xi > x$) nur genügend groß ist. \square

¹Der Satz hat viele Versionen zu unterschiedlichen Fällen. Der einfachste Fall ist: Wenn u und v eine gemeinsame Nullstelle x_o besitzen mit $v'(x_o) \neq 0$, dann ist

$$\lim_{x \rightarrow x_o} \frac{u(x)}{v(x)} = \frac{u'(x_o)}{v'(x_o)}$$

Dies folgt direkt aus der Definition der Differenzierbarkeit:

$$\frac{u(x)}{v(x)} = \frac{u(x) - u(x_o)}{v(x) - v(x_o)} = \frac{u(x) - u(x_o)}{x - x_o} \frac{x - x_o}{v(x) - v(x_o)} \xrightarrow{x \rightarrow x_o} \frac{u'(x_o)}{v'(x_o)}.$$

Statt der Voraussetzung $v'(x_o) \neq 0$ genügt die Annahme, dass $\lim_{x \rightarrow x_o} \frac{u'(x)}{v'(x)}$ existiert; dazu wendet man bereits Satz 5.1 an mit $a = x$, $b = x_o$ oder umgekehrt und erhält für ein ξ zwischen x und x_o :

$$\frac{u(x)}{v(x)} = \frac{u(x) - u(x_o)}{v(x) - v(x_o)} = \frac{u'(\xi)}{v'(\xi)} \frac{x - x_o}{v(x) - v(x_o)} \xrightarrow{x \rightarrow x_o} \frac{u'(x_o)}{v'(x_o)}.$$

Wir wollen nun zwei wichtige Größen diskutieren, die mit einer Kurve im \mathbb{R}^n verbunden sind: die Bogenlänge und das Kurvenintegral. Gegeben sei also eine Kurve, die wir jetzt nicht mehr c , sondern (aus Gründen der physikalischen Notation) x nennen wollen (“Ortsvektor”), genauer $x : [a, b] \rightarrow \mathbb{R}^n$, und wir nehmen zusätzlich an, dass die Ableitung $x' : [a, b] \rightarrow \mathbb{R}^n$ stetig ist; x sei also stetig differenzierbar. Die *Bogenlänge* oder einfach *Länge* der Kurve x ist das “Zeitintegral über die Länge des Geschwindigkeitsvektors”:

$$L(x) = \int_a^b |x'(t)| dt.$$

Beispiel 1: Es seien $x_0, x_1 \in \mathbb{R}^n$ und $x : [0, 1] \rightarrow \mathbb{R}^n$, $x(t) = tx_1 + (1-t)x_0$ die Strecke von x_0 nach x_1 . Dann ist $x'(t) = x_1 - x_0$ für alle t und damit $L(x) = |x_1 - x_0|$. Die Länge der Strecke ist also der Abstand der Endpunkte.

Beispiel 2: $x : [0, \phi] \rightarrow \mathbb{R}^2$, $x(t) = (\cos t, \sin t)$. Dann ist $x'(t) = (-\sin t, \cos t)$ und $|x'(t)| = 1$ für alle t . Somit ist $L(x) = \phi$; der Winkel ϕ ist also die Länge des Kreisbogens von $(1, 0)$ bis $(\cos \phi, \sin \phi)$.

Natürlich sollte die Strecke die kürzeste Verbindung zwischen zwei Punkten sein, d.h. für jede Kurve $x : [a, b] \rightarrow \mathbb{R}^n$ sollte gelten:

$$L(x) \geq |x(b) - x(a)|.$$

Dies ist in der Tat richtig, denn

$$|x(b) - x(a)| \stackrel{1}{=} \left| \int_a^b x'(t) dt \right| \stackrel{2}{\leq} \int_a^b |x'(t)| dt = L(x)$$

Das Gleichheitszeichen 1 gilt nach dem Hauptsatz der Differential- und Integralrechnung, wobei das *Integral über eine vektorwertige Abbildung* als der Vektor der Integrale der Komponentenfunktionen erklärt wird. Die Ungleichung 2 wird im nächste Satz (für $v = x'$) gezeigt:

13. Satz 5.2 Für jede stetige Abbildung $v : [a, b] \rightarrow \mathbb{R}^n$ gilt

$$\int_a^b |v(t)| dt \geq \left| \int_a^b v(t) dt \right|.$$

Beweis. Wir setzen $w = \int_a^b v(t) dt$. Mit Cauchy-Schwarz (Satz 1) ist

$$\begin{aligned} |w|^2 &= \langle w, w \rangle = \left\langle w, \int_a^b v(t) dt \right\rangle = \int_a^b \langle w, v(t) \rangle dt \\ &\leq \int_a^b |w| |v(t)| dt = |w| \cdot \int_a^b |v(t)| dt, \end{aligned}$$

woraus die Behauptung $|w| \leq \int_a^b |v(t)| dt$ folgt. \square

Bemerkung. Wieder scheint hier unsere Intuition zu kurz zu kommen. Zeigt nicht schon der Betrunkene, der hin und hertorkelt und lauter unnütze Wege zur Seite macht, dass die Strecke die kürzeste Verbindung ist? Diese Umwege zur Seite (“Abwege”!) stecken in der Cauchy-Schwarz-Ungleichung. Ist z.B. $w = e_1$ (durch Drehung des Koordinatensystems können wir immer annehmen, dass w ein Vielfaches von e_1 ist), so ist $\langle w, v \rangle = v_1$ und $|w||v| = \sqrt{v_1^2 + v_2^2 + \dots + v_n^2} \geq \sqrt{v_1^2} = |v_1|$. Das, was wir in dieser Abschätzung weglassen, $v_2^2 + \dots + v_n^2$, ist der jeweilige momentane “Abweg” des Betrunkenen.

Der zweite Begriff, den wir diskutieren wollen, ist der des Kurvenintegrals. Gegeben sei dazu eine offene Teilmenge $U^n \subset \mathbb{R}^n$ sowie eine stetige Abbildung $F : U^n \rightarrow \mathbb{R}^n$. Wir deuten diese Abbildung als *Vektorfeld*, indem wir uns für alle $x \in U^n$ den Vektor $F(x)$ im Punkt x angeheftet denken (Abb. 11). Nun sei $x : [a, b] \rightarrow \mathbb{R}^n$ eine stetig

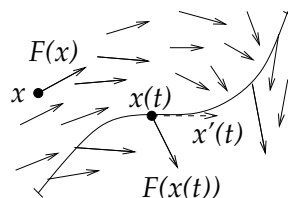


ABBILDUNG 11

differenzierbare Kurve mit Bild $x([a, b]) \subset U^n$. Wir definieren das *Kurvenintegral von F über x* als

$$\int_x \langle F(x), dx \rangle := \int_a^b \langle F(x(t)), x'(t) \rangle dt.$$

Wenn wir F physikalisch als *Kraftfeld* (“force”) und $x(t)$ als Ort zur Zeit t deuten, so gibt das Kurvenintegral bis auf das Vorzeichen die *Arbeit* an, die auf der Bahn x gegen die Kraft F zu leisten ist. Die Arbeit sollte (genau wie die Bogenlänge) sich nicht ändern, wenn wir die Bahn in einem anderen Zeitprogramm (schneller oder langsamer) durchlaufen; dies werden wir gleich sehen.

Wie schon erwähnt, gilt das Hauptinteresse beim Studium einer Kurve $x : I \rightarrow \mathbb{R}^n$ der *Bahn von x* , das ist das Bild $x(I)$ mit ein bißchen zusätzlicher Struktur, die von der Abbildung herkommt; z.B. vergißt man nicht die Durchlauf-Reihenfolge. Die Bahn ändert sich nicht, wenn man eine bijektive stetig differenzierbare Abbildung $\phi : J \rightarrow I$ mit nirgends verschwindender Ableitung (eine umkehrbar stetig differenzierbare Abbildung) vorschaltet und $\tilde{x} = x \circ \phi : J \rightarrow \mathbb{R}^n$ betrachtet.

Eine solche Abbildung ϕ heißt *Parameterwechsel* oder *Umparametrisierung*. Ist zusätzlich $\phi' > 0$, so heißt der Parameterwechsel ϕ *orientiert*. Viele Eigenschaften der Kurve bleiben beim Übergang von x zu $x \circ \phi$ erhalten, z.B. Bogenlänge und Kurvenintegral:

14. Satz 5.3 *Ist $x : [a, b] \rightarrow U^n \subset \mathbb{R}^n$ eine stetig differenzierbare Kurve, $\phi : [c, d] \rightarrow [a, b]$ ein Parameterwechsel und $\tilde{x} = x \circ \phi$, so sind die Längen gleich:*

$$L(\tilde{x}) = L(x).$$

Ist zusätzlich ϕ orientiert und $F : U^n \rightarrow \mathbb{R}^n$ ein stetiges Vektorfeld, so sind die Kurvenintegrale gleich:

$$\int_x \langle F(x), dx \rangle = \int_{\tilde{x}} \langle F(\tilde{x}), d\tilde{x} \rangle.$$

Beweis. Dies folgt aus der Substitutionsregel mit der Substitution $t = \phi(u)$, also $dt = \phi'(u)du$:

$$\begin{aligned} L(\tilde{x}) &= \int_c^d |(x \circ \phi)'(u)| du \\ &= \int_c^d |\phi'(u)| |x'(\phi(u))| du \\ &= \int_a^b |x'(t)| dt, \\ \int_{\tilde{x}} \langle F(\tilde{x}), d\tilde{x} \rangle &= \int_c^d \langle F(\tilde{x}(u)), \tilde{x}'(u) \rangle du \\ &= \int_c^d \phi'(u) \langle F(x(\phi(u))), x'(\phi(u)) \rangle du \\ &= \int_a^b \langle F(x(t)), x'(t) \rangle dt. \end{aligned}$$

□

6. Partielle Ableitungen

Eine Abbildung $f : U^n \rightarrow \mathbb{R}^p$ (wobei $U^n \subset \mathbb{R}^n$ offen ist, wie früher vereinbart) können wir auf verschiedene Weisen auffassen, zum Beispiel als Abbildung von n Veränderlichen $x_1, \dots, x_n \in \mathbb{R}$ oder als Abbildung von einer Veränderlichen $x \in \mathbb{R}^n$. Wir werden in diesem Abschnitt zunächst die erste Auffassung vertreten. Indem wir alle übrigen Variablen festhalten (ihnen einen konstanten Wert geben), können wir aus f wieder eine Abbildung *einer* reellen Variablen x_i für festes $i \in \{1, \dots, n\}$, also eine Kurve in \mathbb{R}^p gewinnen. Sind alle diese Kurven

$x_i \mapsto f(x_1, \dots, x_i, \dots, x_n)$ differenzierbar, so nennen wir f *partiell differenzierbar*. Die Ableitung von $x_i \mapsto f(x_1, \dots, x_i, \dots, x_n)$ heißt *partielle Ableitung nach der i -ten Koordinate* und wird bezeichnet mit $D_i f$ oder $\frac{\partial f}{\partial x_i}$, kurz f_{x_i} .² Etwas genauer: f heißt partiell differenzierbar im Punkt $x \in U$, wenn die Kurve

$$t \mapsto f(x_1, \dots, x_i + t, \dots, x_n) = f(x + te_i)$$

in $t = 0$ differenzierbar ist für alle $i = 1, \dots, n$. (Da U^n offen ist, enthält der Definitionsbereich dieser Abbildung ein Intervall $(-\epsilon, \epsilon)$ für ein $\epsilon > 0$, bitte nachprüfen!) Die Ableitung dieser Kurve in $t = 0$ wird $D_i f(x)$ oder $(\frac{\partial f}{\partial x_i})(x)$ geschrieben. Abb. 12 veranschaulicht die partiellen Ableitungen für $n = 2$, $p = 1$. Die partiellen Ableitungen sind

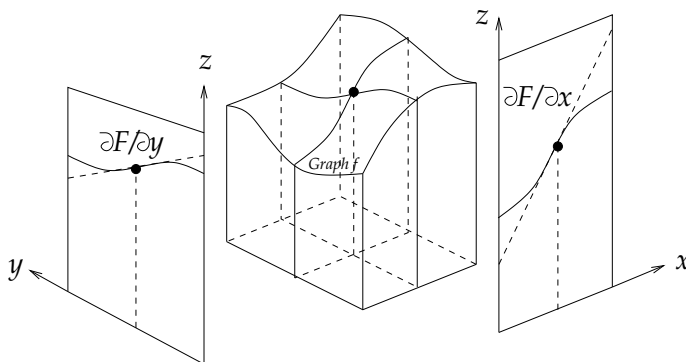


ABBILDUNG 12

wieder Abbildungen $D_i f : U^n \rightarrow \mathbb{R}^p$. Sind diese selbst partiell differenzierbar, so können wir zweite partielle Ableitungen $D_j D_i f$ oder $\frac{\partial^2 f}{\partial x_i \partial x_j}$ bilden, und entsprechend entstehen höhere partielle Ableitungen.

Beispiel $f(x, y) = e^{x^2 y^3}$:

$$\begin{aligned} \frac{\partial f}{\partial x} &= 2xy^3 e^{x^2 y^3}, & \frac{\partial f}{\partial y} &= 3x^2 y^2 e^{x^2 y^3}, \\ \frac{\partial^2 f}{\partial x^2} &= (2y^3 + 4x^2 y^6) e^{x^2 y^3}, & \frac{\partial^2 f}{\partial x \partial y} &= (6xy^2 + 6x^3 y^5) e^{x^2 y^3}, \\ \frac{\partial^2 f}{\partial y \partial x} &= (6xy^2 + 6x^3 y^5) e^{x^2 y^3}, & \frac{\partial^2 f}{\partial y^2} &= (6x^2 y + 9x^4 y^4) e^{x^2 y^3}. \end{aligned}$$

²Die Notationen $D_i f$ und $\frac{\partial f}{\partial x_i}$ haben übrigens etwas unterschiedliche Bedeutung: Das Symbol $\frac{\partial}{\partial x_i}$ ist auf einen Ausdruck $f(x_1, \dots, x_n)$ anzuwenden, in dem eine Variable mit dem Namen x_i vorkommt (kommt sie nicht vor, ist $\frac{\partial f}{\partial x_i} = 0$). Dagegen bezeichnet $D_i f$ die Ableitung der Abbildung f nach der i -ten Variablen unabhängig von deren Namen.

Da es sich bei partiellen Ableitungen um Differentiation nach *einer* Variablen handelt, können wir (jedenfalls für $p = 1$) alle Rechenregeln der eindimensionalen Analysis (Produktregel, Kettenregel usw.) anwenden. Auch der Mittelwertsatz der Differentialrechnung (Spezialfall von Satz 5.1 für $v(t) = t$) gilt:

15. Lemma. *Ist $f : U^n \rightarrow \mathbb{R}$ stetig partiell differenzierbar, so gibt es für alle $x \in U^n$ und alle $t \in \mathbb{R}$ mit $|t| < \epsilon$ (für $\epsilon > 0$ genügend klein) ein τ zwischen 0 und t , so dass*

$$f(x + te_i) - f(x) = t \cdot D_i f(x + \tau e_i).$$

Beweis. Wir wenden den Mittelwertsatz an auf die Funktion $u : (-\epsilon, \epsilon) \rightarrow \mathbb{R}$, $u(t) = f(x + te_i)$. Dann gibt es τ zwischen 0 und t

$$u(t) - u(0) = t \cdot u'(\tau) = t \cdot D_i f(x + \tau e_i). \quad \square$$

Etwas Neues bei partiellen Ableitungen gegenüber der eindimensionalen Analysis ergibt sich daraus, dass $D_i f$ nicht nur von der i -ten, sondern auch von allen übrigen Variablen abhängig ist; dies macht die gemischten zweiten partiellen Ableitungen $D_i D_j f$ möglich. Das einzige interessante neue Resultat über partielle Ableitungen sagt, dass die Reihenfolge der partiellen Ableitungen vertauschbar ist, was wir ja im obigen Beispiel bereits beobachtet haben:

16. Satz 6. *Ist $f : U^n \rightarrow \mathbb{R}^p$ zweimal stetig partiell differenzierbar, d.h. alle zweiten partiellen Ableitungen $D_i D_j f : U^n \rightarrow \mathbb{R}^p$ existieren und sind stetig, so gilt für alle $i, j = 1, \dots, n$:*

$$D_i D_j f = D_j D_i f.$$

Beweis. Da wir jede Komponente von $f = (f_1, \dots, f_p)$ einzeln betrachten können, genügt es, den Fall $p = 1$ zu betrachten. Da wir nur die Vertauschbarkeit von jeweils zwei partiellen Ableitungen zu untersuchen haben, können wir uns außerdem auf den Fall $n = 2$ beschränken. Die Variablen nennen wir nun wieder x (1.Variable) und y (2.Variable). Für einen festen Punkt $(x, y) \in U^2$ und für genügend kleine $|h|, |k|$ betrachten wir die folgende Differenz von Differenzen von Werten von f in den Ecken des Rechtecks $(x, y), (x + h, y), (x, y + k), (x + h, y + k)$:

$$D := \begin{array}{rcc} & f(x + h, y + k) & - & f(x, y + k) \\ - & f(x + h, y) & + & f(x, y) \end{array}$$

Die Kernidee des Beweises besteht darin, dass man D wahlweise als Differenz der beiden Spalten oder als Differenz der beiden Zeilen lesen

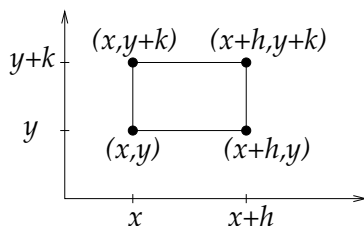


ABBILDUNG 13

kann. Für die erste Auffassung setzen wir $F(x) := f(x, y+k) - f(x, y)$. Nach Mittelwertsatz ist

$$D = F(x+h) - F(x) = h \cdot F'(\xi)$$

für ein ξ zwischen x und $x+h$. Nun ist

$$F'(x) = D_1 f(x, y+k) - D_1 f(x, y),$$

also gilt wiederum nach Mittelwertsatz oder besser nach dem vorstehenden Lemma

$$(1) \quad D = h \cdot (D_1 f(\xi, y+k) - D_1 f(\xi, y)) = hk \cdot D_2 D_1 f(\xi, \eta)$$

für ein η zwischen y und $y+k$. Für die zweite Auffassung von D (als Differenz der Zeilen) setzen wir $G(y) := f(x+h, y) - f(x, y)$ und erhalten ganz analog

$$(2) \quad D = G(y+k) - G(y) = kh \cdot D_1 D_2 f(\tilde{\xi}, \tilde{\eta})$$

für ein (vielleicht anderes) $\tilde{\xi}$ zwischen x und $x+h$ und $\tilde{\eta}$ zwischen y und $y+k$. Lassen wir (h, k) gegen $(0, 0)$ streben, so gehen (ξ, η) und $(\tilde{\xi}, \tilde{\eta})$ beide gegen (x, y) , und damit strebt $D/(hk)$ nach (1) gegen $D_2 D_1 f(x, y)$ und nach (2) gegen $D_1 D_2 f(x, y)$; diese beiden Werte müssen also übereinstimmen. \square

7. Differenzierbarkeit

Die grundlegende Idee der Differentialrechnung ist die *lineare Approximation*: Eine komplizierte Abbildung wird durch eine viel einfachere, nämlich lineare Abbildung angenähert (“approximiert”). In der Umgangssprache kommt diese Idee durch das Gegensatzpaar “glatt” und “rauh” zum Ausdruck: Eine Oberfläche ist glatt, wenn sie keine “Unebenheiten” hat, also lokal durch eine Ebene approximiert werden kann. Im Englischen steht daher das Wort “smooth” (glatt) häufig für “differenzierbar”.

Wie kommt dieses Prinzip in der uns schon bekannten eindimensionalen Analysis zum Ausdruck? Wir wissen: Ist $I \subset \mathbb{R}$ ein offenes Intervall, so ist eine Abbildung $f : I \rightarrow \mathbb{R}$ (oder auch $f : I \rightarrow \mathbb{R}^p$) differenzierbar in einem Punkt $x \in I$, wenn der Limes der Differenzenquotienten,

$$a := \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h}$$

existiert, oder mit anderen Worten, wenn

$$\frac{f(x+h) - f(x) - ah}{h} = \frac{f(x+h) - f(x)}{h} - a \rightarrow 0$$

für $h \rightarrow 0$. (Das a nennen wir dann $f'(x)$.) Setzen wir $f(x+h) - f(x) - ah =: o(h)$ (also $f(x+h) = f(x) + ah + o(h)$), so läßt sich die Differenzierbarkeit in x einfach durch

$$\frac{o(h)}{|h|} \rightarrow 0$$

ausdrücken. Wir haben mit dieser einfachen Umformung eine etwas andere Sichtweise der Differenzierbarkeit gewonnen: Statt zum Limes für $h \rightarrow 0$ überzugehen, arbeiten wir mit einem beliebigen kleinen, aber doch endlichen Zuwachs h und zerlegen $f(x+h)$ additiv in drei Anteile: den Wert $f(x)$ in dem festen Punkt x , von dem wir ausgehen, den linearen Wertezuwachs ah und einen Restzuwachs $o(h)$, der so klein ist, dass er selbst dann noch gegen Null geht, wenn ich ihn vorher mit der großen Zahl $1/|h|$ multipliziere (Abb. 14). Die komplizierte Funktion $h \mapsto f(x+h) - f(x)$ ist also durch die lineare Funktion $h \mapsto ah$ ersetzbar, wenn wir den kleinen Fehler $o(h)$ in Kauf nehmen.

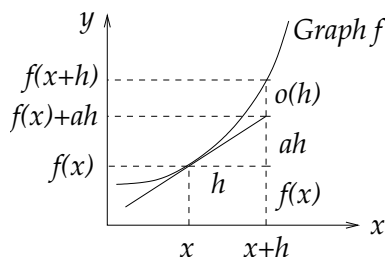


ABBILDUNG 14

Gehen wir nun zu Abbildungen $f : U^n \rightarrow \mathbb{R}^p$ über, so macht der Differenzenquotient keinen Sinn mehr, da der Zuwachs h ja jetzt ein Vektor in \mathbb{R}^n sein muß, wir also nicht mehr durch h dividieren können. In der abgeänderten Form können wir aber die Differenzierbarkeitsdefinition übernehmen, nur dass es jetzt viel mehr lineare Abbildungen gibt: Wir müssen den Ausdruck ah durch Ah ersetzen, wobei A eine

lineare Abbildung von \mathbb{R}^n nach \mathbb{R}^p , also eine $p \times n$ -Matrix $A = ((a_{ij}))$ (mit $i = 1, \dots, p$, $j = 1, \dots, n$) ist, d.h.

$$Ah = \sum_{ij} a_{ij} h_j e_i,$$

wobei h_j die Komponenten von $h \in \mathbb{R}^n$ bezeichnen. Die Menge dieser linearen Abbildungen schreiben wir $\text{Hom}(\mathbb{R}^n, \mathbb{R}^p)$ oder $\mathbb{R}^{p \times n}$. Damit die Matrix-Multiplikationsregel "Zeile mal Spalte" gelten kann, müssen wir übrigens von nun an die Vektoren $x \in \mathbb{R}^n$ als Spalten

$$x = (x_1, \dots, x_n)^T := \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

statt als Zeilen (x_1, \dots, x_n) schreiben, wenn wir genau sein wollen. (Dabei bedeutet " T " die *Transposition*, die aus einer Zeile die entsprechende Spalte macht.)

Wir definieren also: Eine Abbildung $f : U^n \rightarrow \mathbb{R}^p$ heißt (*total*) *differenzierbar im Punkt* $x \in U^n$ (kurz: *db. in* x), wenn es eine lineare Abbildung (Matrix) $A \in \text{Hom}(\mathbb{R}^n, \mathbb{R}^p) = \mathbb{R}^{p \times n}$ und eine Abbildung $o : \mathbb{R}^n \rightarrow \mathbb{R}^p$ mit $o(h)/|h| \rightarrow 0$ für $h \rightarrow 0$ gibt, so dass für alle $h \in \mathbb{R}^n$ mit $x + h \in U^n$ gilt:

$$f(x + h) = f(x) + Ah + o(h). \quad (D)$$

Dabei bedeutet " $o(h)/|h| \rightarrow 0$ für $h \rightarrow 0$ " genau wie in der eindimensionalen Analysis, dass $o(h_k)/|h_k|$ eine Nullfolge ist für jede Nullfolge (h_k) . Dies ist äquivalent zu der Aussage

$$\forall \epsilon > 0 \exists \delta > 0 \forall h \in \mathbb{R}^n [|h| < \delta \Rightarrow \frac{|o(h)|}{|h|} < \epsilon] \quad (o)$$

(bitte überlegen; vgl. die Äquivalenz der beiden Stetigkeitsdefinitionen (1) und (2) in §4). Wir werden den Namen " o " i.f. immer für eine Funktion mit dieser Eigenschaft reservieren; dies ist eine allgemein gebräuchliche Konvention.

Die Matrix A nennen wir die *Ableitung* oder das (*totale*) *Differential* oder die *Jacobimatrix* von f im Punkte x und bezeichnen sie mit $A = Df(x) = Df_x$. In der vorstehenden Diskussion haben wir gesehen, dass speziell für $n = p = 1$, also für $f : I \rightarrow \mathbb{R}$ gilt: $Df(x)h = f'(x)h$ für alle $h \in \mathbb{R}$, also $Df(x) = f'(x)$. Dasselbe gilt natürlich auch für eine Kurve $f : I \rightarrow \mathbb{R}^p$, die ja einfach nur eine Spalte aus p solchen Funktionen f_1, \dots, f_p ist; in diesem Fall ist $f'(x)$ ein Vektor im \mathbb{R}^p , also eine Spalte, und $Df(x)$ ist eine $p \times 1$ -Matrix, also auch eine Spalte. Wir müssen uns

noch den Zusammenhang mit den im letzten Abschnitt eingeführten partiellen Ableitungen klar machen:

17. Satz 7.1 *Ist $f : U^n \rightarrow \mathbb{R}^p$ ist differenzierbar in $x \in U^n$, so ist f dort auch partiell differenzierbar mit*

$$D_i f(x) = Df(x)e_i$$

für $i = 1, \dots, n$.

Beweis. Wir setzen $h = te_i$ mit $t \in \mathbb{R} \setminus \{0\}$. Setzen wir dies in die Differenzierbarkeitsdefinition (D) ein und dividieren durch t , so erhalten wir (mit $A := Df(x)$)

$$\frac{f(x + te_i) - f(x)}{t} = Ae_i + \frac{o(te_i)}{t}.$$

Wir lassen nun $t \rightarrow 0$ gehen. Dann geht der zweite Term rechts gegen Null, da $|o(te_i)/t| = |o(h)|/|h| \rightarrow 0$ (für $h = te_i$), also konvergiert der Differenzenquotient links gegen Ae_i . Damit ist f nach der i -ten Variablen partiell differenzierbar mit $D_i f(x) = Ae_i$. \square

Da Ae_i gerade die i -te Spalte der Matrix A ist (nach dem berühmten Merksatz: *In der i -ten Spalte steht das Bild des i -ten Basisvektors*), können wir aus dem vorstehenden Satz die folgende Merkregel entnehmen: *Die i -te partielle Ableitung ist die i -te Spalte der Jacobimatrix.*

Die Umkehrung des vorstehenden Satzes ist falsch: Aus der partiellen Differenzierbarkeit folgt *nicht* die totale Differenzierbarkeit, wie folgendes Beispiel zeigt:

Beispiel: Es sei $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ definiert durch $f(0,0) = 0$ und

$$f(x,y) = \frac{xy}{x^2 + y^2}$$

für $(x,y) \neq (0,0)$. Da diese Funktion außerhalb der Ursprungs $(0,0)$ rational ist, sind die partielle und totale Differenzierbarkeit nur in $(0,0)$ problematisch (vgl. nächster Abschnitt, §8). Dort ist f partiell differenzierbar mit verschwindenden partiellen Ableitungen, denn für alle $t \neq 0$ ist

$$\frac{f(t,0) - f(0,0)}{t} = \frac{f(0,t) - f(0,0)}{t} = 0.$$

Andererseits ist f in $(0,0)$ aber nicht stetig: Setzen wir z.B. $(x_k, y_k) = (1/k, 1/k)$, so gilt $(x_k, y_k) \rightarrow (0,0)$, aber $f(x_k, y_k) = 1/2$ für alle k , also $f(x_k, y_k) \not\rightarrow f(0,0)$. Dann kann f in $(0,0)$ erst recht nicht (total) differenzierbar sein, wie der folgende Satz zeigt:

18. Satz 7.2 Ist $f : U^n \rightarrow \mathbb{R}^p$ in $x \in U^n$ differenzierbar, so ist f in x stetig.

Beweis. Ist (x_k) eine Folge in U^n mit $x_k \rightarrow x$, so setzen wir $h_k = x_k - x$ und erhalten aus (D) mit $A = Df(x)$:

$$f(x_k) = f(x + h_k) = f(x) + Ah_k + o(h_k) \rightarrow f(x).$$

Hierbei haben wir benutzt, dass die lineare Abbildung $h \mapsto Ah$ stetig ist (als rationale Abbildung), und dass auch $o(h) \rightarrow 0$ für $h \rightarrow 0$; es gilt ja sogar viel stärker $o(h)/|h| \rightarrow 0$. \square

Aus der partiellen Differenzierbarkeit können wir also nicht auf die (totale) Differenzierbarkeit schließen. Wenn wir allerdings zusätzlich die Stetigkeit der partiellen Ableitungen voraussetzen, folgt auch die Differenzierbarkeit. Dies sagt der folgende Satz, der einzige nicht-triviale Satz dieses Abschnittes. (Eine Aussage ist *trivial*, wenn sie unmittelbar aus den Definitionen der vorkommenden Begriffe folgt; sie ist *nicht-trivial*, wenn eine zusätzliche Konstruktion zum Beweis erforderlich ist.)

19. Satz 7.3 Ist $f : U^n \rightarrow \mathbb{R}^p$ stetig partiell differenzierbar (d.h. die partiellen Ableitungen existieren und sind stetig), so ist f auch überall in U^n differenzierbar.

Beweis. Die Abbildung $f = (f_1, \dots, f_p)^T$ ist differenzierbar, wenn die Komponentenfunktionen f_1, \dots, f_p differenzierbar sind (bitte überlegen). Deshalb genügt es, den Fall $p = 1$ zu betrachten. Es sei $x \in U^n$ fest. Wir haben bereits einen Kandidaten für die Ableitung $Df(x)$, nämlich die Matrix A mit den Spalten $Ae_i = D_i f(x)$. Wir müssen also nur zeigen, dass der Rest $r(h) := f(x + h) - f(x) - Ah$ die Eigenschaft $r(h)/|h| \rightarrow 0$ für $h \rightarrow 0$ hat. (Wir schreiben noch nicht $o(h)$ für $r(h)$, da wir die Eigenschaft (*o*) ja noch nicht bewiesen haben.) Es sei also $h \in \mathbb{R}^n$ "genügend klein": Da U^n offen ist, liegt noch ein Ball $B_\epsilon(x)$ ganz in U^n ; wir wählen $|h| < \epsilon$. Die Idee des Beweises ist: Statt in einem Schritt von $f(x)$ zu $f(x + h)$ überzugehen, ändern wir eine Koordinate nach der anderen ab; in jedem Zwischenschritt ändern wir also nur eine Koordinate, so dass wir den Mittelwertsatz anwenden können. Wir streuen also zwischen x und $x + h$ die Punkte $x_{(j)} = x + \sum_{i=1}^j h_i e_i$ für $j = 0, \dots, n$ ein; speziell ist $x_{(0)} = x$ und $x_{(n)} = x + h$, und alle $x_{(j)}$ liegen in $B_\epsilon(x) \subset U^n$. Dann ist

$$f(x + h) - f(x) = \sum_{j=1}^n (f(x_{(j)}) - f(x_{(j-1)}))$$

$$\begin{aligned}
&= \sum_{j=1}^n (f(x_{(j-1)} + h_j e_j) - f(x_{(j-1)})) \\
&= \sum_{j=1}^n h_j D_j f(\xi_{(j)}),
\end{aligned}$$

wobei $\xi_{(j)} = x_{(j-1)} + \tau_j e_j$ für ein τ_j zwischen 0 und h_j ist. Dies gilt nach dem Mittelwertsatz, vgl. das Lemma in §6. Somit erhalten wir

$$\frac{r(h)}{|h|} = \sum_{j=1}^n \frac{h_j}{|h|} (D_j f(\xi_{(j)}) - D_j f(x)) \rightarrow 0,$$

weil $D_j f$ in x stetig ist und $\xi_{(j)} \rightarrow x$ für $h \rightarrow 0$, und weil $|h_j|/|h| \leq 1$, also $h_j/|h|$ beschränkt ist. \square

Wir können die Ergebnisse dieses Abschnittes folgendermaßen zusammenfassen:

$$\begin{array}{ccccc}
\textit{stetig part. db.} & \Rightarrow & \textit{db.} & \Rightarrow & \textit{part. db.} \\
& & \Downarrow & & \\
& & \textit{stetig} & &
\end{array}$$

Keiner dieser Folgerungspfeile ist umkehrbar.

8. Beispiele und Rechenregeln

Wir wollen uns zunächst die Definition der Differenzierbarkeit an Hand einiger Beispiele klarmachen.

Beispiel 1. $f : U^n \rightarrow \mathbb{R}^p$, $f \equiv \textit{const} = c \in \mathbb{R}^p$. Dann ist $f(x+h) - f(x) = 0$ für alle x, h , also ist f differenzierbar mit $Df(x) = 0$ für alle $x \in U^n$. Der Rest $o(h)$ ist in diesem Fall Null.

Beispiel 2. $f : U^n \rightarrow \mathbb{R}^p$, $f(x) = Bx$ für eine Matrix $B \in \mathbb{R}^{p \times n}$, d.h. f ist selbst linear. Dann ist f differenzierbar mit $Df(x) = B$ für alle $x \in U^n$, denn $f(x+h) - f(x) = B(x+h) - Bx = Bh$. Wieder ist das Restglied $o(h) = 0$. Das ist natürlich kein Wunder: Die lineare Approximation einer linearen Abbildung an einer beliebigen Stelle ist die lineare Abbildung selbst.

Beispiel 3. $B = ((b_{ij})) \in \mathbb{R}^{n \times n}$ sei eine *symmetrische* Matrix, d.h. es gilt $B^T = B$ oder $b_{ij} = b_{ji}$ für alle i, j oder $\langle Bx, y \rangle = \langle x, By \rangle$ für alle x, y (da $\langle x, y \rangle = x^T y$). Wir betrachten die Funktion $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $f(x) = \langle Bx, x \rangle = x^T Bx$. Diese Funktion ist differenzierbar mit $Df(x) = 2(Bx)^T$, denn

$$f(x+h) = \langle B(x+h), x+h \rangle = f(x) + 2\langle Bx, h \rangle + \langle Bh, h \rangle.$$

Der mittlere Term rechts, $2\langle Bx, h \rangle = 2(Bx)^T h$ ist linear in h ; wir setzen daher $Ah = 2(Bx)^T h$ oder $A = 2(Bx)^T$. Der letzte Term ist vom Typ $o(h)$, denn nach Cauchy-Schwarz (Satz 1.) gilt

$$\frac{|\langle Bh, h \rangle|}{|h|} = \left| \left\langle Bh, \frac{h}{|h|} \right\rangle \right| \leq |Bh| \rightarrow 0$$

für $h \rightarrow 0$.

Das Schema eines Beweises für Differenzierbarkeit ist in allen Beispielen und auch in den folgenden Sätzen stets dasselbe: Man zerlegt $f(x+h) - f(x)$ additiv in einen linearen Term Ah und einen Restterm $r(h)$, von dem man dann $|r(h)|/|h| \rightarrow 0$ für $h \rightarrow 0$ zeigt. Wir wollen nun einige Rechenregeln kennenlernen, die uns ganze Klassen von weiteren Beispielen verschaffen.

20. Satz 8.1 (*Summen- und Produktregel*) *Es seien $f, g : U^n \rightarrow \mathbb{R}^p$ differenzierbar in $x \in U^n$ und $\alpha \in \mathbb{R}$. Dann gilt*

- (a) $f + g$ und αf sind differenzierbar in x mit

$$D(f + g)_x = Df_x + Dg_x, \quad D(\alpha f)_x = \alpha Df_x.$$

- (b) Ist $p = 1$ (oder $p = 2, \mathbb{R}^2 = \mathbb{C}$) so ist auch fg differenzierbar in x mit

$$D(fg)_x = Df_x g(x) + f(x) Dg_x.$$

Beweis. (a) ist sehr einfach; wir beweisen daher nur (b). Wir setzen $A = Df_x$, $B = Dg_x$; die Restterme seien $o_f(h)$, $o_g(h)$. Dann ist

$$\begin{aligned} & (fg)(x+h) - (fg)(x) \\ &= f(x+h)g(x+h) - f(x)g(x) \\ &= f(x+h)g(x+h) - f(x)g(x+h) + f(x)g(x+h) - f(x)g(x) \\ &= (Ah + o_f(h))g(x+h) + f(x)(Bh + o_g(h)) \\ &= (Ah)g(x) + f(x)Bh + r(h) \end{aligned}$$

mit

$$r(h) = (Ah)(g(x+h) - g(x)) + o_f(h)g(x+h) + f(x)o_g(h).$$

Wir müssen $r(h)/|h| \rightarrow 0$ für $h \rightarrow 0$ zeigen. Für den zweiten und dritten Summanden von $r(h)$ ist das klar, und für den ersten Summanden stimmt es auch, da $|Ah|/|h| \leq \|A\|$ beschränkt ist (Matrix-Norm; cf. §9) und $|g(x+h) - g(x)| \rightarrow 0$ wegen der Stetigkeit von g in x . \square

21. Satz 8.2 (*Kettenregel*) *Es seien $f : U^n \rightarrow \mathbb{R}^p$ und $g : V^p \rightarrow \mathbb{R}^q$ Abbildungen mit $f(U^n) \subset V^p$. Es sei f differenzierbar in $x \in U^n$ und g*

differenzierbar in $f(x) \in V^p$. Dann ist $g \circ f : U^n \rightarrow \mathbb{R}^q$ differenzierbar in x mit

$$D(g \circ f)_x = Dg_{f(x)} \cdot Df_x.$$

Für die partiellen Ableitungen gilt somit:

$$D_i(g \circ f)(x) = \sum_j D_j g(f(x)) \cdot D_i f_j(x).$$

Beweis. Wir setzen $A = Df_x$ und $B = Dg_{f(x)}$; die Restterme seien wieder mit o_f und o_g bezeichnet. Es sei $k = k(h) := f(x+h) - f(x) = Ah + o_f(h)$. Dann ist

$$\begin{aligned} (g \circ f)(x+h) - (g \circ f)(x) &= g(f(x+h)) - g(f(x)) \\ &= g(f(x) + k) - g(f(x)) \\ &= Bk + o_g(k) \\ &= BAh + r(h) \end{aligned}$$

mit $r(h) = Bo_f(h) + o_g(k(h))$. Wieder müssen wir $r(h)/|h| \rightarrow 0$ zeigen. Für den ersten Term von $r(h)$ ist das klar. Für den zweiten erhalten wir:

$$\frac{o_g(k)}{|h|} = \frac{o_g(k)}{|k|} \frac{|k|}{|h|} \leq \frac{o_g(k)}{|k|} \left(\frac{|Ah|}{|h|} + \frac{o_f(h)}{|h|} \right),$$

und dies geht gegen Null, da der erste Faktor rechts gegen 0 geht und vom zweiten Faktor der vordere Term beschränkt ist ($\leq \|A\|$, cf. §9) und der hintere sogar gegen Null strebt. (Dabei müssen wir die Abbildung $k \mapsto o_g(k)/|k|$ für $k = 0$ durch 0 stetig fortsetzen.) \square

Mit diesen Sätzen können wir sehr viele Funktionen und Abbildungen als differenzierbar nachweisen. Zunächst einmal sind die Koordinatenfunktionen $x_i : \mathbb{R}^n \rightarrow \mathbb{R}$, $x \mapsto x_i$ linear, also differenzierbar. Damit sind alle rationalen Funktionen (die durch die vier Grundrechenarten aus den x_i entstehen) differenzierbar; man beachte, dass nach Satz 7.2 mit f auch $1/f$ außerhalb der Nullstellen von f differenzierbar ist mit $D(1/f)_x = -Df_x/f(x)^2$; allgemein gilt $D(g \circ f)_x = g'(f(x)) \cdot Df_x$ falls $f : U^n \rightarrow \mathbb{R}$ und $g : \mathbb{R} \rightarrow \mathbb{R}$. Durch Verkettung mit anderen differenzierbaren Funktionen entstehen viele weitere Beispiele.

9. Exkurs: Die Matrixnorm

Für eine $p \times n$ -Matrix $A = ((a_{ij}))$ definieren wir die *Matrixnorm*

$$\|A\| = \sup_{|x|=1} |Ax|.$$

Wir müssen zunächst sehen, dass diese Zahl endlich ist. Das folgt z.B. durch Vergleich mit der Maximumsnorm oder der euklidischen Norm von A :

22. Satz 9. Für $A = ((a_{ij})) \in \mathbb{R}^{p \times n}$ sei $|A|_0 := \max_{i,j} |a_{ij}|$ und $|A| := \sqrt{\sum_{ij} |a_{ij}|^2}$. Dann gilt

$$|A|_0 \leq \|A\| \leq |A| \leq \sqrt{pn} |A|_0.$$

Beweis. Es sei $|A|_0 = |a_{i_0 j_0}|$. Die erste Ungleichung folgt wegen

$$|a_{i_0 j_0}| = |(Ae_{j_0})_{i_0}| \leq |Ae_{j_0}| \leq \|A\|.$$

Für die zweite Ungleichung müssen wir $|Ax|$ für jedes $x \in \mathbb{R}^n$ mit $|x| = 1$ abschätzen. Da $(Ax)_i = \sum_j a_{ij} x_j = \langle a^i, x \rangle$ ist, wobei $a^i = \sum_j a_{ij} e_j$ den i -ten Zeilenvektor der Matrix A bezeichnet, gilt mit Cauchy-Schwarz und $|x| = 1$:

$$|Ax|^2 = \sum_i ((Ax)_i)^2 = \sum_i \langle a^i, x \rangle^2 \leq \sum_i |a^i|^2 = \sum_{i,j} |a_{ij}|^2 = |A|^2,$$

und mit der früheren Ungleichung $|A| \leq \sqrt{pn} \cdot |A|_0$ (die Dimension ist hier nicht n , sondern pn) folgt die Behauptung. \square

Insbesondere ist also $\|A\|$ eine endliche Zahl ≥ 0 , und für alle $h \in \mathbb{R}^n$ gilt

$$|Ah| \leq \|A\| |h|.$$

Für $h \neq 0$ folgt dies, weil $x = \frac{h}{|h|}$ ein Vektor der Länge Eins ist und damit gilt:

$$\frac{|Ah|}{|h|} = |A(\frac{h}{|h|})| = |Ax| \leq \|A\|.$$

Für $h = 0$ ist die Ungleichung trivial.

Bemerkung. Damit können wir noch einmal die Stetigkeit einer linearen Abbildung $A : \mathbb{R}^n \rightarrow \mathbb{R}^p$ beweisen, denn

$$|A(x+h) - Ax| = |Ah| \leq \|A\| |h| \rightarrow 0$$

für $h \rightarrow 0$. Dies impliziert übrigens, dass das Supremum $\|A\|$ angenommen wird, also ein Maximum ist. Dies folgt aus Satz 4.2 (Existenz des Maximum auf kompakten Mengen) mit der Stetigkeit von A und der Kompaktheit (Abgeschlossenheit und Beschränktheit) der Menge $S = \{x \in \mathbb{R}^n; |x| = 1\}$, der "Einheitssphäre".

Bemerkung. Die Matrixnorm ist tatsächlich eine Norm auf dem Vektorraum $\mathbb{R}^{p \times n}$ im Sinne der Definition in Kap. 1 (letzte Bemerkung). Ist nämlich $\|A\| = 0$, so folgt $Ax = 0$ für alle $x \in \mathbb{R}^n$, also $A = 0$, dies ist (N1). (N2) ist klar, und (N3) folgt aus der Ungleichung $|(A+B)x| = |Ax + Bx| \leq |Ax| + |Bx|$ für alle $x \in \mathbb{R}^n$.

10. Richtungsableitungen und Gradient

Wie berechnet man effektiv die totale Ableitung einer differenzierbaren Abbildung $f : U^n \rightarrow \mathbb{R}^p$ in einem Punkt $x \in U^n$? Natürlich kann man die n partiellen Ableitungen berechnen; wenn man aber nur den Wert $Df(x)v$ für einen Vektor $v \in \mathbb{R}^n$ berechnen will, empfiehlt sich ein anderes Verfahren: Man wähle eine Kurve $c : (-\epsilon, \epsilon) \rightarrow U^n$ mit $c(0) = x$ und $c'(0) = v$, z.B. das Geradenstück $c(t) = x + tv$. Nach der Kettenregel (rückwärts angewendet) ist dann

$$Df(x)v = (f \circ c)'(0) = \left. \frac{d}{dt} f(x + tv) \right|_{t=0},$$

denn $(f \circ c)'(0) = Df(c(0))c'(0)$. Statt n Differentiationen braucht man also nur eine auszuführen. Der Wert $Df(x)v$ wird manchmal auch *Richtungsableitung* von f im Punkt x in Richtung v genannt.

Beispiel 1. Wir nehmen noch einmal das Beispiel 3 aus §8 auf: Es war $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $f(x) = \langle Bx, x \rangle$ für eine symmetrische $n \times n$ -Matrix B . Dann ist $f(x + tv) = \langle Bx, x \rangle + 2t\langle Bx, v \rangle + t^2\langle Bv, v \rangle$, also ist

$$\begin{aligned} Df(x)v &= \left. \frac{d}{dt} \langle B(x + tv), x + tv \rangle \right|_{t=0} \\ &= \left. \frac{d}{dt} (\langle Bx, x \rangle + 2t\langle Bx, v \rangle + t^2\langle Bv, v \rangle) \right|_{t=0} = 2\langle Bx, v \rangle. \end{aligned}$$

Beispiel 2. $f : \mathbb{R}^{n \times n} \rightarrow \mathbb{R}^{n \times n}$, $f(x) = x \cdot x = x^2$ (Matrizenprodukt). Die Abbildung ist differenzierbar, denn für $x = ((x_{ij}))$ sind die Komponenten von $f(x)$ rational, nämlich $f(x)_{ik} = \sum_j x_{ij}x_{jk}$. Für alle $v \in \mathbb{R}^{n \times n}$ ist nun

$$Df(x)v = \left. \frac{d}{dt} (x+tv) \cdot (x+tv) \right|_{t=0} = \left. \frac{d}{dt} (x^2 + t(xv + vx) + t^2v^2) \right|_{t=0} = xv + vx.$$

Für den Rest dieses Abschnittes sei jetzt $p = 1$, also $f : U^n \rightarrow \mathbb{R}$ eine differenzierbare Funktion. In diesem Fall ist $Df(x)$ eine lineare Abbildung von \mathbb{R}^n nach \mathbb{R} , also eine $1 \times n$ -Matrix, nämlich die Zeile

$$Df(x) = (D_1f(x), \dots, D_nf(x)).$$

Die zugehörige Spalte $Df(x)^T \in \mathbb{R}^n$ heißt der *Gradient* von f im Punkte x , geschrieben $\nabla f(x)$. Die Komponenten des Vektors $\nabla f(x)$ sind also die partiellen Ableitungen in x . (Im Beispiel 1 etwa ist $\nabla f(x) = 2(Bx)^{TT} = 2Bx$.) Die Richtungsableitungen sind dann einfach die Skalarprodukte mit dem Gradienten: Für jedes $v \in \mathbb{R}^n$ ist

$$Df(x)v = \nabla f(x)^T v = \langle \nabla f(x), v \rangle.$$

Der Gradient zeigt an jeder Stelle, wo er nicht verschwindet, in die *Richtung größten Anstieges* von f : Vergleichen wir die Richtungsableitungen $Df(x)v$ für alle Einheitsvektoren v , so ist $Df(x)v$ am größten, wenn v und $\nabla f(x)$ gleichgerichtet sind, also für $v_0 = \nabla f(x)/|\nabla f(x)|$, denn nach Cauchy-Schwarz ist (mit $|v| = 1$)

$$Df(x)v = \langle \nabla f(x), v \rangle \leq |\nabla f(x)| = Df(x)v_0,$$

und die Ungleichung ist strikt für $v \neq v_0$.

Die Funktion, mit der wir im Alltag wohl am meisten zu tun haben, ist die *geographische Höhenfunktion*, die jedem Punkt auf der Erde ihre Höhe über dem Meeresspiegel zuordnet. Die Erde dürfen wir uns in unserer Nähe als im Wesentlichen eben, also als eine offene Teilmenge $U^2 \subset \mathbb{R}^2$ vorstellen (jede Landkarte ist ja auch ein ebenes Stück Papier), und die Höhe ist jeweils eine Zahl; die geographische Höhenfunktion ist also eine Funktion $h : U^2 \rightarrow \mathbb{R}$. Auf dem Graphen dieser Funktion trampeln wir ständig herum; es ist nämlich das Gelände. Der Gradient der Höhenfunktion zeigt also in die Richtung, in der das Gelände am steilsten ansteigt. Auf der Landkarte ist das die Richtung senkrecht zu den Höhenlinien; das werden wir in Kap. 15 noch genauer verstehen. Die geographische Höhenfunktion wird uns auch weiterhin zur Veranschaulichung der mehrdimensionalen Analysis gute Dienste leisten.

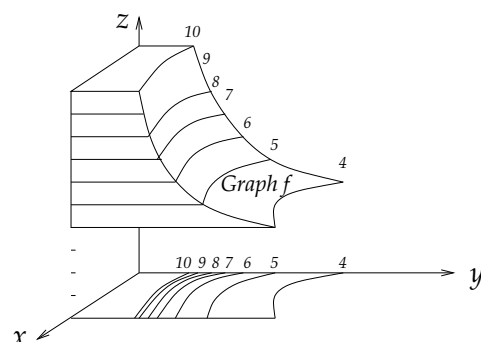


ABBILDUNG 15

Ist eine Funktion $f : U^n \rightarrow \mathbb{R}$ stetig differenzierbar, so ist der Gradient ein stetiges Vektorfeld auf U^n . Für ein solches Vektorfeld sind die Kurvenintegrale (vgl. §6) besonders einfach: Sie hängen gar nicht von der Kurve ab, sondern nur von deren Anfangs- und Endpunkt:

23. Satz 10.1 *Für jede Kurve $x : [a, b] \rightarrow U^n$ gilt:*

$$\int_x \langle \nabla f(x), dx \rangle = f(x(b)) - f(x(a)).$$

Beweis.

$$\int_a^b \langle \nabla f(x(t)), x'(t) \rangle dt = \int_a^b (f \circ x)'(t) dt = f(x(b)) - f(x(a)).$$

□

Dieser einfache Satz ist für die Physik von großer Bedeutung. Wenn ein stetiges Vektorfeld $F : U^n \rightarrow \mathbb{R}^n$ als *Kraftfeld* interpretiert wird, so haben wir das Kurvenintegral $\int_x \langle F(x), dx \rangle$ als die *Arbeit* interpretiert, die längs des Weges x gegen die Kraft F zu leisten ist. Ist nun F ein Gradientenfeld, also $F = \nabla f$ für eine stetig differenzierbare Funktion f auf U^n , so ist diese Arbeit unabhängig von der Wahl des Weges durch die Funktion f gegeben. Solche Kräfte vom Typ $F = \nabla f$ heißen in der Physik *konservativ*; die zugehörige Funktion f heißt *Potential*. Beispiele solcher Kräfte sind die Schwerkraft oder die elektro- und magnetostatischen Kräfte. Man kann in einem solchen Kraftfeld keine Arbeit oder Energie gewinnen: Was man auf dem Hinweg (beim Herunterfallen) vielleicht gewinnt, muß man auf dem Rückweg (beim Hochklettern) wieder aufwenden, egal, welchen Rückweg man nimmt. Die Arbeit oder Energie bleibt erhalten; man gewinnt oder verliert nichts (“konservativ” = “erhaltend”). Es gibt also kein Gravitations- oder Magnetkraftwerk (Perpetuum mobile). Gegenteiliges lese man nach bei “Jim Knopf und die Wilde 13”.

Wie sieht man aber einem Vektorfeld $F : U^n \rightarrow \mathbb{R}^n$ an, ob es ein Gradientenvektorfeld $F = \nabla f$ für irgendeine (unbekannte) Funktion f ist oder nicht? Oder hat vielleicht jedes Vektorfeld diese Eigenschaft? Zumindest, wenn F selbst stetig differenzierbar ist, sieht man sofort, dass dies nicht der Fall ist; es gibt nämlich die folgende notwendige Bedingung:

24. Satz 10.2 *Ist $F = (F_1, \dots, F_n)^T : U^n \rightarrow \mathbb{R}^n$ ein stetig differenzierbares Vektorfeld, das Gradient einer Funktion $f : U^n \rightarrow \mathbb{R}$ ist, also $F = \nabla f$, so gilt für $i, j = 1, \dots, n$:*

$$D_i F_j = D_j F_i.$$

Beweis. Dies folgt direkt aus Satz 6 (Vertauschbarkeit der Ableitungen), denn

$$D_i F_j = D_i D_j f = D_j D_i f = D_j F_i.$$

□

Bemerkung. Für ein beliebiges stetig differenzierbares Vektorfeld $F : U^n \rightarrow \mathbb{R}^n$ heißt die Matrix $A = ((a_{ij}))$ mit den Komponenten $a_{ij} = D_i F_j - D_j F_i$ die *Rotation* von F , geschrieben $\text{rot } F$. Satz 10.2 sagt also $\text{rot } \nabla f = 0$. Für $n = 3$ hat $A = \text{rot } F$ wegen $a_{ij} = -a_{ji}$ nur drei unabhängige Komponenten; diese werden zu dem *Vektor* $\text{rot } F := (a_{23}, a_{31}, a_{12})^T \in \mathbb{R}^3$ zusammengefaßt. Beide Sätze 10.1 und 10.2 besitzen unter gewissen Voraussetzungen an U^n eine Umkehrung, vgl. [Wüst], Kap. 18.4.

11. Der Schrankensatz

Wie schnell ändern sich die Werte einer stetig differenzierbaren Abbildung f , wenn wir das Argument variieren? Wie in der eindimensionalen Analysis (Mittelwertsatz) wird die Größe dieser Änderung durch eine Schranke für die Ableitung von f gegeben. Da die Ableitung $Df(x)$ für jedes x jetzt eine Matrix ist, benötigen wir die Matrixnorm $\|Df_x\|$ (vgl. §9). Wir führen noch eine Bezeichnung ein: Für zwei Punkte $x, y \in \mathbb{R}^n$ sei $[x, y]$ die *Strecke* zwischen x und y :

$$[x, y] = \{ty + (1-t)x; t \in [0, 1]\}.$$

25. Satz 11. Schrankensatz: *Es sei $f : U^n \rightarrow \mathbb{R}^p$ eine stetig differenzierbare Abbildung, und es gebe eine Konstante L mit $\|Df(x)\| \leq L$ für alle $x \in U^n$. Dann gilt für alle $x, y \in U^n$ mit $[x, y] \subset U^n$:*

$$|f(y) - f(x)| \leq L|y - x|. \quad (*)$$

Beweis. Es sei $c : [0, 1] \rightarrow U^n$, $c(t) = ty + (1-t)x$. Nach dem Hauptsatz der Differential- und Integralrechnung und der Kettenregel ist

$$f(y) - f(x) = \int_0^1 (f \circ c)'(t) dt = \int_0^1 Df_{c(t)}(y - x) dt,$$

also (vgl. Satz 5.2 und §9)

$$\begin{aligned} |f(y) - f(x)| &\leq \int_0^1 |Df_{c(t)}(y - x)| dt \\ &\leq \int_0^1 \|Df_{c(t)}\| |y - x| dt \\ &\leq L|y - x|. \end{aligned}$$

□

Die Menge $U^n \subset \mathbb{R}^n$ heißt *konvex*, wenn für je zwei Punkte $x, y \in U^n$ die Strecke $[x, y]$ ganz in U^n enthalten ist. In diesem Fall gilt (*) für alle $x, y \in U^n$. Die Eigenschaft nennt man *Lipschitz-Stetigkeit mit Lipschitz-Konstante L*; sie ist eine Verschärfung der gleichmäßigen Stetigkeit.

Auf ähnliche Weise können wir auch zeigen, dass eine stetig differenzierbare Abbildung $f : U^n \rightarrow \mathbb{R}^p$ mit $Df_x = 0$ für alle $x \in U^n$ konstant ist, falls je zwei Punkte in U^n durch eine stetig differenzierbare Kurve in U^n verbunden werden können (was der Fall ist, wenn U^n wegweise zusammenhängend ist). Sind nämlich $x, y \in U^n$ und ist $c : [a, b] \rightarrow U^n$ differenzierbar mit $c(a) = x$, $c(b) = y$, so ist $f(y) - f(x)$ das Integral über $(f \circ c)'(t) = Df_{c(t)}c'(t) = 0$, also ist $f(x) = f(y)$. Die Umkehrung, dass eine konstante Abbildung Ableitung Null hat, haben wir schon früher gesehen (§8, Beispiel 1).

12. Exkurs: Der Banachsche Fixpunktsatz

Eine der Differentiationsregeln, die wir aus der eindimensionalen Analysis gewohnt sind, fehlt noch, nämlich die Differenzierbarkeit und die Ableitung der Umkehrabbildung. Dies ist unser nächstes Ziel. Um die Umkehrabbildung zu einer Abbildung f überhaupt definieren zu können, müssen wir eine Gleichung lösen, nämlich die Gleichung $y = f(x)$ nach x auflösen. In diesem Abschnitt wollen wir einen sehr allgemeinen Satz kennenlernen, der ein Lösungsverfahren für diese und viele andere Gleichungen an die Hand gibt.

26. Satz 12. (Banachscher Fixpunktsatz) *Es sei $D \subset \mathbb{R}^n$ und $T : D \rightarrow \mathbb{R}^n$ eine Kontraktion, d.h. es gebe eine Konstante $q \in (0, 1)$, so dass für alle $x, x' \in D$ gilt:*

$$|Tx' - Tx| \leq q|x' - x|.$$

Ferner sei $K \subset D$ eine in \mathbb{R}^n abgeschlossene Teilmenge, die unter T invariant ist, also $T(K) \subset K$. Dann besitzt T in K genau einen Fixpunkt, also ein $x^ \in K$ mit $Tx^* = x^*$, und zwar ist x^* Limes jeder Folge (x_k) in K mit beliebigem Anfangspunkt $x_0 \in K$ und $x_{k+1} = Tx_k$ für alle $k \geq 0$.*

Beweis. Wir zeigen zunächst, dass so eine Folge (x_k) konvergiert. Es gilt nämlich für alle k

$$|x_{k+1} - x_k| = |Tx_k - Tx_{k-1}| \leq q|x_k - x_{k-1}|.$$

Wendet man dies noch einmal auf die rechte Seite an, so erhält man

$$|x_{k+1} - x_k| \leq q^2|x_{k-1} - x_{k-2}|$$

und durch Induktion schließlich

$$|x_{k+1} - x_k| \leq q^k a$$

mit $a := |x_1 - x_0| = |Tx_0 - x_0|$. Dies zeigt, dass (x_k) eine Cauchyfolge ist, denn für alle $k \leq m$ ist

$$(*) \quad |x_m - x_k| \leq \sum_{j=k}^{m-1} |x_{j+1} - x_j| \leq \sum_{j=k}^{m-1} q^j a \leq \frac{q^k}{1-q} a,$$

und dies geht gegen 0 für $k \rightarrow \infty$. (Man erinnere sich an die geometrische Reihe: $\sum_{j=k}^{m-1} q^j \leq \sum_{j=k}^{\infty} q^j = q^k \sum_{j=0}^{\infty} q^j = q^k / (1-q)$.) Also konvergiert (x_k) gegen einen Punkt $x^* \in K$ (da K abgeschlossen). Da T Lipschitz-stetig, also erst recht stetig ist, gilt

$$x^* = \lim x_k = \lim x_{k+1} = \lim Tx_k = T(\lim x_k) = Tx^*,$$

und damit ist x^* Fixpunkt von T . Zu zeigen bleibt noch, dass es nur einen Fixpunkt gibt. Sind nämlich $x_1^*, x_2^* \in K$ mit $Tx_1^* = x_1^*$ und $Tx_2^* = x_2^*$, so ist

$$|x_1^* - x_2^*| = |Tx_1^* - Tx_2^*| \leq q|x_1^* - x_2^*|,$$

woraus wegen $q < 1$ sofort $|x_1^* - x_2^*| = 0$, also $x_1^* = x_2^*$ folgt. \square

Der Satz gibt uns also ein Iterationsverfahren zur Lösung der Gleichung $Tx = x$ an die Hand: Wir starten mit einem beliebigen Wert $x_0 \in K$ und berechnen sukzessiv $x_1 = Tx_0$, $x_2 = Tx_1$, $x_3 = Tx_2$ usw. Gewöhnlich ist keiner dieser Werte die gesuchte Lösung x^* , aber wegen (*) haben wir im k -ten Iterationsschritt die Fehlerabschätzung $|x_k - x^*| \leq q^k / (1-q)$, die für große k immer besser wird; wir können den gesuchten Wert x^* durch x_k also immer genauer eingrenzen.

Beispiel. Es sei $T : [\frac{3}{2}, \infty) \rightarrow \mathbb{R}$, $T(x) = 1 + \frac{1}{x}$. Auf D gilt aus Monotoniegründen $|T'(x)| = 1/x^2 \leq 4/9$. Nach dem Schrankensatz ist also T eine Kontraktion mit $q = 4/9$. Wir wählen $K = [\frac{3}{2}, 2]$. Da T monoton fallend ist und die Randwerte $T(\frac{3}{2}) = \frac{5}{3}$ und $T(2) = \frac{3}{2}$ beide in K liegen, wird K in sich abgebildet. Das eben beschriebene Iterationsverfahren mit $x_0 = \frac{3}{2}$ liefert $x_1 = \frac{5}{3}$, $x_2 = \frac{8}{5}$, $x_3 = \frac{13}{8}$ usw., und allgemein $x_k = f_{k+1}/f_k$, wobei $f_{-2} = f_{-1} = 1$, $f_0 = 2$, $f_1 = 3$, ..., $f_{k+1} = f_k + f_{k-1}$ die Reihe der *Fibonacci-Zahlen* ist; bitte durch Induktion nachweisen! Die Fibonacci-Quotienten (x_k) konvergieren also gegen die positive Lösung der Gleichung $x = 1 + \frac{1}{x}$, diese ist bekanntlich das goldene Schnittverhältnis $x^* = (1 + \sqrt{5})/2$. Wir erhalten die Fehlerabschätzung $|x_k - x^*| < \frac{9}{5}(\frac{4}{9})^k$, in Wirklichkeit läßt sich die Fehlerschranke nach jedem Iterationsschritt noch verbessern.

Bemerkung. Der Beweis von Satz 12 zeigt, dass man \mathbb{R}^n durch einen beliebigen Banachraum (vgl. Bemerkung §4) ersetzen kann. Wir wollen eine wichtige Anwendung davon zeigen. Gegeben sei eine stetige Abbildung $F : \mathbb{R}^n \times \mathbb{R} \rightarrow \mathbb{R}^n$, die im zweiten Argument *uniform* Lipschitz-stetig ist, d.h. es gebe eine Konstante L , so dass

$$|F(a, t) - F(b, t)| = L \cdot |a - b|$$

für alle $a, b \in \mathbb{R}^n$ und alle $t \in \mathbb{R}$. Gesucht ist eine Kurve $x : [-t_0, t_0] \rightarrow \mathbb{R}^n$ mit $0 < t_0 < 1/L$.

$$(1) \quad x'(t) = F(x(t), t), \quad x(0) = x_0$$

für einen gegebenen Punkt $x_0 \in \mathbb{R}^n$ (*gewöhnliche Differentialgleichung mit Anfangsbedingung*). Dann gilt: *Es gibt genau eine solche Kurve.* Zunächst formen wir (1) mit dem Hauptsatz der Differential- und Integralrechnung um zu

$$(2) \quad x(t) = x_0 + \int_0^t F(x(s), s) ds$$

für alle $t \in [-t_0, t_0] =: I$. Die rechte Seite von (2) können wir zunächst für eine beliebige stetige Funktion $x : I \rightarrow \mathbb{R}^n$ definieren: Wir setzen

$$Tx(t) = x_0 + \int_0^t F(x(s), s) ds$$

für alle $t \in I$. Damit haben wir eine andere stetige Abbildung $Tx : I \rightarrow \mathbb{R}^n$ definiert. Bezeichnen wir den Vektorraum aller (beschränkten) stetigen Abbildungen $x : I \rightarrow \mathbb{R}^n$ mit V , so haben wir damit eine Abbildung $T : V \rightarrow V$ definiert, und die Lösungen von (2) sind genau die "Fixpunkte" von T , also die stetigen Abbildungen $x \in V$ mit $x = Tx$. Wir erinnern uns, dass V mit der Supremumsnorm $\|x\| = \sup_{t \in I} |x(t)|$ ein Banachraum ist (vgl. Bemerkung zu §4). Wir zeigen nun, dass T eine Kontraktion auf V ist: Für alle $x, \tilde{x} \in V$ und für alle $t \in I$ gilt nämlich

$$\begin{aligned} |Tx(t) - T\tilde{x}(t)| &\leq \int_0^t |F(x(s), s) - F(\tilde{x}(s), s)| ds \\ &\leq L \cdot \int_0^t |x(s) - \tilde{x}(s)| ds \\ &\leq L \cdot t_0 \cdot \|x - \tilde{x}\|, \end{aligned}$$

also $\|Tx - T\tilde{x}\| \leq q \cdot \|x - \tilde{x}\|$ mit $q = Lt_0 < 1$. Damit besitzt T einen eindeutigen Fixpunkt und also (2) und (1) eine eindeutige Lösung. Dies ist der Grundgedanke des Existenz- und Eindeutigkeitsatzes für gewöhnliche Differentialgleichungen (vgl. z.B. [Forster II]). Natürlich benötigt man die uniforme Lipschitzbedingung an F nicht

auf ganz $\mathbb{R}^n \times R$, sondern nur in einer Umgebung von $(x_0, 0)$, die von der Lösungskurve $(t, x(t))$ auch wirklich erreicht wird; es wird dann nur schwieriger, das Definitionsintervall anzugeben.

13. Der Umkehrsatz

Eine stetig differenzierbare Abbildung heißt auch C^1 -Abbildung (“C” steht für “continuous” = stetig); allgemeiner sprechen wir von einer C^k -Abbildung, wenn alle partiellen Ableitungen bis zur Ordnung k existieren und stetig sind. Eine Abbildung $f : U^n \rightarrow V^p$ heißt (C^1 -)Diffeomorphismus, wenn f eine umkehrbar stetig differenzierbare Abbildung ist, d.h. wenn f umkehrbar (invertierbar) ist und sowohl f als auch seine Umkehrabbildung $f^{-1} : V^p \rightarrow U^n$ C^1 -Abbildungen sind. Das Kunstwort “Diffeomorphismus” ist ein etwas merkwürdiges Gebilde. Es ist dem Wort “Homöomorphismus” nachgebildet, was “Gestalterhaltende” (nämlich umkehrbar stetige) Abbildung bedeutet, und mit der Silbe “difeo-” verquickt, die an “differenzierbar” erinnern soll. Nicht gerade ästhetisch, hat sich aber eingebürgert.

Für so einen Diffeomorphismus gilt also

$$f^{-1} \circ f = id_{U^n}, \quad f \circ f^{-1} = id_{V^p},$$

und nach der Kettenregel (Satz 8.2) gilt damit für alle $x \in U^n$ und $y = f(x)$:

$$D(f^{-1})_y \cdot Df_x = I_n, \quad Df_x \cdot D(f^{-1})_y = I_p$$

wobei $I = I_n$ die Einheitsmatrix in \mathbb{R}^n bezeichnet. Also ist Df_y^{-1} die inverse Matrix zu Df_x , d.h.

$$D(f^{-1})_y = (Df_x)^{-1}. \quad (*)$$

Insbesondere ist Df_x also eine invertierbare Matrix, und natürlich muß dann $p = n$ gelten. Wir haben damit gesehen: Ist $f : U^n \rightarrow V^n$ ein Diffeomorphismus, so ist Df_x umkehrbar für jedes $x \in U^n$, und es gilt (*). Der folgende Satz behauptet die Umkehrung davon: Ist Df_x invertierbar, so ist f nahe x ein Diffeomorphismus. Der Beweis ist recht lang und kann zunächst zurückgestellt werden; die Tatsache selbst aber spielt im Weiteren eine große Rolle.

27. Satz 13. (Umkehrsatz) Es sei $f : U^n \rightarrow \mathbb{R}^n$ eine C^1 -Abbildung, und für ein $x_0 \in U^n$ sei $Df(x_0)$ invertierbar. Dann ist f nahe x_0 ein Diffeomorphismus, d.h. es gibt offene Umgebungen $X^n \subset U^n$ von x_0 und Y^n von $f(x_0)$, so dass $f : X^n \rightarrow Y^n$ ein Diffeomorphismus ist.

Beweis. Der Beweis hat vier Teile: Wir werden zunächst (a) mit Hilfe des Banachschen Fixpunktsatzes eine Abbildung g mit $f \circ g = id$ konstruieren, dann als Zwischenschritt (b) deren Lipschitz-Stetigkeit und (c) die stetige Differenzierbarkeit beweisen. Schließlich schränken wir die Definitionsbereiche von f und g so ein, dass g die Umkehrabbildung zu f wird (d).

Um die Bezeichnung zu vereinfachen, nehmen wir $x_0 = 0$ und $f(x_0) = 0$ an. Dies können wir erreichen, indem wir $f(x)$ durch $\tilde{f}(x) := f(x + x_0) - f(x_0)$ ersetzen, also jeweils eine Translation davor und danach anwenden; dann ist $D\tilde{f}_0 = Df_{x_0}$ invertierbar.

(a) Um die Umkehrabbildung zu konstruieren, müssen wir zu jedem y nahe 0 genau ein x nahe 0 finden mit

$$(1) \quad y = f(x),$$

wir müssen also die Gleichung (1) nach x auflösen. Wir setzen $A = Df_0$ und $B = A^{-1}$. Wäre f linear, d.h. $f = A$, so wäre die Lösung $x = By$. Die Hauptidee des Beweises ist: Die Gleichung $x = By$ stimmt bis auf einen kleinen Fehler, auf den wir das Fixpunktverfahren von §12 anwenden. Statt $f = A$ haben wir $f = A + r$ mit einem kleinen Rest r ; durch Anwenden von B auf die Gleichung $y = f(x) = Ax + r(x)$ erhalten wir statt $x = By$ die zu (1) äquivalente Gleichung

$$(2) \quad x = By - Br(x) =: T_y(x).$$

Dies ist eine Fixpunktgleichung für die Abbildung $T_y : U^n \rightarrow \mathbb{R}^n$,

$$T_y(x) = By - Br(x) = By - Bf(x) + x.$$

Diese Abbildung ist C^1 mit $DT_y(x) = I - B \cdot Df(x)$. Insbesondere ist $DT_y(0) = I - BA = 0$. Wegen der Stetigkeit der Abbildung $x \mapsto DT_y(x)$ (genauer: Stetigkeit der partiellen Ableitungen von T_y , also der Spalten der Matrix DT_y) ist $DT_y(x)$ immer noch fast 0 für x nahe 0. Genauer: Es gibt ein $\delta > 0$, so dass

$$\|DT_y(x)\| \leq 1/2 \quad \text{für } |x| \leq 2\delta$$

(vgl. § 9). Nach dem Schrankensatz (§11) ist $T_y = By + B \circ r$ also eine Kontraktion mit Kontraktionsfaktor $q = 1/2$ auf der abgeschlossenen Kugel $K = K_{2\delta}(0)$ (wobei stets $K_r(x) := \{x' \in \mathbb{R}^n; |x' - x| \leq r\}$ die *abgeschlossene Kugel um x mit Radius r* ist); genauer gesagt hat die Abbildung $B \circ r$ bereits diese Kontraktionseigenschaft. Um den Banachschen Fixpunktsatz anwenden zu können, müssen wir noch $T_y(K) \subset K$ zeigen. Da $|T_y(0)| = |By| \leq \|B\||y|$, erhalten wir mit $\|B\| =: \beta$

$$|T_y(0)| < \delta \quad \text{für } |y| < \frac{\delta}{\beta}.$$

Daraus ergibt sich für $|x| \leq 2\delta$ nach Dreiecksungleichung

$$(*) \quad |T_y(x)| \leq |T_y(x) - T_y(0)| + |T_y(0)| < \frac{1}{2}|x - 0| + \delta \leq 2\delta,$$

also $T_y(K) \subset K$. Nach dem Banachschen Fixpunktsatz finden wir für jedes $y \in Y_0^n := B_{\delta/\beta}(0)$ genau einen Fixpunkt $x \in K$ von T_y , also genau eine Lösung von (2) und damit von (1). Wegen (*) gilt sogar $|x| = |T_y(x)| < 2\delta$, also ist $x \in X_0^n := B_{2\delta}(0)$. Wenn wir dieses $x =: g(y)$ setzen, so haben wir damit eine Abbildung $g : Y_0^n \rightarrow X_0^n$ konstruiert mit

$$(3) \quad f(g(y)) = y \quad \forall y \in Y_0^n.$$

(b) Es seien $y, y' \in Y_0^n$ und $x, x' \in X_0^n$ ihre Bilder unter g , also die Fixpunkte von T_y und $T_{y'}$. Dann gilt

$$\begin{aligned} |x - x'| &= |T_y(x) - T_{y'}(x')| \\ &= |B(y - y') - Br(x) + Br(x')| \\ &\leq \beta \cdot |y - y'| + \frac{1}{2}|x - x'| \end{aligned}$$

wegen der oben erwähnten Kontraktionseigenschaft der Abbildung $B \circ r$. Daraus folgt

$$\frac{1}{2}|x - x'| \leq \beta \cdot |y - y'|,$$

und damit ist g Lipschitz-stetig mit Lipschitz-Konstante 2β .

(c) Um die Differenzierbarkeit von g zu zeigen, müssen wir die Differenz $k := g(y+h) - g(y)$ für jedes $y \in Y_0^n$ und für genügend kleines $|h|$ berechnen. Wir setzen $x := g(y)$. Mit (3) und der Differenzierbarkeit von f in x folgt

$$(4) \quad h = f(x+k) - f(x) = Df(x)k + o(k).$$

Wenn δ genügend klein gewählt wurde, ist $Df(x)$ invertierbar für jedes $x \in X_0^n$. (Denn die Determinante einer Matrix ist rational in den Koeffizienten der Matrix, also ist $x \mapsto \det Df(x)$ stetig und ungleich Null in $x_0 = 0$, also ungleich Null auch noch auf einem kleinen Ball um 0.) Deshalb erhalten wir durch Anwenden der Matrix $B_x := Df(x)^{-1}$ auf die obige Gleichung (4)

$$(5) \quad g(y+h) - g(y) = k = B_x h - B_x o(k),$$

und für den letzten Term rechts gilt

$$\frac{|B_x o(k)|}{|h|} \leq \|B_x\| \frac{|o(k)|}{|h|} \leq \|B_x\| \frac{|o(k)|}{|k|} \frac{|k|}{|h|} \leq \|B_x\| \frac{|o(k)|}{|k|} \cdot 2\beta,$$

denn wegen (b) (Lipschitzstetigkeit von g) ist $|k| = |g(y+h) - g(y)| \leq 2\beta \cdot |h|$. Damit haben wir $B_x o(k)/|h| \rightarrow 0$ für $h \rightarrow 0$ und damit nach (5) die Differenzierbarkeit von g in y gezeigt, und die Ableitung $Dg(y) = Df(g(y))^{-1}$ ist stetig, da g und Df stetig sind.

(d) Wir setzen nun $X^n = f^{-1}(Y_0^n) \cap X_0^n$ und $Y^n = g^{-1}(X^n) \subset Y_0^n$. Diese Mengen X^n und Y^n sind offene Umgebungen von 0 in \mathbb{R}^n (denn Urbilder offener Mengen unter stetigen Abbildungen und Schnitte offener Mengen sind offen). Für jedes $x \in X^n$ ist $f(x) \in Y_0^n$ und daher $x' := g(f(x)) \in X_0^n \subset K$. Dann gilt $f(x') = f(x) =: y$, also sind x und x' beides Fixpunkte von T_y . Also gilt $x' = x$ und damit die noch fehlende Beziehung (vgl. (3))

$$(6) \quad g(f(x)) = x \quad \forall x \in X^n,$$

somit $g = f^{-1}$. Damit ist $Y^n = f(X^n)$ und $f : X^n \rightarrow Y^n$ ein Diffeomorphismus. \square

Bemerkung. 1. Die Stetigkeit der Ableitung von f hat im obigen Beweis eine große Rolle gespielt. In der Tat wird die Behauptung des Satzes falsch, wenn wir diese Voraussetzung weglassen, wie folgendes Beispiel zeigt. Es sei $f : \mathbb{R} \rightarrow \mathbb{R}$ mit $f(0) = 0$ und

$$f(x) = \frac{1}{2}x + x^2 \sin \frac{1}{x}$$

für $x \neq 0$ (Abb. 16). Da $|x^2 \sin(1/x)|/|x| \leq |x| \rightarrow 0$ für $x \rightarrow 0$, ist f

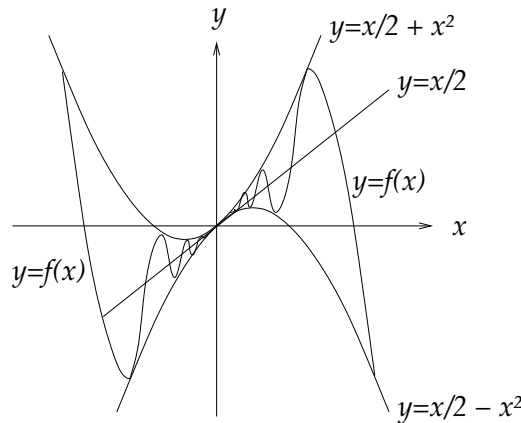


ABBILDUNG 16

in 0 differenzierbar mit $f'(0) = 1/2 \neq 0$, die Ableitung ist also in 0

invertierbar. In jedem Punkt $x \neq 0$ ist f natürlich auch differenzierbar mit

$$f'(x) = \frac{1}{2} + 2x \sin \frac{1}{x} - \cos \frac{1}{x},$$

Für $x \rightarrow 0$ geht der mittlere Term rechts gegen 0, und der letzte Term schwankt zwischen 1 und -1 , so dass f' unendlich oft sein Vorzeichen wechselt. Daher ist f in keiner Umgebung von 0 monoton und also auch nicht invertierbar.

Bemerkung. 2. Hat eine C^1 -Abbildung $f : U^n \rightarrow \mathbb{R}^n$ invertierbare Ableitung Df_x an jeder Stelle $x \in U^n$, so ist f in einer Umgebung jedes Punktes ein Diffeomorphismus (*lokaler Diffeomorphismus*). Für $n \geq 2$ folgt aber nicht, dass f injektiv ist (anders als für $n = 1$, wo dies aus der Monotonie folgt). Zm Beispiel sei $n = 2$ und $U^2 = \mathbb{R}^2 \setminus \{0\} = \mathbb{C} \setminus \{0\}$. Wir betrachten die Abbildung $f : U^2 \rightarrow \mathbb{C}$, $f(z) = z^2$, oder reell geschrieben, $f(x, y) = (x^2 - y^2, 2xy)$. Diese Funktion ist sogar *komplex differenzierbar*, d.h. für alle $z \in U^2$ existiert

$$\lim_{h \rightarrow 0} \frac{f(z+h) - f(z)}{h} =: f'(z),$$

(d.h. für jede Nullfolge (h_k) in \mathbb{C} konvergiert die Folge $((f(z+h_k) - f(z))/h_k)$ gegen die komplexe Zahl $f'(z)$). Damit ist

$$f(z+h) - f(z) - f'(z) \cdot h = o(h)$$

mit $o(h)/|h| \rightarrow 0$, also ist f auch (total) differenzierbar mit $Df(z)h = f'(z)h$. In unserem Fall ist wie im Reellen (und mit derselben Begründung) $f'(z) = 2z \neq 0$, also ist $Df(z)$ invertierbar für alle $z \in U^2$ (was wir natürlich auch "reell" hätten sehen können). Aber f ist nicht injektiv, denn $f(z) = f(-z)$ für alle $z \in U^2$.

14. Der Satz über implizite Abbildungen

Der wichtigste Punkt im Umkehrsatz im letzten Abschnitt war das Lösen einer Gleichung, nämlich $f(x) = y$. Mit Hilfe des Umkehrsatzes können wir Diffeomorphismen konstruieren, die uns wiederum helfen, allgemeinere Gleichungen vom Typ $f(x, y) = 0$ zu lösen: Durch Anwenden eines geeigneten Diffeomorphismus werden wir die Gleichung nämlich in eine ganz einfache überführen. "Lösen" heißt wieder, die Gleichung nach einer der beiden Variablen aufzulösen, d.h. die Lösungsmenge der Gleichung $f(x, y) = 0$ in der Form $y = g(x)$, also als *Graph* einer Abbildung g zu schreiben. Die durch die Gleichung $f(x, y) = 0$ ausgedrückte Abhängigkeit der Variablen y von x nennt man auch *implizit*, die durch $y = g(x)$ gegebene *explizit*.

Die gegebene Abbildung f hängt von zwei Variablen x, y ab, die aber selber vektorwertig sein dürfen; wir haben also eine Abbildung $f : U^n \times V^p \rightarrow \mathbb{R}^q$ vorliegen. Hierzu zunächst einige Bezeichnungen. Wir sprechen auch in diesem Fall von *partiellen Ableitungen von f* , wenn wir jeweils eine der beiden Variablen konstant halten und nur die andere variieren. Genauer: Für jedes $x \in U^n$ und $y \in V^p$ erklären wir Abbildungen $f_x : V^p \rightarrow \mathbb{R}^q$ und $f^y : U^n \rightarrow \mathbb{R}^q$, indem wir $f_x(y) := f^y(x) := f(x, y)$ setzen, und definieren

$$D_1 f(x, y) := D(f^y)(x) \in \mathbb{R}^{q \times n}, \quad D_2 f(x, y) := D(f_x)(y) \in \mathbb{R}^{q \times p}.$$

Natürlich kann es hier Verwechslungen mit den partiellen Ableitungen nach den ersten zwei Koordinaten geben; durch die vorher angegebene Zerlegung des Definitionsbereichs wird aber deutlich, welche partiellen Ableitungen wir meinen. Weiterhin benötigen wir die beiden Projektionen von $\mathbb{R}^n \times \mathbb{R}^p$ (oder offenen Teilmengen davon) nach \mathbb{R}^n und \mathbb{R}^p , die wir pr_1 und pr_2 nennen wollen, also

$$pr_1(x, y) = x, \quad pr_2(x, y) = y.$$

Beispiel 1. Leider läßt sich die implizite Abhängigkeit $f(x, y) = 0$ nicht immer in eine explizite $y = g(x)$ umwandeln. Betrachten wir dazu ein einfaches Beispiel, die Kreisgleichung in der Ebene: $f(x, y) := x^2 + y^2 - 1 = 0$. In der Nähe des Punktes $(0, 1)$ auf der y -Achse läßt sich diese Gleichung nach y auflösen, nämlich durch $y = \sqrt{1 - x^2}$, aber in der Nähe des Punktes $(1, 0)$ auf der x -Achse ist dies nicht möglich: Der Kreis kann dort nicht Graph einer differenzierbaren Funktion sein, denn die Tangente ist senkrecht (Abb. 17). Analytisch wird dieses Faktum durch $D_2 f(1, 0) = 0$ ausgedrückt (siehe nächster §). Um dieses Problem zu vermeiden, setzen wir im folgenden Satz die Umkehrbarkeit von $D_2 f$ voraus; insbesondere müssen wir $q = p$ annehmen.

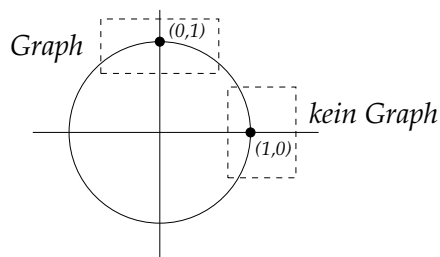


ABBILDUNG 17

28. Satz 14.1 *Es sei $f : U^n \times V^p \rightarrow \mathbb{R}^p$ stetig differenzierbar und $(x_0, y_0) \in U^n \times V^p$ mit $f(x_0, y_0) = 0$ und $D_2 f(x_0, y_0)$ invertierbar.*

Dann gibt es offene Umgebungen $A^{n+p} \subset U^n \times V^p$ von (x_0, y_0) und $B^{n+p} \subset \mathbb{R}^{n+p}$ von $(x_0, 0)$ sowie einen Diffeomorphismus $G : B^{n+p} \rightarrow A^{n+p}$ mit

$$f \circ G = pr_2.$$

29. Satz 14.2 (Satz über implizite Abbildungen) *Unter denselben Voraussetzungen gibt es offene Umgebungen X^n von x_0 und A^{n+p} von (x_0, y_0) und eine C^1 -Abbildung $g : X^n \rightarrow \mathbb{R}^p$, so dass $f^{-1}(0) \cap A^{n+p} = \text{Graph } g$, oder mit anderen Worten: Für alle $(x, y) \in A^{n+p}$ gilt:*

$$f(x, y) = 0 \iff y = g(x).$$

Beweis. Wir beweisen beide Sätze auf einmal. Die Voraussetzungen (Invertierbarkeit einer Ableitung) erinnern uns an den Umkehrsatz. Um ihn anwenden zu können, erweitern wir f so, dass die ganze Ableitung invertierbar wird. Dazu setzen wir

$$F = (pr_1, f) : U^n \times V^p \rightarrow \mathbb{R}^n \times \mathbb{R}^p,$$

also $F(x, y) = (x, f(x, y))$, oder besser (unter Beachtung der Konvention, dass Vektoren im $\mathbb{R}^n, \mathbb{R}^p, \mathbb{R}^{n+p}$ eigentlich keine Zeilen, sondern Spalten sind):

$$F \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ f(x, y) \end{pmatrix}.$$

Dann ist

$$DF(x, y) = \begin{pmatrix} I & 0 \\ D_1f(x, y) & D_2f(x, y) \end{pmatrix}.$$

An der Stelle (x_0, y_0) ist diese Matrix invertierbar, da $D_2f(x_0, y_0)$ invertierbar ist. Nach dem Umkehrsatz (Satz 13) gibt es daher offene Umgebungen A^{n+p} von (x_0, y_0) und B^{n+p} von $F(x_0, y_0) = (x_0, 0)$, so dass $F : A^{n+p} \rightarrow B^{n+p}$ Diffeomorphismus ist. Für die Umkehrabbildung $G = F^{-1} : B^{n+p} \rightarrow A^{n+p}$ gilt $F \circ G = id$, $G \circ F = id$, also mit der Definition $F = (pr_1, f)$:

$$(1) \quad (pr_1(G(u, v), f(G(u, v)))) = (u, v),$$

$$(2) \quad G(x, f(x, y)) = (x, y)$$

für alle $(u, v) \in B^{n+p}$, $(x, y) \in A^{n+p}$. Die zweite Komponente von (1) zeigt $f(G(u, v)) = v$, also $f \circ G = pr_2$, was in 14.1 behauptet wurde. Setzen wir $g(x) := pr_2(G(x, 0))$, so sagt die zweite Komponente von (2) für alle $(x, y) \in A^{n+p}$:

$$g(x) = y \iff G(x, 0) = (x, y) \iff f(x, y) = 0,$$

was die Behauptung von 14.2 ist. \square

30. Korollar. Für die Abbildung g aus Satz 14.2 gilt:

$$g(x_0) = y_0, \quad Dg(x_0) = -D_2f(x_0, y_0)^{-1} \cdot D_1f(x_0, y_0).$$

Beweis. Da $(x_0, y_0) \in A^{n+p}$ und $f(x_0, y_0) = 0$, ist $g(x_0) = y_0$ nach 14.2. Die zweite Behauptung folgt aus der Kettenregel: Wir wissen, dass g differenzierbar ist mit $f(x, g(x)) = 0$. Setzen wir $h(x) := (x, g(x))$, genauer

$$h(x) = \begin{pmatrix} x \\ g(x) \end{pmatrix},$$

so ergibt sich $f(h(x)) = 0$ und daraus nach der Kettenregel:

$$\begin{aligned} 0 = Df(h(x)) \cdot Dh(x) &= (D_1f(h(x)), D_2f(h(x))) \cdot \begin{pmatrix} I \\ Dg(x) \end{pmatrix} \\ &= D_1f(h(x)) + D_2f(h(x)) \cdot Dg(x). \end{aligned}$$

Durch Anwenden von $D_2f(h(x))^{-1}$ erhalten wir also

$$Dg(x) = -D_2f(h(x))^{-1} \cdot D_1f(h(x)).$$

Einsetzen von $h(x_0) = (x_0, y_0)$ ergibt die zweite Behauptung. \square

Im Beispiel 1 konnten wir die Gleichung $f(x, y) = 0$ “per Hand” nach y auflösen. Der Satz über implizite Abbildungen ist natürlich nur in komplizierteren Situationen nützlich, wenn einem nicht genügend Standard-Rechenoperationen (wie die vier Grundrechenarten, Wurzelziehen, elementare Funktionen wie Exponentialfunktion, Logarithmus usw.) zur Verfügung stehen, um die Gleichung explizit aufzulösen. Dann sagt der Satz, dass unter den gegebenen Voraussetzungen trotzdem eine C^1 -Abbildung g mit $f(x, g(x)) = 0$ definiert wird, und mit dem Korollar können wir sogar die Ableitung $Dg(x_0)$ berechnen.

Beispiel 2. Gegeben seien die Gleichungen

$$\begin{aligned} x^3 + y^3 + z^3 &= 7, \\ xy + yz + zx &= -2. \end{aligned}$$

Eine Lösung ist $(x_0, y_0, z_0) = (2, -1, 0)$. Kann man die Gleichungen in der Nähe dieses Punktes nach (y, z) auflösen, die Lösungen also in der Form $y = g_1(x)$, $z = g_2(x)$ angeben? Wir definieren dazu (*linke Seite*) – (*rechte Seite*) = $f(x, y, z)$, genauer: Wir definieren $f : \mathbb{R} \times \mathbb{R}^2 \rightarrow \mathbb{R}^2$,

$$f(x, (y, z)) = \begin{pmatrix} x^3 + y^3 + z^3 - 7 \\ xy + yz + zx + 2 \end{pmatrix}.$$

Die Rolle von y in Satz 14.2 spielt jetzt (y, z) ; die partielle Ableitung D_2f bezüglich der Aufspaltung $\mathbb{R}^3 = \mathbb{R} \times \mathbb{R}^2$ ist also

$$D_2f = \left(\frac{\partial f}{\partial y}, \frac{\partial f}{\partial z} \right) = \begin{pmatrix} 3y^2 & 3z^2 \\ x+z & y+x \end{pmatrix}.$$

Insbesondere ist

$$D_2f(x_0, (y_0, z_0)) = \begin{pmatrix} 3 & 0 \\ 2 & 1 \end{pmatrix}$$

eine invertierbare Matrix. Nach 14.2 existiert also ein offenes Intervall I um $x_0 = 2$ und eine C^1 -Abbildung $g : I \rightarrow \mathbb{R}^2$ mit $f(x, g(x)) = 0$ (genauer: $f^{-1}(0) \cap A^3 = \text{Graph } g$ für eine offene Umgebung A^3 von $(2, -1, 0)$). Mit dem Korollar von 14.2 können wir die Ableitung von g berechnen: Da für die Inversion von 2×2 -Matrizen bekanntlich gilt:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

(nachrechnen!), erhalten wir

$$D_2f(x_0, (y_0, z_0))^{-1} = \frac{1}{3} \begin{pmatrix} 1 & 0 \\ -2 & 3 \end{pmatrix},$$

und ferner

$$D_1f(x_0, (y_0, z_0)) = \frac{\partial f}{\partial x}(2, -1, 0) = \left. \begin{pmatrix} 3x^2 \\ y+z \end{pmatrix} \right|_{(2,-1,0)} = \begin{pmatrix} 12 \\ -1 \end{pmatrix}.$$

Somit ist nach dem Korollar

$$Dg(2) = -\frac{1}{3} \begin{pmatrix} 1 & 0 \\ -2 & 3 \end{pmatrix} \begin{pmatrix} 12 \\ -1 \end{pmatrix} = \begin{pmatrix} -4 \\ 9 \end{pmatrix}.$$

Die Lösungsmenge der Gleichung ist somit nahe $(2, -1, 0)$ Graph einer Kurve $g : I \rightarrow \mathbb{R}^2$ (mit Werten in der yz -Ebene) mit

$$g(2) = \begin{pmatrix} -1 \\ 0 \end{pmatrix}, \quad g'(2) = \begin{pmatrix} -4 \\ 9 \end{pmatrix}.$$

Übrigens hätten wir auch in diesem Fall versuchen können, die Gleichungen explizit aufzulösen, indem wir aus der zweiten Gleichung $y = -(2 + zx)/(x + z)$ berechnen und in die erste Gleichung substituieren. Dies ergibt eine Gleichung 6. Grades zwischen x und z . Wenn wir es schaffen, diese nach z aufzulösen, können wir das so gewonnene $z = g_2(x)$ in die Gleichung $y = -(2 + zx)/(x + z)$ einsetzen und erhalten $y = g_1(x)$. Viel Spaß!

Für Leute, die dazu verdammt sind, Übungs- und Klausuraufgaben stellen oder rechnen zu müssen, ist Satz 14.2 zweifellos schrecklich wichtig. Für die Mathematik aber ist der Satz 14.1 viel aufschlußreicher. Er

sagt ja, dass die fast beliebig komplizierte Abbildung $f : U^n \times V^p \rightarrow \mathbb{R}^p$ “bis auf einen Diffeomorphismus” (nämlich G) dasselbe ist wie die einfachste überhaupt denkbare Abbildung: die Projektion (die ja einfach im Weglassen von Variablen besteht). Die verschiedenen *Niveaumengen* von f (das sind die Urbilder von Punkten, also $f^{-1}(a)$ für $a \in \mathbb{R}^p$) werden dabei durch den Diffeomorphismus $F = G^{-1}$ auf offene Mengen in den zueinander parallelen affinen Teilräumen $\mathbb{R}^n \times \{a\} \subset \mathbb{R}^{n+p}$ (nämlich auf $B^{n+p} \cap (\mathbb{R}^n \times \{a\})$) abgebildet; offene Mengen in affinen Teilräumen werden wir manchmal *affine Mengen* nennen (Abb. 18). (Ein *affiner Teilraum* des \mathbb{R}^N ist nach Definition ein verschobener Untervektorraum, also eine Menge $V + a = \{v + a; v \in V\}$, wobei $V \subset \mathbb{R}^N$ ein Untervektorraum oder *linearer Teilraum* ist.) Wir sehen hier ein weiteres Prinzip der Analysis, das uns auch weiter begleiten wird: Durch Anwenden geeigneter Diffeomorphismen lassen sich Situationen drastisch vereinfachen und dadurch verstehen.

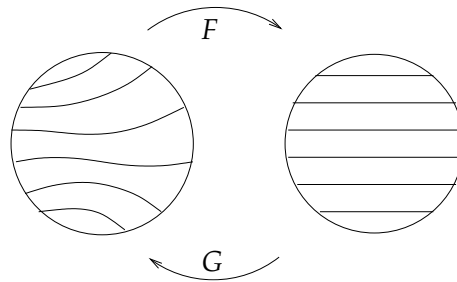


ABBILDUNG 18

Die in 14.1 gemachten Voraussetzungen an f sind allerdings noch etwas unbefriedigend, weil durch die spezielle Aufspaltung $\mathbb{R}^{n+p} = \mathbb{R}^n \times \mathbb{R}^p$ den letzten p Koordinaten des \mathbb{R}^{n+p} eine Sonderrolle zufällt; davon werden wir uns im folgenden Abschnitt befreien.

15. Mannigfaltigkeiten und reguläre Niveaumengen

Eine Teilmenge $M \subset \mathbb{R}^n$ heißt *m-dimensionale Mannigfaltigkeit* ($m \leq n$), wenn sie überall lokal diffeomorph zu einer m -dimensionalen affinen Menge ist, genauer: Zu jedem Punkt $x \in M$ gibt es offene Umgebungen U^m von x und V^m von 0 in \mathbb{R}^m und einen Diffeomorphismus $\Phi : U^m \rightarrow V^m$ mit $\Phi(x) = 0$, so dass

$$\Phi(M \cap U^m) = \mathbb{R}^m \cap V^m.$$

(Abb. 19). Dabei fassen wir \mathbb{R}^m für $m \leq n$ in üblicher Weise als Unterraum von \mathbb{R}^n auf, nämlich als $\{x \in \mathbb{R}^n; x_{m+1} = \dots = x_n = 0\}$.

Natürlich können wir in der obigen Definition \mathbb{R}^m auch durch einen dazu parallelen affinen Teilraum $\mathbb{R}^m + a = \{x + a; x \in \mathbb{R}^m\}$ ersetzen, falls $\Phi(x) = a \neq 0$. Der Diffeomorphismus Φ heißt *Karte* oder *Koordinatensystem* für M um den Punkt x . (Bei “Karte” ist an “Landkarte” gedacht: Eine Landkarte bildet ein Stück Erdoberfläche auf ein ebenes Stück Papier ab.)

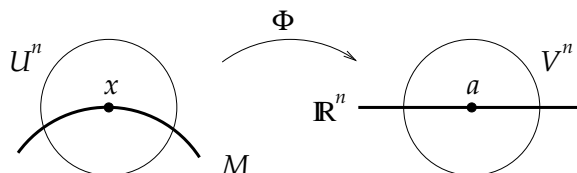


ABBILDUNG 19

Ein *Tangentialvektor* von M in einem Punkt $x \in M$ ist ein Tangentenvektor $c'(0)$ einer (differenzierbaren) Kurve $c : (-\epsilon, \epsilon) \rightarrow M \subset \mathbb{R}^n$ mit $c(0) = x$. Die Menge aller Tangentenvektoren in x bildet den *Tangentialraum* $T_x M$ von M in x (Abb. 20). Anschaulich denke man dabei an die Tangente einer Kurve oder die Tangentialebene einer Fläche. (Der Tangentialraum wird gewöhnlich in den Punkt x parallelverschoben gezeichnet, so dass der Nullpunkt von $T_x M$ nicht in 0 , sondern in x liegt.)

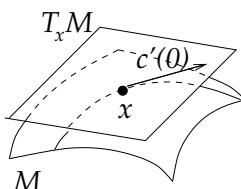


ABBILDUNG 20

31. Satz 15.1 *Ist $M \subset \mathbb{R}^n$ eine m -dimensionale Mannigfaltigkeit, so ist $T_x M \subset \mathbb{R}^n$ ein m -dimensionaler linearer Teilraum, und zwar gilt: $T_x M = D(\Phi^{-1})_{\Phi(x)} \mathbb{R}^m$ für jede Karte Φ um x .*

Beweis. Es sei $a = \Phi(x)$ und $\Psi = \Phi^{-1} : V^n \rightarrow U^n$. Ist $v = c'(0) \in T_x M$ für eine Kurve $c : (-\epsilon, \epsilon) \rightarrow M \cap U^n$ mit $c(0) = x$, und ist $\Phi : U^n \rightarrow V^n$ eine Karte, so ist $\Phi \circ c$ eine Kurve in \mathbb{R}^m , und nach Kettenregel ist $v = (\Psi \circ \Phi \circ c)'(0) = D\Psi(a)w$ mit $w = (\Phi \circ c)'(0) \in \mathbb{R}^m$. Ist umgekehrt $w \in \mathbb{R}^m$ gegeben, so ist $c(t) = \Psi(a + tw)$ eine Kurve in M mit $c(0) = x$ und daher ist $D\Psi_a w = c'(0) = D\Psi(a)w \in T_x M$. \square

Die wichtigsten Beispiele von Mannigfaltigkeiten sind bestimmte Niveaumengen von Abbildungen: Gegeben sei eine C^1 -Abbildung $f : U^n \rightarrow \mathbb{R}^p$ mit $p \leq n$. Ein Punkt $x \in U^n$ heißt *regulärer Punkt* für f , wenn die lineare Abbildung $Df(x) : \mathbb{R}^n \rightarrow \mathbb{R}^p$ surjektiv ist, wenn also die Matrix $Df(x)$ den Rang p hat. Besteht eine Niveaumenge

$$f^{-1}(a) = \{x \in \mathbb{R}^n; f(x) = a\}$$

für ein $a \in \mathbb{R}^p$ nur aus regulären Punkten für f , so heißt $f^{-1}(a)$ eine *reguläre Niveaumenge*.

32. Satz 15.2 *Ist $M = f^{-1}(a)$ reguläre Niveaumenge einer C^1 -Abbildung $f : U^n \rightarrow \mathbb{R}^p$, so ist M eine Mannigfaltigkeit der Dimension $m = n - p$.*

Beweis. Es sei $x \in M$. Da $Df(x)$ den Rang p hat, sind p Spalten von $Df(x)$ linear unabhängig; durch Ummummern der Koordinaten können wir annehmen, dass es die letzten p Spalten sind. Wir spalten nun den \mathbb{R}^n auf als $\mathbb{R}^m \times \mathbb{R}^p$. Da U^n offen ist, gibt es in U^n eine kleinere offene Umgebung von x von der Gestalt $U_1^n = V^m \times W^p$ (siehe Abb. 21). Wir schränken f auf U_1^n ein. Die partielle Ableitung $D_2 f(x)$ bezüglich

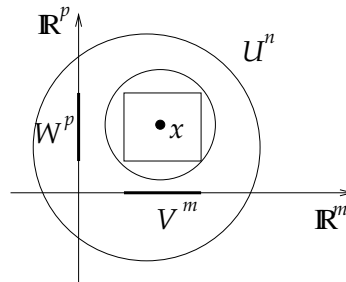


ABBILDUNG 21

dieser Aufspaltung besteht aus den letzten p Spalten von $Df(x)$. Diese sind linear unabhängig, also ist $D_2 f(x)$ invertierbar. Nach Satz 14.1 gibt es einen Diffeomorphismus $F : A^n \rightarrow B^n$, auf einer noch kleineren Umgebung $A^n \subset U_1^n$ von x definiert, so dass $f \circ G = pr_2$ für $G := F^{-1}$. Dann wird M durch F auf eine m -dimensionale affine Menge abgebildet, nämlich

$$F(M \cap A^n) = (\mathbb{R}^m \times \{a\}) \cap B^n,$$

denn $x' \in M \cap A^n \Leftrightarrow f(GF x') = a \Leftrightarrow pr_2(F x') = a \Leftrightarrow F x' \in \mathbb{R}^m \times \{a\}$. \square

Reguläre Niveaumengen $f^{-1}(a)$ können wir auch mit Hilfe der Gradienten der Komponentenfunktionen von f kennzeichnen: Für eine C^1 -Abbildung $f = (f_1, \dots, f_p)^T : U^n \rightarrow \mathbb{R}^p$ sind die Ableitungen der Komponenten $Df_i = (\nabla f_i)^T$ die Zeilen von Df . Also ist $Df(x)$ genau dann surjektiv (d.h. hat Rang p), wenn die Vektoren $\nabla f_1(x), \dots, \nabla f_p(x)$ linear unabhängig sind. Diese sind außerdem senkrecht zum Tangentialraum (Abb. 22):

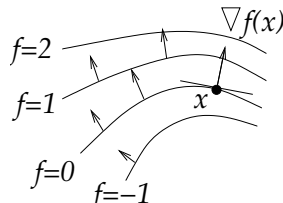


ABBILDUNG 22

33. Lemma *Ist $M = f^{-1}(a)$ reguläre Niveaumenge und $x \in M$, so gilt $\nabla f_i(x) \perp T_x M$ für $i = 1, \dots, p$.*

Beweis. Für jedes $v \in T_x M$ gibt es eine Kurve $c : (-\epsilon, \epsilon) \rightarrow M$ mit $c(0) = x$, $c'(0) = v$. Da $M = f^{-1}(a)$, ist $f(c(t)) = a$ und damit $f_i(c(t)) = a_i = \text{const}$ für alle $t \in (-\epsilon, \epsilon)$. Die Ableitung von $f_i \circ c$ ist also Null, und damit ist

$$0 = \left. \frac{d}{dt} \right|_0 f_i(c(t)) = Df_i(x)v = \langle \nabla f_i(x), v \rangle.$$

□

Wir wissen jetzt also, dass in jedem Punkt $x \in M$ die Gradienten $\nabla f_1(x), \dots, \nabla f_p(x)$ linear unabhängig sind und senkrecht auf $T_x M$ stehen. Da $\dim T_x M = n - p$, hat das orthogonale Komplement $(T_x M)^\perp$ die Dimension p , also bilden $\nabla f_1(x), \dots, \nabla f_p(x)$ eine Basis von $(T_x M)^\perp$. Für jede Mannigfaltigkeit $M \subset \mathbb{R}^n$ und jeden Punkt $x \in M$ heißt der Raum $N_x M := (T_x M)^\perp$ der *Normalraum von M in x* (Abb. 23). Wir haben damit gezeigt:

34. Satz 15.3 *Es sei $f = (f_1, \dots, f_p)^T : U^n \rightarrow \mathbb{R}^p$ eine C^1 -Abbildung und $M = f^{-1}(a)$ ein reguläres Urbild. Dann bilden die Gradienten $\nabla f_1(x), \dots, \nabla f_p(x)$ für jedes $x \in M$ eine Basis des Normalraumes $N_x M$.*

□

Die Dimension $p = n - m$ des Normalraums einer Mannigfaltigkeit M nennt man die *Codimension* von M . Mannigfaltigkeiten der Codimension eins heißen *Hyperflächen*. Der Inhalt des vorstehenden Satzes

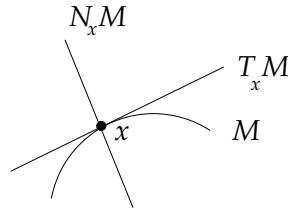


ABBILDUNG 23

für $p = 1$ wird manchmal etwas verkürzt auch so ausgedrückt: *Der Gradient einer Funktion steht senkrecht auf seinen Niveau-Hyperflächen.*

Beispiel 1. Jede offene Teilmenge $U^n \subset \mathbb{R}^n$ ist n -dimensionale Mannigfaltigkeit, und zwar reguläre Niveaumenge für die konstante Funktion $f : U^n \rightarrow \mathbb{R}^0 = \{0\}$.

Beispiel 2. Eine reguläre Niveaumenge einer Abbildung $f : U^n \rightarrow \mathbb{R}^n$ (also $p = n$) ist eine diskrete Punktmenge in U^n ; dies ist eine 0-dimensionale Mannigfaltigkeit. (Eine Menge $M \subset \mathbb{R}^n$ heißt *diskret*, wenn es zu jedem $x \in M$ einen Ball $B_\epsilon(x)$ gibt, der x als einzigen Punkt von M enthält: $M \cap B_\epsilon(x) = \{x\}$.)

Beispiel 3. Es sei $M = \partial B_R(0) = S_R = \{x \in \mathbb{R}^n; |x| = R\}$ für ein $R > 0$. Diese Menge heißt *Sphäre* vom Radius R (für $n = 2$ *Kreislinie*, für $n = 3$ *Kugelfläche*). Es gilt $M = f^{-1}(R^2)$ für $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $f(x) = \langle x, x \rangle$. Die Funktion $f(x) = \langle Bx, x \rangle$ (B symmetrische Matrix) haben wir schon in §8, Beispiel 3 und in §10, Beispiel 2 kennengelernt und $\nabla f(x) = 2Bx$ berechnet. Hier liegt der Spezialfall $B = I$ vor, also $\nabla f(x) = 2x \neq 0$ für $|x| > 0$. Also ist M eine reguläre Niveau-Hyperfläche, und der "Ortsvektor" x steht senkrecht auf $T_x M$ (was wir ja schon seit Urzeiten wissen: die Kreistangente ist senkrecht zum Radius).

Beispiel 4. Der *Rotationstorus* $T_{R,r}$ mit Radien R und r (mit $R > r > 0$) ist die Menge aller Punkte im \mathbb{R}^3 , die von der Kreislinie mit Radius R in der xy -Ebene den Abstand r haben (Abb. 24). Bezeichnen wir mit $\rho := \sqrt{x^2 + y^2}$ den Abstand eines Punktes (x, y, z) von der z -Achse, so ist der Torus $T_{R,r}$ die Lösungsmenge der Gleichung

$$(*) \quad (\rho - R)^2 + z^2 = r^2.$$

Also ist $T_{R,r} = f^{-1}(r^2)$ mit $f : \mathbb{R}^3 \setminus z\text{-Achse} \rightarrow \mathbb{R}$, $f = (\rho - R)^2 + z^2$. Dabei werden $\rho : (x, y, z) \mapsto \sqrt{x^2 + y^2}$ und $z : (x, y, z) \mapsto z$ als Funktionen auf \mathbb{R}^3 aufgefaßt (ebenso x und y); z ist überall und ρ auf $\mathbb{R}^3 \setminus z\text{-Achse}$ differenzierbar. Da die z -Achse eine abgeschlossene

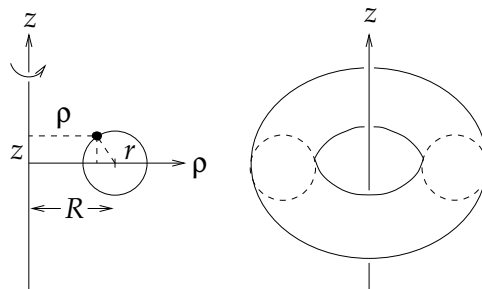


ABBILDUNG 24

Teilmenge von \mathbb{R}^3 ist (nämlich $\{(0, 0, z); z \in \mathbb{R}\}$), ist $U^3 := \mathbb{R}^3 \setminus z$ -Achse offen. Es ist

$$\frac{\partial \rho}{\partial x} = \frac{x}{\rho}, \quad \frac{\partial \rho}{\partial y} = \frac{y}{\rho}, \quad \frac{\partial \rho}{\partial z} = 0,$$

denn nach Kettenregel ist $D_i \sqrt{u} = D_i u / (2\sqrt{u})$ für jede Funktion u . Somit erhalten wir

$$\nabla \rho(x, y, z) = \frac{1}{\rho}(x, y, 0)^T, \quad \nabla z = (0, 0, 1)^T,$$

also wieder nach Kettenregel

$$\nabla f = 2(\rho - R)\nabla \rho + 2z\nabla z = 2\frac{\rho - R}{\rho} \begin{pmatrix} x \\ y \\ 0 \end{pmatrix} + 2z \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

Dieser Vektor wird Null nur dann, wenn $z = 0$ und außerdem entweder $\rho = R$ oder $x = y = 0$. Der Schnitt von $T_{R,r}$ mit der xy -Ebene $\{z = 0\}$ besteht aber nur aus den beiden Kreisen $\rho = R \pm r$, also gilt dort weder $\rho = R$ noch $\rho = 0$. Also verschwindet der Gradient von f nirgends auf $T_{R,r} = f^{-1}(r^2)$ und somit ist $T_{R,r}$ eine reguläre Niveaumenge, insbesondere also eine 2-dimensionale Mannigfaltigkeit, eine *Fläche*.

Bemerkung. 1. Wir können die Gleichung (*) noch etwas umformen, um die Quadratwurzel zu beseitigen, die in ρ steckt: (*) $\Leftrightarrow 2R\rho = \rho^2 + z^2 + R^2 - r^2 \Leftrightarrow$

$$(**) \quad 4R^2(x^2 + y^2) = ((x^2 + y^2 + z^2 + R^2 - r^2)^2).$$

Hier ist *Rechte Seite* – *Linke Seite* ein Polynom 4. Ordnung, d.h. es treten höchstens vierfache Produkte der Koordinatenfunktionen x, y, z auf (siehe nächster §). Solche Flächen nennt man *Flächen 4. Ordnung* oder *Quartiken*. (Flächen 3. Ordnung heißen *Kubiken*, Flächen 2. Ordnung *Quadriken*.) Allgemeiner ist eine Hyperfläche k -ter Ordnung in \mathbb{R}^n eine reguläre Niveaumenge $M = f^{-1}(0)$, wobei $f : \mathbb{R}^n \rightarrow \mathbb{R}$ ein

Polynom k -ter Ordnung auf \mathbb{R}^n ist, d.h. es treten höchstens k -fache Produkte der Koordinatenfunktionen x_1, \dots, x_n auf. Eine solche Hyperfläche schneidet eine Gerade $g : \mathbb{R} \rightarrow \mathbb{R}^n$, $g(t) = x + tv$ höchstens k mal, da $f \circ g : \mathbb{R} \rightarrow \mathbb{R}$ ein Polynom von höchstens k -ter Ordnung in der Variablen t ist und also höchstens k Nullstellen hat, d.h. es gibt höchstens k Zahlen $t_i \in \mathbb{R}$ mit $g(t_i) \in M$. Für den Torus sehen wir ja auch anschaulich, dass jede Gerade höchstens viermal schneidet (Abb. 25)

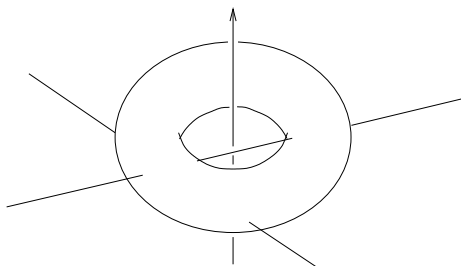


ABBILDUNG 25

Bemerkung. 2. Einen linearen Unterraum $V \subset \mathbb{R}^n$ können wir auf zwei verschiedenen Weisen beschreiben: Als Kern (Niveaumenge von 0) oder als Bild einer linearen Abbildung. Ebenso können Mannigfaltigkeiten entweder als Niveaumenge oder als Bild einer i.a. nichtlinearen differenzierbaren Abbildung beschrieben werden. Lokal gibt es ja Diffeomorphismen $\Phi : U^n \rightarrow V^n$ mit $\Phi(M \cap U^n) = \mathbb{R}^m \cap V^n$. Spalten wir den \mathbb{R}^n auf als $\mathbb{R}^n = \mathbb{R}^m \times \mathbb{R}^{n-m}$ und nennen die Projektion auf den zweiten Faktor \mathbb{R}^{n-m} wieder pr_2 , so ist $M \cap U^n$ reguläre Niveaumenge der Abbildung $f = pr_2 \circ \Phi$, nämlich $M = f^{-1}(0)$. Setzen wir andererseits $\Psi = \Phi^{-1}$ und schränken Ψ auf $V^m := \mathbb{R}^m \cap V^n$ ein, so erhalten wir eine differenzierbare Abbildung $\psi : V^m \rightarrow \mathbb{R}^n$, deren Bild genau $M \cap U^n$ ist. Als Einschränkung eines Diffeomorphismus auf einen Unterraum ist ψ injektiv, und auch $D\psi(u) : \mathbb{R}^m \rightarrow \mathbb{R}^n$ ist eine injektive lineare Abbildung für alle $u \in V^m$. Eine solche Abbildung heißt *lokale Parametrisierung* von M .

16. Lokale Extrema (mit Nebenbedingungen)

Ein *lokales Maximum* (bzw. *lokales Minimum*) einer Funktion $f : D \rightarrow \mathbb{R}$ mit $D \subset \mathbb{R}^n$ ist ein Punkt $x_0 \in D$ so dass $f(x) \leq f(x_0)$ (bzw. $f(x) \geq f(x_0)$) für alle $x \in D \cap B_\epsilon(x_0)$ für ein $\epsilon > 0$. Lokale Maxima und Minima zusammen nennt man auch *lokale Extrema*.

35. Satz 16. Es sei $f : U^n \rightarrow \mathbb{R}$ differenzierbar, $M \subseteq U^n$ eine Mannigfaltigkeit und $x_0 \in M$ sei ein lokales Extremum von $f|_M$ (Abb. 26). Dann gilt $\nabla f(x_0) \in N_{x_0}M$, insbesondere $\nabla f(x_0) = 0$ für $M = U^n$. Ist $M = g^{-1}(a)$ reguläre Niveaumenge einer C^1 -Abbildung $g = (g_1, \dots, g_p)^T : U^n \rightarrow \mathbb{R}^p$, so gibt es (eindeutige) Zahlen $\lambda_1, \dots, \lambda_p \in \mathbb{R}$, genannt Lagrange-Multiplikatoren, so dass

$$\nabla f(x_0) = \sum_{i=1}^p \lambda_i \nabla g_i(x_0).$$

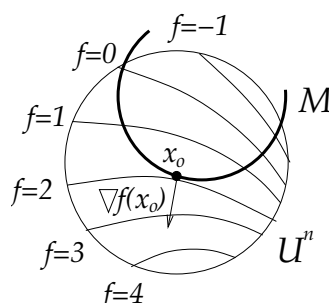


ABBILDUNG 26

Beweis. Wir dürfen annehmen, dass x_0 lokales *Maximum* von f ist (sonst gehen wir zu $-f$ über). Wir wollen zeigen, dass $\nabla f(x_0) \perp T_{x_0}M$. Dazu sei $v \in T_{x_0}M$ beliebig; es gibt also eine Kurve $c : (-\delta, \delta) \rightarrow M$ mit $c(0) = x_0$, $c'(0) = v$. Wir dürfen annehmen, dass das Bild von c in $B_\epsilon(x_0)$ liegt; sonst verkleinern wir δ noch etwas. Dann ist $f(c(t)) \leq f(x_0) = f(c(0))$ für alle t , also hat die differenzierbare Funktion $f \circ c : (-\delta, \delta) \rightarrow \mathbb{R}$ ein Maximum bei $t = 0$. Also ist dort die Ableitung Null und nach Kettenregel ist $0 = (f \circ c)'(0) = \langle \nabla f(x_0), v \rangle$. Wir erhalten also $\nabla f(x_0) \perp v$ für alle $v \in T_{x_0}M$, und damit ist $\nabla f(x_0) \in N_{x_0}M$. Ist $M = g^{-1}(a)$ reguläre Niveaumenge, so ist $\nabla g_1(x_0), \dots, \nabla g_p(x_0)$ eine Basis von $N_{x_0}M$ (vgl. Satz 15.3), also ist $\nabla f(x_0) \in N_{x_0}M$ eine eindeutige Linearkombination dieser Vektoren. \square

Die Einschränkung der Funktion f auf die Mannigfaltigkeit M nennt man auch *Nebenbedingung*. Die Punkte $x \in M$, in denen $\nabla f(x) \perp T_xM$ gilt, heißen *kritische Punkte* der Funktion $f|_M$. Falls $M = U^n$, wenn es also keine Nebenbedingung gibt, bedeutet dies $\nabla f(x) = 0$. Satz 16 sagt also, dass lokale Extrema kritische Punkte sind. Die Umkehrung stimmt natürlich nicht, wie wir schon aus der eindimensionalen Analysis wissen: Die Ableitung kann verschwinden, ohne dass ein lokales

Extremum vorliegt (z.B. in Sattelpunkten). Wir werden dies in §19 weiter diskutieren.

Der Satz 16 wird besonders nützlich in Verbindung mit einem Existenzsatz für Extrema, wie z.B. Satz 4.2 (Existenz von Extrema stetiger Funktionen auf kompakten Mengen). Mit einem reinen Existenzsatz wie diesem ist allein nicht viel anzufangen: Was nützt es der Polizei, wenn sie weiß, dass es einen Mörder geben muß, solange sie keinen Anhaltspunkt hat, wer es ist? Solch einen Anhaltspunkt (Indiz) liefert Satz 16: Das Extremum muß einer der kritischen Punkte sein.

Beispiel. Es sei $B \in \mathbb{R}^{n \times n}$ eine symmetrische Matrix (d.h. $B^T = B$) und $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $f(x) = \langle Bx, x \rangle$. Weiterhin sei $g : \mathbb{R}^n \rightarrow \mathbb{R}$, $g(x) = \langle x, x \rangle$ und $M = S = g^{-1}(1)$ die Einheitssphäre. Wir suchen die Minima von $f|_S$. Da S kompakt ist (beschränkt und abgeschlossen), nimmt $f|_S$ nach Satz 4.2 ein Minimum $f(x_1)$ für ein $x_1 \in S$ an. Dort gilt $\nabla f(x_1) = \lambda \nabla g(x_1)$ für ein $\lambda \in \mathbb{R}$. Da $\nabla f(x) = 2Bx$ und $\nabla g(x) = 2x$ (vgl. §10), folgt

$$Bx_1 = \lambda x_1,$$

d.h. x_1 ist Eigenvektor zum Eigenwert λ . Das Minimum wird also bei einem Eigenvektor von B angenommen; genauer gesagt sind die kritischen Punkte von $f|_S$ genau die Eigenvektoren von B . Insbesondere haben wir damit gesehen, dass eine symmetrische Matrix einen reellen Eigenwert besitzt.

Bemerkung. Wir können auf diese Weise sogar eine volle Orthonormalbasis aus Eigenvektoren von B konstruieren und damit einen bekannten Satz aus der Linearen Algebra beweisen: Symmetrische (selbstadjungierte) Endomorphismen haben nämlich die Eigenschaft, dass das orthogonale Komplement jedes invarianten Unterraums selbst invariant ist: Ist $V \subset \mathbb{R}^n$ ein Unterraum mit $B(V) \subset V$ und ist $V^\perp := \{x \in \mathbb{R}^n; x \perp V\}$, so gilt auch $B(V^\perp) \subset V^\perp$, denn für alle $w \in V^\perp$ und alle $v \in V$ gilt $\langle Bw, v \rangle = \langle w, Bv \rangle = 0$, da $Bv \in V$. Insbesondere ist also für einen Eigenvektor x_1 die Hyperebene $H_1 = x_1^\perp$ invariant unter B . Wir können daher B als einen symmetrischen Endomorphismus auf dem Vektorraum $H_1 \cong \mathbb{R}^{n-1}$ auffassen und mit demselben Argument wie eben einen Eigenvektor x_2 von B in H_1 finden. Dann ist auch $H_2 = \{x_1, x_2\}^\perp \cong \mathbb{R}^{n-2}$ invariant unter B , und wir finden einen Eigenvektor $x_3 \in H_2$, usw. So erhalten wir per Induktion n zueinander senkrechte Eigenvektoren x_1, \dots, x_n , die wir auf Einheitslänge normieren können, so dass sie eine Orthonormalbasis aus Eigenvektoren bilden.

Eine kleine formale Schwierigkeit bei diesem Beweis besteht allerdings darin, dass wir die Analysis nur für Funktionen auf \mathbb{R}^n , nicht aber für Funktionen auf einem linearen Teilraum (wie H_1) oder gar

einem abstrakten endlich dimensionalen \mathbb{R} -Vektorraum mit Skalarprodukt entwickelt haben. Dies ist jedoch ohne Mühe möglich; durch Wahl einer Orthonormalbasis wird alles auf die Theorie im \mathbb{R}^n zurückgeführt.

Alternativ kann man die Einschränkung von $f(x) = \langle Bx, x \rangle$ auf die Teilräume H_i als weitere Nebenbedingungen behandeln: Im zweiten Schritt (Einschränkung auf $H_1 = x_1^\perp$) suchen wir also das Minimum von f unter den zwei Nebenbedingungen

$$\begin{aligned} g_1(x) &= \langle x, x_1 \rangle = 0, \\ g_2(x) &= \langle x, x \rangle = 1, \end{aligned}$$

mit $\nabla g_1(x) = x_1$, $\nabla g_2(x) = 2x$. Nach Satz 16 gilt in einem lokalen Minimum x_2 von $f|_{H_1 \cap S}$ die Bedingung $\nabla f(x_2) = \lambda_1 \nabla g_1(x_2) + \lambda_2 \nabla g_2(x_2)$, also

$$2Bx_2 = \lambda_1 x_1 + 2\lambda_2 x_2. \quad (*)$$

Mit $x_2 \in H_1$ ist $Bx_2 \in H_1$. Deshalb darf auch die rechte Seite von $(*)$ keine Komponente in $H_1^\perp = \mathbb{R}x_1$ haben, also muß $\lambda_1 = 0$ gelten. Damit ist x_2 ein zweiter Eigenvektor, senkrecht zu x_1 , usw.

17. Exkurs über Polynome

Unser nächstes Ziel ist die Übertragung des Satzes von Taylor auf Funktionen mehrerer Veränderlicher. Der Satz von Taylor in einer Veränderlichen sagt, wie sich mehrfach differenzierbare Funktionen durch Polynome approximieren lassen, und dasselbe sagt auch der Satz von Taylor in n Veränderlichen. Dazu müssen wir zunächst Polynome in n Variablen definieren. Ein *Polynom* $p : \mathbb{R}^n \rightarrow \mathbb{R}$ ist eine Abbildung der Form

$$p(h) = a + \sum_i a_i h_i + \sum_{i,j} a_{ij} h_i h_j + \dots + \sum_{i_1, \dots, i_m} a_{i_1 \dots i_m} h_{i_1} \dots h_{i_m}$$

für alle $h = (h_1, \dots, h_n) \in \mathbb{R}^n$, wobei $a, a_i, a_{ij}, \dots, a_{i_1 \dots i_m} \in \mathbb{R}$ die *Koeffizienten* des Polynoms sind. Die höchste vorkommende Anzahl von Faktoren der Unbestimmten h_i (hier m) heißt der *Grad* des Polynoms. Hat jeder Summand die gleiche Anzahl von Faktoren, etwa q , also

$$p(h) = \sum_{i_1, \dots, i_q} a_{i_1 \dots i_q} h_{i_1} \dots h_{i_q}.$$

so heißt p *homogenes Polynom vom Grad q* oder *q -Form*. Eine Null-Form ist eine Konstante. Eine q -Form für $q = 1, 2, 3, 4$ heißt auch *Linearform*, *quadratische Form*, *kubische Form*, *quartische Form*; Linearformen sind dasselbe wie \mathbb{R} -wertige lineare Abbildungen. Jedes Polynom p vom Grad m schreibt sich also als $p = \sum_{q=0}^m p_q$, wobei p_q eine q -Form

ist. Demnach ist ein Polynom p genau dann homogen vom Grad q , wenn für alle $\lambda \in \mathbb{R}$ und für alle $h \in \mathbb{R}^n$ gilt:

$$p(\lambda h) = \lambda^q p(h).$$

Differenziert man übrigens diese Gleichung nach λ bei $\lambda = 1$, so ergibt sich die *Gleichung von Euler*

$$\langle h, \nabla p(h) \rangle = q \cdot p(h).$$

Bemerkung. Die homogenen Polynome vom Grad q über \mathbb{R}^n bilden offensichtlich einen \mathbb{R} -Vektorraum $P(n, q)$ mit Basis

$$\{p_{i_1 \dots i_q}; 1 \leq i_1 \leq \dots \leq i_q \leq n\}.$$

wobei $p_{i_1 \dots i_q}(h) := h_{i_1} \dots h_{i_q}$. Die Dimension dieses Raumes ist also die Anzahl der (schwach) monoton wachsenden Folgen (i_1, \dots, i_q) mit Werten in $\{1, \dots, n\}$. Wie groß ist deren Anzahl? Wir können jeder solchen Folge (i_1, \dots, i_q) die *streng* monoton wachsende Folge $(j_1 = i_1 + 1, \dots, j_q = i_q + q)$ mit Werten in $\{2, \dots, n + q\}$ zuordnen, und umgekehrt: Für jede streng monoton wachsenden Folge (j_1, \dots, j_q) in $\{2, \dots, n + q\}$ ist (i_1, \dots, i_q) mit $i_k = j_k - k$ eine schwach monoton wachsende Folge in $\{1, \dots, n\}$. Es gibt also genauso viele Folgen vom Typ (i_1, \dots, i_q) wie vom Typ (j_1, \dots, j_q) . Die Anzahl der streng monoton wachsenden Folgen (j_1, \dots, j_q) in $\{2, \dots, n + q\}$ ist aber gleich der Anzahl der q -elementigen Teilmengen von $\{2, \dots, n + q\}$ (denn jede Teilmenge läßt sich auf eindeutige Weise nach der Größe ordnen), also ist die gesuchte Anzahl $\binom{n+q-1}{q}$. Wir haben also gesehen:

$$\dim P(n, q) = \binom{n+q-1}{q}.$$

18. Höhere Ableitungen und Satz von Taylor

Es sei $f : U^n \rightarrow \mathbb{R}$ eine m -mal stetig partiell differenzierbare (C^m) Funktion. Wir werden die Taylorformel für diese Funktion von n Veränderlichen aus der für eine Veränderliche folgendermaßen herleiten: Es sei $x \in U^n$ und für ein $h = (h_1, \dots, h_n) \in \mathbb{R}^n$ sei die Strecke $[x, x + h]$ ganz in U^n enthalten, d.h. die Abbildung $g : I \rightarrow \mathbb{R}^n$, $g(t) = x + th$ hat ihre Werte in U^n , wobei I ein offenes Intervall ist, das $[0, 1]$ umfaßt. Dann ist die Funktion $f \circ g : I \rightarrow \mathbb{R}$ nach Kettenregel differenzierbar mit Ableitung

$$(f \circ g)'(t) = Df(g(t))h = \sum_i D_i f(g(t))h_i,$$

also

$$(f \circ g)' = \sum_i D_i f \circ g \cdot h_i. \quad (1)$$

Da $D_i f$ und g wiederum differenzierbar sind, ist es auch $(f \circ g)'$ mit Ableitung $(f \circ g)'' = \sum_i ((D_i f) \circ g)' \cdot h_i$, und da $((D_i f) \circ g)' = \sum_j (D_j (D_i f) \circ g) \cdot h_j$ nach (1) (wobei f durch $D_i f$ zu ersetzen ist), folgt weiter

$$(f \circ g)'' = \sum_{i,j} (D_j D_i f) \circ g \cdot h_j h_i.$$

Durch Induktion erhält man also die q -te Ableitung für alle $q \leq m$:

$$(f \circ g)^{(q)} = \sum_{i_1, \dots, i_q} (D_{i_1} \dots D_{i_q} f) \circ g \cdot h_{i_1} \dots h_{i_q}.$$

Zur Abkürzung setzen wir

$$D^q f_x(h) := \sum_{i_1, \dots, i_q} D_{i_1} \dots D_{i_q} f(x) \cdot h_{i_1} \dots h_{i_q}.$$

$D^0 f_x$ ist also die Konstante $f(x)$, $D^1 f_x$ die Linearform Df_x (die Ableitung) und $D^2 f_x(h) = \sum_{i,j} D_i D_j f(x) h_i h_j$ eine quadratische Form, genannt die *Hessesche Form*. Allgemein ist $D^q f_x$ eine q -Form (s. §17.). Mit dieser Abkürzung gilt also:

$$(f \circ g)^{(q)}(t) = D^q f_{x+th}(h). \quad (2)$$

36. Satz 18. (*Taylorische Formel*) Ist $f : U^n \rightarrow \mathbb{R}$ eine C^m -Funktion und $[x, x+h] \subset U^n$, so gilt:

$$\begin{aligned} f(x+h) &= f(x) + D^1 f_x(h) + \frac{1}{2} D^2 f_x(h) + \dots \\ &\quad + \frac{1}{(m-1)!} D^{m-1} f_x(h) + \frac{1}{m!} D^m f_\xi(h) \\ &= \sum_{q=0}^{m-1} \frac{1}{q!} D^q f_x(h) + \frac{1}{m!} D^m f_\xi(h) \end{aligned}$$

für ein $\xi \in [x, x+h]$.

Beweis. Wie vorher sei $g : I \rightarrow U^n$, $g(t) = x + th$ mit $I \supset [0, 1]$. Nach dem Satz von Taylor (mit Lagrange-Restglied) für Funktionen einer Variable, angewandt auf $f \circ g$, gilt

$$(f \circ g)(1) = \sum_{q=1}^{m-1} \frac{1}{q!} (f \circ g)^{(q)}(0) \cdot 1^q + \frac{1}{m!} (f \circ g)^{(m)}(\tau) \cdot 1^m$$

für ein $\tau \in (0, 1)$. Setzen wir $\xi = g(\tau) = x + \tau h$, so erhalten wir aus (2):

$$f(x+h) = (f \circ g)(1) = \sum_{q=0}^{m-1} \frac{1}{q!} D^q f_x(h) + \frac{1}{m!} D^m f_\xi(h).$$

□

Das Polynom $T_m(h) = \sum_{q=0}^m \frac{1}{q!} D^q f_x(h)$ heißt auch *Taylorpolynom der Ordnung m von f in x* . Eine C^∞ -Funktion f heißt *analytisch*, wenn für jedes $x \in U^n$ ein $\epsilon > 0$ existiert, so dass für jedes feste $h \in \mathbb{R}^n$ mit $|h| < \epsilon$ gilt: $|f(x+h) - T_m(h)| \rightarrow 0$ für $m \rightarrow \infty$. In diesem Fall ist die *Taylorreihe* $\sum_q D^q f_x(h)/q!$ konvergent und gleich $f(x+h)$ für $|h| < \epsilon$.

37. Korollar. *Ist $f : U^n \rightarrow \mathbb{R}$ eine C^m -Funktion und ist $[x, x+h] \subset U^n$, so gilt*

$$f(x+h) = T_m(h) + o(|h|^m),$$

d.h. $\lim_{h \rightarrow 0} (f(x+h) - T_m(h))/|h|^m = 0$.

Beweis. Nach dem vorstehenden Satz ist für ein $\xi \in [x, x+h]$

$$\begin{aligned} f(x+h) &= T_{m-1}(h) + \frac{1}{m!} D^m f_\xi(h) \\ &= T_m(h) + \frac{1}{m!} (D^m f_\xi(h) - D^m f_x(h)). \end{aligned}$$

Also ist

$$\frac{|f(x+h) - T_m(h)|}{|h|^m} \leq \frac{1}{m!} \sum_{i_1 \dots i_m} a_{i_1 \dots i_m} b_{i_1 \dots i_m}$$

mit

$$\begin{aligned} a_{i_1 \dots i_m} &= |D_{i_1} \dots D_{i_m} f(\xi) - D_{i_1} \dots D_{i_m} f(x)|, \\ b_{i_1 \dots i_m} &= \frac{|h_{i_1}|}{|h|} \dots \frac{|h_{i_m}|}{|h|}. \end{aligned}$$

Wegen der Stetigkeit von $D_{i_1} \dots D_{i_m} f$ geht $a_{i_1 \dots i_m}$ gegen Null für $h \rightarrow 0$ während der zweite Faktor $b_{i_1 \dots i_m} \leq 1$ beschränkt bleibt, also folgt die Behauptung. □

Das Korollar läßt sich sofort auf \mathbb{R}^p -wertige Abbildungen

$$f = (f_1, \dots, f_p)$$

übertragen, indem wir es auf die einzelnen Komponenten f_i anwenden. Das Taylorpolynom T_m von f ist dann \mathbb{R}^p -wertig.

Bemerkung. Satz 18 selbst läßt sich in der vorliegenden Form nicht auf \mathbb{R}^p -wertige Abbildungen $f = (f_1, \dots, f_p)$ übertragen, da das $\xi \in [x, x+h]$ im Lagrange-Restglied nicht für alle Komponenten f_i gleich gewählt werden kann. Der Satz hätte sich aber übertragen lassen, wenn wir statt des Lagrangerestgliedes das Taylorrestglied

$$f(x+h) - T_{m-1}(h) = \frac{1}{(m-1)!} \int_0^1 (1-t)^{m-1} D^m f_{x+th}(h) dt$$

gewählt hätten.

19. Lokale Extrema: Bedingungen höherer Ordnung

Wir nennen eine q -Form $p : \mathbb{R}^n \rightarrow \mathbb{R}$ *positiv definit* bzw. *positiv semidefinit*, wenn $p(h) > 0$ bzw. $p(h) \geq 0$ für alle $h \neq 0$. Wir schreiben dafür einfach $p > 0$ bzw. $p \geq 0$. Entsprechend definiert man *negativ (semi-)definit*. Wegen der Homogenitätsgleichung $p(\lambda \cdot h) = \lambda^q p(h)$ können nur Formen von geradem Grad positiv oder negativ semidefinit sein (wenn sie nicht Null sind). Für $q = 2$ und $n = 2$ sind z.B. die quadratischen Formen $p(x, y) = x^2 + y^2$ oder $p(x, y) = x^2 - 2xy + 2y^2$ positiv definit. (Für die letztere Form folgt dies durch quadratische Ergänzung: $x^2 - 2xy + 2y^2 = (x-y)^2 + y^2 > 0$, es sei denn dass $x-y = y = 0$, also $(x, y) = (0, 0)$.) Für die positive Definitheit einer quadratischen Form im \mathbb{R}^n gilt das folgende nützliche *Determinantenkriterium*: $p(h) = \sum_{i,j} a_{ij} h_i h_j$ mit $a_{ij} = a_{ji}$ ist positiv definit genau dann, wenn alle *Hauptminoren* der Matrix (a_{ij}) positiv sind, d.h. für $k = 1, \dots, n$ muß gelten:

$$\det(a_{ij})_{1 \leq i, j \leq k} > 0.$$

38. Satz 19. $f : U^n \rightarrow \mathbb{R}$ sei m -mal stetig differenzierbar (C^m) mit $m \geq 2$ und es sei $x \in U^n$.

- (a) Ist x ein lokales Minimum von f , so gibt es ein $r \in \{2, \dots, m\}$ mit $D^r f_x \geq 0$ und $D^q f_x = 0$ für alle $q = 1, \dots, r-1$.
- (b) Gibt es ein $r \in \{2, \dots, m\}$ mit $D^r f_x > 0$ und $D^q f_x = 0$ für alle $q \in \{1, \dots, r-1\}$, so ist x lokales Minimum von f .

Entsprechendes gilt für lokale Maxima.

Beweis. Zu (b): Nach Voraussetzung gilt $T_r(h) = f(x) + \frac{1}{r!} D^r f_x(h)$ für das Taylorpolynom r -ter Ordnung von f in x . Nach dem Korollar in §18 und der Homogenität von $D^r f_x$ gilt also

$$\frac{f(x+h) - f(x)}{|h|^r} = \frac{1}{r!} D^r f_x \left(\frac{h}{|h|} \right) + \frac{o(|h|^r)}{|h|^r}. \quad (*)$$

Es sei $S = \{h \in \mathbb{R}^n; |h| = 1\}$ die Einheitssphäre. Da S kompakt und $D^r f_x|_S > 0$, nimmt $D^r f_x|_S$ ein Minimum $M > 0$ an. Der erste Term der rechten Seite von (*) ist also für alle h durch $\frac{1}{r!}M > 0$ nach unten beschränkt, während der zweite Term für $h \rightarrow 0$ gegen 0 geht. Ist daher $|h|$ genügend klein, also $0 < |h| < \epsilon$, so ist die rechte Seite von (*) positiv und damit $f(x+h) > f(x)$, d.h. f besitzt in x ein lokales Minimum. Ersetzen wir f durch $-f$, so erhalten wir die entsprechende Aussage für ein lokales Maximum.

Zu (a): Es sei x ein lokales Minimum für f . Dann ist $D^1 f_x = 0$ (vgl. §16). Wenn alle Ableitungen $D^1 f_x, \dots, D^m f_x$ verschwinden, so ist nichts zu zeigen. Es sei also $D^r f$ die niedrigste Ableitung, die in x nicht verschwindet. Es sei $h \in \mathbb{R}^n$ so, dass $D^r f_x(h) \neq 0$. Da f in x ein lokales Minimum besitzt, nimmt die Funktion

$$\phi : (-\epsilon, \epsilon) \rightarrow \mathbb{R}, \quad \phi(t) = f(x + th)$$

bei $t = 0$ ein Minimum an, falls $\epsilon > 0$ genügend klein ist. Nach Kettenregel ist $\phi^{(q)}(0) = D^q f_x(h)$ (vgl. Gleichung (2) in §18); insbesondere ist $\phi^{(q)}(0) = 0$ für $1 \leq q < r$ und $\phi^{(r)}(0) \neq 0$. Wäre $\phi^{(r)}(0) < 0$, so wäre (ähnlich wie in (b))

$$\frac{\phi(t) - \phi(0)}{t^r} = \frac{1}{r!}\phi^{(r)}(0) + \frac{o(t^r)}{t^r} < 0$$

für genügend kleine $t > 0$, und $t = 0$ könnte nicht lokales Minimum von ϕ sein. Also ist $D^r f_x(h) = \phi^{(r)}(0) > 0$, was zu zeigen war. \square

Am häufigsten wird der Fall $r = 2$ des vorstehenden Satzes angewendet. Um lokale Extrema einer Funktion $f : U^n \rightarrow \mathbb{R}$ zu entdecken, müssen also zunächst die Nullstellen von ∇f bestimmt werden. An diesen Stellen x wird sodann die Hessesche $D^2 f_x$ berechnet. Ist $D^2 f_x$ (bzw. $-D^2 f_x$) positiv definit, so liegt ein lokales Minimum (bzw. Maximum) vor. Faustregel: Wenn es nach jeder Seite hin abwärts geht (Hesseform negativ definit), ist man auf einem Gipfel (Maximum). Ist die *Hessematrix* $D^2 f(x) := (D_i D_j f(x))$ *indefinit*, d.h. weder positiv noch negativ definit, aber nicht ausgeartet, also $\det(D^2 f(x)) \neq 0$, so liegt ein *Sattelpunkt* vor, d.h. in manchen Richtungen steigt die Funktion an, in anderen fällt sie ab.

Beispiel: $f : \mathbb{R}^2 \rightarrow \mathbb{R}$, $f(x, y) = x^4 - x^2 + y^2$. Dann ist

$$\nabla f(x, y) = \begin{pmatrix} 4x^3 - 2x \\ 2y \end{pmatrix}, \quad D^2 f(x, y) = \begin{pmatrix} 12x^2 - 2 & 0 \\ 0 & 2 \end{pmatrix}.$$

Also ist $\nabla f(x, y) = 0$ genau dann, wenn $y = 0$ und $x(2x^2 - 1) = 0$; die Nullstellen liegen also bei $(0, 0)$ und $(\pm 1/\sqrt{2}, 0)$, und dort erhalten wir

$$D^2 f(0, 0) = \begin{pmatrix} -2 & 0 \\ 0 & 2 \end{pmatrix}, \quad D^2 f(\pm 1/\sqrt{2}, 0) = \begin{pmatrix} 4 & 0 \\ 0 & 2 \end{pmatrix}.$$

Bei $(\pm 1/\sqrt{2}, 0)$ liegen also lokale Minima und bei $(0, 0)$ ein Sattel, und $f(\pm 1/\sqrt{2}, 0) = -1/4$. Dieser Wert ist gleichzeitig das absolute Minimum von f , denn er ist das absolute Minimum von $x \mapsto x^4 - x^2$ (warum?).

Bemerkung. Wir können uns in diesem Fall leicht das Niveaulinienbild und den Graphen von f klarmachen: Die Niveaulinie $f^{-1}(c)$ für beliebiges $c \in \mathbb{R}$ ist die Menge der (x, y) mit $y^2 = c - x^4 + x^2$, also der Graph der Funktionen $\pm \sqrt{u_c(x)}$ mit $u_c(x) = c - x^4 + x^2$. Die Quadratwurzel ist nur dort definiert, wo $u_c(x) \geq 0$ ist. Für $c < -1/4$ ist das nirgends der Fall, für $c \in (-1/4, 0)$ zerfällt die Menge dieser x in zwei getrennte Intervalle, und nur für $c > 0$ ist sie ein zusammenhängendes Intervall. Für $c > 0$ besteht $f^{-1}(c)$ daher aus einer einzigen geschlossenen Kurve, für $c \in (-1/4, 0)$ dagegen aus zwei getrennten geschlossenen Kurven um die beiden lokalen Minima $(\pm 1/\sqrt{2}, 0)$, die bei $c = 0$ zu der 8-förmigen Kurve $f^{-1}(0)$ mit der Gleichung $y^2 = x^2 - x^4$, einer sogenannten *Lemniskate*, zusammenwachsen (Abb. 27).

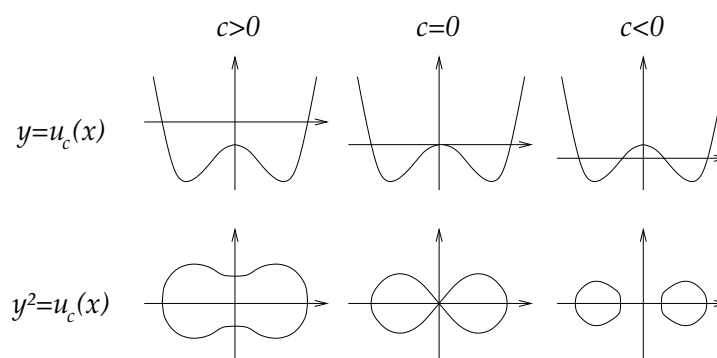


ABBILDUNG 27

Obwohl die Niveaumenge $f^{-1}(0)$ keine Mannigfaltigkeit ist (wegen des Kreuzungspunktes in $(0, 0)$), läßt sie sich durch eine reguläre Kurve parametrisieren, nämlich durch $t \mapsto (x(t), y(t))$ mit

$$x(t) = \sin t, \quad y(t) = \frac{1}{2} \sin 2t$$

(Abb. 28). In der Tat gilt ja $\sin 2t = 2 \sin t \cos t$, also $y(t)^2 = \sin^2 t \cos^2 t = \sin^2 t(1 - \sin^2 t) = x(t)^2 - x(t)^4$. Die Kurven

$$t \mapsto (a \cdot \sin(k(t - t_1)), b \cdot \sin(l(t - t_2)))$$

für teilerfremde ganze Zahlen k, l und $a, b, t_1, t_2 \in \mathbb{R}$ bezeichnet man als *Lissajou'sche Figuren*.

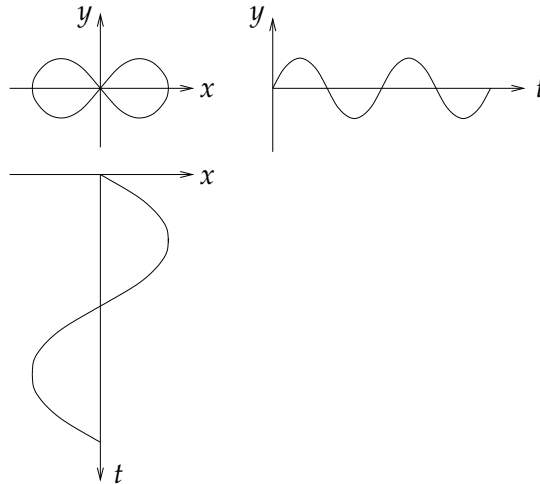


ABBILDUNG 28

20. Koordinaten

Diffeomorphismen spielen in der Analysis eine besondere Rolle: Mit ihrer Hilfe lassen sich komplizierte Situationen auf einfache zurückführen. Dies haben wir bereits an einigen Beispielen gesehen: Eine C^1 -Abbildung $f : U^{n+p} \rightarrow \mathbb{R}^p$ mit surjektiver Ableitung $Df_x : \mathbb{R}^{n+p} \rightarrow \mathbb{R}^p$ ist nahe x “bis auf Diffeomorphismen” eine Projektion (Satz. 14.1). Mannigfaltigkeiten sind nach Definition lokal affine Mengen “bis auf Diffeomorphismen”. Eine C^∞ -Funktion $f : U^n \rightarrow \mathbb{R}$ ist nahe einem kritischen Punkt (Nullstelle von ∇f) mit nicht-ausgearteter Hesseform “bis auf Diffeomorphismen” eine quadratische Form (vgl. [Milnor]) u.a.m.

Ein Diffeomorphismus $\Phi : U^n \rightarrow V^n \subset \mathbb{R}^n$ läßt sich in seine Komponenten zerlegen: $\Phi = (\Phi_1, \dots, \Phi_n)$; die n Komponentenfunktionen $\Phi_i : U^n \rightarrow \mathbb{R}$ heißen auch *Koordinaten*, denn jeder Punkt $x \in U^n$ ist durch Angabe der n reellen Zahlen $\Phi_1(x), \dots, \Phi_n(x)$ eindeutig bestimmt, ebenso gut wie durch die üblichen kartesischen Koordinaten x_1, \dots, x_n (Abb. 29).

Daher wird ein Diffeomorphismus $\Phi : U^n \rightarrow V^n$ auch als *Koordinatensystem* von U^n bezeichnet, die Umkehrabbildung $\Phi^{-1} : V^n \rightarrow U^n$ heißt

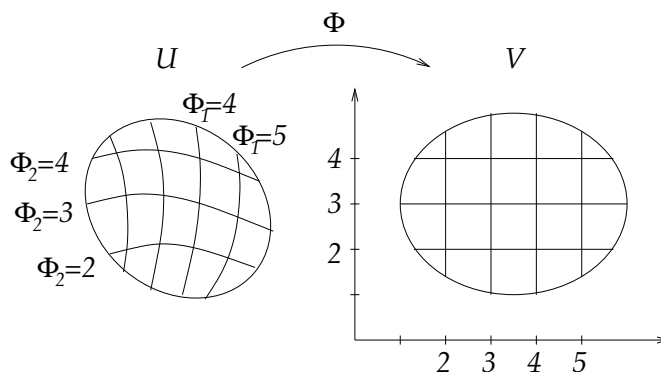


ABBILDUNG 29

Parametrisierung von U^n . Beispiele von (nicht-kartesischen) Koordinatensystemen haben wir bereits in der Linearen Algebra kennengelernt, nämlich invertierbare lineare Abbildungen (Basiswechsel) oder etwas allgemeiner invertierbare affine Abbildungen $\Phi(x) = Ax + b$ für eine invertierbare Matrix $A \in \mathbb{R}^{n \times n}$ und einen beliebigen Vektor $b \in \mathbb{R}^n$. Für die Analysis interessanter sind nicht affine (“krummlinige”) Koordinatensysteme. (Die Bezeichnung “krummlinig” ist mißverständlich, denn es gibt nicht-affine Koordinatensysteme mit gradlinigen Koordinatenlinien, z.B. eine perspektivische Abbildung einer Ebene.)

Wir wollen einige gebräuchliche Koordinatensysteme in Dimension $n = 2$ und $n = 3$ diskutieren. Sie haben leider einen Schönheitsfehler: Angegeben wird nicht Φ , sondern die Umkehrung $\Psi = \Phi^{-1}$, also die Parametrisierung, aber diese ist nur ein *lokaler* Diffeomorphismus, also global gar nicht umkehrbar, solange nicht Bild und Urbild auf die richtigen offenen Teilmengen zurechtgestutzt worden sind. Dazu wollen wir kurz über lokale Diffeomorphismen reden.

Exkurs über lokale Diffeomorphismen

Eine C^1 -Abbildung $f : U^n \rightarrow \mathbb{R}^n$ heißt *lokaler Diffeomorphismus*, wenn Df_x für alle $x \in U^n$ invertierbar ist. Der Umkehrsatz (§13) sagt, dass es dann zu jedem $x \in U^n$ offene Umgebungen U_x^n von x und V_x^n von $f(x)$ gibt, so dass $f : U_x^n \rightarrow V_x^n$ ein Diffeomorphismus ist. Daraus folgt aber nicht, dass f insgesamt injektiv ist, denn weit entfernte Punkte können immer noch dasselbe Bild haben. Aber ein lokaler Diffeomorphismus ist immerhin eine *offene* Abbildung, d.h. die Bilder offener Mengen sind offen. Ist nämlich $A^n \subset U^n$ eine offene Menge, so ist $f(A^n)$ offen, denn mit jedem Punkt $f(x)$ enthält $f(A^n)$ auch die eben erwähnte Umgebung $V_x^n = f(U_x^n)$, wobei wir U_x^n in A^n zu wählen haben (was keine Einschränkung ist, da auch A^n ja eine offene Umgebung von

x ist; wir können also einfach mit A^n schneiden.) Ist noch zusätzlich f auf A^n injektiv, so ist $f : A^n \rightarrow B^n = f(A^n)$ ein Diffeomorphismus: Da $f : A^n \rightarrow B^n$ bijektiv ist, gibt es eine (eindeutige) Umkehrabbildung, die mit den lokalen Umkehrabbildungen auf $V_x^n = f(U_x^n)$ (für $x \in A^n$) übereinstimmen muß und somit auch C^1 ist. Wir haben also gezeigt:

39. Satz. *Ist ein lokaler Diffeomorphismus $f : U^n \rightarrow \mathbb{R}^n$ auf einer offenen Teilmenge $A^n \subset U^n$ injektiv, so ist $B^n = f(A^n)$ offen und $f|_{A^n} : A^n \rightarrow B^n$ ist ein Diffeomorphismus.*

(Ende des Exkurses)

Beispiel 1: Polarkoordinaten in \mathbb{R}^2

Wir betrachten die Abbildung $\Psi : [0, \infty) \times \mathbb{R} \rightarrow \mathbb{R}^2 = \mathbb{C}$,

$$\Psi(r, \phi) = \begin{pmatrix} r \cos \phi \\ r \sin \phi \end{pmatrix} = r \cdot e^{i\phi}$$

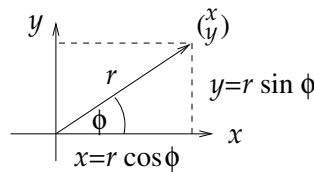


ABBILDUNG 30

(Abb. 30). Ψ ist lokaler Diffeomorphismus auf $(0, \infty) \times \mathbb{R} \subset \mathbb{R}^2$, denn

$$D\Psi(r, \phi) = \begin{pmatrix} \cos \phi & -r \sin \phi \\ \sin \phi & r \cos \phi \end{pmatrix}$$

ist invertierbar, da

$$\det D\Psi(r, \phi) = r(\cos^2 \phi + \sin^2 \phi) = r \neq 0.$$

Ψ ist aber nicht injektiv, da $\Psi(r, \phi) = \Psi(r, \phi + 2\pi)$. Schränken wir aber Ψ ein auf die offene Teilmenge

$$V_\alpha^2 = (0, \infty) \times (\alpha - \pi, \alpha + \pi)$$

für beliebiges $\alpha \in \mathbb{R}$, so wird Ψ injektiv und nach dem vorstehenden Satz also ein Diffeomorphismus auf die "geschlitzte Ebene"

$$U_\alpha^2 = \Psi(V_\alpha^2) = \mathbb{R}^2 \setminus \{t \cdot (\cos \alpha, \sin \alpha)^T; t \leq 0\}.$$

Die Einschränkung spielt keine große Rolle, da der durch α bestimmte Schlitz ja stets passend gewählt werden kann. Nun ist die Umkehrabbildung

$$\Psi^{-1} = (r, \phi) : U_\alpha^2 \rightarrow V_\alpha^2$$

ein Koordinatensystem von U_α^2 . Die Koordinaten r und ϕ heißen *Polarkoordinaten* (Abb. 31).

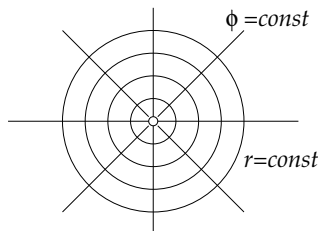


ABBILDUNG 31

Beispiel 2: Zylinderkoordinaten in \mathbb{R}^3

Die Zylinderkoordinaten in $\mathbb{R}^3 = \mathbb{R}^2 \times \mathbb{R}$ sind die Erweiterung der Polarkoordinaten der Ebene durch die z -Koordinate. Man geht wieder aus von der Umkehrabbildung

$$\Psi = (x, y, z)^T : (0, \infty) \times \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}^3 \setminus \{z\text{-Achse}\},$$

wobei die drei Variablen jetzt ρ, ϕ, z heißen, mit

$$x = \rho \cos \phi, \quad y = \rho \sin \phi, \quad z = z$$

(Abb. 32). Die Einschränkung auf $V_\alpha^2 \times \mathbb{R}$ liefert einen Diffeomor-

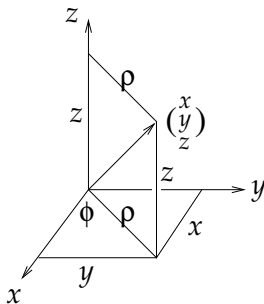


ABBILDUNG 32

phismus auf $U_\alpha^2 \times \mathbb{R}$; die Umkehrabbildung $\Psi^{-1} = (\rho, \phi, z)$ ist also ein Koordinatensystem auf $U_\alpha^2 \times \mathbb{R} \subset \mathbb{R}^3$. Diese Koordinaten heißen *Zylinderkoordinaten*, da die Flächen $\{\rho = \text{const}\}$ Zylinder um die z -Achse sind.

Beispiel 3: Kugelkoordinaten in \mathbb{R}^3

Diese sind gegeben durch die Abbildung

$$\Psi = (x, y, z)^T : (0, \infty) \times (0, \pi) \times \mathbb{R} \rightarrow \mathbb{R}^3 \setminus \{z\text{-Achse}\}$$

mit den Variablen r, θ, ϕ , wobei

$$\begin{aligned} x &= r \sin \theta \cos \phi, \\ y &= r \sin \theta \sin \phi, \\ z &= r \cos \theta \end{aligned}$$

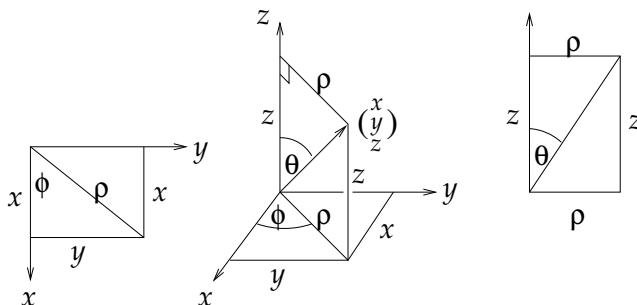


ABBILDUNG 33

(Abb. 33). Somit ist

$$D\Psi(r, \theta, \phi) = \begin{pmatrix} \sin \theta \cos \phi & r \cos \theta \cos \phi & -r \sin \theta \sin \phi \\ \sin \theta \sin \phi & r \cos \theta \sin \phi & r \sin \theta \cos \phi \\ \cos \theta & -r \sin \theta & 0 \end{pmatrix}$$

Die drei partiellen Ableitungen (die Spalten dieser Matrix) stehen senkrecht aufeinander (nachrechnen!), also ist der Betrag der Determinante das Produkt ihrer Längen, somit $|\det D\Psi| = r^2 \sin \theta \neq 0$. Also ist Ψ ein lokaler Diffeomorphismus, und seine Einschränkung auf $(0, \infty) \times (0, \pi) \times (\alpha - \pi, \alpha + \pi)$ für beliebiges $\alpha \in \mathbb{R}$ ist Diffeomorphismus. Die Umkehrung $\Psi^{-1} = (r, \theta, \phi)$ ist das *Kugelkoordinatensystem*, so genannt, weil die Flächen $\{r = \text{const}\}$ Kugelflächen (Sphären) sind.

III. Integration

21. Flächeninhalt, Volumen, Maß

Der Begriff Volumen (Rauminhalt) oder sein 2-dimensionales Analog, der Flächeninhalt, ist vielleicht der komplizierteste mathematische Begriff unserer Alltagssprache. Man versuche einmal zu erklären, was das Wort “ein Liter” bedeutet! Das Erstaunliche an diesem Begriff ist, dass hier die Größe (Quantität) eines Raumstücks unabhängig von seiner Form erklärt wird. Wir sind damit nur deshalb so vertraut, weil wir mit inkompressiblen Flüssigkeiten umzugehen gewohnt sind. “Inkompressibel” heißt, dass jeder noch so kleine Teil der Flüssigkeit sein Volumen beibehält. Wenn wir die Form der Flüssigkeit verändern, denken wir sie uns aus beliebig kleinen Teilen zusammengesetzt, die zwar umgeschichtet werden, aber ihre Größe dabei nicht verändern. Der Begriff des Volumens gehört also sozusagen von Geburt an in die Lehre von der Zusammensetzung des beliebig Kleinen, die *Infinitesimalrechnung*. Das wird noch deutlicher, wenn wir uns erinnern, wie wir den Flächeninhalt einer krummlinig begrenzten Figur definieren: Entweder nähern wir sie durch einbeschriebene Quadrate oder Rechtecke an, die zum Rand hin immer kleiner werden, um die Figur möglichst genau auszuschöpfen; wir müssen also den Flächeninhalt von schließlich unendlich vielen immer kleiner werdenden Quadraten oder Rechtecken aufaddieren (Abb. 34).

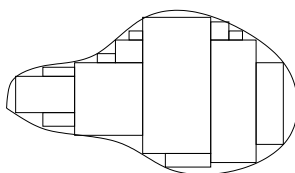


ABBILDUNG 34

Oder wir überdecken sie durch Rechtecke, die ebenfalls am Rand immer kleiner und immer mehr werden, um eine möglichst genaue Überdeckung zu erzielen (Abb. 35), und addieren deren Flächeninhalte auf.

Volumen- und Flächenberechnungen gehörten von Beginn an zu den Aufgaben der Mathematik (Feldeinteilung nach Nilschwemmen im alten Ägypten). Viele Einzelergebnisse (Volumina spezieller Körper wie Kugel, Kegel, Polyeder) waren in die Antike bekannt, besonders Archimedes ist hier zu erwähnen. Dennoch gibt es eine umfassende Theorie

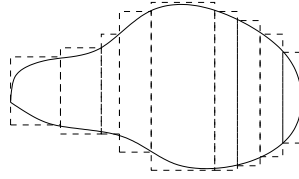


ABBILDUNG 35

erst seit dem 19. Jahrhundert (Riemann, Lebesgue). Volumenberechnung (in beliebigen Dimensionen) ist im wesentlichen dasselbe wie *Integration*: Man kann Integration einer Funktion von n Veränderlichen entweder als eine Art gewichtetes n -dimensionales Volumen ansehen, wobei das Volumen jedes kleinen Raumstückchens mit dem dortigen Wert der Funktion multipliziert wird, oder als Volumen (mit Vorzeichen) in der Dimension $n + 1$ (Integral = Fläche unter der Kurve). Wir wollen zunächst den Riemannsches Integralbegriff kennenlernen, da er der anschaulichen Vorstellung am besten entspricht. Die spätere Lebesgue'sche Erweiterung dieses Integralbegriffs wird noch einmal einen weiteren Schritt ins Unendliche erforderlich machen.

Wie nicht anders zu erwarten, ist der Ausgangspunkt das Volumen von Quadern bzw. der Flächeninhalt von Rechtecken. Ein (*achsenparalleler*) *Quader* $Q \subset \mathbb{R}^n$ ist ein Produkt von n kompakten Intervallen

$$Q = I_1 \times \dots \times I_n$$

mit $I_k = [a_k, b_k]$. Das *Maß*³ von Q ist das Produkt der Intervall-Längen:

$$\mu(Q) = (b_1 - a_1) \cdot \dots \cdot (b_n - a_n).$$

kompakte Teilintervalle $I'_k \subset I_k$ für $k = 1, \dots, n$ definieren *Teilquader* $I'_1 \times \dots \times I'_n$. Eine *Zerlegung* von Q ist eine endliche Menge Z von Teilquadern $A \subset Q$ mit folgenden Eigenschaften:

- (a) $A^\circ \cap B^\circ = \emptyset$ für alle $A, B \in Z$,
- (b) $\bigcup_{A \in Z} A = Q$,

wobei $A^\circ = A \setminus \partial A$ das *Innere* von A bezeichnet: Ist $A = [c_1, d_1] \times \dots \times [c_n, d_n]$, so ist $A^\circ = (c_1, d_1) \times \dots \times (c_n, d_n)$. Mit $\bigcup_{A \in Z} A$ bezeichnen wir die Vereinigung aller Teilquader der Zerlegung Z :

$$\bigcup_{A \in Z} A = \{x \in \mathbb{R}^n; \exists_{A \in Z} x \in A\}.$$

Eine andere Zerlegung Z' von Q heißt *Verfeinerung* der Zerlegung Z , wenn sich Z' disjunkt in Teilmengen Z'_A , $A \in Z$, zerlegen läßt,

³1-dimensional: *Länge*, 2-dimensional: *Flächeninhalt*, 3-dimensional: *Volumen*)

so dass Z'_A eine Zerlegung von A ist, für alle $A \in Z$. Im Fall $n = 1$ ist eine Zerlegung Z eines Intervalls $I = [a, b]$ durch eine Folge von Unterteilungspunkten

$$a = t_0 < t_1 < \dots < t_N = b$$

gegeben; die zugehörige Zerlegung ist $Z = \{[t_{i-1}, t_i], i = 1, \dots, N\}$.

Die wichtigste Eigenschaft von Zerlegungen ist die Erhaltung des Maßes:

40. Lemma 21. *Ist Z eine Zerlegung eines Quaders $Q \in \mathbb{R}^n$, so gilt*

$$\mu(Q) = \sum_{A \in Z} \mu(A).$$

Beweis. Der Satz ist klar in Dimension 1: Dann ist $Q = I = [a, b]$ und $Z = \{[t_{i-1}, t_i], i = 1, \dots, N\}$, somit

$$\sum_{A \in Z} \mu(A) = \sum_{i=1}^N (t_i - t_{i-1}) = t_N - t_0 = b - a = \mu(Q).$$

Der Satz ist auch in Dimension n offensichtlich, wenn Z eine *Produktzerlegung* oder *kartesische Zerlegung* von $Q = I_1 \times \dots \times I_n$ ist, d.h. es gibt Zerlegungen Z_k der einzelnen Intervalle I_k , $k = 1, \dots, n$, so dass

$$Z = \{J_1 \times \dots \times J_n; J_k \in Z_k, k = 1, \dots, n\}.$$

Auf Grund des eindimensionalen Falles ist nämlich

$$\sum_{A \in Z} \mu(A) = \sum_{J_1 \in Z_1} \dots \sum_{J_n \in Z_n} \mu(J_1) \cdot \dots \cdot \mu(J_n) = \mu(I_1) \dots \mu(I_n) = \mu(Q).$$

Jede beliebige Zerlegung Z eines Quaders $Q = I_1 \times \dots \times I_n$ läßt sich aber zu einer Produktzerlegung Z' verfeinern, indem wir als Unterteilungspunkte des Intervalls I_k die k -ten Koordinaten der Eckpunkte aller $A \in Z$ definieren (Abb. 36).

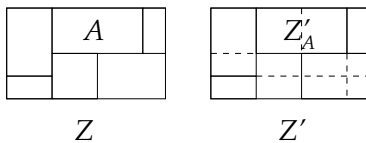


ABBILDUNG 36

“Verfeinerung” bedeutete, dass sich Z' in disjunkte Teilmengen Z'_A unterteilen lassen, die für jedes $A \in Z$ Zerlegungen von A sind, und zwar

in diesem Fall wieder Produktzerlegungen. Da der Satz für Produktzerlegungen bereits bewiesen ist, erhalten wir einerseits

$$\mu(Q) = \sum_{A' \in Z'} \mu(A') = \sum_{A \in Z} \sum_{A' \in Z'_A} \mu(A')$$

und andererseits für alle $A \in Z$

$$\mu(A) = \sum_{A' \in Z'_A} \mu(A').$$

Durch Einsetzen der zweiten Gleichung in die erste folgt die Behauptung. \square

22. Integrierbarkeit

Nun sei eine beschränkte Funktion $f : Q \rightarrow \mathbb{R}$ auf einem Quader $Q \subset \mathbb{R}^n$ gegeben. "Beschränkt" bedeutet, dass es eine Konstante C gibt mit

$$f(x) \leq C \quad \forall x \in Q.$$

Für jede Zerlegung Z von Q definieren wir die *Obersumme* und die *Untersumme* von f bezüglich der Zerlegung Z folgendermaßen:

$$\begin{aligned} \Sigma^* f &:= \sum_{A \in Z} \sup_A f \cdot \mu(A), \\ \Sigma_* f &:= \sum_{A \in Z} \inf_A f \cdot \mu(A), \end{aligned}$$

wobei $\sup_A f$ bzw. $\inf_A f$ das Supremum bzw. Infimum der Menge $f(A) = \{f(x); x \in A\} \subset \mathbb{R}$ bezeichnet (Abb. 37).

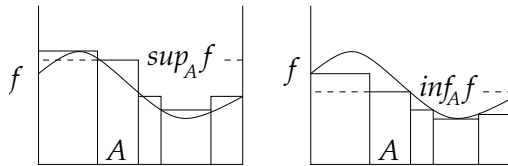


ABBILDUNG 37

41. Lemma 22.1 *Ist Z' eine Verfeinerung der Zerlegung Z von Q , so gilt*

$$\begin{aligned} \Sigma_{*Z} f &= \sum_{A \in Z} \inf_A f \cdot \mu(A) \leq \sum_{A' \in Z'} \inf_{A'} f \cdot \mu(A') = \Sigma_{*Z'} f, \\ \Sigma_Z^* f &= \sum_{A \in Z} \sup_A f \cdot \mu(A) \geq \sum_{A' \in Z'} \sup_{A'} f \cdot \mu(A') = \Sigma_Z^* f. \end{aligned}$$

Beweis. Dies folgt aus Lemma 21.: Ist $A' \subset A$, so ist $\sup_{A'} f \leq \sup_A f$ und daher

$$\begin{aligned} \sum_{A' \in \mathfrak{Z}'} \sup_{A'} f \cdot \mu(A') &= \sum_{A \in \mathfrak{Z}} \sum_{A' \in \mathfrak{Z}'_A} \sup_{A'} f \cdot \mu(A') \\ &\leq \sum_{A \in \mathfrak{Z}} \sum_{A' \in \mathfrak{Z}'_A} \sup_A f \cdot \mu(A') \\ &= \sum_{A \in \mathfrak{Z}} \sup_A f \cdot \mu(A), \end{aligned}$$

und die andere Ungleichung folgt entsprechend. \square

Je feiner also die Unterteilung wird, desto kleiner wird die Ober-
summe und desto größer die Untersumme, und diese Zahlen rücken
daher mit zunehmender Feinheit immer näher aneinander. Wir gehen
deshalb zum Infimum der Obersummen und zum Supremum der Un-
tersummen über alle Zerlegungen über und erhalten das *Oberintegral*
und das *Unterintegral* von f :

$$\begin{aligned} \int^* f &:= \inf_{Z \in \mathfrak{Z}(Q)} \Sigma_Z^* f = \inf_{Z \in \mathfrak{Z}(Q)} \sum_{A \in Z} \sup_A f \cdot \mu(A), \\ \int_* f &:= \sup_{Z \in \mathfrak{Z}(Q)} \Sigma_Z f = \sup_{Z \in \mathfrak{Z}(Q)} \sum_{A \in Z} \inf_A f \cdot \mu(A), \end{aligned}$$

sobei $\mathfrak{Z}(Q)$ die Menge aller Zerlegungen von Q bezeichnet. Wenn wir
den Bereich angeben wollen, schreiben wir $\int_Q^* f$ und $\int_{*Q} f$ statt $\int^* f$
und $\int_* f$.

Ein wichtiger Sonderfall sind solche Funktionen, die nur die Werte
0 und 1 annehmen: Ist $T \subset Q$ eine beliebige Teilmenge, so heißt die
Funktion $1_T : Q \rightarrow \mathbb{R}$ mit

$$1_T(x) = 1 \Leftrightarrow x \in T, \quad 1_T(x) = 0 \Leftrightarrow x \notin T,$$

die *charakteristische Funktion* oder *Indikatorfunktion* von T (sie wird
oft auch mit χ_T bezeichnet). Wir nennen $\mu^*(T) := \int^* 1_T$ das *äußere*
Maß und $\mu_*(T) = \int_* 1_T$ das *innere Maß* der Teilmenge T . Da die Funk-
tion 1_T nur die Werte 0 und 1 annimmt, gilt dasselbe für $\sup_A 1_T$ und
 $\inf_A 1_T$, und zwar

$$\begin{aligned} \sup_A 1_T = 1 &\Leftrightarrow A \cap T \neq \emptyset, \\ \inf_A 1_T = 1 &\Leftrightarrow A \subset T. \end{aligned}$$

Deshalb vereinfacht sich die Definition des Ober- und Unterintegrals
von 1_T folgendermaßen:

42. Lemma 22.2 Für jede Teilmenge T eines Quaders $Q \subset \mathbb{R}^n$ gilt:

$$\begin{aligned}\mu^*(T) &= \int^* 1_T = \inf_{Z \in \mathfrak{Z}(Q)} \sum_{A \in Z_+(T)} \mu(A), \\ \mu_*(T) &= \int_* 1_T = \sup_{Z \in \mathfrak{Z}(Q)} \sum_{A \in Z_-(T)} \mu(A),\end{aligned}$$

wobei wir für jede Zerlegung $Z \in \mathfrak{Z}(Q)$ setzen:

$$\begin{aligned}Z_+(T) &= \{A \in Z; A \cap T \neq \emptyset\}, \\ Z_-(T) &= \{A \in Z; A \subset T\}.\end{aligned}\quad \square$$

Wir erhalten also für das innere und äußere Maß genau die Definition, die wir oben anschaulich skizziert haben (vgl. Abb. 34 und 35).

Eine beschränkte Funktion $f : Q \rightarrow \mathbb{R}$ auf einem Quader $Q \subset \mathbb{R}^n$ heißt *(Riemann)-integrierbar*, wenn

$$\int_* f = \int^* f =: \int f$$

(*Riemann-Integral*). Da wir zu je zwei Zerlegungen von Q eine gemeinsame Verfeinerung finden, können wir die Integrierbarkeit mit Hilfe von Lemma 22.1 (“Verfeinerung verbessert den Wert der Ober- und Unter-summe”) auch so ausdrücken: Zu jedem $\epsilon > 0$ gibt es eine Zerlegung $Z \in \mathfrak{Z}(Q)$ mit

$$\sum_{A \in Z} (\sup_A f - \inf_A f) \mu(A) < \epsilon. \quad (1)$$

Wenn die charakteristische Funktion einer Teilmenge $T \subset Q$ integrierbar ist, wenn also inneres und äußeres Maß übereinstimmen, $\mu_*(T) = \mu^*(T) =: \mu(T)$, so heißt die Menge T *messbar (im Sinne von Riemann oder Jordan)*, und $\mu(T)$ heißt das (*Jordan-*)*Maß* von T . Nach (1) ist T also genau dann messbar, wenn

$$\sum_{A \in Z_+(T) \setminus Z_-(T)} \mu(A) < \epsilon \quad (2)$$

für eine geeignete Zerlegung $Z \in \mathfrak{Z}(Q)$, wobei $Z_+(T)$ wie früher definiert die $A \in Z$ enthält, die T schneiden und $Z_-(T)$ diejenigen A , die ganz in T enthalten sind.

43. Lemma 22.3 Für beschränkte Funktionen $f, g : Q \rightarrow \mathbb{R}$ und Konstanten $\alpha > 0$ gilt:

- 0.) $\int^*(-f) = -\int_* f$
- 1.) $\int^*(\alpha f) = \alpha \int^* f$,
- 2.) $\int^*(f + g) \leq \int^* f + \int^* g$,

- 3.) $f \leq g \Rightarrow \int^* f \leq \int^* g$,
 4.) $\forall Z \in \mathfrak{Z}(Q) \int_Q^* f = \sum_{A \in Z} \int_A^* f$.

Entsprechendes gilt für das Unterintegral $\int_* f$, wobei in 2. die Ungleichung umgedreht wird, d.h. \leq wird durch \geq ersetzt.

Beweis. 0.) gilt, weil die Multiplikation mit -1 alle Ungleichungszeichen umdreht. 1.) ist klar: Wenn alle Werte von f mit $\alpha > 0$ multipliziert werden, dann auch das Supremum. 2.) folgt wegen

$$\sup_A (f + g) \leq \sup_A f + \sup_A g.$$

3.) ist klar: Je größer die Werte, desto größer ihr Supremum.

Um 4.) zu zeigen, sei eine Zerlegung Z von Q gegeben. Für jedes $A \in Z$ und jede Zerlegung Z_A von A gilt dann

$$\begin{aligned} \sum_{A \in Z} \Sigma_{Z_A}^* f &= \sum_{A \in Z} \sum_{B \in Z_A} \sup_B f \mu(B) \\ &= \sum_{B \in Z'} \sup_B f \mu(B) \\ &= \Sigma_{Z'}^* f \geq \int_Q^* f, \end{aligned}$$

wobei $Z' = \bigcup_{A \in Z} Z_A$ die Vereinigung aller Teilzerlegungen Z_A ist; dies ist eine Zerlegung von Q , eine Verfeinerung von Z . Da die Ungleichung für $\Sigma_{Z_A}^* f$ für alle Zerlegungen Z_A zutrifft, gilt sie auch noch für $\int_A^* f$, das Infimum über alle Z_A , also $\sum_{A \in Z} \int_A^* f \geq \int_Q^* f$.

Zum Beweis der umgekehrten Ungleichung gehen wir von einer Zerlegung Z' aus, deren Obersumme bereits um höchstens ϵ vom Oberintegral abweicht. Da zwei Zerlegungen eine gemeinsame Verfeinerung besitzen und Verfeinerungen den Fehler $\Sigma^* - \int^*$ nur verkleinern, dürfen wir annehmen, dass Z' bereits eine Verfeinerung von Z ist, d.h. Z' ist Vereinigung von Zerlegungen Z_A für alle $A \in Z$, $Z' = \bigcup_{A \in Z} Z_A$. Dann gilt

$$\int^* f \approx_\epsilon \Sigma_{Z'} f = \sum_{A \in Z} \Sigma_{Z_A} f \geq \sum_{A \in Z} \int_A^* f$$

und damit $\int^* f \geq \sum_{A \in Z} \int_A^* f$.⁴

Die entsprechenden Beziehungen für \int_* gelten wegen 0.). □

Wenn f integrierbar ist, $\int^* f = \int_* f$, dann gelten immer beide Ungleichungen und wir erhalten den folgenden Satz:

⁴Wir sagen $a \approx_\epsilon b \iff |a - b| \leq \epsilon$.

44. Satz 22.4 Für integrierbare Funktionen $f, g : Q \rightarrow \mathbb{R}$ und $\alpha \in \mathbb{R}$ gilt:

- 1.) αf ist integrierbar und $\int(\alpha f) = \alpha \int f$,
- 2.) $f + g$ ist integrierbar und $\int(f + g) = \int f + \int g$,
- 3.) $f \leq g \Rightarrow \int^* f \leq \int^* g$,
- 4.) $\forall Z \in \mathfrak{Z}(Q) \int_Q^* f = \sum_{A \in Z} \int_A^* f$.

Eigenschaft 3.) heißt *Monotonie* des Integrals, 2.) die *Bereichsadditivität*, und die Eigenschaften 1.) und 2.) sind die *Linearität* des Integrals. Damit ist der Raum der Riemann-integrierbaren Funktionen

$$R(Q) = \{f : Q \rightarrow \mathbb{R}; f \text{ Riemann-integrierbar}\}$$

ein Vektorraum, ein Unterraum des Raums aller Funktionen von Q nach \mathbb{R} .

45. Satz 22.5 (Mittelwertsatz der Integralrechnung) Ist $Q \subset \mathbb{R}^n$ ein abgeschlossener Quader, $f : Q \rightarrow \mathbb{R}$ stetig und $g \in R(Q)$ mit $g \geq 0$, dann gibt es $x_0 \in Q$ mit

$$\int fg = f(x_0) \int g.$$

Beweis. O.E. sei $g \not\equiv 0$, also $\int g > 0$. Da Q kompakt ist, nimmt f Maximum und Minimum auf Q an. Da $g \geq 0$, folgt für alle $x \in Q$:

$$g(x) \min f \leq g(x)f(x) \leq g(x) \max f$$

Wegen der Monotonieeigenschaft folgt die entsprechende Ungleichungskette für die Integrale, also

$$\min f \leq \frac{\int(gf)}{\int g} \leq \max f.$$

Nach dem Zwischenwertsatz für stetige Funktionen (Satz 4.1) gibt es dann ein $x_0 \in Q$ mit $\int(gf)/\int g = f(x_0)$. \square

23. Klassen integrierbarer Funktionen

46. Satz 23.1 Ist $f : Q \rightarrow \mathbb{R}$ eine stetige Funktion auf einem Quader $Q \subset \mathbb{R}^n$, so ist f integrierbar.

Beweis. Da Q abgeschlossen, also kompakt ist, so ist f sogar gleichmäßig stetig (Satz 4.3), d.h. für vorgegebenes $\epsilon > 0$ existiert ein $\delta > 0$, so dass für alle $x, x' \in Q$ mit $|x - x'| < \delta$ gilt: $|f(x') - f(x)| < \epsilon$. Wir wählen dann eine Zerlegung Z von Q , die so fein ist, dass jedes $A \in Z$ einen Durchmesser $< \delta$ hat, also $|x - x'| < \delta$ für alle $x, x' \in A$ für alle

$A \in Z$. Dann unterscheiden sich die Werte in zwei Punkten von A um höchstens ϵ , also gilt auch $\sup_A f - \inf_A f \leq \epsilon$ und damit (mit der Maß-Additivität 21.1)

$$\sum_{A \in Z} (\sup_A f - \inf_A f) \mu(A) \leq \epsilon \mu(Q).$$

Da wir ϵ beliebig klein vorgeben können, haben wir damit die Integrierbarkeit bewiesen. \square

Es gibt noch eine zweite interessante Klasse integrierbarer Funktionen: Indikatorfunktionen 1_T von messbaren Teilmengen $T \subset Q$. Das sind genau diejenigen T , deren Rand eine *Nullmenge* ist. Eine Teilmenge N eines Quaders $Q \subset \mathbb{R}^n$ heißt (Riemannsche oder Jordansche)⁵ *Nullmenge*, wenn $\mu^*(N) = 0$, d.h. wenn es zu jedem $\epsilon > 0$ eine Zerlegung $Z \in \mathfrak{Z}(Q)$ gibt, so dass die Quader $A \in Z$, die N treffen, alle zusammen ein Maß kleiner als ϵ haben:

$$\sum_{A \in Z_+(N)} \mu(A) \leq \epsilon.$$

Eine Nullmenge N ist insbesondere messbar mit $\mu(N) = 0$, denn $0 \leq \mu_*(N) \leq \mu^*(N) = 0$.

47. Satz 23.2 *Eine Teilmenge T eines Quaders $Q \subset \mathbb{R}^n$ ist messbar genau dann, wenn ihr Rand ∂T eine Nullmenge ist.*

Beweis. T ist messbar $\iff 1_T \in R(Q) \iff$ für jedes $\epsilon > 0$ gibt es eine Zerlegung $Z \in \mathfrak{Z}(Q)$ mit

$$\begin{aligned} \epsilon &\geq \sum_{A \in Z} \left(\sup_A f - \inf_A f \right) \mu(A) \\ &= \sum_{A \in Z_+(T)} \mu(A) - \sum_{A \in Z_-(T)} \mu(A) \\ &= \sum_{A \in Z_+(T) \setminus Z_-(T)} \mu(A) \\ &= \sum_{A \in Z; A \cap T \neq \emptyset, A \not\subset T} \mu(A) \\ &\stackrel{*}{=} \sum_{A \in Z_+(\partial T)} \mu(A) - \sum_{A \in Z_+(\partial T); A \cap T = \emptyset} \mu(A) - \sum_{A \in Z_+(\partial T); A \subset T} \mu(A). \end{aligned}$$

Die letzte Gleichheit $\stackrel{*}{=}$ gilt, weil jeder Quader $A \in Z$, der T schneidet ohne ganz in T enthalten zu sein, den Rand von T treffen muss. Allerdings könnte es noch andere Quader $A \in Z$ geben, die zwar ∂T treffen,

⁵Camille Jordan, 1838 (bei Lyon) - 1922 (Paris)

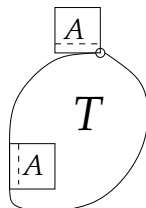


ABBILDUNG 38

aber T entweder gar nicht treffen oder ganz darin enthalten sind; deren Maß muss noch abgezogen werden (zweiter und dritter Term rechts von $\stackrel{*}{=}$). Durch weitere Verfeinerung von Z können wir aber erreichen, dass diese Terme beliebig klein werden, kleiner als ϵ (Abb. 38). Wir können sie also vernachlässigen: T ist messbar $\iff \sum_{A \in Z_+(\partial T)} \mu(A) \leq 3\epsilon$ für genügend feine Zerlegungen Z , d.h. ∂T ist eine Nullmenge. \square

Die einfachsten Beispiele für messbare Mengen sind natürlich Quader, deren Rand ja mit "flachen" Quadern überdeckt werden kann, die also in einer Richtung beliebig dünn sind und damit beliebig kleines Maß haben (Abb. 39).

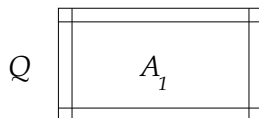


ABBILDUNG 39

Weitere integrierbare Funktionen erhalten wir z.B. durch Limesbildung:

48. Satz 23.3 (Stetigkeit des Integrals) Ist $f_k : Q \rightarrow \mathbb{R}$, $k \in \mathbb{N}$, eine Folge von Funktionen in $R(Q)$, die gleichmäßig gegen $f : Q \rightarrow \mathbb{R}$ konvergiert (" $f_k \xrightarrow{glm} f$ "), so ist $f \in R(Q)$ mit $\int f_k \rightarrow \int f$.

Beweis. Für genügend großes $k \in \mathbb{N}$ ist $|f_k - f| \leq \epsilon$, also $f_k - \epsilon \leq f \leq f_k + \epsilon$ und damit nach Satz 24.2:

$$\int f_k - \epsilon \cdot \mu(Q) \leq \int_* f \leq \int^* f \leq \int f_k + \epsilon \cdot \mu(Q).$$

Da $\epsilon > 0$ beliebig klein gewählt werden kann, folgt $\int_* f = \int^* f$, also $f \in R(Q)$, und $\int f_k \rightarrow \int f$. \square

Auch durch Rechenoperationen kann man neue integrierbare Funktionen gewinnen. Für die Addition integrierbarer Funktionen haben wir dies bereits gesehen; nicht ganz so einfach ist es für die Multiplikation:

49. Satz 24.4 *Ist $f, g \in \mathbf{R}(Q)$, so ist $f \cdot g \in \mathbf{R}(Q)$.*

Beweis. Wir nehmen zunächst $g \geq 0$ und $f \geq 0$ an. Ist Z eine Zerlegung von Q , so gilt für alle $A \in Z$:

$$\begin{aligned} \sup_A(fg) - \inf_A(fg) &\leq \sup_A f \sup_A g - \inf_A f \inf_A g \\ &\leq \sup_A f(\sup_A g - \inf_A g) + (\sup_A f - \inf_A f) \inf_A g \\ &\leq C_f(\sup_A g - \inf_A g) + C_g(\sup_A f - \inf_A f) \end{aligned}$$

wobei C_f und C_g obere Schranken für f und g sind. Ist die Zerlegung Z genügend fein, so ist (vgl. §22, (1))

$$\sum_{A \in Z} (\sup_A f - \inf_A f) \mu(A) < \epsilon, \quad \sum_{A \in Z} (\sup_A g - \inf_A g) \mu(A) < \epsilon,$$

und damit

$$\sum_{A \in Z} (\sup_A(fg) - \inf_A(fg)) \mu(A) < \epsilon(C_f + C_g),$$

also ist fg integrierbar.

Ebenso ist fg integrierbar, falls $g \geq 0$ und $f \leq 0$. Eine beliebige Funktion $f \in \mathbf{R}(Q)$ zerlegen wir als $f = f_+ + f_-$ mit $f_+ \geq 0$ und $f_- \leq 0$, wobei

$$f_+(x) = \max(f(x), 0), \quad f_-(x) = \min(f(x), 0).$$

Die Summanden f_+ und f_- sind wieder integrierbar, wie das nachfolgende Lemma zeigt, und somit ist auch $fg = f_+g + f_-g$ integrierbar für $g \geq 0$ und ebenso für $g \leq 0$ und schließlich durch entsprechende Zerlegung $g = g_+ + g_-$ auch für beliebige $g \in \mathbf{R}(Q)$. \square

50. Lemma. *Ist $f \in \mathbf{R}(Q)$, so ist auch $f_+, f_- \in \mathbf{R}(Q)$, wobei $f_+ = \max(f, 0)$, $f_- = \min(f, 0)$.*

Beweis. Weil f integrierbar ist, gibt es eine Zerlegung Z von Q mit

$$\sum_{A \in Z} (\sup_A f - \inf_A f) \mu(A) < \epsilon.$$

Wir zeigen dieselbe Ungleichung für f_+ und f_- anstelle von f . Da $f_+ \geq f$, gilt stets $\inf_A f_+ \geq \inf_A f$. Ferner ist $\sup_A f_+$ nur dann echt größer als $\sup_A f$, wenn $f|_A < 0$. In diesem Fall ist aber $\inf_A f_+ = \sup_A f_+ = 0$. In jedem Fall gilt also

$$\sup_A f_+ - \inf_A f_+ \leq \sup_A f - \inf_A f,$$

so dass die obige Abschätzung auf f_+ zutrifft. Daher ist f_+ Riemann-integrierbar. Da $f_- = -(-f)_+$, ist auch f_- Riemann-integrierbar. \square

Damit können wir eine Funktion $f \in \mathcal{R}(Q)$ auch über eine beliebige Jordan-messbar Teilmenge $T \subset Q$ integrieren, indem wir setzen:

$$\int_T f := \int (f \cdot 1_T).$$

Nach dem obigen Satz 24.4 ist $f \cdot 1_T$ integrierbar, denn $1_T \in \mathcal{R}(Q)$ und $1_T \geq 0$.

24. Nullmengen

Nullmengen sind für die Integrationstheorie sehr wichtig. Zum einen werden wir sehen, dass wir Funktionen auf Nullmengen beliebig (beschränkt) abändern können, ohne den Wert des Integrals zu ändern. Zum anderen: Wenn wir uns mit Volumenberechnung beschäftigen, sollte die erste Frage lauten: Welchen Teilmengen (“Körpern”) können wir überhaupt ein vernünftiges Volumen oder Maß zuordnen? Wir haben bereits eine Antwort gefunden: genau dann, wenn der Rand eine Nullmenge ist. Damit können wir aber erst dann etwas anfangen, wenn wir Nullmengen erkennen können.

Zunächst können wir das Kriterium für Nullmengen ein bißchen vereinfachen:

51. Satz 24.1 *Eine beschränkte Teilmenge $N \subset \mathbb{R}^n$ ist eine Nullmenge (in jedem Quader $Q \supset N$) genau dann, wenn es zu jedem $\epsilon > 0$ endlich viele offene Quader A_1, \dots, A_r gibt, so dass*

- (a) $N \subset A_1 \cup \dots \cup A_r$,
- (b) $\sum_{i=1}^r \mu(A_i) < \epsilon$.

Man braucht also nicht länger zu fordern, dass die A_i zu einer Zerlegung eines Quaders $Q \supset N$ gehören; insbesondere braucht nicht $A_i \cap A_j = \emptyset$ zu gelten.

Beweis. Wir wählen einen Quader $Q = I_1 \times \dots \times I_n \supset N$ und unterteilen das Intervall I_k durch die k -ten Koordinaten der Eckpunkte aller A_1, \dots, A_r . Damit erhalten wir eine Produktzerlegung Z von Q , die Unterteilungen Z_i von allen A_i als Teilmengen enthält, und

$$\sum_{A \in Z_+(N)} \mu(A) = \sum_{i=1}^r \sum_{A \in (Z_i)_+(N)} \mu(A) \leq \sum_{i=1}^r \mu(A_i) < \epsilon.$$

Damit haben wir das Kriterium des Satzes auf die frühere spezielle Definition von Nullmengen zurückgeführt. Die andere Richtung ist trivial. \square

Bemerkung. Wir können jede Nullmenge durch Würfel W_1, \dots, W_s mit gleicher Kantenlänge l und Gesamtmaß $\sum_{i=1}^s \mu(W_i) = s \cdot l^n < \epsilon$ überdecken. (Ein *Würfel* ist ein Quader mit gleichen Kantenlängen.) Haben wir nämlich eine Überdeckung durch beliebige Quader A_1, \dots, A_n mit Gesamtmaß $< \epsilon$, so vergrößern wir alle Kantenlängen um einen beliebig kleinen Faktor zu rationalen Zahlen; das Gesamtmaß wird dabei unwesentlich größer, sagen wir $< 2\epsilon$. Diese endlich vielen rationalen Zahlen haben einen gemeinsamen Teiler der Form $1/p$ (wobei p z.B. das kleinste gemeinsame Vielfache aller Nenner ist); die neuen, etwas größeren Quader A'_1, \dots, A'_r lassen sich also durch Würfel mit der Kantenlänge $1/p$ zerlegen, die nach dem Zerlegungssatz 21.1 dasselbe Gesamtmaß haben, was zu zeigen war.

Einige Nullmengen kennen wir schon, z.B. die Seiten von Quadern. Der folgende Satz sagt uns, wie wir Nullmengen verbiegen können:

52. Satz 24.2 *Ist $N \subset \mathbb{R}^n$ eine Nullmenge und $F : U^n \rightarrow \mathbb{R}^n$ Lipschitzstetig auf einer offenen Umgebung $U^n \supset N$, so ist $F(N)$ auch eine Nullmenge.*

Beweis. Nach der obigen Bemerkung können wir annehmen, dass N durch Würfel W_1, \dots, W_s mit Kantenlänge l und Gesamtmaß $\sum_{i=1}^s \mu(W_i) = s \cdot l^n < \epsilon$ überdeckt wird. Alle W_i sollen nichtleeren Schnitt mit U^n haben, es gibt also einen Punkt $w_i \in W_i \cap U^n$. Dann gilt

$$F(W_i) \subset B_{\frac{1}{2}L'l}(F(w_i)) \subset W_{L'l}(F(w_i)),$$

wobei $L' = 2\sqrt{n}L$ und L die Lipschitz-Konstante von F ist, und $W_r(x)$ den Würfel mit Kantenlänge r und Mittelpunkt x bezeichnet (Abb. 40). Die Würfel $W'_i = W_{L'l}(F(w_i))$ überdecken $F(N)$ mit Gesamtmaß

$$\sum_{i=1}^s \mu(W'_i) = s \cdot (L'l)^n = (L')^n \cdot \epsilon.$$

Damit ist $F(N)$ eine Nullmenge. □

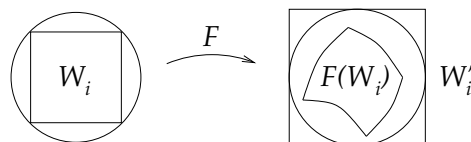


ABBILDUNG 40

53. Satz 24.3 *Jede kompakte Mannigfaltigkeit $M \subset \mathbb{R}^n$ mit $\dim M = m < n$ ist eine Nullmenge.*

Beweis. Jeder Punkt $x \in M$ besitzt eine offene Umgebung U^n in \mathbb{R}^n , auf der ein C^1 -Diffeomorphismus $\Phi_x : U^n \rightarrow V^n$ definiert ist mit $\Phi_x(x) = 0$ und $\Phi_x(M \cap U^n) \subset \mathbb{R}^m \cap V^n$ (Abb. 41). Die offene Umgebung V^n von 0 enthält einen abgeschlossenen Ball $K_r(0) = \overline{B_r(0)}$, auf dem die (stetige) Ableitung der Umkehrfunktion $F_x = \Phi_x^{-1}$ beschränkt sein muß: $\|DF_x(v)\| < L$ für alle $v \in K_r(0)$. Wir setzen $U_x = F_x(B_r(0))$; dies ist eine offene Umgebung von x und $F_x : B_r(0) \rightarrow U_x$ ist nach dem Schrankensatz (Satz 11) Lipschitz-stetig mit Lipschitz-Konstante L . Somit ist $M \cap U_x = F_x(\mathbb{R}^m \cap B_r(0))$ nach dem vorigen Satz eine Nullmenge in \mathbb{R}^n , denn $\mathbb{R}^m \cap B_r(0)$ ist in \mathbb{R}^n offensichtlich eine Nullmenge ist (da $m < n$). Jeder Punkt einer m -dimensionalen Mannigfaltigkeit besitzt also eine Umgebung U_x , deren Schnitt mit M eine Nullmenge ist.

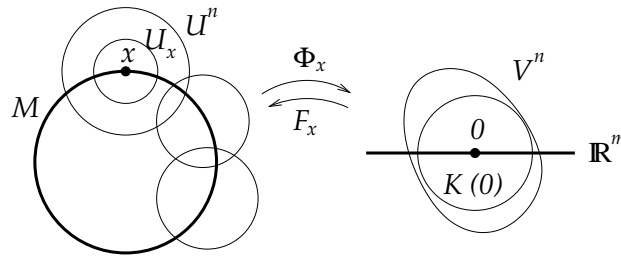


ABBILDUNG 41

Wir zeigen jetzt, dass eine kompakte Mannigfaltigkeit M bereits durch endlich viele solcher Umgebungen überdeckt wird. Dazu brauchen wir eine Eigenschaft kompakter Mengen, die wir bisher noch nicht benutzt haben (vgl. nachstehende Bemerkung): Jede offene Überdeckung einer kompakten Menge $K \subset \mathbb{R}^n$ besitzt eine endliche Teilüberdeckung. Eine *offene Überdeckung* von K ist eine Menge \mathcal{U} , deren Elemente offene Mengen sind, deren Vereinigung K enthält. Da die Menge $\{U_x; x \in M\}$ eine offene Überdeckung von M bildet, liegt M schon in der Vereinigung von endlich vielen Elementen dieser Menge und ist damit eine endliche Vereinigung von Nullmengen, also selbst Nullmenge. \square

Bemerkung. Eine Teilmenge $K \subset \mathbb{R}^n$ ist genau dann kompakt, wenn jede offene Überdeckung von K eine endliche Teilüberdeckung besitzt. Ist K nämlich nicht kompakt, so gibt es eine Folge (x_k) in K (mit paarweise verschiedenen x_k) ohne Häufungspunkt in K . Wir wählen zu jedem x_k eine offene Umgebung U_k , die kein anderes Folgenglied enthält und setzen $U_0 = \mathbb{R}^n \setminus X$, wobei X den Abschluß der Menge $\{x_k; k \in \mathbb{N}\}$ bezeichnet. Dann ist $\{U_0, U_1, U_2, \dots\}$ eine offene Überdeckung von K ohne endliche Teilüberdeckung.

Ist dagegen K kompakt und \mathcal{U} eine offene Überdeckung, so zeigen wir durch Widerspruch, dass \mathcal{U} eine endliche Teilüberdeckung besitzt. Da K beschränkt ist, gibt es einen abgeschlossenen Würfel $W \subset \mathbb{R}^n$, der K enthält. Durch Halbieren der Seitenlänge unterteilen wir W in 2^n abgeschlossene Teilwürfel. Wenn \mathcal{U} keine endliche Teilüberdeckung von K besitzt, finden wir unter diesen Teilwürfeln wenigstens einen, genannt W_1 , so dass \mathcal{U} auch keine endliche Teilüberdeckung von $K \cap W_1$ besitzt. Halbieren wir dessen Seiten erneut, so gibt es wieder einen unter den 2^n Teilwürfel von W_1 , nennen wir ihn W_2 , so dass \mathcal{U} auch keine endliche Teilüberdeckung von $K \cap W_2$ besitzt, usf. Auf diese Weise erzeugen wir eine Folge ineinander enthaltener Würfel W_k mit jeweils halbierten Kantenlänge, so dass \mathcal{U} keine endliche Teilüberdeckung $K \cap W_k$ besitzt, für alle $k \in \mathbb{N}$. Dies ist aber unmöglich: Wählen wir für jedes k einen Punkt $x_k \in K \cap W_k$, so bilden die x_k eine Cauchyfolge, die gegen einen Punkt x konvergiert, der in allen W_k und außerdem in K liegt, da W_k und K abgeschlossen sind. Dieser Punkt $x \in K$ liegt in einer der offenen Überdeckungsmengen, etwa in $U_0 \in \mathcal{U}$. Da der Durchmesser der W_k gegen Null strebt, liegen die W_k schließlich in einem beliebig kleinen Ball $B_\epsilon(x) \subset U_0$. Die ein-elementige Teilmenge $\{U_0\}$ von \mathcal{U} überdeckt also bereits $K \cap W_k$, Widerspruch!

25. Der Satz von Fubini

Mit unserer bisherigen Integrationstheorie können wir praktisch noch wenig anfangen; es ist sehr umständlich, Integrale mit Ober- und Untersummen zu berechnen. In *einer* Variablen, $n = 1$, steht uns der ‘‘Hauptsatz der Differential- und Integralrechnung’’ zur Verfügung: Ist $f : I \rightarrow \mathbb{R}$ stetig, so ist $F(x) = \int_a^x f$ eine Stammfunktion für jedes $a \in I$, d.h. F ist differenzierbar mit $F' = f$. Dabei war definiert: $\int_a^x f := \int_{[a,x]} f$ falls $a \leq x$ und $\int_a^x f := -\int_{[x,a]} f$ falls $a > x$. Erinnerung an den Beweis: Für $h \neq 0$ gibt es ein x_h zwischen x und $x + h$ mit

$$\frac{F(x+h) - F(x)}{h} = \frac{1}{h} \int_x^{x+h} f = f(x_h)$$

nach dem Mittelwertsatz der Integralrechnung (24.5 mit $g \equiv 1$). Für $h \rightarrow 0$ geht $x_h \rightarrow x$ und damit $f(x_h) \rightarrow f(x)$, also ist F in x differenzierbar mit Ableitung $F'(x) = f(x)$.

Um diesen Satz auch zur Berechnung mehrdimensionaler Integrale verwenden zu können, müssen wir diese auf eindimensionale Integrale zurückführen. Das leistet der Satz von Fubini. Dazu betrachten wir Quader $X \subset \mathbb{R}^p$ und $Y \subset \mathbb{R}^q$. Dann ist $X \times Y$ ein Quader in \mathbb{R}^{p+q} . Für jede Funktion $f : X \times Y \rightarrow \mathbb{R}$ definieren durch Festhalten jeweils

einer Variablen $x \in X$ oder $y \in Y$ die Funktionen $f_x : Y \rightarrow \mathbb{R}$ und $f^y : X \rightarrow \mathbb{R}$,

$$f_x(y) = f^y(x) = f(x, y).$$

54. Satz 25 (Fubini) *Es sei $f \in \mathcal{R}(X \times Y)$. Für $x \in X$ und $y \in Y$ sei $F(y) := \int_X f^y$ und $G(x) := \int_Y f_x$, wobei jedes Integral nach Belieben als Ober- oder Unterintegral zu interpretieren ist, falls f^y oder f_x nicht integrierbar sind. Die so definierten Funktionen $F : Y \rightarrow \mathbb{R}$ und $G : X \rightarrow \mathbb{R}$ sind dann integrierbar mit*

$$\int_{X \times Y} f = \int_Y F = \int_X G.$$

Kurz, aber etwas unpräzise:

$$\int_{X \times Y} f = \int_Y \left(\int_X f^y(x) dx \right) dy = \int_X \left(\int_Y f_x(y) dy \right) dx.$$

Beweis. (Vgl. [Aulbach]) Wir beweisen nur die erste Gleichung; die zweite folgt analog (oder mit der einfachsten Form des Transformationsatzes, siehe nächster Abschnitt). Es sei Z eine Zerlegung von $X \times Y$ mit

$$\sum_{A \in Z} (\sup_A f - \inf_A f) \mu(A) < \epsilon. \quad (*)$$

Nach eventueller Verfeinerung dürfen wir annehmen, dass Z eine Produktzerlegung ist, dass es also Zerlegungen Z_X von X und Z_Y von Y gibt, so dass

$$Z = \{A = B \times C; B \in Z_X, C \in Z_Y\}.$$

Dann ist

$$\sum_{A \in Z} \inf_A f \cdot \mu(A) = \sum_{C \in Z_Y} \left(\sum_{B \in Z_X} \inf_{B \times C} f \cdot \mu(B) \right) \cdot \mu(C).$$

Betrachten wir die innere Summe: Für jedes $C \in Z_Y$ und alle $y \in C$ gilt:

$$\sum_{B \in Z_X} \inf_{B \times C} f \cdot \mu(B) \leq \sum_{B \in Z_X} \inf_B f^y \cdot \mu(B) \leq \int_{*X} f^y.$$

Da dies für *alle* $y \in C$ gilt, können wir zum Infimum über alle $y \in C$ übergehen und erhalten:

$$\sum_{B \in Z_X} \inf_{B \times C} f \cdot \mu(B) \leq \inf_{y \in C} F_-(y)$$

mit $F_-(y) := \int_{*X} f^y$. Summieren wir über alle C , so ergibt sich:

$$\sum_{A \in Z} \inf_A f \cdot \mu(A) \leq \sum_{C \in Z_Y} \inf_{y \in C} F_-(y) \cdot \mu(C) \leq \int_{*Y} F_-.$$

Auf dieselbe Weise erhalten wir:

$$\sum_{A \in Z} \sup_A f \cdot \mu(A) \geq \int_Y^* F_+$$

mit $F_+(y) := \int_X^* f^y$. Nach (*) unterscheiden sich die Obersumme und die Untersumme von $\int f$ höchstens um ϵ , daher erhalten wir die Ungleichungskette

$$\int_{X \times Y} f - \epsilon \leq \int_{*Y} F_- \leq \int_Y^* F_+ \leq \int_{X \times Y} f + \epsilon.$$

Alle diese Terme unterscheiden sich also um höchstens 2ϵ . Da

$$\int_{*Y} F_- \leq \int_Y^* F_- \leq \int_Y^* F_+, \quad \int_{*Y} F_- \leq \int_{*Y} F_- \leq \int_Y^* F_+,$$

unterscheiden sich Unter- und Oberintegral von F_- ebenso wie von F_+ um höchstens ϵ , also sind F_- und F_+ integrierbar, und ihre Integrale sind beide gleich $\int_{X \times Y} f$. \square

Beispiel 1: $X = [0, 1/2]$, $Y = [0, 2\pi]$ und $f(x, y) = x \cdot \cos xy$. Dann ist

$$\begin{aligned} \int_{X \times Y} f &= \int_{x=0}^{1/2} \left(\int_{y=0}^{2\pi} x \cos xy dy \right) dx \\ &= \int_0^{1/2} ([\sin xy]_{y=0}^{y=2\pi}) dx = \int_0^{1/2} \sin(2\pi x) dx \\ &= \left[-\frac{1}{2\pi} \cos(2\pi x) \right]_{x=0}^{x=1/2} = 1/\pi. \end{aligned}$$

Berechnet man das Integral in der umgekehrten Reihenfolge, also

$$\int_{X \times Y} f = \int_{y=0}^{2\pi} \left(\int_{x=0}^{1/2} x \cos(xy) dx \right) dy,$$

so hat man in diesem Fall mehr Arbeit. Man muß dann zunächst das innere Integral durch partielle Integration berechnen und erhält:

$$\int_0^{1/2} x \cos xy dx = \frac{1}{2y} \sin \frac{y}{2} - \frac{1}{y^2} (1 - \cos \frac{y}{2}) = \frac{d}{dy} \left(\frac{1}{y} (1 - \cos \frac{y}{2}) \right).$$

Führt man nun das äußere Integral aus, so ergibt sich

$$\int_{X \times Y} f = \left[\frac{1}{y} (1 - \cos \frac{y}{2}) \right]_0^{2\pi} = 1/\pi,$$

denn die Potenzreihe $\frac{1}{y} (1 - \cos \frac{y}{2}) = \frac{y}{2^2 \cdot 2!} - \frac{y^3}{2^4 \cdot 4!} + \dots$ hat den Wert Null für $y = 0$.

55. Korollar 1: Cavalierisches Prinzip. Es sei $M \subset \mathbb{R}^n = \mathbb{R}^{n-1} \times \mathbb{R}$ eine Jordan-messbare Menge. Für jedes $y \in \mathbb{R}$ sei

$$M^y := \{x \in \mathbb{R}^{n-1}; (x, y) \in M\} \subset \mathbb{R}^{n-1}$$

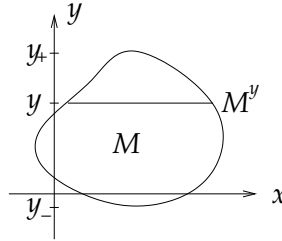


ABBILDUNG 42

(Abb. 42). Dann ist

$$\mu(M) = \int_{y_-}^{y_+} \mu_*(M^y) dy = \int_{y_-}^{y_+} \mu^*(M^y) dt,$$

wobei y_- das Infimum und y_+ das Supremum aller $y \in \mathbb{R}$ mit $M^y \neq \emptyset$ ist.

Beweis. M liegt in einem Quader $X \times Y \subset \mathbb{R}^{n-1} \times \mathbb{R}$; wir können $Y = [y_-, y_+]$ wählen. Der Satz ist nun ein Spezialfall des Satzes von Fubini für $f = 1_M$ mit $f^y = 1_{M^y}$ für jedes $y \in \mathbb{R}$. \square

56. Korollar 2: Cavalierisches Vergleichsprinzip Sind $M, \tilde{M} \subset \mathbb{R}^n = \mathbb{R}^{n-1} \times \mathbb{R}$ Jordan-messbar mit $\mu^*(M^y) = \mu^*(\tilde{M}^y)$ für alle $y \in \mathbb{R}$, so ist $\mu(M) = \mu(\tilde{M})$. \square

Die meisten Anwendungen des Cavalierischen Prinzips benutzen, dass das Maß unter Translationen invariant (ungeändert) bleibt und bei zentrischen Streckungen um einen Faktor $\alpha > 0$ (also bei Anwenden der Abbildung $x \mapsto \alpha x : \mathbb{R}^n \rightarrow \mathbb{R}^n$) mit α^n zu multiplizieren ist. Dies ist sofort einsichtig, da Quader diese Eigenschaft haben und das Maß mit Überdeckungen und Ausschöpfungen durch Quader definiert ist. Wir werden im nächsten Abschnitt darauf in allgemeinerem Zusammenhang zurückkommen.

Beispiel 2: Es sei $M \subset \mathbb{R}^n = \mathbb{R}^{n-1} \times \mathbb{R}$ messbar und $\Phi : \mathbb{R}^{n-1} \times \mathbb{R} \rightarrow \mathbb{R}^{n-1} \times \mathbb{R}$ sei eine *Scherung*, d.h.

$$\Phi(x, y) = (x + \xi(y), y)$$

für eine stetig differenzierbare Abbildung $\xi : \mathbb{R} \rightarrow \mathbb{R}^{n-1}$. Dann ist $\Phi(M)$ wieder messbar (Φ ist Lipschitz-stetig auf kompakten konvexen

Mengen und erhält daher Nullmengen), und $\mu(\Phi(M)) = \mu(M)$ nach Korollar 2, denn $\Phi(M)^y = M^y + \xi(y)$ für alle $y \in \mathbb{R}$, d.h. $\Phi(M)^y$ und M^y unterscheiden sich nur durch die Translation mit dem Vektor $\xi(y)$ und haben daher gleiches Maß.

Bemerkung. Das Cavalierische Vergleichsprinzip (Korollar 2) war bereits in der Antike bekannt; Archimedes berechnete damit das Kugelvolumen. Dabei war M die Halbkugel vom Radius r und \tilde{M} der Kreiszyylinder mit Radius r und Höhe r , aus dem ein auf der Spitze stehender Kreiskegel mit Radius r und Höhe r ausgebohrt ist (Abb. 43).

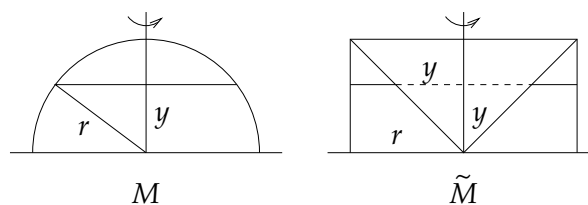


ABBILDUNG 43

Für $0 \leq y \leq r$ ist M^y ein Kreis mit Radius $\sqrt{r^2 - y^2}$ und \tilde{M}^y ein Kreisring mit äußerem Radius r und innerem Radius y ; beide haben den Flächeninhalt $\pi(r^2 - y^2)$. Das Zylindervolumen Grundfläche \times Höhe und das Kegelvolumen $\frac{1}{3}$ Grundfläche \times Höhe (siehe unten) konnte Archimedes, somit ergab sich als Volumen der Halbkugel $\pi r^2 \cdot r - \frac{1}{3} \pi r^2 \cdot r = \frac{2}{3} \pi r^3$.

Woher konnte man in der Antike das Volumen eines Kegels mit beliebiger Grundfläche? Jede Grundfläche läßt sich aus Dreiecken zusammensetzen oder jedenfalls approximieren; es war also nur das Volumen eines Kegels über einem Dreieck (eines *Tetraeders*) zu berechnen. Aus dem Cavalierischen Vergleichsprinzip war bekannt, dass man die Spitze eines Kegels auf gleicher Höhe beliebig verschieben kann, ohne das Volumen zu verändern (Beispiel 2 und Abb. 44).

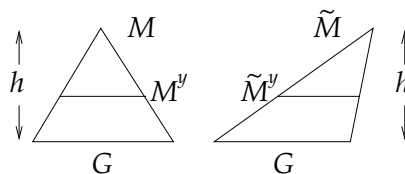


ABBILDUNG 44

Alle Kegel gleicher Höhe über einem festen Dreieck haben also dasselbe Volumen. Der Zylinder über dem Dreieck (*Prisma*) hat das Volumen

Grundfläche \times Höhe und läßt sich folgendermaßen in drei volumengleiche Tetraeder zerlegen (Abb. 45): Sind a, b, c die Eckpunkte der Bodenfläche und A, B, C die der Deckfläche des Prismas, so sind die die drei Tetraeder durch die Eckpunkte (a, b, c, C) , (a, b, B, C) sowie (a, A, B, C) gegeben.

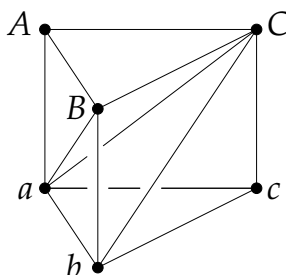


ABBILDUNG 45

Die ersten beiden Tetraeder können als Kegel mit Spitze in a über den kongruenten in derselben Ebene liegenden Grundflächen (b, c, C) und (b, C, B) aufgefaßt werden, und ebenso können der zweite und der dritte Tetraeder als Kegel mit Spitze C über den kongruenten Grundflächen (a, b, B) und (a, B, A) in derselben Ebene aufgefaßt werden. Alle drei Tetraeder haben also dasselbe Volumen, und da sie eine Zerlegung des Prismas bilden, ist ihr Volumen (speziell das des Kegels (a, b, c, C) über der Grundseite (a, b, c)) gleich $\frac{1}{3}$ Grundseite \times Höhe. Das Argument läßt sich übrigens leicht auf beliebige Dimension n verallgemeinern.

Beispiel 3: Volumen des Kegels. Es sei $G \subset \mathbb{R}^{n-1}$ eine Jordanmessbare Teilmenge (“Grundseite”) und $s = (x_0, h) \in \mathbb{R}^{n-1} \times \mathbb{R}$ (“Spitze”) mit $h > 0$ (“Höhe”). Der Kegel mit Spitze s über der Grundseite G ist die Menge

$$M = \bigcup_{x \in G} [x, s] = \{t \cdot s + (1-t)x; x \in G, t \in [0, 1]\} \subset \mathbb{R}^n,$$

wobei wir $x \in \mathbb{R}^{n-1}$ mit $(x, 0) \in \mathbb{R}^n$ identifiziert haben. Der Rand von M besteht aus der Grundseite G und dem Kegel über ∂G (“Kegelmantel”). Beides sind Nullmengen in \mathbb{R}^n : Die Grundseite ist eine beschränkte Teilmenge von \mathbb{R}^{n-1} und daher in einem Quader in \mathbb{R}^n mit beliebig kleiner Höhe enthalten, und der Kegelmantel ist Bild der kompakten Nullmenge $\partial G \times [0, 1] \subset \mathbb{R}^{n-1} \times \mathbb{R}$ unter der differenzierbaren Abbildung $\Phi : \mathbb{R}^{n-1} \times \mathbb{R} \rightarrow \mathbb{R}^n$, $\Phi(x, t) = ts + (1-t)x$. Also ist M Jordan-messbar. Die “Ordinate” (= Projektion auf den zweiten

Faktor) eines beliebigen Punktes $t\binom{x_0}{h} + (1-t)\binom{x_0}{0} \in M$ ist $y = t \cdot h$, also gilt für alle $y \in [0, h]$:

$$M^y = \left\{ \frac{y}{h}x_0 + \left(1 - \frac{y}{h}\right)x; x \in G \right\}.$$

Bis auf die Translation um den Vektor $\frac{y}{h}x_0$ ist M^y die zentrische Streckung von G um den Faktor $\alpha = 1 - \frac{y}{h}$, also ist nach Cavalieri

$$\mu(M) = \int_0^h \mu(M^y) = \mu(G) \int_0^h \left(1 - \frac{y}{h}\right)^{n-1} dy = \mu(G) \cdot h \cdot \frac{1}{n}.$$

Das Volumen des n -dimensionalen Kegels über der Grundseite $G \subset \mathbb{R}^{n-1}$ mit Höhe h ist also $\frac{1}{n}\mu(G)h$.

Beispiel 4: Volumen der Kugel. Wir betrachten die n -dimensionale Kugel vom Radius $r > 0$,

$$K_r^n := \{x \in \mathbb{R}^n; |x| \leq r\}.$$

Diese ist Jordan-messbar, da $\partial K_r^n = S_r$ eine kompakte Mannigfaltigkeit ist, die $(n-1)$ -dimensionale Sphäre. Da K_r^n durch zentrische Streckung mit dem Faktor r aus der Einheitskugel K_1^n entsteht, gilt

$$\mu(K_r^n) = r^n \mu(K_1^n), \quad (*)$$

wir brauchen also nur das Volumen der Einheitskugel zu berechnen. K_1^1 ist das Intervall $[-1, 1]$ mit Länge 2, also $\mu(K_1^1) = 2$. Von jetzt an sei $M = K_1^n$ für $n \geq 2$. Wir zerlegen wieder $\mathbb{R}^n = \mathbb{R}^{n-1} \times \mathbb{R}$. Dann ist

$$x \in M^y \iff (x, y) \in M \iff |x|^2 + y^2 \leq 1,$$

also ist $M^y = K_r^{n-1}$ mit $r = \sqrt{1 - y^2}$ für $-1 \leq y \leq 1$ (Abb. 46). Nach Cavalieri und Gleichung (*) ist daher

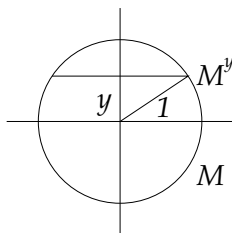


ABBILDUNG 46

$$\mu(K_1^n) = \mu(K_1^{n-1}) \int_{-1}^1 (\sqrt{1 - y^2})^{n-1} dy.$$

Das Integral wird mit der Substitution $y = \sin u$, also $dy = \cos u \cdot du$ umgeformt:

$$\int_{-1}^1 (\sqrt{1-y^2})^{n-1} dy = \int_{-\pi/2}^{\pi/2} \cos^n u du = 2 \int_0^{\pi/2} \sin^n u du.$$

Dieses Integral

$$A_n = \int_0^{\pi/2} \sin^n u du$$

wird auf die bekannte Weise (vgl.z.B. [Forster I], S. 147) durch partielle Integration rekursiv berechnet: Man setzt $\sin^n = \sin^{n-1} \sin = f \cdot g'$ mit $f = \sin^{n-1}$ und $g = -\cos$, dann ist $\int f g' = f g - \int f' g$. Da $f g = -\sin^{n-1} \cos$ an den Grenzen 0 und $\pi/2$ verschwindet und

$$-f'g = (n-1) \sin^{n-2} \cos^2 = (n-1)(\sin^{n-2} - \sin^n),$$

erhalten wir $A_n = (n-1)(A_{n-2} - A_n)$ und damit die Rekursionsformel

$$A_n = \frac{n-1}{n} A_{n-2}; \quad A_0 = \frac{\pi}{2}, \quad A_1 = 1,$$

also $A_2 = \frac{1}{2} \frac{\pi}{2}$, $A_3 = \frac{2}{3}$, $A_4 = \frac{3}{4} \frac{1}{2} \frac{\pi}{2}$, $A_5 = \frac{4}{5} \frac{2}{3}$ usw. Wir haben berechnet:

$$\mu(K_1^n) = \mu(K_1^{n-1}) \cdot 2A_n,$$

also erhalten wir $\mu(K_1^1) = 2$ (Länge von $[-1, 1]$), $\mu(K_1^2) = 2 \frac{\pi}{2} = \pi$, $\mu(K_1^3) = \pi \cdot \frac{4}{3}$, $\mu(K_1^4) = \pi \cdot \frac{4}{3} \cdot \frac{3}{4} \frac{\pi}{2} = \frac{\pi^2}{2}$, $\mu(K_1^5) = \pi^2 \frac{4}{5} \frac{2}{3} = \frac{8}{15} \pi^2$ usw.

57. Korollar 3. (Integration über Teilmengen) *Es sei Q ein Quader im $\mathbb{R}^n = \mathbb{R}^{n-1} \times \mathbb{R}$ und $M \subset Q$ Jordan-messbar. Dann gilt für jede Funktion $g \in \mathcal{R}(Q)$ mit den Bezeichnungen aus Korollar 1:*

$$\int_M g = \int_{y_-}^{y_+} \left(\int_{M^y} g(x, y) dx \right) dy$$

Beweis. Wir haben die Funktion $f = g \cdot 1_M$ über den Quader $Q = X \times [y_-, y_+]$ zu integrieren. Nach Fubini ist

$$\int f = \int_{y_-}^{y_+} \left(\int_X f^y(x) dx \right) dy.$$

Dabei ist $f^y = g^y \cdot 1_{M^y}$, und da $M^y \subset X$, ist das innere Integral gleich $\int_{M^y} g^y$, was zu zeigen war. \square

Beispiel 5: Berechne $\int_M xy d(xy)$ für

$$M = \{(x, y); x, y \in [0, a], x^2 + y^2 \geq a^2\}.$$

(Abb. 47). Es gilt also für festes $y \in [0, a]$:

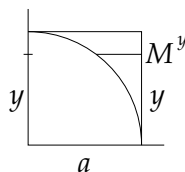


ABBILDUNG 47

$$(x, y) \in M \iff \sqrt{a^2 - y^2} \leq x \leq a,$$

also ist $M^y = [\sqrt{a^2 - y^2}, a]$. Die Funktion $g(x, y) = xy$ ist stetig auf \mathbb{R}^2 , und M liegt in dem "Quader" $[0, a] \times [0, a]$. Nach Korollar 3 ist $\int_M g = \int_0^a (\int_{M^y} g^y) dy$ mit

$$\int_{M^y} g^y = \int_{\sqrt{a^2 - y^2}}^a xy dx = yx^2/2 \Big|_{x=\sqrt{a^2 - y^2}}^{x=a} = y^3/2.$$

Somit ist $\int_M g = \int_0^a \frac{y^3}{2} dy = a^4/8$.

Beispiel 6: ("Integral ist Fläche unter dem Graphen") Es sei X ein Quader in \mathbb{R}^{n-1} und $f, g \in R(X)$ mit $f \leq g$. Die Menge

$$M = \{(x, y) \in \mathbb{R}^{n-1} \times \mathbb{R}; f(x) \leq y \leq g(x)\}$$

("Zwischenraum zwischen den Graphen von f und g ") liegt in $X \times Y$ für ein kompaktes Intervall Y , und ∂M ist eine Nullmenge (Integrierbarkeit von f und g). Wir berechnen $\mu(M)$ nach Fubini. Ist $h := 1_M$, so ist $h_x = 1_{[f(x), g(x)]}$ und $\int_Y h_x = g(x) - f(x)$, also ist

$$\mu(M) = \int_{X \times Y} h = \int_X \left(\int_Y h_x \right) dx = \int_X g - \int_X f.$$

Speziell für $f = 0$ ergibt sich also, dass $\int g$ das Volumen unter dem Graphen von g ist (für $g \geq 0$).

26. Der Transformationssatz

Ein wichtiges Hilfsmittel zur Berechnung eindimensionaler Integrale ist die *Substitutionsregel*:

$$\int_a^b (f \circ \Phi \cdot \Phi') = \int_{\Phi(a)}^{\Phi(b)} f$$

oder ausführlicher:

$$\int_a^b f(\Phi(x)) \Phi'(x) dx = \int_{\Phi(a)}^{\Phi(b)} f(u) du,$$

wo für den komplizierten Ausdruck $\Phi(x)$ eine neue Variable u “substituiert” (= eingetauscht, ersetzt) wird, um das Integral zu vereinfachen; z.B. ist $\int_a^b \sin x \cos x dx = \int_{\sin a}^{\sin b} u du = \frac{1}{2}(\sin^2 b - \sin^2 a)$, wobei wir $u = \sin x$ (und $du = \cos x dx$) substituiert haben. Die Substitutionsregel folgt unmittelbar aus dem “Hauptsatz der Differential- und Integralrechnung” (Berechnung von Integralen durch Stammfunktionen): Ist F eine Stammfunktion von f , so ist nach Kettenregel $F \circ \Phi$ eine Stammfunktion von $f \circ \Phi \cdot \Phi'$.

Das Ziel dieses Abschnittes ist die Herleitung einer analogen Formel für n -dimensionale Integrale, wobei Φ ein C^1 -Diffeomorphismus ist und Φ' durch die Determinante der Jacobimatrix ersetzt wird:⁶

58. Satz 26.1 (Transformationssatz) *Es seien X und U offene Teilmengen des \mathbb{R}^n und $\Phi : X \rightarrow U$ ein C^1 -Diffeomorphismus. Weiterhin sei $f : U \rightarrow \mathbb{R}$ eine stetige Funktion und $M \subset X$ eine kompakte messbare Menge. Dann gilt*

$$\int_M (f \circ \Phi \cdot |\det D\Phi|) = \int_{\Phi(M)} f. \quad (1)$$

oder ausführlicher

$$\int_{x \in M} f(\Phi(x)) |\det D\Phi_x| dx = \int_{u \in \Phi(M)} f(u) du.$$

Der Beweis für höhere Dimension n ist schwieriger, da der “Hauptsatz der Differential- und Integralrechnung” nur für $n = 1$ gilt. Beide Seiten von (1) sind definiert, denn $\Phi(M)$ ist wieder messbar, da $\Phi(\partial M) = \partial\Phi(M)$ eine Nullmenge ist (Φ ist Lipschitz; vgl. 23.2). Wir werden den Diffeomorphismus lokal durch affine Abbildungen (lineare Abbildung plus Konstante) approximieren und den Satz auf eine Tatsache der linearen Algebra zurückführen, die wir im folgenden Kapitel beweisen werden: Der Betrag der Determinante einer linearen Abbildung ist die Volumenveränderung dieser Abbildung:

59. Satz 26.2 (Determinante als Volumenänderung) *Es sei $M \subset \mathbb{R}^n$ eine messbare Menge und $\Phi : \mathbb{R}^n \rightarrow \mathbb{R}^n$, $\Phi(x) = Ax + a$ eine invertierbare affine Abbildung. Dann ist auch $\Phi(M)$ messbar mit*

$$\mu(\Phi(M)) = |\det A| \cdot \mu(M). \quad (2)$$

Beweis von Satz 26.1:

⁶Übung: Wieso ist die obige eindimensionale Substitutionsregel ein Spezialfall dieser Formel (1)? Es gibt zwei Unterschiede, die sich gegenseitig aufheben.

Im Folgenden bezeichnet C eine beliebige (wechselnde) Konstante, die nur von M und der Lipschitzkonstante L der Abbildungen abhängen darf. Wir denken uns ein beliebig kleines $\epsilon > 0$ vorgegeben. Mit $a \approx b$ meinen wir $|a - b| < C\epsilon$. Außerdem benötigen wir eine genügend kleine Konstante $\delta > 0$, die noch zu wählen sein wird.

Wir dürfen uns auf den Fall beschränken, dass M ein Würfel ist, denn da M messbar ist, können wir M bis auf eine Menge von beliebig kleinem Maß als disjunkte Vereinigung von endlich vielen Würfeln darstellen, für die wir die Integralformel einzeln nachweisen und aufsummieren. Es sei also von jetzt an $M = W$ ein abgeschlossener Würfel mit Kantenlänge D . Durch Unterteilung aller Kanten von W in k Stücke der Länge $D/k = \delta$ definieren wir eine Zerlegung Z von W in $k^n = D^n/\delta^n$ kongruente Würfel (Abb. 48), die so klein sein sollen, dass $\Phi|_A$ "fast affin" ist für jeden Unterteilungswürfel $A \in Z$, d.h. $\Phi|_A$ wird "genügend genau" (siehe Lemma 26.4) durch sein Taylorpolynom erster Ordnung

$$\Phi^A(x) = \Phi(x_A) + D\Phi_{x_A}(x - x_A)$$

approximiert, wobei x_A der Mittelpunkt von A ist; auf Φ^A können wir dann Satz 26.2 anwenden.

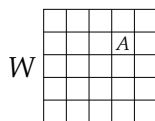


ABBILDUNG 48

60. Lemma 26.3 *Zu jedem $\epsilon > 0$ gibt es $\delta > 0$ mit der folgenden Eigenschaft: Ist die Kantenlänge der Unterteilungswürfel A nicht größer als δ , dann gilt für jedes $A \in Z$ und jedes $x \in A$*

$$|\Phi(x) - \Phi^A(x)| < \epsilon \cdot |x - x_A|.$$

Dies folgt aus dem Satz von Taylor und der gleichmäßigen Stetigkeit von $D\Phi$, siehe Beweis weiter unten. Wenn $\sup_{\Phi A} f - \inf_{\Phi A} f < \epsilon$, dann ist $\sup_{\Phi A} f \approx_\epsilon f(\Phi x_A) \approx_\epsilon \inf_{\Phi A} f$ und

$$f \approx_\epsilon \bar{f} := \sum_A f(\Phi x_A) 1_{\Phi A}.$$

Dies sind eigentlich zwei Ungleichungen: $f \leq \bar{f} + \epsilon$ und $\bar{f} \leq f + \epsilon$. Die Monotonie des Integrals über ΦW ergibt $\int f \leq \int \bar{f} + \epsilon'$ und $\int \bar{f} \leq \int f + \epsilon'$ und damit $\int f \approx_{\epsilon'} \int \bar{f}$ mit $\epsilon' = \epsilon \mu(\Phi W)$. Da

$$\int \bar{f} = \sum_{A \in Z} f(\Phi x_A) \int 1_{\Phi A} = \sum_{A \in Z} f(\Phi x_A) \mu(\Phi A),$$

erhalten wir:

$$\begin{aligned}
\int_{\Phi W} f &\approx \sum_{A \in Z} f(\Phi x_A) \mu(\Phi A) \\
&\approx^* \sum_{A \in Z} f(\Phi x_A) \mu(\Phi^A A) \\
&= \sum_{A \in Z} f(\Phi x_A) |\det D\Phi_{x_A}| \mu(A) \\
&\approx \int_W f \cdot |\det D\Phi|.
\end{aligned}$$

In der letzten Zeile benutzen wir, dass für die Funktion $g = f \cdot |\det D\Phi|$ wieder $\sup_A g - \inf_A g < \epsilon$ gilt. In der vorletzten Zeile ist Satz 26.2 (Determinante als Volumenfaktor) eingegangen. Die Beziehung \approx^* (drittletzte Zeile) wird im nachfolgenden Lemma 26.4 gerechtfertigt; der Fehler $|\mu(\Phi^A A) - \mu(\Phi A)|$ muss so klein sein, dass auch nach Aufsummieren der k^n Summanden der Gesamtfehler immer noch kleiner als ϵ ist. Damit ist der Beweis von Satz 26.1 abgeschlossen. \square

61. Lemma 26.4 *Es seien Φ und $\tilde{\Phi}$ umkehrbar Lipschitz-stetig, definiert in einer Umgebung des Würfels A mit Kantenlänge δ , und es gelte*

$$|\Phi - \tilde{\Phi}| < \epsilon \delta$$

auf A . Dann gilt

$$|\mu(\Phi A) - \mu(\tilde{\Phi} A)| < C \epsilon \delta^n$$

wobei die Konstante C nur von der (gemeinsamen) Lipschitzkonstanten L der Abbildungen Φ , $\tilde{\Phi}$ und ihrer Inversen abhängt.

Beweis. Wir werden $\mu(\tilde{\Phi} A) \leq \mu(\Phi A) + C \epsilon \delta^n$ zeigen; da Φ und $\tilde{\Phi}$ gleichberechtigt sind, können wir auch die Rollen von Φ und $\tilde{\Phi}$ vertauschen; aus beiden Ungleichungen zusammen erhalten wir die Behauptung.

Für jede Teilmenge $T \subset \mathbb{R}^n$ und jedes $\rho > 0$ definieren wir:

$$B_\rho(T) = \{x \in \mathbb{R}^n; \exists t \in T | x - t| < \rho\};$$

$B_\rho(T)$ enthält also die Punkte, die in T liegen oder nicht weiter als ρ von T entfernt sind.

Nun sehen wir

$$\tilde{\Phi} A \subset B_\rho(\Phi A)$$

mit $\rho := \epsilon \delta$, denn für jedes $y = \tilde{\Phi} x \in \tilde{\Phi} A$ ist $\Phi x \in \Phi A$ und $|\tilde{\Phi} x - \Phi x| < \delta \epsilon$. Anwenden von Φ^{-1} vergrößert Bälle höchstens um die Lipschitzkonstante L , also

$$\Phi^{-1} B_\rho(\Phi A) \subset B_{\rho'}(A). \quad B_\rho(\Phi A) \subset \Phi B_{\rho'}(A)$$

mit $\rho' = \rho L$, und insgesamt erhalten wir

$$\tilde{\Phi}A \subset B_\rho(\Phi A) \subset \Phi B_{\rho'}(A).$$

Andererseits

$$\begin{aligned} \mu(\Phi B_{\rho'}(A)) - \mu(\Phi A) &= \mu(\Phi(B_{\rho'}(A) \setminus A)) \\ &\leq L^n(\mu(B_{\rho'}A) - \mu(A)) \\ &= L^n((\delta + \rho')^n - \delta^n) \\ &= L^n\delta^n((1 + L\epsilon)^n - 1) \\ &\leq L^n\delta^n C' L\epsilon \end{aligned}$$

und somit $\mu(\tilde{\Phi}A) - \mu(\Phi A) \leq \mu(\Phi B_{\rho'}(A)) - \mu(\Phi A) \leq C\delta^n\epsilon$ mit $C = L^{n+1}C'$. \square

Beweis von Lemma 26.3: Für jedes $a \in W$ lautet die Taylorformel der Ordnung 0 für jede Komponente Φ_i der vektorwertigen Abbildung $\Phi = (\Phi_1, \dots, \Phi_n)^T$:

$$\begin{aligned} \Phi_i(x) &= \Phi_i(a) + (D\Phi_i)_{\tilde{a}}(x - a) \\ &= \Phi_i(a) + (D\Phi_i)_a(x - a) + ((D\Phi_i)_{\tilde{a}} - (D\Phi_i)_a)(x - a) \end{aligned}$$

für ein $\tilde{a} \in [x, a]$ = Strecke von x nach a . Da $D\Phi_i$ auf W gleichmäßig stetig ist, folgt

$$\|(D\Phi_i)_{\tilde{a}} - (D\Phi_i)_a\| \leq \epsilon$$

wann immer $|x - a|$ und damit auch $|x - \tilde{a}|$ kleiner als $\delta' = \sqrt{n}\delta$ ist. Insbesondere folgt für $x \in A$ und $a = x_A$ für jedes $A \in Z$:

$$|\Phi(x) - \Phi^A(x)| \leq \epsilon|x - x_A|. \quad \square$$

Beispiel 1: Polarkoordinaten Zu berechnen sei $\int_K xy \, dx dy$, wobei K ein Viertelkreis ist:

$$K = \{(x, y) \in \mathbb{R}^2; x, y \geq 0, x^2 + y^2 \leq a^2\}.$$

Wir benutzen die Substitution $x = r \cos \phi$, $y = r \sin \phi$, also die Polarkoordinaten $\Phi : (0, \infty) \times (-\pi, \pi) \rightarrow \mathbb{R}^2 \setminus (\mathbb{R}_- \times \{0\})$,

$$\Phi(r, \phi) = \begin{pmatrix} r \cos \phi \\ r \sin \phi \end{pmatrix}, \quad D\Phi_{(r, \phi)} = \begin{pmatrix} \cos \phi & -r \sin \phi \\ \sin \phi & r \cos \phi \end{pmatrix}$$

mit $\det D\Phi_{(r, \phi)} = r(\cos^2 \phi + \sin^2 \phi) = r > 0$. Die Menge K liegt zwar nicht ganz im Wertebereich von Φ , aber $K' := K \setminus \{0\}$, was keinen Unterschied macht, da $\{0\}$ eine Nullmenge ist. Wir benutzen dieselbe Küchenregel wie bei der eindimensionalen Substitutionsregel:

$$\begin{aligned} (x, y) &= \Phi(r, \phi) = (r \cos \phi, r \sin \phi), \\ dx dy &= |\det D\Phi_{(r, \phi)}| dr d\phi = r dr d\phi, \end{aligned}$$

$$K = \Phi(\{(r, \phi); 0 \leq r \leq a, 0 \leq \phi \leq \frac{\pi}{2}\})$$

also gilt

$$\begin{aligned} \int_K xy \, dx dy &= \int_{r=0}^a \int_{\phi=0}^{\pi/2} r \cos(\phi) r \sin(\phi) r dr d\phi \\ &= \int_0^a r^3 dr \cdot \int_0^{\pi/2} \cos \phi \sin \phi d\phi \\ &= \frac{a^4}{4} \cdot \frac{1}{2} \end{aligned}$$

(man verwende das Eingangsbeispiel von §26 oder $\int \cos \phi \sin \phi d\phi = \frac{1}{2} \int \sin 2\phi d\phi = -\frac{1}{4} \cos 2\phi$).

Beispiel 2: Radiale Funktionen und Gaußverteilung Besonders einfach wird die Transformation in Polarkoordinaten, wenn man über einen Kreis $K_a = \{(x, y); x^2 + y^2 \leq a^2\}$ oder einen Kreisring integriert und der Integrand f eine *radiale Funktion* ist, also nur von der Radiuskoordinate r abhängt, d.h. $f(x, y) = g(\sqrt{x^2 + y^2}) = g(r)$ für eine stetige Funktion $g: \mathbb{R}_+ \rightarrow \mathbb{R}$. Dann ist nämlich einfach

$$\int_{K_a} f = \int_{\phi=0}^{2\pi} \int_{r=0}^a g(r) r dr d\phi = 2\pi \int_0^a g(r) r dr. \quad (*)$$

Eine hübsche und überraschende Anwendung hiervon ist die Berechnung des Integrals der Funktion $f(x) = e^{-x^2}$, die in der Wahrscheinlichkeitstheorie eine große Rolle spielt (*Gauß'sche Normalverteilung*, deren Graph auf dem alten 10-Mark-Schein zu sehen war). Statt $I = \int_{-\infty}^{\infty} e^{-x^2} dx$ berechnet man I^2 :

$$\begin{aligned} I^2 &= \int_{-\infty}^{\infty} e^{-x^2} dx \cdot \int_{-\infty}^{\infty} e^{-y^2} dy \\ &= \int_{\mathbb{R}^2} e^{-x^2} e^{-y^2} dx dy \\ &= \lim_{a \rightarrow \infty} \int_{K_a} e^{-(x^2+y^2)} dx dy, \end{aligned}$$

wobei der Satz von Fubini einmal anders herum verwendet wurde. Nach der Vorbemerkung (bzw. dem Transformationssatz) ist nun

$$\int_{K_a} e^{-(x^2+y^2)} dx dy = 2\pi \int_0^a e^{-r^2} r dr = \pi(1 - e^{-a^2})$$

(mit der Substitution $u = r^2$, $du = 2r dr$ ist $\int e^{-r^2} r dr = \frac{1}{2} \int e^{-u} du = -\frac{1}{2} e^{-u} = -\frac{1}{2} e^{-r^2}$). Also ist $I^2 = \lim_{a \rightarrow \infty} \pi(1 - e^{-a^2}) = \pi$ und damit $I = \sqrt{\pi}$.

Das Integral I hängt übrigens auch mit der *Gamma-Funktion* $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$ zusammen, die die Fakultät interpoliert (vgl. [Forster I], S. 156): Mit der Substitution $x^2 = t$, also $dt = 2x dx$ oder $dx = \frac{1}{2} t^{-1/2} dt$ gilt nämlich

$$I = 2 \int_0^\infty e^{-x^2} dx = \int_0^\infty t^{-1/2} e^{-t} dt = \Gamma(1/2).$$

Die Gleichung (*) zur Berechnung des Integrals radialer Funktionen im \mathbb{R}^2 läßt sich auf beliebige Dimensionen verallgemeinern:

62. Satz 26.5 (Integral radialer Funktionen) *Ist $g : [0, a] \rightarrow \mathbb{R}$ stetig und $f : K_a^n \rightarrow \mathbb{R}$, $f(x) = g(|x|)$, so gilt mit $\alpha_n := \mu(K_1^n)$:*

$$\int_{K_a^n} f = n \cdot \alpha_n \int_0^a r^{n-1} g(r) dr.$$

Beweis. Mit g ist auch $h : [0, a] \rightarrow \mathbb{R}$, $h(r) = r^{n-1} g(r)$ (sogar gleichmäßig) stetig. Es gibt also eine Zerlegung $0 = r_0 < r_1 < \dots < r_N = a$ von $I = [0, a]$ mit $I_j := [r_{j-1}, r_j]$, so dass für $j = 1, \dots, N$

$$\left(\sup_{I_j} g - \inf_{I_j} g \right) < \epsilon, \quad \left(\sup_{I_j} h - \inf_{I_j} h \right) < \epsilon. \quad (1)$$

Es sei $S_j = K_{r_j} \setminus B_{r_{j-1}}$ der Kreisring mit äußerem Radius r_j und innerem Radius r_{j-1} . Da $\sup_{I_j} g = \sup_{S_j} f$ und $\inf_{I_j} g = \inf_{S_j} f$, ist

$$\sum_j \inf_{I_j} g \cdot 1_{S_j} \leq f \leq \sum_j \sup_{I_j} g \cdot 1_{S_j}$$

und daher

$$\sum_j \inf_{I_j} g \cdot \mu(S_j) \leq \int_K f \leq \sum_j \sup_{I_j} g \cdot \mu(S_j),$$

mit $K := K_a^n$, also

$$\int_K f \approx_{\epsilon \mu(K)} \sum_j \inf_{I_j} g \cdot \mu(S_j). \quad (2)$$

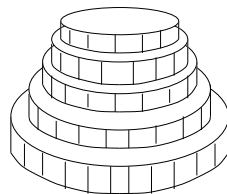


ABBILDUNG 49

(Abb. 49). Dabei ist $\mu(S_j) = \alpha_n(r_j^n - r_{j-1}^n)$. Nach dem Mittelwertsatz der Differentialrechnung, angewandt auf die Funktion $r \mapsto r^n$, ist

$$\frac{r_j^n - r_{j-1}^n}{r_j - r_{j-1}} = n\tilde{r}_j^{n-1}$$

für ein $\tilde{r}_j \in (r_{j-1}, r_j)$, also ist

$$\mu(S_j) = n\alpha_n\tilde{r}_j^{n-1}\mu(I_j).$$

Wegen (1) gilt

$$\inf_{I_j} g \cdot \tilde{r}_j^{n-1} \approx_{\epsilon_1} g(\tilde{r}_j) \cdot \tilde{r}_j^{n-1} = h(\tilde{r}_j) \approx_{\epsilon} \inf_{I_j} h$$

mit $\epsilon_1 := \epsilon a^{n-1}$. Also ergibt sich aus (2)

$$\int_{K_a^n} f \approx_{\epsilon_2\mu(K)} n\alpha_n \sum_j \inf_{I_j} h \mu(I_j) \approx_{\epsilon_3 a} n\alpha_n \int_0^a h$$

mit $\epsilon_2 := \epsilon_1 + \epsilon$ und $\epsilon_3 := n\alpha_n\epsilon$. Da ϵ und damit ϵ_2, ϵ_3 beliebig klein gewählt werden können, folgt die Behauptung. \square

Bemerkung. Der vorstehende Satz kann auch als Spezialfall des Transformationsatzes aufgefaßt werden, wobei die Transformation die Abbildung $\Phi : S \times (0, a) \rightarrow \mathbb{R}^n$, $\Phi(x, t) = tx$ mit $S = \{x \in \mathbb{R}^n; |x| = 1\}$ ist. Allerdings ist der Definitionsbereich von Φ nicht mehr eine offene Teilmenge des \mathbb{R}^n , sondern eine n -dimensionale Mannigfaltigkeit, nämlich $S \times (0, 1)$; der Transformationsatz kann aber auf diese Situation verallgemeinert werden (siehe übernächster Abschnitt).

27. Lineare Invarianz des Maßes

Mit \mathbf{K} bezeichnen wir die Menge aller kompakten messbaren Teilmengen von \mathbb{R}^n . Dann ist das *Maß* definiert als eine Abbildung

$$\mu : \mathbf{K} \rightarrow \mathbb{R}_+.$$

Ist $A : \mathbb{R}^n \rightarrow \mathbb{R}^n$ eine lineare Abbildung ($A \in \mathbb{R}^{n \times n}$), so ist mit jedem $K \in \mathbf{K}$ auch $AK \in \mathbf{K}$, wobei $AK = \{Ax; x \in K\}$. Wir wollen zeigen, dass $\mu(AK) = |\det A| \mu(K)$ für jedes $K \in \mathbf{K}$. Dazu kennzeichnen wir die Abbildung μ durch ihre Eigenschaften bis auf eine Konstante und zeigen, dass dieselben Eigenschaften auch für das durch A verzerrte Maß μ_A gelten:

$$\mu_A : \mathbf{K} \rightarrow \mathbb{R}_+, \quad \mu_A(K) := \mu(AK)$$

Dann müssen μ und μ_A bis auf einen positiven konstanten Faktor gleich sein, und diesen Faktor identifizieren wir als $|\det A|$.

Als kennzeichnende Eigenschaften kann man z.B. die folgenden wählen (vgl. Skriptum Analysis 1, Kap. 23):

- (1) **Verschiebungsinvarianz:** $\mu(T_v K) = \mu(K)$ für alle $K \in \mathbb{K}$ und $v \in \mathbb{R}^n$, wobei $T_v(x) = x + v$ die Verschiebung um v ist.
- (2) **Monotonie:** Sind $K, L \in \mathbb{K}$ mit $K \subset L$, so ist $\mu(K) \leq \mu(L)$,
- (3) **Additivität:** Ist $K, L \in \mathbb{K}$ und $A^\circ \cap B^\circ = \emptyset$, so ist $K \cup L \in \mathbb{K}$ und $\mu(K \cup L) = \mu(K) + \mu(L)$.

Die drei Eigenschaften ergeben sich aus den Eigenschaften des Riemann-Integrals oder ganz elementar aus der Definition der Maßes durch Zählen der (enthaltenen oder schneidenden) Würfeln. Aus dem letzteren Argument (Zählen der Würfel) sieht man auch, dass diese drei Eigenschaften das Maß bis auf eine Konstante (z.B. das Maß des Einheitswürfels) kennzeichnen: Für jede Abbildung $\tilde{\mu} : \mathbb{K} \rightarrow \mathbb{R}_+$ mit diesen drei Eigenschaften gibt es eine Konstante α mit

$$\tilde{\mu}(K) = \alpha \cdot \mu(K)$$

für alle $K \in \mathbb{K}$.

Wir sehen schnell ein, dass auch μ_A (wie oben definiert) die drei Eigenschaften erfüllt: Das ist klar für (2) und (3), und auch (1) ist erfüllt, da

$$\mu_A(T_v K) \stackrel{1}{=} \mu(AT_v K) \stackrel{2}{=} \mu(T_{Av} AK) \stackrel{3}{=} \mu(AK) \stackrel{4}{=} \mu_A(K),$$

wobei $\stackrel{1}{=}$ und $\stackrel{4}{=}$ nach Definition von μ_A gelten, $\stackrel{3}{=}$ wegen der Translationsinvarianz von μ und $\stackrel{2}{=}$ wegen

$$A(T_v(x)) = A(x + v) = Ax + Av = T_{Av}(Ax).$$

Es bleibt, für μ_A die Konstante D zu berechnen. Wir betrachten zunächst zwei Spezialfälle:

1. A ist Diagonalmatrix, $A = \text{diag}(\lambda_1, \dots, \lambda_n) = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}$. Dann gilt $Ae_i = \lambda_i e_i$ und

$$A([0, 1]^n) = [0, |\lambda_1|] \times \dots \times [0, |\lambda_n|],$$

also $\alpha = \mu(A([0, 1]^n)) = |\lambda_1| \cdot \dots \cdot |\lambda_n| = |\det A|$.

2. A ist orthonormale Matrix, $A^T A = I$. Da $\alpha = \mu_A(K)/\mu(K)$ für jedes $K \in \mathbb{K}$ mit $\mu(K) \neq 0$, wählen wir für K die Einheitskugel, $K_1 = \{x \in \mathbb{R}^n; |x| \leq 1\}$. Da $|Ax|^2 = Ax \cdot Ax = A^T Ax \cdot x = x \cdot x = |x|^2$, gilt $AK_1 = K_1$ und daher $\alpha = 1$, mit anderen Worten: *Das Maß ist invariant unter orthonormalen Abbildungen ("Drehungen")*. Andererseits ist auch $|\det A| = 1$, denn $1 = \det A^T A = \det A^T \det A = (\det A)^2$.

Jede Matrix A lässt sich aber als Produkt von Matrizen der Typen 1 und 2 schreiben, das ist die sog. *Singulärwertzerlegung*, die man wie folgt einsieht: Die Matrix $A^T A$ ist symmetrisch und besitzt deshalb eine Orthogonalbasis (b_1, \dots, b_n) aus Eigenvektoren: $A^T A b_i = \lambda_i b_i$. Für $i \neq j$ folgt $0 = \lambda_i b_i \cdot b_j = (A^T A b_i) \cdot b_j = (A b_i) \cdot (A b_j)$, was zeigt, dass die Vektoren $A b_i$ senkrecht aufeinander stehen und die auf Einheitslänge normierten Vektoren $c_i := A b_i / |A b_i|$ demnach eine Orthonormalbasis bilden. Somit ist $A b_i = \mu_i c_i$ mit $\mu_i^2 = |A b_i|^2 = (A^T A b_i) \cdot b_i = \lambda_i$. In Matrixschreibweise: $AB = CD$ mit $D = \text{diag}(\mu_1, \dots, \mu_n)$ oder $A = CDB^{-1} = \text{Orthonormal mal Diagonal mal Orthonormal}$ (denn mit B ist auch $B^{-1} = B^T$ orthonormal). Da B^{-1} und C das Maß erhalten (Typ 2), ist $\alpha = |\det D|$, und $|\det A| = |\det C| |\det D| |\det B^{-1}| = |\det D| = \alpha$, womit die Behauptung von Satz 26.2 vollständig bewiesen ist.

Alternativ dazu kann man eine Matrix auch als Produkt von Elementarmatrizen darstellen und gelangt zum selben Ergebnis.

28. Die Integralsätze von Gauß und Stokes

Satz 28.

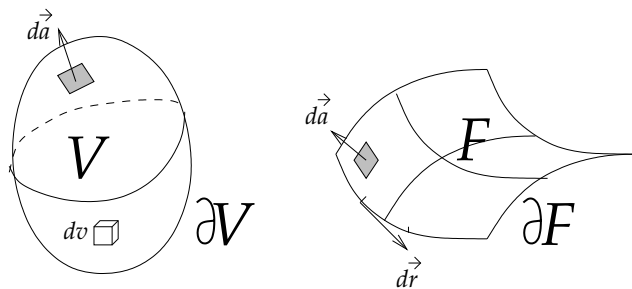
$$\begin{aligned} (1) \quad \int_V \text{div } \vec{E} \, dv &= \int_{\partial V} \vec{E} \cdot d\vec{a} \\ (2) \quad \int_F \text{rot } \vec{E} \cdot d\vec{a} &= \int_{\partial F} \vec{E} \cdot d\vec{r} \end{aligned}$$

Dabei ist $\vec{E} = (E_1, E_2, E_3)^T$ ein Vektorfeld, d.h. eine differenzierbare Abbildung $\vec{E} : \mathbb{R}_o^3 \rightarrow \mathbb{R}^3$, und

$$\begin{aligned} \text{div } \vec{E} &= \vec{\nabla} \cdot \vec{E} = D_1 E_1 + D_2 E_2 + D_3 E_3 \\ \text{rot } \vec{E} &= \vec{\nabla} \times \vec{E} = (D_2 E_3 - D_3 E_2, D_3 E_1 - D_1 E_3, D_1 E_2 - D_2 E_1)^T \end{aligned}$$

(*Divergenz* und *Rotation* von X), wobei $\vec{\nabla} = (D_1, D_2, D_3)^T$ den Vektor der partiellen Ableitungen nach den drei Variablen x_1, x_2, x_3 bezeichnet. Die Integrationsbereiche sind ein Raumgebiet V , seine umgebende Randfläche ∂V , ein beliebiges Flächenstück F und dessen Randkurve ∂F . Mit dv bezeichnen wir das Volumen eines kleinen Abschnittes von V (*Volumenelement*), mit $d\vec{a}$ den Flächeninhalt eines kleinen Abschnittes der Fläche ∂V oder F (*Flächenelement*), multipliziert mit dem Einheitsvektor \vec{N} , der auf diesem Flächenstück senkrecht steht (*Normalenvektor*), und mit $d\vec{r}$ die Länge eines kleinen Abschnittes der Randkurve ∂F (*Längenelement*), multipliziert mit dem Einheitsvektor

\vec{t} tangential an die Kurve ∂F .⁷ Die Flächen- und Kurvenabschnitte müssen so klein gewählt sein, dass das Flächen- bzw. Linienelement als eben bzw. gerade angesehen werden kann. Mit dem Malpunkt \cdot ist das Skalarprodukt bezeichnet. Die Integranden $\operatorname{div} \vec{E} dv$, $\vec{E} \cdot d\vec{a}$, $\operatorname{rot} \vec{E} \cdot d\vec{a}$ und $\vec{E} \cdot d\vec{r}$ sind also reelle Zahlen, und das Integral ist als Summe über diese Zahlen zu verstehen.



Der Beweis beider Sätze beruht auf zwei Ideen:

1. Die Integrationsbereiche lassen sich in beliebig kleine Teile zerlegen; das Gesamtintegral ist die Summe der Teilintegrale.
2. In kleinen Bereichen wird das Vektorfeld \vec{E} durch eine affine Abbildung (lineare Abbildung + Konstante) angenähert.

Die erste Idee lässt sich leicht auf die linken Seiten der beiden Gleichungen anwenden; das ganze Raumgebiet V bzw. die ganze Fläche F wird in kleine Teile zerlegt, Würfel (bei V) bzw. Polygone (bei F). Die Idee kann aber auch auf die rechten Seiten angewandt werden: Wenn wir über die Ränder der kleinen Würfel oder Polygone integrieren, so kommt abgesehen von den Außenseiten jede Seite in *zwei* benachbarten Würfeln oder Polygonen vor, über sie wird also zweimal integriert, aber mit unterschiedlichen Vorzeichen, deshalb heben die Integrale über die inneren Seiten sich gegenseitig auf, wenn man alles aufsummiert, und es bleiben nur die Integrale über die Außenseiten übrig.

Die zweite Idee haben wir in der Vorlesung bereits ausgeführt: Wenn $|\vec{h}|$ genügend klein ist, so gilt

$$(3) \quad \vec{E}(\vec{r} + \vec{h}) = \vec{E}(\vec{r}) + A\vec{h} + \vec{o}(\vec{h})$$

⁷Es gibt zwei mögliche Orientierungen von \vec{N} und \vec{t} , die man so festlegt: Auf ∂V soll \vec{N} nach außen weisen, und das Flächenstück F soll "auf der linken Seite" des Randes liegen, genauer: Wenn \vec{n} ein nach innen weisender Tangentenvektor von F ist, dann soll $(\vec{t}, \vec{n}, \vec{N})$ eine rechtshändige Basis des \mathbb{R}^3 sein.

wobei $A = Df_{\vec{r}}$ die 3×3 -Matrix ist, deren Spalten die drei partiellen Ableitungen von \vec{E} in \vec{r} sind, und $|\vec{o}(\vec{h})|/|\vec{h}|$ strebt für $\vec{h} \rightarrow 0$ gegen 0.⁸ In den kleinen Teilbereichen können wir $\vec{o}(\vec{h})$ einfach vernachlässigen und $\vec{E}(\vec{r} + \vec{h}) = \vec{E}(\vec{r}) + A\vec{h}$ setzen.⁹

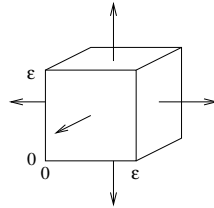
Zum Beweis von (1) ersetzen wir also das Raumgebiet V durch einen kleinen Würfel mit unterem Eckpunkt \vec{r}_o und Kantenlänge ϵ ,

$$W = \{\vec{r}_o + \vec{h}; \vec{h} = (h_1, h_2, h_3)^T, 0 \leq h_1, h_2, h_3 \leq \epsilon\}.$$

Wir wollen die rechte Seite $\int_{\partial W} \vec{E} \cdot d\vec{a}$ berechnen. Dabei ist

$$\int_{\partial W} \vec{E}(\vec{r}_o + \vec{h}) d\vec{a} \approx \int_{\partial W} (\vec{E}_o + A\vec{h}) d\vec{a}$$

mit $\vec{h} := \vec{r} - \vec{r}_o$. Das Integral über den konstanten Vektor $\vec{E}_o = \vec{E}(\vec{r}_o)$ ist Null, denn die Integrale über die Seitenflächen des Würfels treten immer paarweise mit unterschiedlichem Vorzeichen auf.



Es bleibt also $A\vec{h}$ über die Seiten des Würfels zu integrieren. Die Koeffizienten von A seien $a_{ij} = A\vec{e}_j \cdot \vec{e}_i$. Für die Seite $h_1 = \epsilon$ mit (nach außen weisendem) Normalvektor \vec{e}_1 ist der Integrand gleich $a_{11}\epsilon + a_{12}h_2 + a_{13}h_3$, denn $A\vec{h} \cdot \vec{e}_1 = h_1 A\vec{e}_1 \cdot \vec{e}_1 = a_{11}h_1 + a_{12}h_2 + a_{13}h_3$ mit $h_1 = \epsilon$. Für die Seite $h_1 = 0$ dagegen ist der Integrand $-(a_{12}h_2 + a_{13}h_3)$. Die Würfelseiten haben Flächeninhalt ϵ^2 ; die beiden Teilintegral ergeben also zusammen $\epsilon^3 a_{11}$. Ebenso ergeben die Beiträge der beiden anderen Seitenflächenpaare $\epsilon^3 a_{22}$ und $\epsilon^3 a_{33}$. Also erhalten wir insgesamt

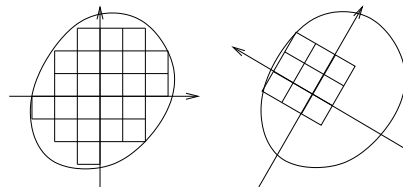
$$\int_{\partial W} \vec{E}(\vec{r}_o + \vec{h}) d\vec{a} \approx \epsilon^3 (a_{11} + a_{22} + a_{33}).$$

⁸Der Rest $\vec{o}(\vec{h})$ ist so klein, dass er selbst nach Multiplikation mit der großen Zahl $1/|\vec{h}|$ noch gegen Null geht für $\vec{h} \rightarrow 0$.

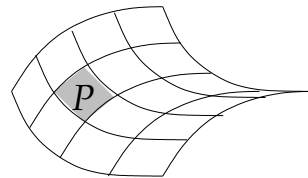
⁹Warum können wir die Integranden in den kleinen Teilbereichen nicht einfach als konstant annehmen? Für die linken Seiten wäre das in Ordnung, aber nicht für die rechten Seiten. Im Fall von (1) zerlegt man das Raumgebiet V etwa in Würfel mit Kantenlänge ϵ und Volumen ϵ^3 ; deren Anzahl ist $N \approx \text{vol}(V)/\epsilon^3$. Die Oberfläche jedes einzelnen Würfels ist $6\epsilon^2$, ihre Gesamtoberfläche also $6N\epsilon^2 = C/\epsilon$. Der Gesamtfehler beim Integrieren geht also genau dann gegen Null für $\epsilon \rightarrow 0$, wenn der Fehler des Integranden $o(\epsilon)$ ist mit $o(\epsilon)/\epsilon \rightarrow 0$. Bei (2) ist es ganz analog.

Der Faktor ϵ^3 ist das Volumen von W , der Ausdruck $a_{11} + a_{22} + a_{33}$ (die *Spur* der Matrix $A = (D_1\vec{E}, D_2\vec{E}, D_3\vec{E})$) ist nach Definition die *Divergenz* von \vec{E} im Punkt \vec{r}_o . Wenn wir alle Teilintegrale über die kleinen Würfel, aus denen V zusammengesetzt ist, aufaddieren, erhalten wir das Raumintegral über die Divergenz, d.h. die linke Seite von (1).

Man könnte gegen diesen Beweis einwenden, dass sich das gegebene Raumgebiet V in den meisten Fällen gar nicht gut in achsenparallele Würfel zerlegen lässt, weil deren Oberflächen nicht am Rand von V anliegen. Um diesem Einwand zu begegnen, zerlegt man das Vektorfeld \vec{E} in eine endliche Summe von Vektorfeldern \vec{E}_α , die alle am Rand entweder ganz verschwinden oder nur in einem kleinen Teil des Randes, der als eben angesehen werden kann, ungleich Null sind.¹⁰ Beide Seiten von (1) ändern sich nicht bei Drehungen des Koordinatensystems, man kann also die Koordinaten für jedes Teilvektorfeld an das betreffende Randstück anpassen.



Um (2) zu beweisen, unterteilen wir die krumme Fläche F annähernd in kleine ebene Polygone P :



Wieder gehen wir von der rechten Seite der gesuchten Gleichung (2) aus und ersetzen die ganze Fläche F durch eines der kleinen Parallelogramme P , wir wollen also $\int_{\partial P} \vec{E} \cdot d\vec{r}$ berechnen. Einer der Eckpunkte von P möge \vec{r}_o sein, und weil P klein ist, können wir dort $\vec{E}(\vec{r}_o + \vec{h}) \approx \vec{E}_o + A\vec{h}$ annehmen. Das Integral über den konstanten Vektor $\vec{E}_o = \vec{E}(\vec{r}_o)$ verschwindet, weil dieser ein Gradient ist, nämlich $\vec{E}_o = \vec{\nabla} f$ mit $f(\vec{r}) = \vec{E}_o \cdot \vec{r}$.

¹⁰Ein Beispiel für die Wahl eines solchen Summanden von \vec{E} ist $\vec{E}_\alpha(\vec{r}) = f_\alpha(\vec{r})\vec{E}(\vec{r})$, wobei f_α eine Funktion ist, die überall außerhalb einer kleinen offenen Menge verschwindet. Durch Addition mehrerer solcher Vektorfelder, die auf unterschiedlichen Mengen ungleich Null sind, kann man leicht \vec{E} zurückgewinnen.

Allgemein ist der *Gradient* $\vec{\nabla} f$ einer Funktion $f : \mathbb{R}_o^3 \rightarrow \mathbb{R}$ der Vektor der partiellen Ableitungen, $\vec{\nabla} f = (D_1 f, D_2 f, D_3 f)^T$. Für jede Kurve $\vec{c} : [a, b] \rightarrow \mathbb{R}^3$ ist $(f \circ \vec{c})'(t) = \vec{\nabla} f_{\vec{c}(t)} \cdot \vec{c}'(t)$ nach Kettenregel: äußere mal innere Ableitung. Damit ist $\int_a^b \vec{\nabla} f_{\vec{c}(t)} \cdot \vec{c}'(t) dt = \int_a^b (f \circ \vec{c})'(t) dt = f(\vec{c}(b)) - f(\vec{c}(a))$, und wenn die Kurve \vec{c} sich schließt, $\vec{c}(b) = \vec{c}(a)$, dann ist diese Differenz Null. Insbesondere ist $\int_{\partial P} \vec{\nabla} f \cdot d\vec{r} = 0$. In unserem Fall $f(\vec{r}) = \vec{E}_o \cdot \vec{r}$ ist $D_k f = \vec{E}_o \cdot \vec{e}_k$ und somit $\vec{\nabla} f = \vec{E}_o$.

Es bleibt $\int_{\partial P} A \vec{h} \cdot d\vec{r}$ mit $h = \vec{r} - \vec{r}_o$ zu berechnen. Jede Matrix A lässt sich in einen symmetrischen und einen antisymmetrischen Anteil zerlegen: Wir setzen $A_+ = A + A^T$ und $A_- = A - A^T$; offensichtlich gilt $A_+^T = A_+$ und $A_-^T = -A_-$ und $A = \frac{1}{2}(A_+ + A_-)$. Wir bemerken zunächst, dass der symmetrische Anteil beim Integrieren verschwindet:

$$\int_{\partial P} A_+ \vec{h} \cdot d\vec{r} = 0.$$

Der Grund dafür ist wieder, dass das Vektorfeld $\vec{F}(\vec{x}) = A_+ \vec{x}$ ein *Gradientenfeld* ist, der Gradient der Funktion $f(\vec{x}) = A \vec{x} \cdot \vec{x} = \sum_{ij} a_{ij} x_i x_j$, denn $D_k f = \sum_j a_{kj} x_j + \sum_i a_{ik} x_i = \sum_j (a_{kj} + a_{jk}) x_j$ und $\vec{\nabla} f = A_+ \vec{x}$.¹¹

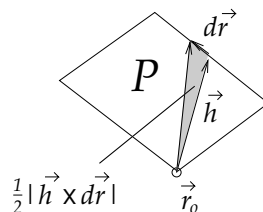
Es bleibt also nur der A_- -Anteil, und dieser ist nach Definition das Kreuzprodukt mit der Rotation von \vec{E} an der Stelle \vec{r}_o ,¹²

$$A_- \vec{h} \cdot d\vec{r} = (\text{rot } \vec{E} \times \vec{h}) \cdot d\vec{r} = \text{rot } \vec{E} \cdot (\vec{h} \times d\vec{r}) = |\vec{h} \times \partial \vec{r}| \text{rot } \vec{E} \cdot \vec{N},$$

wobei \vec{N} der Einheitsnormalenvektor auf dem ebenen Flächenstück P ist. Das Skalarprodukt $\text{rot } \vec{E} \cdot \vec{N}$ ist konstant auf P und $\int_{\partial P} |\vec{h} \times \partial \vec{r}|$ ist der doppelte Flächeninhalt $2A(P)$ (siehe Figur).

¹¹Alternative: $f(\vec{x} + \vec{h}) = A(\vec{x} + \vec{h}) \cdot (\vec{x} + \vec{h}) = (A\vec{x} + A\vec{h}) \cdot (\vec{x} + \vec{h}) = A\vec{x} \cdot \vec{x} + A\vec{x} \cdot \vec{h} + A\vec{h} \cdot \vec{x} + A\vec{h} \cdot \vec{h} = f(\vec{x}) + B\vec{h} + o(\vec{h})$ mit dem linearen Anteil $B\vec{h} = A\vec{x} \cdot \vec{h} + A\vec{h} \cdot \vec{x} = A\vec{x} \cdot \vec{h} + \vec{h} \cdot A^T \vec{x} = (A + A^T)\vec{x} \cdot \vec{h} = A_+ \vec{x} \cdot \vec{h}$ sowie dem Rest $o(\vec{h}) = A\vec{h} \cdot \vec{h}$. Da $\vec{\nabla} f = Df^T = B^T$, folgt $\vec{\nabla} f \cdot \vec{h} = B\vec{h} = A_+ \vec{x} \cdot \vec{h}$ und daher $\vec{\nabla} f = A_+ \vec{x}$.

¹²Das Kreuzprodukt mit einem Vektor $\vec{v} = (a, b, c)^T$, d.h. die lineare Abbildung $\vec{h} \mapsto \vec{v} \times \vec{h}$ hat die Matrix mit den Spalten $\vec{v} \times \vec{e}_1 = \begin{pmatrix} 0 \\ c \\ -b \end{pmatrix}$, $\vec{v} \times \vec{e}_2 = \begin{pmatrix} -c \\ 0 \\ a \end{pmatrix}$ und $\vec{v} \times \vec{e}_3 = \begin{pmatrix} b \\ -a \\ 0 \end{pmatrix}$; die Matrix ist also $\begin{pmatrix} 0 & -c & b \\ c & 0 & -a \\ -b & a & 0 \end{pmatrix}$, aber *jede* antisymmetrische 3×3 -Matrix ($a_{ji} = -a_{ij}$) ist von dieser Form.



Der antisymmetrisch Anteil von A ist $\frac{1}{2}A_-$, daher ist

$$\int_{\partial P} \vec{E} \cdot d\vec{r} = \frac{1}{2} \int_{\partial P} A_- \vec{h} \cdot d\vec{r} = A(P) \operatorname{rot} \vec{E}(\vec{r}_0) \cdot \vec{N}.$$

Aufaddieren dieser Werte über alle kleinen Polygone, aus denen die Fläche F zusammengesetzt ist, ergibt die linke Seite von (2).

29. Integration über Mannigfaltigkeiten und Divergenzsatz

Für ein C^1 -Vektorfeld $v = (v_1, \dots, v_n)^T : U^n \rightarrow \mathbb{R}^n$ ist die *Divergenz* die stetige Funktion $\operatorname{div} v : U^n \rightarrow \mathbb{R}$,

$$\operatorname{div} v(x) = \sum_{i=1}^n \frac{\partial v_i(x)}{\partial x_i} = \operatorname{Spur} Dv_x.$$

63. Satz 29 (Divergenzsatz) *Es sei $G \subset \mathbb{R}^n$ offen mit glattem Rand, d.h. ∂G sei ein C^1 -Hyperfläche, und $\bar{G} = G \cup \partial G$ sei kompakt. Dann gilt für jedes C^1 -Vektorfeld v auf einer Umgebung von \bar{G} :*

$$\int_G \operatorname{div} v \, d^n x = \int_{\partial G} \langle v, N \rangle d^{n-1} x,$$

wobei $N : \partial G \rightarrow \mathbb{R}^n$ das nach außen weisende Einheitsnormalenfeld ist (Abb. 50).

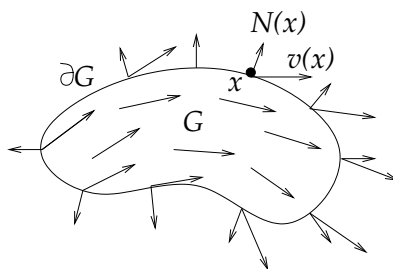


ABBILDUNG 50

Auf der rechten Seite der gewünschten Gleichung steht etwas Neues, nämlich das *Oberflächenintegral* über die Hyperfläche ∂G . Wie ist

dieses definiert? Zunächst sei $M \subset \mathbb{R}^n$ eine *m-dimensionale parametrisierte Mannigfaltigkeit* für $m < n$, d.h. es gibt einen Diffeomorphismus $\Psi : U^m \rightarrow V^n$ mit $\Psi(U^m \cap \mathbb{R}^m) = M$ (das ist i.a. nur lokal erfüllt; vgl. §§15 und 20). Die Abbildung $\psi : U^m := U^m \cap \mathbb{R}^m \rightarrow M \subset \mathbb{R}^n$ heißt *Parametrisierung* (Abb. 51); Parametrisierungen sind die Umkehrfunktionen von Karten (s. §15).

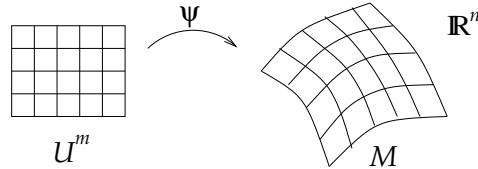


ABBILDUNG 51

Für eine stetige Funktion mit kompaktem Träger $f : M \rightarrow \mathbb{R}$ definiert man (analog zum Transformationssatz)

$$\int_M f := \int_M f(x) d^m x := \int_{U^m} (f(\psi(u)) | \det D\psi_u | d^m u. \quad (1)$$

(Wir schreiben " $d^m x$ " ebenso wie " $d^m u$ ", um deutlich zu machen, dass es sich um ein m -dimensionales Integral handelt. Als Teilmenge von \mathbb{R}^n betrachtet wäre M ja eine Nullmenge!) Der Ausdruck $| \det D\psi_u |$ in (1) ist allerdings zunächst gar nicht definiert, denn $D\psi_u : \mathbb{R}^m \rightarrow \mathbb{R}^n$ ist eine $n \times m$ -Matrix, und die Determinante ist nur für $n \times n$ -Matrizen definiert. Allerdings können wir doch noch wenigstens den *Absolutbetrag der Determinante* für eine beliebige $n \times m$ -Matrix A definieren, indem wir nämlich setzen:

$$| \det A | := \sqrt{\det A^T A}.$$

$A^T A$ ist eine $m \times m$ -Matrix, also ist die rechte Seite definiert;

$$\det A^T A = \det ((\langle Ae_i, Ae_j \rangle))$$

heißt die *Gram'sche Determinante von A*. Insbesondere ist also

$$| \det D\psi_u | = \sqrt{\det ((\langle \frac{\partial \psi(u)}{\partial u_i}, \frac{\partial \psi(u)}{\partial u_j} \rangle))}.$$

Der Transformationssatz läßt sich sofort auf diese Situation übertragen (Übung): Ist $\Phi : M \rightarrow \tilde{M}$ ein Diffeomorphismus von parametrisierten Mannigfaltigkeiten M und \tilde{M} (d.h. $\tilde{\psi}^{-1} \circ \Phi \circ \psi$ ist Diffeomorphismus, wobei ψ und $\tilde{\psi}$ die Parametrisierungen von M und \tilde{M} sind), so gilt für

jede stetige Funktion $f : \tilde{M} \rightarrow \mathbb{R}$ mit kompaktem Träger:

$$\int_{\tilde{M}} f = \int_M ((f \circ \Phi) |\det D\Phi|),$$

wobei $|\det D\Phi_x|$ für $D\Phi_x : T_x M \rightarrow T_{\Phi(x)} \tilde{M}$ wieder mit Hilfe der Determinante der linearen Abbildung $(D\Phi_x)^T \circ D\Phi_x : T_x M \rightarrow T_x M$ definiert wird, also

$$|\det D\Phi_x| := \sqrt{\det((D\Phi_x)^T \circ D\Phi_x)}.$$

Eine andere Möglichkeit der Berechnung (die hieraus sofort folgt), ist es, die lineare Abbildung $D\Phi_x$ als Matrix bezüglich Orthonormalbasen (e_i) von $T_x M$ und (\tilde{e}_j) von $T_{\Phi(x)} \tilde{M}$ darzustellen:

$$|\det D\Phi_x| = |\det (\langle D\phi_x e_i, \tilde{e}_j \rangle)|.$$

Hat man nun eine *beliebige* m -dimensionale Mannigfaltigkeit M und eine stetige Funktion $f : M \rightarrow \mathbb{R}$ mit kompaktem Träger vorliegen, dann läßt sich f stets in eine Summe $f = f_1 + \dots + f_k$ zerlegen, so dass jedes f_j stetig ist und seinen Träger in einer relativ offenen Teilmenge M_j von M hat, die eine *parametrisierte* Mannigfaltigkeit ist (vgl. z.B. [Forster III], S. 138f). Damit kann man das Integral über f als Summe der Integrale über die f_j definieren. Ist M kompakt, so läßt sich insbesondere die Konstante 1 integrieren; $\mu_m(M) := \int_M 1 d^m x$ heißt das *m -dimensionale Maß* von M ; ist $m = n - 1$, heißt das $(m - 1)$ -dimensionale Maß auch *Oberfläche*.

In den meisten Fällen ist die Zerlegung von M in der Praxis überflüssig, da M bereits selber bis auf eine μ_m -Nullmenge eine parametrisierte Mannigfaltigkeit ist.

Für den Beweis von Satz 29 benutzen wir ohne Beweis die Tatsache, dass jedes Vektorfeld einen *Fluss* besitzt. Der *Fluss* eines C^1 -Vektorfeldes $v : U^n \rightarrow \mathbb{R}^n$ ist eine C^1 -Abbildung $(x, t) \mapsto \phi_t(x)$, definiert auf einer Umgebung von $U^n \times \{0\}$ in $U^n \times \mathbb{R}$, mit

$$\phi_0(x) = x, \quad \frac{d}{dt} \phi_t(x) = v(\phi_t(x)) \quad (2)$$

für alle $(x, t) \in U^n \times \mathbb{R}$, für die $\phi_t(x)$ definiert ist (Abb. 52). Man erhält den Fluss $(\phi_t)_{t \in \mathbb{R}}$ durch Lösen der Differentialgleichung (2) für die Kurve $t \mapsto \phi_t(x)$ (vgl. Bemerkung in §12); allerdings muß man noch die C^1 -Abhängigkeit vom Anfangswert x nachweisen (vgl. z.B. [Heuser], §13).

Wegen des Eindeutigkeitsatzes für Lösungen von Differentialgleichungen gilt $\phi_t \circ \phi_s = \phi_{t+s}$ überall, wo beide Seiten definiert sind; insbesondere ist ϕ_t Diffeomorphismus mit Umkehrabbildung $\phi_t^{-1} = \phi_{-t}$. Auch

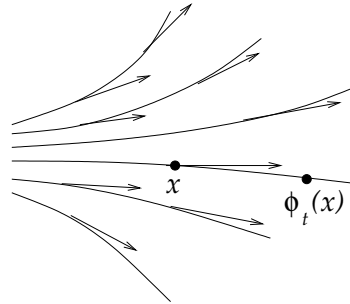


ABBILDUNG 52

wenn wir diese Tatsache hier nicht beweisen, ist sie doch für die Vorstellung sehr hilfreich: Man sollte sich ein Vektorfeld als das Geschwindigkeitsvektorfeld eines (realen) Flusses vorstellen; $v(x)$ gibt also an jeder Stelle x Richtung und Geschwindigkeit des Flusses an, und $t \mapsto \phi_t(x)$ beschreibt den Ort zur Zeit t eines vom Fluss mitgeschwemmten Teilchens, das zur Zeit 0 am Ort x war.

Beweisskizze von Satz 29: Es sei $(\phi_t)_{t \in \mathbb{R}}$ der Fluss von v . Da v in einer Umgebung der kompakten Menge \bar{G} definiert ist, ist ϕ_t für kleine $|t|$ (sagen wir: für $|t| < \epsilon$) auf ganz \bar{G} definiert; wir setzen $G_t := \phi_t(G)$. Wir dürfen ohne Beschränkung der Allgemeinheit annehmen, dass $v|_{\partial G}$ nach außen weist, also $\langle v, N \rangle > 0$ und damit $G_t \supset G$ (Abb. 53), denn wir können ein beliebiges Vektorfeld v zerlegen als $v = v' - v''$, wobei $v'|_{\partial G}$ und $v''|_{\partial G}$ nach außen weisen.

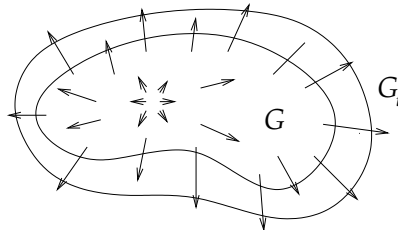


ABBILDUNG 53

Jedes G_t ist Jordan-messbar, da der Rand eine kompakte Hyperfläche und damit eine Nullmenge in \mathbb{R}^n ist (vgl. 23.2). Wir berechnen nun $\frac{d}{dt}\mu(G_t)|_{t=0}$ auf zwei Weisen:

- (1) Wir berechnen das Maß von $G_t = \phi_t(G)$ mit Hilfe des Transformationssatzes,
- (2) wir berechnen das Maß des “Überschusses” $G_t \setminus G$, der durch den Rand ∂G nach außen geflossen ist.

Deutet man (1) als das Volumen, das durch den expandierenden Fluss neu entstanden ist und (2) als dasjenige, das durch ∂G abfließt, so läßt sich der Beweis “hydrodynamisch” interpretieren: *Was im Inneren von G neu entsteht, fließt durch den Rand von G ab.* Wir betrachten nun die beiden Berechnungsarten im Einzelnen.

Zu (1): Nach dem Transformationssatz ist $\mu(\Phi_t(G)) = \int_G |\det D\phi_t|$. Die Determinante ist automatisch positiv (wir können also die Betragsstriche weglassen), da $\det D(\phi_0)_x = \det I = 1$ und $t \mapsto \det D(\phi_t)_x$ stetig ist und nicht Null werden kann. Bei der Berechnung der Ableitung nach t dürfen wir Integral und Ableitung vertauschen, d.h. unter dem Integral differenzieren, was wir auch nicht bewiesen haben (Übung!). Die Determinante der Matrix $D\phi_t$ kann als eine Art Produkt der Spaltenvektoren $D_1\phi_t, \dots, D_n\phi_t$ aufgefaßt und deshalb nach der Produktregel differenziert werden:

$$\begin{aligned} & \left. \frac{d}{dt} \right|_0 \det(D_1\phi_t(x), \dots, D_n\phi_t(x)) \\ &= \sum_{i=1}^n \det \left(D_1\phi_t(x), \dots, \frac{d}{dt} D_i\phi_t(x), \dots, D_n\phi_t(x) \right) \Big|_{t=0} \\ &= \sum_{i=1}^n \det(e_1, \dots, D_i v(x), \dots, e_n) \\ &= \text{Spur } Dv_x = \text{div } v(x). \end{aligned}$$

Somit erhalten wir:

$$\left. \frac{d}{dt} \right|_0 \mu(G_t) = \int_G \left. \frac{d}{dt} \right|_0 \det(D\phi_t) = \int_G \text{div } v.$$

Dies ist die linke Seite der gesuchten Gleichung.

Zu (2): Wir können $\mu(G_t \setminus G)$ mit Hilfe des Transformationssatzes für die Abbildung

$$\Phi : \partial G \times (-\epsilon, \epsilon) \rightarrow \mathbb{R}^n, \quad \Phi(x, t) = \phi_t(x)$$

berechnen, da $G \setminus G_t = \Phi_t(\partial G \times [0, t])$ für $0 < t < \epsilon$. Dabei ist $|\det D\Phi_{(x,0)}|$ zu berechnen für die lineare Abbildung $D\Phi_{(x,0)} : T_x \partial G \oplus \mathbb{R} \rightarrow \mathbb{R}^n$. Wir wählen dazu eine Orthonormalbasis e_1, \dots, e_{n-1} von $T_x \partial G$ und ergänzen sie durch $e_n := N(x)$ zu einer Orthonormalbasis von \mathbb{R}^n . Dann ist $D\Phi_{(x,0)} e_i = e_i$ für $i = 1, \dots, n-1$ (da $\Phi|_{\partial G \times \{0\}} = \phi_0|_{\partial G}$) und

$$\left. \frac{\partial \Phi(x, t)}{\partial t} \right|_{t=0} = \left. \frac{d}{dt} \right|_0 \phi_t(x) = v(x),$$

der Einheitsvektor des Faktors \mathbb{R} in der direkten Summe $T_x \partial G \oplus \mathbb{R}$ (häufig als $\frac{\partial}{\partial t}$ bezeichnet) wird also durch $D\Phi_{(x,0)}$ auf $v(x)$ abgebildet.

Also ist

$$|\det D\Phi_{(x,0)}| = |\det(e_1, \dots, e_{n-1}, v(x))| = \langle v(x), e_n \rangle,$$

und wir erhalten

$$\begin{aligned} \frac{d}{dt} |_0 \mu(G_t) &= \frac{d}{dt} |_0 \mu(G_t \setminus G) \\ &= \frac{d}{dt} |_0 \int_0^t \int_{\partial G} |\det D\Phi_{(x,t)}| d^{n-1}x dt \\ &= \int_{\partial G} |\det D\Phi_{(x,0)}| d^{n-1}x \\ &= \int_{\partial G} \langle v(x), N(x) \rangle d^{n-1}x. \end{aligned}$$

Dies ist die rechte Seite der gesuchten Gleichung.

Beispiel: Oberfläche der Kugel. Es sei $G = B_a = \{x \in \mathbb{R}^n; |x| < a\}$ und $v = id : \mathbb{R}^n \rightarrow \mathbb{R}^n$, d.h. $v(x) = x$ für alle $x \in \mathbb{R}^n$. Dann ist $Dv_x = I$ für alle x , also gilt $\operatorname{div} v = \operatorname{Spur} I = n$. Somit ist die linke Seite des Divergenzsatzes:

$$\int_G \operatorname{div} v d^n x = n \cdot \mu(G) = na^n \alpha_n,$$

wobei wie früher $\alpha_n = \mu(B_1^n) = \mu(K_1^n)$ das Maß des Einheitsballes bezeichnet. Andererseits ist

$$N(x) = \frac{1}{a} x$$

das nach außen weisende Einheitsnormalenfeld auf der Kugel $S_a = \partial B_a$, also ist $\langle v(x), N(x) \rangle = \frac{1}{a} \langle x, x \rangle = a^2/a = a$ für alle $x \in S_a$. Somit ergibt die rechte Seite des Divergenzsatzes

$$\int_{\partial G} \langle v, N \rangle d^{n-1}x = a \cdot \mu_{n-1}(S_a),$$

wir erhalten also aus dem Divergenzsatz die Oberfläche der Kugel vom Radius a :

$$\mu_{n-1}(S_a) = n \cdot a^{n-1} \cdot \alpha_n.$$

In §25, Beispiel 3 hatten wir einige der α_n berechnet; die Oberfläche der Einheitskugel in \mathbb{R}^n ist demnach $2 \cdot \alpha_2 = 2\pi$ für $n = 2$, $3 \cdot \alpha_3 = 4\pi$ für $n = 3$, $4 \cdot \alpha_4 = 2\pi^2$ für $n = 4$ und $5 \cdot \alpha_5 = 8\pi^2/3$ für $n = 5$.

(Eine andere Idee, Volumen und Oberfläche der Kugel miteinander in Verbindung zu bringen, ist die Zerlegung der Kugel in Kugelschalen von Radius r und Dicke dr , deren Volumen mit $\mu_{n-1}(S_1) \cdot dr$ angesetzt wird. Damit ergibt sich $\mu(B_a) = \int_0^a \mu_{n-1}(S_r) dr = \mu_{n-1}(S_1) \int_0^a r^{n-1} dr =$

$\mu_{n-1}(S_r)a^n/n = \mu_{n-1}(S_a)a/n$. Hierzu ist das Cavalierische Prinzip auf Mannigfaltigkeiten, nämlich äquidistante Hyperflächenfamilien zu verallgemeinern.)

Bemerkung. Der Divergenzsatz läßt sich mit demselben Beweis auf Mannigfaltigkeiten übertragen:

64. Divergenzsatz auf Mannigfaltigkeiten: Es sei $M \subset \mathbb{R}^N$ eine n -dimensionale Mannigfaltigkeit und $G \subset M$ eine relativ offene Teilmenge mit glattem relativen Rand, d.h. $\partial_M G \subset M$ ist eine $(n-1)$ -dimensionale C^1 -Mannigfaltigkeit, und $\bar{G} = G \cup \partial_M G$ sei kompakt. Dann gilt für jedes tangentielle C^1 -Vektorfeld v auf M :

$$\int_G \operatorname{div} v \, d^n x = \int_{\partial_M G} \langle v, n \rangle d^{n-1} x,$$

wobei $n : \partial_M G \rightarrow \mathbb{R}^N$ das nach außen wiesende Einheitsnormalenfeld ist, d.h. $n(x) \in T_x M$ und $n(x) \perp T_x \partial G$.

In dieser Form umfaßt der Divergenzsatz alle anderen Integralsätze, z.B. den klassischen Satz von Stokes:

65. Satz von Stokes Es sei $F \subset \mathbb{R}^3$ eine Fläche mit Einheitsnormalenfeld $N : F \rightarrow \mathbb{R}^3$, $G \subset F$ eine relativ offene Teilmenge, deren (relativer) Rand $\partial_F G$ in F eine eindimensionale Mannigfaltigkeit ist, und $G \cup \partial_F G$ sei kompakt. Es sei $v : U^3 \rightarrow \mathbb{R}^3$ ein C^1 -Vektorfeld, definiert auf einer offenen Umgebung von F . Dann gilt

$$\int_G \langle \operatorname{rot} v(x), N(x) \rangle d^2 x = \int_{\partial_F G} \langle v(x), dx \rangle.$$

Dabei bedeutet die rechte Seite das Kurvenintegral von v über die Kurve $\partial_F G$, also $\int_I \langle v(\gamma(t)), \gamma'(t) \rangle dt$ für eine Parametrisierung $\gamma : I \rightarrow \partial_F G$ von $\partial_F G$, für die G in Laufrichtung links von γ liegt, wenn man von "oben" (von der Richtung, in die N zeigt) auf die Fläche sieht (Abb.54).

Die linke Seite enthält die *Rotation* von v ; dies ist das stetige Vektorfeld

$$\operatorname{rot} v = \nabla \times v = \begin{pmatrix} \partial_2 v_3 - \partial_3 v_2 \\ \partial_3 v_1 - \partial_1 v_3 \\ \partial_1 v_2 - \partial_2 v_1 \end{pmatrix}.$$

(Ist v ein Vektorfeld in \mathbb{R}^n für beliebige Dimension n , so ist $\operatorname{rot} v$ eine schiefsymmetrische Matrix, nämlich $\operatorname{rot} v = Dv - (Dv)^T$ mit den Komponenten $(\operatorname{rot} v)_{ij} = \partial_i v_j - \partial_j v_i$, aber in Dimension $n = 3$ gibt das Kreuzprodukt einen Isomorphismus $x \mapsto A_x$, $A_x v := x \times v$ zwischen \mathbb{R}^3 und den schiefsymmetrischen 3×3 -Matrizen.) Der Stokes'sche Satz ist

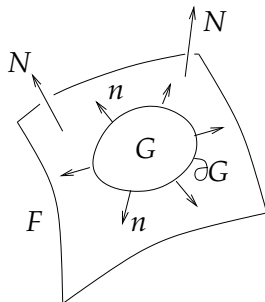


ABBILDUNG 54

der Divergenzsatz auf F für das tangentielle Vektorfeld $w = v \times N$ auf F . Ausgedrückt in der symbolischen Sprache der Physiker, die $\nabla = (\partial_1, \partial_2, \partial_3)^T$ wie einen Vektor in \mathbb{R}^3 behandeln, ist nämlich

$$\operatorname{div} w = \langle \nabla, w \rangle = \langle \nabla, v \times N \rangle = \langle \nabla \times v, N \rangle = \langle \operatorname{rot} v, N \rangle.$$

Auf der anderen Seite ist die Abbildung $w \mapsto w \times N$ die 90-Grad-Drehung nach rechts (von "oben" gesehen) in der Ebene $T_x F = N(x)^\perp$. Ist also $\gamma : [0, L] \rightarrow \partial_F G$ eine Parametrisierung von $\partial_F G$ nach Bogenlänge, so ist $\gamma'(t)$ ein Einheitsvektor in $T_{\gamma(t)} F$ tangential zu $\partial_F G$ und $n(t) := \gamma'(t) \times N(\gamma(t)) \in T_{\gamma(t)} F$ ist ein Einheitsvektor tangential zu F und senkrecht zu $\partial_F G$, der bei der vorgeschriebenen Orientierung von γ nach rechts, also in das Äußere von G weist. Längs γ gilt:

$$\langle v, \gamma' \rangle = \langle v \times N, \gamma' \times N \rangle = \langle w, n \rangle,$$

und daher folgt der Satz von Stokes aus dem Divergenzsatz.

Der klassische Satz von Stokes besitzt eine sehr elegante Verallgemeinerung auf beliebige Dimensionen, die äquivalent zum Divergenzsatz auf Mannigfaltigkeiten ist. Dabei wird die Rotation, die ja eigentlich eine antisymmetrische Bilinearform ist, durch eine antisymmetrische Multilinearform (*Differentialform*) ersetzt, vgl. [Forster III] oder [Spivak].

Anhang: Das Lebesgue-Integral

A1. Mängel des Riemannsches Integralbegriffs

Das Riemannsches Integral ist völlig ausreichend, solange man es nur mit der Integration *einer* Funktion zu tun hat. Probleme gibt es erst mit Funktionsfolgen: Nur bei gleichmäßiger Konvergenz wissen wir, dass Limes und Integral vertauschen. Das ist in vielen Situationen nicht ausreichend. Die einfachste solche Situation haben wir bereits bei uneigentlichen Integralen wie $\int_0^1 (1/\sqrt{x})dx$ oder $\int_1^\infty x^2 dx$ angetroffen: Das erste dieser beiden Integrale läßt sich z.B. als $\lim_{k \rightarrow \infty} \int f_k$ deuten für die Funktionenfolge $f_k : (0, 1) \rightarrow \mathbb{R}$ mit $f_k(x) = 1/\sqrt{x}$ für $1/k < x < 1$ und $f_k(x) = 0$ für $0 < x \leq 1/k$. Die Limesfunktion $f(x) = 1/\sqrt{x}$ ist nicht beschränkt und daher nicht Riemann-integrierbar, und die Konvergenz $f_k \rightarrow f$ ist nicht gleichmäßig; dennoch definieren wir ein “uneigentliches” Integral für f mit dem Wert $\int_0^1 f = \lim \int_0^1 f_k = \frac{1}{2}\sqrt{x}|_0^1 = 1/2$.

Man könnte meinen, dass es nur die Unbeschränktheit der Funktion oder des Integrationsbereiches ist, die hier Probleme macht, aber das ist nicht so, wie das nächste Beispiel zeigt, in dem wir die charakteristische Funktion einer offenen Teilmenge des Intervalls $(0, 1)$ betrachten: Es sei $\{q_1, q_2, \dots\}$ eine Abzählung der rationalen Zahlen in $(0, 1)$ und

$$U_k = \bigcup_{j=1}^k (q_j - \epsilon/2^j, q_j + \epsilon/2^j) \cap (0, 1)$$

$$U = \bigcup_{j=1}^{\infty} (q_j - \epsilon/2^j, q_j + \epsilon/2^j) \cap (0, 1)$$

Dann ist U_k Jordan-messbar bzw. $f_k = 1_{U_k}$ Riemann-integrierbar mit

$$\int f_k = \mu(U_k) \leq \sum_{j=1}^k 2\epsilon/2^j \leq 2\epsilon.$$

Aber $f = 1_U = \lim f_k$ ist nicht Riemann-integrierbar bzw. $U = \bigcup U_k$ nicht Jordan-messbar, denn da U dicht in $(0, 1)$ ist (U enthält ja alle rationalen Zahlen in $(0, 1)$), ist das äußere Maß $\int^* f = \mu^*(U) = 1$, während für das innere Maß gilt:

$$\int_* f = \mu_*(U) = \lim \mu(U_k) \leq 2\epsilon.$$

Beweis: Um das innere Maß zu berechnen, müssen wir U von innen durch endlich viele kompakte Intervalle ausschöpfen. Ist A solch ein kompaktes Intervall mit $A \subset U$, so gibt aber bereits ein $k \in \vec{N}$ mit

$A \subset U_k$, denn andernfalls hätte die Folge $(A \setminus U_k)$ von immer kleiner werdenden kompakten Mengen einen nichtleeren Durchschnitt, also $A \setminus U \neq \emptyset$ und somit $A \not\subset U$. Eine beliebige endliche Vereinigung von kompakten Intervallen $A_i \subset U$ ist daher bereits in einem der U_k enthalten und damit $\mu(\bigcup_i A_i) \leq \mu(U_k)$. Da wir $\mu_*(U)$ durch $\mu(\bigcup_i A_i)$ approximieren, folgt $\mu_*(U) = \lim \mu(U_k)$.

Beiden beschriebenen Situationen ist gemeinsam, dass die Funktionsfolge (f_k) *monoton* gegen f konvergiert (" $f_k \nearrow f$ ") und die Folge der Integrale $(\int f_k)$ beschränkt und natürlich auch monoton ist ($f_{k+1} \geq f_k \Rightarrow \int f_{k+1} \geq \int f_k$) und deshalb konvergiert. In einer solchen Situation würden wir gerne auch die Limesfunktion $f = \lim f_k$ integrierbar nennen mit $\int f = \lim \int f_k$. Um dies zu erreichen, muß der Begriff der Integrierbarkeit ausgeweitet werden.

Monotone Konvergenz sollte allerdings auch nicht das entscheidende Kriterium für Integrierbarkeit sein. Die Monotonie geht durch eine geringe Abänderung der f_k bereits verloren, ohne dass sich dies auf die Konvergenz auswirken muß. Das richtige Kriterium ist, anschaulich gesprochen, dass zwischen den Graphen von f_{k+1} und f_k , besser noch zwischen den Graphen von f_{k+m} und f_k (für beliebige m) ein sehr kleines Volumen liegt, wenn k groß wird. In Formeln heißt das:

$$\forall \epsilon > 0 \exists N \forall k, l \geq N \int |f_k - f_l| < \epsilon.$$

Eine solche Folge von integrierbaren Funktionen (f_k) werden wir L^1 -*Cauchyfolge* nennen, und der ausgeweitete Integralbegriff von Lebesgue wird gerade so gemacht sein, dass die Limesfunktionen solcher Folgen, $f = \lim f_k$, auch integrierbar sind mit $\int f = \lim \int f_k$. Für alle Theorien, die sich mit Funktionenräumen beschäftigen, zum Beispiel Wahrscheinlichkeitstheorie (Raum der Wahrscheinlichkeitsverteilungen), Physik (Raum der Wellenfunktionen) oder Variationsrechnung (Raum der Funktionen, auf denen ein gegebenes Funktional minimiert werden soll), ist dieser erweiterte Integralbegriff unentbehrlich.

Wir werden sehen, dass wir zunächst sogar einen Schritt zurück gehen und statt aller Riemann-integrierbaren Funktionen nur die stetigen mit kompaktem Träger betrachten dürfen, also die stetigen Funktionen auf \mathbb{R}^n , die außerhalb einer kompakten Menge Null sind. (Der *Träger* einer Funktion ist der Abschluß der Menge, auf der die Funktion nicht Null ist.) Der Vektorraum aller stetigen Funktionen mit

kompaktem Träger auf \mathbb{R}^n wird mit $C_0(\mathbb{R}^n)$ bezeichnet. Die Lebesgue-integrierbaren Funktionen (darunter alle Riemann-integrierbaren) werden genau die Limiten von L^1 -Cauchyfolgen in $C_0(\mathbb{R}^n)$ sein. Der Vorteil dieser Einschränkung ist, dass solche L^1 -Cauchyfolgen tatsächlich richtige Cauchyfolgen für eine Norm auf $C_0(\mathbb{R}^n)$ sind, nämlich für die L^1 -Norm

$$\|f\|_1 := \int |f|.$$

Man zeige als Übung, dass dies eine Norm auf $C_0(\mathbb{R}^n)$ ist (vgl. §1)! Wir werden nun zunächst abstrakt erklären, wie man einen normierten Vektorraum oder allgemeiner einen sog. *Metrischen Raum* durch Hinzunahme aller “Limiten” von Cauchyfolgen *vervollständigen* kann, so wie man die reellen Zahlen durch Limiten von Cauchyfolgen rationaler Zahlen gewinnt. Wir werden zeigen, dass zu jedem abstrakten Limes einer L^1 -Cauchyfolge eine im wesentlichen eindeutig bestimmte Funktion gehört, gegen die eine Teilfolge der Cauchyfolge “fast überall” punktweise konvergiert, und dass insbesondere alle Riemann-integrierbaren Funktionen (mit dem bisherigen Integralwert) dazugehören. Schließlich werden wir sehen, welche Grenzwertsätze in dieser neuen Integrations-theorie gelten. Die hier gegebene Darstellung lehnt sich stark an [Alt] an.

A2. Vervollständigung metrischer Räume

Um das Problem der Vervollständigung möglichst einfach und unabhängig von allen speziellen Eigenschaften des Integrals zu beschreiben, führen wir einen sehr abstrakten Begriff ein. (Der Sinn der Abstraktion ist ja überhaupt, dass von allen speziellen Eigenschaften, die für das ins Auge gefaßte Phänomen unwesentlich sind und die Sicht nur verstellen, abgesehen werden kann.) Eine *Metrik* oder *Abstandsfunktion* auf einer Menge X ist eine Abbildung $d : X \times X \rightarrow \mathbb{R}$ mit folgenden Eigenschaften: Für alle $x, y, z \in X$ gilt:

- 1. (Positivität) $d(x, y) \geq 0$ mit $d(x, y) = 0 \iff x = y$,
- 2. (Symmetrie) $d(y, x) = d(x, y)$,
- 3. (Dreiecksungleichung) $d(x, z) \leq d(x, y) + d(y, z)$.

Diese drei Axiome enthalten alle Eigenschaften, die ein vernünftiger Abstandsbegriff wie z.B. der Abstand im \mathbb{R}^n , $d(x, y) = |x - y|$, haben sollte. Ist X ein Vektorraum über \mathbb{R} mit einer Norm $\|\cdot\|$, so ist $d(x, y) = \|x - y\|$ eine Metrik auf X . Statt $d(x, y)$ wollen wir in Zukunft wieder $|x - y|$ schreiben; das Zeichen “-” meint nicht notwendig

eine Differenz, sondern der ganze Ausdruck $|x - y|$ soll nur den Funktionswert $d(x, y)$ beschreiben. Eine Menge X mit einer Metrik d auf X heißt auch *metrischer Raum*.

Die Begriffe *Cauchyfolge* und *konvergente Folge* sind wie bisher definiert: Ist $(x_k)_{k \in \vec{N}}$ (kurz: (x_k)) eine *Folge* in X , d.h. eine Abbildung $k \mapsto x_k : \vec{N} \rightarrow X$, so definiert man

$$\begin{aligned} (x_k) \text{ Cauchyfolge} &\iff \forall \epsilon > 0 \exists N \forall k, l \geq N |x_k - x_l| < \epsilon, \\ x_k \rightarrow x &\iff \forall \epsilon > 0 \exists N \forall k \geq N |x_k - x| < \epsilon. \end{aligned}$$

Konvergente Folgen sind Cauchyfolgen, aber nicht notwendig umgekehrt. Wenn alle Cauchyfolgen konvergieren, so heißen die Metrik und der metrische Raum *vollständig*.

Ist X nicht vollständig, so wollen wir einen größeren metrischen Raum \bar{X} konstruieren, der vollständig ist, die *Vervollständigung* von X . Das einfachste Modell dafür ist $X = \mathbb{Q}$ (rationale Zahlen) und $\bar{X} = \mathbb{R}$. Eine reelle Zahl läßt sich z.B. durch einen unendlichen Dezimalbruch $n+0, a_1 a_2 a_3 \dots$ beschreiben. Dieser stellt eine Cauchyfolge von endlichen Dezimalbrüchen, also rationalen Zahlen $q_k = n+0, a_1 \dots a_k$ dar. Jede reelle Zahl läßt sich also als Cauchyfolge rationaler Zahlen beschreiben, und in diesem Sinne könnten wir \mathbb{R} als die Menge der Cauchyfolgen in \mathbb{Q} bezeichnen. Allerdings läßt sich dieselbe reelle Zahl auf viele verschiedene Weisen durch Cauchyfolgen rationaler Zahlen beschreiben: Dual-, Trial-, Dezimal-, Hexadezimalzahlen usw., andere unendliche Reihen, Kettenbrüche u.a.m. Daher müssen wir alle diejenigen Cauchyfolgen miteinander "identifizieren" (als "gleich" ansehen), die dieselbe reelle Zahl beschreiben; dies ist für zwei Cauchyfolgen (x_k) und (y_k) genau dann der Fall, wenn $|x_k - y_k| \rightarrow 0$. Mit anderen Worten, Elemente von \bar{X} sind nicht einzelne Cauchyfolgen, sondern *Äquivalenzklassen* von Cauchyfolgen unter der Äquivalenzrelation

$$(x_k) \sim (y_k) \iff |x_k - y_k| \rightarrow 0. \quad (1)$$

In einem beliebigen metrischen Raum X machen wir es ebenso und definieren \bar{X} als die Menge aller Äquivalenzklassen von Cauchyfolgen in X unter der durch (1) definierten Äquivalenzrelation. Man könnte zwei solche Cauchyfolgen *kofinal* (= "mit gemeinsamem Zielpunkt") nennen. Für zwei Cauchyfolgen $\bar{x} = (x_k)$ und $\bar{y} = (y_k)$ in X definieren wir nun einen "Abstand"

$$|\bar{x} - \bar{y}| := \lim |x_k - y_k|. \quad (2)$$

Nach (1) ist $\bar{x} \sim \bar{y} \iff |\bar{x} - \bar{y}| = 0$, und es ist sofort zu sehen, dass wir damit auf \bar{X} eine Metrik definiert haben (Übung). Der ursprüngliche metrische Raum X kann als Teilmenge von \bar{X} aufgefaßt werden, indem

wir jedes $x \in X$ als Äquivalenzklasse der konstanten Folge (x_k) mit $x_k = x$ für alle k auffassen. Nach (2) ist der Abstand von zwei Punkten $x, y \in X$ derselbe in X und \bar{X} , daher ist \bar{X} wirklich eine Erweiterung von X .

66. Satz: \bar{X} ist vollständiger metrischer Raum.

Beweis. Wir zeigen zunächst, dass jede "alte" Cauchyfolge (x_k) mit $x_k \in X$ in \bar{X} konvergiert, nämlich gegen ihre eigene Äquivalenzklasse $\bar{x} = [(x_k)]$. Dazu müssen wir den Abstand $|x_k - \bar{x}|$ in \bar{X} berechnen, d.h. jedes einzelne x_k als konstante Folge auffassen und die Definition (2) anwenden: Da (x_k) Cauchyfolge ist, gilt

$$|x_k - \bar{x}| = \lim_{l \rightarrow \infty} |x_k - x_l| \xrightarrow{k \rightarrow \infty} 0$$

und somit $x_k \rightarrow \bar{x}$ in \bar{X} .

Wir müssen aber viel mehr zeigen, nämlich dass *alle* Cauchyfolgen in \bar{X} , also auch die vielen neu hinzugekommenen, konvergent sind. Es sei also (\bar{x}_k) eine Cauchyfolge in \bar{X} . Jedes Folgenglied \bar{x}_k repräsentiert eine Cauchyfolge $(x_{kj})_{j \in \bar{N}}$ in X . Eine beliebige Cauchyfolge (y_j) in X hat nun folgende Eigenschaft: Sie konvergiert zwar vielleicht nicht wirklich, aber doch "beinahe", d.h. für jedes $\epsilon > 0$ gibt es ein $y \in X$ mit der Eigenschaft

$$\exists_N \forall_{j \geq N} |y_j - y| < \epsilon; \quad (3)$$

nach Definition einer Cauchyfolge können wir z.B. $y = y_m$ für genügend großes m wählen. Fassen wir die Cauchyfolge (y_k) in X wieder als Repräsentant eines Elements $\bar{y} \in \bar{X}$ auf, so können wir (3) auch so ausdrücken (vgl. (2)):

$$|\bar{y} - y| \leq \epsilon.$$

Somit können wir also auch für unsere $\bar{x}_k \in \bar{x}$ eine $x_k \in X$ finden mit

$$|\bar{x}_k - x_k| \leq 1/k \quad (4)$$

für alle $k \in \bar{N}$. Die Folgen (\bar{x}_k) und (x_k) in dem metrischen Raum \bar{X} sind damit kofinal, d.h. $|\bar{x}_k - x_k| \rightarrow 0$, und insbesondere ist mit (\bar{x}_k) auch (x_k) eine Cauchyfolge in \bar{X} . Da aber (x_k) sogar in X liegt, ist (x_k) in \bar{X} konvergent, wie wir eingangs gesehen haben, nämlich $x_k \rightarrow \bar{x}$, wobei $\bar{x} \in \bar{X}$ die Äquivalenzklasse von (x_k) ist. Da die Folgen (\bar{x}_k) und (x_k) kofinal sind, konvergiert auch (\bar{x}_k) gegen \bar{x} , wie man sofort sieht (Übung); dies war zu zeigen. \square

Bemerkung. Ist X ein normierter Vektorraum, also $|x - y| = \|x - y\|$, so gilt dasselbe für \bar{X} (Übung). In diesem Fall ist also \bar{X} ein vollständiger normierter Vektorraum (*Banachraum*).

A3. Limesfunktionen von L^1 -Cauchyfolgen

Eine ϵ -Menge für ein beliebiges $\epsilon > 0$ ist eine Teilmenge $S \subset \mathbb{R}^n$, die in der Vereinigung von höchstens abzählbar vielen Quadern A_1, A_2, \dots mit "Gesamtmaß" $\sum_j \mu(A_j) \leq \epsilon$ enthalten ist. Eine *Nullmenge* (im Sinne von Lebesgue) in \mathbb{R}^n ist eine ϵ -Menge für *jedes* $\epsilon > 0$. Der einzige Unterschied zur früher definierten Riemannschen Nullmenge (vgl. §23) ist, dass jetzt nicht nur endliche, sondern auch abzählbar unendliche Überdeckungen durch Quader mit kleinem Gesamtmaß zugelassen sind. Natürlich sind endliche Vereinigungen von Nullmengen wieder Nullmengen, aber jetzt eben auch abzählbar unendliche Vereinigungen: Sind N_1, N_2, \dots Lebesgue-Nullmengen, so ist auch $N = \bigcup_{k=1}^{\infty} N_k$ eine Lebesgue-Nullmenge. Für jedes N_k finden wir nämlich eine Überdeckung durch Quader A_{k1}, A_{k2}, \dots mit $\sum_j \mu(A_{kj}) \leq \epsilon/2^k$; alle A_{kj} zusammen überdecken N , und es gilt

$$\sum_{k,j} \mu(A_{kj}) \leq \sum_k \epsilon/2^k \leq \epsilon.$$

Da jeder Punkt natürlich eine Nullmenge bildet, sind also alle abzählbaren Teilmengen von \mathbb{R}^n Nullmengen, z.B. die Menge der Punkte mit rationalen Koordinaten. Auch $\mathbb{R}^{n-1} \subset \mathbb{R}^n$ und alle m -dimensionalen Mannigfaltigkeiten $M \subset \mathbb{R}^n$ mit $m < n$ sind Nullmengen: Man überdecke z.B. \mathbb{R}^{n-1} mit abzählbar vielen $n-1$ -dimensionalen Quadern B_1, B_2, \dots mit $n-1$ -dimensionalem Maß Eins und mache daraus n -dimensionale Quader

$$A_j = B_j \times (-\epsilon/2^j, \epsilon/2^j) \subset \mathbb{R}^{n-1} \times \mathbb{R} = \mathbb{R}^n$$

mit $\mu(A_j) = 2\epsilon/2^j$; das Gesamtmaß aller A_j ist dann 2ϵ .

Eine Funktion $f : \mathbb{R}^n \rightarrow \mathbb{R}$ heißt *Lebesgue-integrierbar*, wenn f *fast überall*, d.h. außerhalb einer beliebigen Nullmenge, punktwieser Limes einer L^1 -Cauchyfolge (f_k) in $C_0(\mathbb{R}^n)$ ist. Für eine L^1 -Cauchyfolge (f_k) ist die Folge der Integrale $(\int f_k)$ eine Cauchyfolge in \mathbb{R} (also konvergent), denn für genügend großes k, l ist

$$\left| \int f_k - \int f_l \right| \leq \int |f_k - f_l| = \|f_k - f_l\|_1 < \epsilon.$$

Wir können daher einer Lebesgue-integrierbaren Funktion $f \doteq \lim f_k$ (" \doteq " heißt "fast überall gleich") den Integralwert

$$\int f := \lim \int f_k$$

zuweisen; wir müssen allerdings zeigen, dass dieser Wert wohldefiniert ist, also nur von f und nicht von der Folge (f_k) abhängt. Da $|\int f_k| - |\int f_l|$

$\leq |f_k - f_l|$, ist auch $(|f_k|)$ eine L^1 -Cauchyfolge, und daher ist $|f| \doteq \lim |f_k|$ Lebesgue-integrierbar. Dasselbe gilt für $|f_k - f|$ für alle k , und es folgt $\int |f_k - f| = \lim_l \int |f_k - f_l| \leq \epsilon$, also

$$\int |f_k - f| \rightarrow 0.$$

Beispiel: Es sei (f_k) eine monoton wachsende Folge in $C_0(\mathbb{R}^n)$ mit beschränkter Integralfolge $(\int f_k)$ und es gelte $f_k \nearrow f$. Dann ist (f_k) eine L^1 -Cauchyfolge und somit f Lebesgue-integrierbar, denn da $(\int f_k)$ monoton und beschränkt, also Cauchyfolge ist, gilt für genügend große $k \leq l$:

$$\|f_k - f_l\|_1 = \int |f_k - f_l| = \int (f_l - f_k) = \int f_l - \int f_k < \epsilon.$$

Lebesgue-integrierbare Funktionen werden wir auch L^1 -Funktionen nennen. Diese bilden offensichtlich einen Vektorraum, den wir mit $L(\mathbb{R}^n)$ bezeichnen wollen.

Wir werden nun das Ergebnis von §A2 auf den normierten Vektorraum $X := C_0(\mathbb{R}^n)$ mit der L^1 -Norm $\|f\|_1 = \int |f|$ anwenden. Wir wollen aber Äquivalenzklassen von L^1 -Cauchyfolgen in X durch L^1 -Funktionen ersetzen. Dies ist möglich, wenn wir drei Dinge nachgewiesen haben:

- 1. Jede L^1 -Cauchyfolge in $C_0(\mathbb{R}^n)$ besitzt eine fast überall punktweise konvergente Teilfolge,
- 2. Sind (f_k) und (g_k) äquivalente (= kofinale) L^1 -Cauchyfolgen in $C_0(\mathbb{R}^n)$ mit Teilfolgen, die fast überall punktweise gegen Funktionen f und g konvergieren, so sind f und g fast überall gleich.
- 3. Ist (f_k) eine Cauchyfolge, die fast überall gegen Null konvergiert, so gilt $\|f_k\| \rightarrow 0$.

Wir werden dann zwei L^1 -Funktionen f und g als gleich ansehen, wenn sie fast überall gleich sind, sich also nur auf einer Nullmenge unterscheiden. Dann können wir nach 1. und 2. jeder Äquivalenzklasse von L^1 -Cauchyfolgen bijektiv die Limesfunktion einer Teilfolge zuordnen und erhalten so einen Isomorphismus von $\overline{C_0(\mathbb{R}^n)}$ (= Vervollständigung von $C_0(\mathbb{R}^n)$ bezüglich der L^1 -Norm) auf $L^1(\mathbb{R}^n) := L(\mathbb{R}^n) / \doteq$. Die Norm auf $\overline{C_0(\mathbb{R}^n)}$ geht dabei über in die L^1 -Norm $\|f\|_1 = \int |f|$ auf $L^1(\mathbb{R}^n)$, und somit ist $L^1(\mathbb{R}^n)$ mit der L^1 -Norm vollständig.

Die Argumente zu 1., 2. und 3. beruhen auf dem folgenden Lemma, das Riemann-integrierbare Funktionen g mit kleiner L^1 -Norm kennzeichnet: Die Menge, auf der $|g|$ groß ist, muß klein sein. Ist $T \subset \mathbb{R}^n$, so sei $\mathbb{C}T := \mathbb{R}^n \setminus T$ die *Komplementmenge*.

67. Lemma 1. *Ist $g : \mathbb{R}^n \rightarrow \mathbb{R}$ Riemann-integrierbar mit Träger in einem Quader Q und mit $\int |g| < \delta$, so gibt es zu eine Teilmenge $T \subset Q$, die endliche Vereinigung von Teilquadern ist, so dass gilt:*

$$\mu(T) \leq \sqrt{\delta} \quad \text{und} \quad |g| \leq \sqrt{\delta} \quad \text{auf } \mathbb{C}T$$

Beweis. Zu jedem $\epsilon > 0$ gibt es eine Zerlegung Z des Quaders Q , so dass

$$0 \leq \sum_{A \in Z} c_A \mu(A) - \int |g| < \epsilon,$$

wobei $c_A := \sup_A |g| \geq 0$ für jedes $A \in Z$. Nach Voraussetzung ist $\int |g| < \delta$. Wir können daher ϵ so klein wählen, dass auch noch $\sum_{A \in Z} c_A \mu(A) < \delta$. Nun sei

$$Z_\delta = \{A \in Z; c_A > \sqrt{\delta}\} \quad \text{und} \quad T = \bigcup \{A; A \in Z_\delta\}.$$

T ist also die Vereinigung aller $A \in Z$ mit $c_A > \sqrt{\delta}$. Außerhalb T gilt offensichtlich $|g| \leq \sqrt{\delta}$, und wegen

$$\delta > \sum_{A \in Z} c_A \mu(A) \geq \sum_{A \in Z_\delta} c_A \mu(A) \geq \sum_{A \in Z_\delta} \sqrt{\delta} \cdot \mu(A) = \sqrt{\delta} \cdot \mu(T)$$

folgt $\mu(T) \leq \sqrt{\delta}$. □

Bemerkung. Wir können diesen Beweis auch so deuten, dass wir $|g|$ durch die *Treppenfunktion*

$$t = \sum_{A \in Z} c_A 1_A$$

mit $|g| \leq t$ und $\int t \leq \delta$ ersetzen (denn $\int t$ ist der Ausdruck $\sum_{A \in Z} c_A \mu(A)$). Für

$$T = \{x \in \mathbb{R}^n; t(x) > \sqrt{\delta}\}$$

gilt dann $|g| \leq t \leq \sqrt{\delta}$ auf $\mathbb{C}T$ und

$$\delta > \int t \geq \int_T t \geq \sqrt{\delta} \cdot \mu(T),$$

also $\mu(T) < \sqrt{\delta}$.

68. Satz 1. Ist (g_j) eine Folge in $C_0(\mathbb{R}^n)$ mit $\|g_j\|_1 \xrightarrow{j \rightarrow \infty} 0$, so gibt es eine Teilfolge (g_{j_k}) , die außerhalb einer ϵ -Menge (für beliebiges $\epsilon > 0$) gleichmäßig und außerhalb einer Nullmenge punktweise gegen Null konvergiert.

Beweis. Wir wählen die Teilfolge g_{j_k} so, dass $\|g_{j_k}\|_1 < \delta_k := 1/2^{2k}$. Nach Lemma 1 ist $|g_{j_k}| < \sqrt{\delta_k} = 1/2^k$ außerhalb einer Menge T_k , die endliche Vereinigung von Quadern ist und Maß $\mu(T_k) < \sqrt{\delta_k} = 1/2^k$ hat. Setzen wir nun $S_l = \bigcup_{k \geq l+1} T_k$, so wird S_l von abzählbar vielen Quadern mit Gesamtmaß

$$\sum_{k \geq l+1} \mu(T_k) \leq \sum_{k \geq l+1} 1/2^k = 1/2^l$$

überdeckt, und für jedes $x \in \mathbb{C}S_l$ gilt $|g_{j_k}(x)| < 1/2^k$ für alle $k > l$. Somit konvergiert g_{j_k} sogar gleichmäßig auf $\mathbb{C}S_l$. Setzen wir $N = \bigcap_{l \in \vec{N}} S_l$, so ist N Nullmenge, da $N \subset S_l$ für alle $l \in \vec{N}$. Jedes $x \in \mathbb{C}N$ liegt im Komplement von mindestens einem S_l , daher gilt $|g_{j_k}| \leq 1/2^k \rightarrow 0$ für $k > l$. (Die Konvergenz auf $\mathbb{C}N$ ist allerdings nicht gleichmäßig, da das l von x abhängt.) \square

Bemerkung. Leider folgt aus $\|g_j\|_1 \rightarrow 0$ nicht, dass die *ganze* Folge (g_j) fast überall punktweise konvergiert, wie folgendes Gegenbeispiel zeigt, das ich Jens Heber verdanke: Wir betrachten die (divergente) harmonische Reihe $s_j = \sum_{k=2}^j 1/k$ für $j \geq 2$ und setzen $s_1 = 0$. Wir definieren nun $g_j : [0, 1] \rightarrow \{0, 1\}$ als die charakteristische Funktion einer Teilmenge $I_j \subset [0, 1]$, die folgende Vereinigung von ein oder zwei abgeschlossenen Intervallen ist: Es sei $[s_j]$ die größte ganze Zahl, die kleiner oder gleich s_j ist. Da $s_j - 1 < s_{j-1} < s_j$, gilt stets einer der folgenden zwei Fälle:

$$\begin{aligned} \text{Fall (a) :} & \quad [s_j] \leq s_{j-1} \leq s_j, \\ \text{Fall (b) :} & \quad [s_j] - 1 < s_{j-1} < [s_j] \leq s_j. \end{aligned}$$

Wir setzen dann

$$\begin{aligned} I_j &= [s_{j-1} - [s_j], s_j - [s_j]] \quad \text{im Fall (a),} \\ &= [0, s_j - [s_j]] \cup [s_{j-1} - ([s_j] - 1), 1] \quad \text{im Fall (b).} \end{aligned}$$

Die lückenlos aneinandergrenzenden Intervalle $[s_{j-1}, s_j]$ sind auf diese Weise maßerhaltend auf das Einheitsintervall transplantiert worden (Abb. 55). Es gilt

$$\int |g_j| = \int g_j = \int 1_{[s_{j-1}, s_j]} = s_j - s_{j-1} = 1/j \rightarrow 0,$$

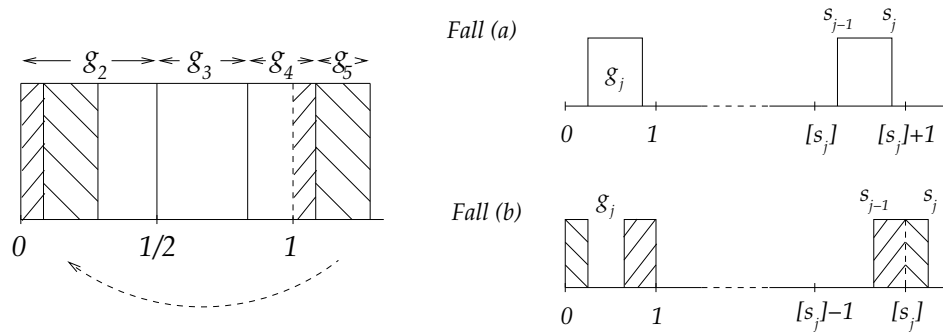


ABBILDUNG 55

aber jeder Punkt $x \in [0, 1]$ liegt in unendlich vielen der I_j , so dass $(g_j(x))$ für unendlich viele Indizes j den Wert Eins hat und daher nicht gegen Null konvergiert.

Das Beispiel hat allerdings noch den Schönheitsfehler, dass die Funktionen g_j nicht stetig sind. Dies läßt sich aber leicht beheben: Ist $[a_j, b_j]$ eine der (zwei möglichen) Zusammenhangskomponenten von I_j , so ersetzen wir dort die charakteristische Funktion g_j durch die stetige Funktion

$$\begin{aligned} \tilde{g}_j(x) &= (x - a_j)/\epsilon_j \quad \text{für } a_j \leq x \leq a_j + \epsilon_j \\ &= 1 \quad \text{für } a_j + \epsilon_j \leq x \leq b_j - \epsilon_j \\ &= (b_j - x)/\epsilon_j \quad \text{für } b_j - \epsilon_j \leq x \leq b_j \end{aligned}$$

(Abb. 56), wobei $\epsilon_j = \epsilon/2^j$ für beliebig kleines $\epsilon > 0$. Das Gesamtmaß aller Intervalle, auf denen sich irgendein \tilde{g}_j von g_j unterscheidet, ist somit höchstens $\sum_j 4 \cdot \epsilon/2^j = 4\epsilon$. Das Komplement dieser Menge ist also sicher keine Nullmenge, und dort gilt nach wie vor, dass die Folge $(\tilde{g}_j(x)) = (g_j(x))$ keine Nullfolge ist.

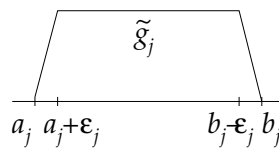


ABBILDUNG 56

69. Korollar Es seien (f_i) und (\tilde{f}_j) kofinale L^1 -Cauchyfolgen in $C_0(\mathbb{R}^n)$ (d.h. $\|\tilde{f}_k - f_k\|_1 \rightarrow 0$), für die jeweils eine Teilfolge fast überall punktweise konvergiert: $f_{i_k} \rightarrow f$ auf $\mathbb{C}N$ und $\tilde{f}_{j_k} \rightarrow \tilde{f}$ auf $\mathbb{C}\tilde{N}$ für Nullmengen N, \tilde{N} . Dann gilt $\tilde{f} = f$.

Beweis. Wir setzen $g_k = \tilde{f}_{j_k} - f_{i_k}$. Dann ist (g_k) eine L^1 -Nullfolge, denn

$$\|g_k\|_1 \leq \|\tilde{f}_{j_k} - f_{i_k}\|_1 + \|f_{i_k} - f_{j_k}\|_1 \rightarrow 0.$$

Nach Satz 1 gibt es also eine Nullmenge \bar{N} und eine Teilfolge (g_{k_l}) mit $g_{k_l} \rightarrow 0$ auf $\mathbb{C}\bar{N}$. Da

$$g_{k_l} = \tilde{f}_{j_{k_l}} - f_{i_{k_l}}$$

und die rechte Seite auf $\mathbb{C}(\bar{N} \cup N)$ gegen $\tilde{f} - f$ konvergiert, gilt $\tilde{f} = f$ auf $\mathbb{C}(\bar{N} \cup \tilde{N} \cup N)$. \square

70. Satz 2. *Jede L^1 -Cauchyfolge (f_j) in $C_0(\mathbb{R}^n)$ besitzt eine Teilfolge, die außerhalb einer ϵ -Menge (für beliebiges $\epsilon > 0$) gleichmäßig und außerhalb einer Nullmenge punktweise gegen eine Funktion f konvergiert.*

Beweis. Folgende Vorbemerkung (Erinnerung) könnte nützlich sein: Jede Folge (a_k) in \mathbb{R} kann ich bekanntlich auch als Partialsummenfolge einer Reihe schreiben: $a_k = a_0 + \sum_{j=1}^k d_j$ mit $d_k = a_k - a_{k-1}$ für alle $k \geq 1$. Besitzt $\sum_j d_j$ eine konvergente Majorante, also $|d_j| \leq c_j$ mit $\sum c_j < \infty$, z.B. $c_j = 1/2^j$, so ist die Reihe $\sum_j d_j$ und damit die Folge (a_k) konvergent.

Der Beweis selbst ist sehr ähnlich zu dem von Satz 1. Wir finden eine Teilfolge (f_{j_k}) , so dass für die Differenzen $d_k = f_{j_k} - f_{j_{k-1}}$ gilt:

$$\|d_k\|_1 \leq \delta_k := 1/2^{2k}.$$

Nach Lemma 1 folgt wieder $|d_k| \leq \sqrt{\delta_k} = 1/2^k$ auf $\mathbb{C}T_k$, wobei T_k endliche Vereinigung von Quadern ist mit $\mu(T_k) \leq \sqrt{\delta_k} = 1/2^k$. Nach der Vorbemerkung ist somit (f_{j_k}) gleichmäßig konvergent auf $\mathbb{C}S_l$ mit $S_l = \bigcup_{k>l} T_k$. Somit konvergiert (f_{j_k}) punktweise auf $\bigcup_l (\mathbb{C}S_l) = \mathbb{C}N$, wobei $N = \bigcap_l S_l$ Nullmenge ist (vgl. Beweis von Satz 1). \square

71. Satz 3. *Ist (f_k) eine L^1 -Cauchyfolge in $C_0(\mathbb{R}^n)$, so dass f_k fast überall gegen Null geht, so gilt $\|f_k\|_1 \rightarrow 0$.*

Bemerkung. Das Problem des Beweises ist folgendes: Selbst wenn $|f_k|$ klein ist außerhalb einer immer kleineren Menge, könnte doch $\int |f_k|$, anschaulich gesprochen die Fläche unter dem Graphen von $|f|$, aus zwei Gründen groß bleiben (Abb. 57):

- (a) Der Quader Q_k , in dem der Träger von f_k liegt, könnte immer größer werden,
- (b) Die Werte von $|f_k|$ auf der Restmenge könnten immer größer werden.

Beide Probleme werden im folgenden Beweis gelöst durch die Voraussetzung, dass $\int |f_k - f_l|$ beliebig klein wird.

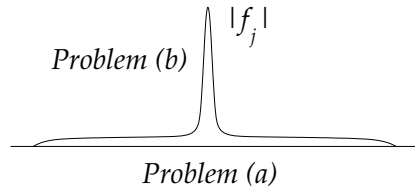


ABBILDUNG 57

Beweis. Wie im Beweis von Satz 1 ersetzen wir $|f_k|$ durch eine Treppenfunktion t_k , die diesmal allerdings von unten approximiert: Da $|f_k|$ Riemann-integrierbar, gibt es eine Zerlegung Z des Quaders Q_k , der den Träger von f_k enthält, so dass für die Treppenfunktion

$$t_k = \sum_{A \in Z} \inf_A |f| \cdot 1_A$$

gilt:

$$0 \leq t_k \leq |f_k| \text{ und } 0 \leq \int |f_k| - \int t_k \leq \epsilon_k$$

für eine Nullfolge (ϵ_k) . Es genügt also, $\int t_k \rightarrow 0$ zu zeigen. Wir wählen ein festes, genügend großes k . Für alle $l > k$ zerlegen wir $\int f_l$ in die Integrale über Q_k (wo der Träger von t_k liegt) und über $\mathbb{C}Q_k$. In dem Integral über $\mathbb{C}Q_k$ können wir den Integranden t_l durch $t_l - t_k$ ersetzen, da t_k dort verschwindet. Also erhalten wir

$$\int t_l = \int_{Q_k} t_l + \int_{\mathbb{C}Q_k} (t_l - t_k) \leq \int_{Q_k} t_l + \|t_l - t_k\|_1.$$

Da der zweite Term $\|t_l - t_k\|_1$ für genügend großes k beliebig klein wird, genügt es, den ersten Term $\int_{Q_k} t_l$ weiter abzuschätzen. Der Bereich Q_k ist jetzt fest, d.h. von t_l unabhängig, wodurch das Problem (a) in der vorstehenden Bemerkung gelöst ist. Wir betrachten nun die Menge G_l , auf der t_l groß ist: Für fest vorgegebenes $\epsilon > 0$ sei

$$G_l = \{x; t_l(x) \geq \epsilon\}.$$

Dann ist $t_l < \epsilon$ auf $Q \setminus G_l$ und daher

$$\int_{Q_k} t_l = \int_{Q_k \cap G_l} t_l + \int_{Q_k \setminus G_l} t_l \leq \int_{Q_k \cap G_l} t_l + \epsilon \cdot \mu(Q_k).$$

Der zweite Term ist beliebig klein, wenn ϵ nur genügend klein gewählt wurde. Den ersten Term schätzen wir ab durch

$$\int_{G_l} t_l \leq \int_{G_l} t_k + \|t_l - t_k\|_1 \leq \sup t_k \cdot \mu(G_l) + \epsilon,$$

womit auch das Problem (b) der einleitenden Bemerkung gelöst wurde. Der Beweis ist beendet, wenn wir $\mu(G_l) \rightarrow 0$ für $l \rightarrow \infty$ zeigen können; dies ist die Aussage des folgenden Lemmas. \square

72. Lemma 2. *Es sei (t_k) eine Folge von Treppenfunktionen, die eine L^1 -Cauchyfolge ist und fast überall gegen Null geht. Für beliebiges $\epsilon > 0$ sei $G_k = \{x; t_k(x) \geq \epsilon\}$. Dann gilt (nach Übergang zu einer Teilfolge) $\mu(G_k) \rightarrow 0$ für $k \rightarrow \infty$.*

Beweis. Nach Übergang zu einer Teilfolge dürfen wir $\|t_k - t_{k-1}\|_1 \leq 1/2^{2k}$ annehmen. Daraus folgt nach Lemma 1:

$$|t_k - t_{k-1}| \leq 1/2^k \text{ auf } \mathbb{C}T_k \text{ mit } \mu(T_k) \leq 1/2^k.$$

Wir halten k fest und setzen für alle $l > k$

$$T^l := \bigcup_{j=k+1}^l T_j.$$

Dann gilt für alle $l > k$:

$$\mu(T^l) \leq \sum_{j=k+1}^l 1/2^j < 1/2^k.$$

Nach Dreiecksungleichung gilt für alle $x \in \mathbb{R}^n$

$$t_k(x) \leq t_l(x) + |t_l(x) - t_k(x)|. \quad (*)$$

Da $t_l \rightarrow 0$ außerhalb einer Nullmenge N , geht der erste Term rechts gegen Null für $l \rightarrow \infty$ für alle $x \in \mathbb{C}N$, und für den zweiten Term haben wir die Abschätzung

$$|t_l - t_k| \leq \sum_{j=k+1}^l |t_j - t_{j-1}| \leq \sum_{j=k+1}^l 1/2^j < 1/2^k$$

außerhalb der Vereinigung aller T^l . Die Nullmenge N können wir durch Quader A_1, A_2, \dots mit Gesamtmaß $\sum \mu(A_j) < 1/2^k$ überdecken. Setzen wir

$$A^l = \bigcup_{j=1}^l A_j \text{ und } B^l = A^l \cup T^l,$$

so gilt $\mu(B^l) \leq \mu(A^l) + \mu(T^l) < 2/2^k$. Dann ist $t_k(x) < 1/2^k$ außerhalb von $B^\infty := \bigcup_l B^l$; wir müssen nur den Index l in $(*)$ (in Abhängigkeit von x) genügend groß wählen. Ist k so groß, dass $\epsilon > 1/2^k$, so muß daher $G_k \setminus B^\infty = \emptyset$ gelten. Andererseits ist

$$G_k \setminus B^\infty = \bigcap_l (G_k \setminus B^l),$$

und da die Mengenfolge $(B^l)_l$ monoton wachsend ist, d.h. $B^l \subset B^{l+1}$, ist die Komplementfolge $(G_k \setminus B^l)_l$ monoton fallend. Wir dürfen annehmen, dass G_k kompakt ist und alle B^l offen sind, indem wir eventuell alle Quader von G_k zu kompakten Quadern verkleinern und die von T_j und A_j zu offenen Quadern vergrößern, ohne das Maß wesentlich zu ändern. Dann ist also $G_k \setminus B^l$ kompakt. Da $G_k \setminus B^\infty = \emptyset$, muß auch $G_k \setminus B^l = \emptyset$ gelten für genügend großes l , denn sonst hätte die monoton fallende Folge kompakter Mengen $(G_k \setminus B^l)_l$ einen Durchschnitt. Also gilt $G_k \subset B^l$ für genügend großes l , und damit

$$\mu(G_k) \leq \mu(B^l) < 2/2^k.$$

□

Als Folgerung erhalten wir die Eindeutigkeit des Lebesgue-Integrals: Ist $f \doteq \lim f_k \doteq \lim g_k$ für zwei L^1 -Cauchyfolgen (f_k) und (g_k) in $C_0(\mathbb{R}^n)$, so folgt $\lim(f_k - g_k) \doteq 0$ und damit $\|f_k - g_k\|_1 \rightarrow 0$ nach dem obigen Satz, also sind (f_k) und (g_k) kofinal bezüglich der L^1 -Norm, und es folgt $\lim \int f_k = \lim \int g_k$. Insbesondere ist $\|f\|_1 = \int |f|$ für jedes $f \in L(\mathbb{R}^n)$ erklärt, und $\|f\| = 0 \iff f \doteq 0$. Somit definiert $\|\cdot\|_1$ eine Norm auf $L^1(\mathbb{R}^n) = L(\mathbb{R}^n)/\doteq$.

Wir können die bisherigen Ergebnisse i.w. in folgendem Satz zusammenfassen:

73. Satz 4. *Die Abbildung, die jeder L^1 -Cauchyfolge in $C_0(\mathbb{R}^n)$ die Limesfunktion einer fast überall konvergenten Teilfolge zuordnet, definiert einen linearen Norm-erhaltenden Isomorphismus der Vervollständigung $\overline{C_0(\mathbb{R}^n)}$ von $C_0(\mathbb{R}^n)$ auf $L^1(\mathbb{R}^n)$. Insbesondere ist $L^1(\mathbb{R}^n)$ vollständig, und der L^1 -Limes einer L^1 -Cauchyfolge in $C_0(\mathbb{R}^n)$ ist die Limesfunktion einer fast überall konvergenten Teilfolge.*

Beweis. Nach Satz 2 gibt es stets eine fast überall konvergente Teilfolge. Die Wohldefiniertheit (Unabhängigkeit von der Wahl der L^1 -Cauchyfolge unter allen kofinalen) und die Linearität folgen aus dem Korollar zu Satz 1: Jede Teilfolge einer Cauchyfolge (f_k) ist kofinal zu (f_k) , und wenn $f \doteq \lim f_k$ und $g \doteq \lim g_k$, so ist $f + g \doteq \lim(f_k + g_k)$ und $\alpha f \doteq \lim \alpha f_k$ für $\alpha \in \mathbb{R}$. Die Injektivität folgt aus Satz 3. Die Surjektivität und die Norm-Treue ergibt sich aus der Definition von $L(\mathbb{R}^n)$ und des Integrals auf $L(\mathbb{R}^n)$: Eine Lebesgue-integrierbare Funktion f ist fast überall Limes einer L^1 -Cauchyfolge (f_k) , und da auch $(|f_k|)$ eine L^1 -Cauchyfolge in $C_0(\mathbb{R}^n)$ ist, gilt

$$\|f\|_1 = \int |f| = \lim \int |f_k| = \lim \|f_k\|_1 = \|\bar{f}\|_1,$$

wobei \bar{f} wie in §A2 die Äquivalenzklasse der Cauchyfolge (f_k) als Element der Vervollständigung $\overline{C_0(\mathbb{R}^n)}$ bezeichnet. \square

74. Satz 5. (*Monotonie des Lebesgue-Integrals*) Sind $f, g \in L(\mathbb{R}^n)$ mit $f \geq g$ fast überall, so gilt $\int f \geq \int g$.

Beweis. Wir dürfen annehmen, dass $g \equiv 0$ ist, indem wir nötigenfalls f durch $f - g$ ersetzen. Es sei (f_k) eine Cauchyfolge in $C_0(\mathbb{R}^n)$ mit $f_k \rightarrow f$ fast überall. Dann sind die Funktionen

$$\tilde{f}_k = \max(f_k, 0) = \frac{1}{2}(f_k + |f_k|)$$

immer noch in $C_0(\mathbb{R}^n)$ und bilden wegen $|\tilde{f}_k - \tilde{f}_l| \leq |f_k - f_l|$ auch eine L^1 -Cauchyfolge, und da $f \geq 0$ außerhalb einer Nullmenge, gilt dort auch $|\tilde{f}_k - f| \leq |f_k - f| \rightarrow 0$. Nach Satz 3 folgt $\|\tilde{f}_k - f\|_1 \rightarrow 0$ und damit $\int f = \lim \int \tilde{f}_k \geq 0$, denn aus $\tilde{f}_k \geq 0$ folgt $\int \tilde{f}_k \geq 0$ (Eigenschaft des Riemann-Integrals). \square

A4. Riemann- und Lebesgue-Maß

Eine Teilmenge $M \subset \mathbb{R}^n$ heißt *Lebesgue-messbar*, wenn die charakteristische Funktion 1_M Lebesgue-integrierbar ist, und $\mu_L(M) := \int 1_M$ heißt das *Lebesgue-Maß* von M . Wir wollen in diesem Abschnitt einige Lebesgue-messbare Mengen kennenlernen, insbesondere die beschränkten offenen und abgeschlossenen Mengen, und zeigen, dass das Lebesgue-Maß mit dem Riemann- oder Jordan-Maß übereinstimmt, wenn dieses definiert ist. Wir werden dazu die charakteristische Funktion einer beschränkten abgeschlossenen Menge A oder offenen Menge U im \mathbb{R}^n monoton durch $C_0(\mathbb{R}^n)$ -Funktionen approximieren (vgl. Abb. 58).

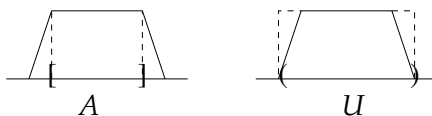


ABBILDUNG 58

Hierzu führen wir die *Abstandsfunktion von einer abgeschlossenen Menge* ein. Ist $A \subset \mathbb{R}^n$ eine nichtleere abgeschlossene Teilmenge, so sei $d_A : \mathbb{R}^n \rightarrow \mathbb{R}$,

$$d_A(x) = \inf_{a \in A} |x - a|.$$

Das Infimum ist in Wahrheit ein Minimum, also $d_A(x) = |x - a_x|$ für ein $a_x \in A$, denn für festes $x \in \mathbb{R}^n$ nimmt die stetige Funktion

$a \mapsto |x - a|$ auf der kompakten Menge $A \cap K_R(x)$ ein Minimum an, wobei $R = |x - a_0|$ für irgend ein $a_0 \in A$; dieser minimale Wert ist auch das Minimum von $|x - a|$ für alle $a \in A$, denn für $a \in A \setminus K_R(x)$ ist $|x - a|$ ja ohnehin größer als $|x - a_0|$. Der Punkt $a_x \in A$, wo das Minimum angenommen wird, heißt der *zu x nächstliegende Punkt in A* .

Die so definierte Funktion d_A ist stetig, sogar Lipschitz-stetig mit Lipschitz-Konstante Eins (*schwach kontrahierend*), denn für alle $x, y \in \mathbb{R}^n$ ist

$$\begin{aligned} d_A(x) &\leq |x - a_y| \leq |x - y| + |y - a_y| = |x - y| + d_A(y), \\ d_A(y) &\leq |y - a_x| \leq |y - x| + |x - a_x| = |y - x| + d_A(x), \end{aligned}$$

wobei a_x und a_y die zu x und y nächstliegenden Punkte in A sind. Die beiden Ungleichungen zusammen ergeben die gewünschte Lipschitz-Abschätzung

$$|d_A(x) - d_A(y)| \leq |x - y|$$

für alle $x, y \in \mathbb{R}^n$. Offensichtlich gilt außerdem

$$d_A(x) = 0 \iff x \in A.$$

Ist nun $\chi = 1_A$ die charakteristische Funktion einer beschränkten abgeschlossenen Menge A , so approximieren wir diese durch die Funktionen

$$1_k^+ = \alpha_k \circ d_A,$$

wobei $\alpha_k : \mathbb{R} \rightarrow \mathbb{R}$ die Funktion mit $\alpha_k \equiv 1$ auf $(-\infty, 0]$ und $\alpha_k \equiv 0$ auf $[1/k, \infty)$ sowie $\alpha_k(x) = 1 - kx$ für $x \in [0, 1/k]$ (s. Abb. 59).

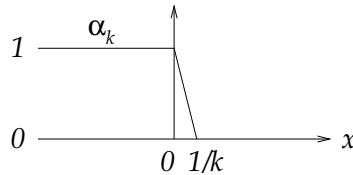


ABBILDUNG 59

Da α_k und d_A stetig sind, ist auch 1_k^+ stetig, und alle 1_k^+ sind Null außerhalb der Menge $A_1 := \{x \in \mathbb{R}^n; d_A(x) \leq 1\}$; diese Menge ist abgeschlossen und beschränkt, nämlich im Ball $B_{R+1}(0)$ enthalten, wenn $A \subset B_R(0)$. Also ist $1_k^+ \in C_0(\mathbb{R}^n)$. Außerdem gilt $1_k^+ \searrow \chi$, denn für jedes $x \in \mathbb{R}^n$ sind entweder alle $1_k^+(x)$ Null oder alle Eins oder der Wert $1_k^+(x)$ springt ab irgendeinem k von Eins auf Null (nämlich falls $0 < d_A(x) < 1$). Die Integralfolge $\int 1_k^+$ ist uniform beschränkt durch das Maß eines Quaders, der A_1 enthält, also ist (wegen der Monotonie)

(1_k^+) eine L^1 -Cauchyfolge in $C_0(\mathbb{R}^n)$ (vgl. S. 118) und somit $\chi \in L(\mathbb{R}^n)$ mit

$$\mu_L(A) = \int \chi = \lim \int 1_k^+.$$

Ist A zusätzlich Jordan-messbar, so ist der Rand ∂A eine Jordan-Nullmenge. Nur in der Nähe des Randes unterscheiden sich aber die Riemann-integrierbaren Funktionen 1_k^+ und χ , nämlich auf der Menge

$$B_k = \{x \in \mathbb{R}^n; 0 < d_A(x) < 1/k\}.$$

Überdecken wir die Nullmenge ∂A mit offenen Quadern Q_1, \dots, Q_N mit Gesamtmaß $\mu(Q_1) + \dots + \mu(Q_N) < \epsilon$, so überdecken diese Quader auch B_k für genügend großes k . Andernfalls gäbe es nämlich für jedes $k \in \mathbb{N}$ einen Punkt $x_k \in B_k$, der in keinem der Q_i enthalten ist; die Folge (x_k) läge in der kompakten Menge A_1 und besäße daher eine konvergente Teilfolge $x_{k_j} \rightarrow x$. Auch der Limespunkt x wäre in keinem der Q_i enthalten, da $\mathbb{R}^n \setminus Q_i$ abgeschlossen ist, aber andererseits wäre $d_A(x) = \lim d_A(x_{k_j}) = 0$, also wäre $x \in A$ und damit $x \in \partial A$, denn x ist ja Limes einer Folge in $\mathbb{R}^n \setminus A$. Dann wäre x aber doch in einem der Q_i enthalten, da diese ja ∂A überdecken, Widerspruch! Da $0 \leq 1_k^+ - \chi \leq 1$, erhalten wir für die Riemann-Integrale

$$\left| \int 1_k^+ - \int \chi \right| = \int |1_k^+ - \chi| \leq \sum_{i=1}^N \mu(Q_i) < \epsilon$$

und damit konvergiert $\int 1_k^+$ auch gegen das Riemann-Integral von χ , womit $\mu(A) = \mu_L(A)$ gezeigt ist.

Ist andererseits $U \subset \mathbb{R}^n$ eine offene und beschränkte Teilmenge, so benutzen wir die Abstandsfunktion von der Komplementmenge $\mathbb{C}U = \mathbb{R}^n \setminus U$, die ja abgeschlossen ist. Die Approximation der charakteristischen Funktion $\chi = 1_U$ ist diesmal

$$1_k^- = \beta_k \circ d_{\mathbb{C}U}$$

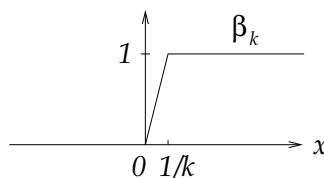


ABBILDUNG 60

mit $\beta_k := 1 - \alpha_k$ (Abb. 60). Diese Funktion ist Null außerhalb von U , da dort $d_{\mathbb{C}U}$ verschwindet, und ist Eins auf einer etwas kleineren Teilmenge

von U . Wieder ist $1_k^- \in C_0(\mathbb{R}^n)$, und diesmal gilt $1_k^- \nearrow \chi$, wobei die Integralfolge durch das Maß jedes Quaders, der U enthält, beschränkt ist. Also ist χ wieder Limes einer L^1 -Cauchyfolge in $C_0(\mathbb{R}^n)$ und somit Lebesgue-integrierbar mit $\mu_L(U) = \int \chi = \lim \int 1_k^-$, und mit derselben Begründung wie vorher gilt $\mu(U) = \mu_L(U)$, falls U Jordan-messbar und damit ∂U eine Jordan-Nullmenge ist. Wir haben also gezeigt:

75. Satz. *Jede beschränkte offene oder abgeschlossene Menge $M \subset \mathbb{R}^n$ ist Lebesgue-messbar, und falls M auch Jordan-messbar ist, stimmen das Lebesgue-Maß und das Jordan-Maß überein.*

Eine unmittelbare Folgerung ist, dass alle Quader auch Lebesgue-messbar sind mit demselben Maß, denn jeder Quader unterscheidet sich von seinem Abschluß nur durch eine Jordan-Nullmenge, die erst recht eine Lebesgue-Nullmenge ist; die charakteristischen Funktionen des Quaders und seines Abschlusses sind also “fast überall” gleich und haben daher gleiches Lebesgue-Integral. Damit sehen wir auch, dass alle Treppenfunktionen (endliche Linearkombinationen von charakteristischen Funktionen von Quadern) Lebesgue-integrierbar mit dem bekannten Integralwert sind. Weil alle Riemann-integrierbaren Funktionen sich durch Treppenfunktionen (im L^1 -Sinn) approximieren lassen, folgt daraus auch, dass jede Riemann-integrierbare Funktion auch Lebesgue-integrierbar mit demselben Integralwert ist, wie wir im nächsten Abschnitt noch genauer sehen werden. Eine andere Folgerung, die wir ebenfalls in allgemeinerem Rahmen sehen werden, ist die Lebesgue-messbarkeit von unendlichen Vereinigungen von Quadern, wie sie in den Beweisen von §A3 immer wieder vorkamen.

Bemerkung. Wir können den Satz leicht noch etwas verbessern: Ist $M \subset \mathbb{R}^n$ beschränkt und abgeschlossen oder offen, so gilt

$$\begin{aligned} \mu_L(M) &= \mu^*(M) \quad \text{falls } M \text{ abgeschlossen,} \\ \mu_L(M) &= \mu_*(M) \quad \text{falls } M \text{ offen,} \end{aligned}$$

wobei $\mu^*(M)$ das äußere und $\mu_*(M)$ das innere Maß bezeichnen (vgl. §21, S. 77f). Dies sieht man so: Ist M eine beliebige beschränkte Menge (in einem Quader Q), so konstruieren wir zu jedem $\epsilon > 0$ jeweils eine offene und eine abgeschlossene Jordan-messbare Menge T_+ und T_- (endliche Vereinigung von Quadern) mit

$$T_+ \supset M \supset T_-, \quad \mu(T_+) \approx_\epsilon \mu^*(M), \quad \mu(T_-) \approx_\epsilon \mu_*(M)$$

folgendermaßen: Zu einer genügend feinen Zerlegung Z von Q wählen wir T_+ und T_- in der Form

$$T_+ = \bigcup_{P \in Z_+(M)} P, \quad T_- = \bigcup_{P \in Z_-(M)} P$$

wobei P_+ ein etwas größerer offener Quader mit $\bar{P} \subset P_+$ und P_- ein etwas kleinerer abgeschlossener Quader mit $P_- \subset P$ ist. Ist nun M abgeschlossen, so nimmt die Abstandsfunktion $d_{\mathbb{C}T_+}$ auf der kompakten Menge M ein Minimum $\delta > 0$ an. Für $1/k < \delta$ ist daher

$$1_{T_+} \geq 1_k^+ \geq 1_M$$

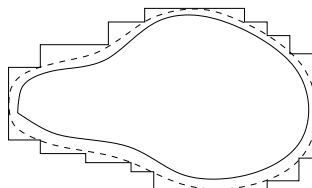


ABBILDUNG 61

(Abb. 61) und somit gilt

$$\mu^*(M) = \lim_{k \rightarrow \infty} \int 1_k^+ = \mu_L(M).$$

Ist M dagegen offen, so nimmt $d_{\mathbb{C}M}$ auf der kompakten Menge T_- einen Minimalwert $\delta > 0$ an, also gilt für $1/k < \delta$

$$1_{T_-} \leq 1_k^- \leq 1_M$$

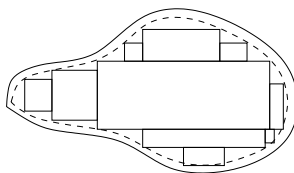


ABBILDUNG 62

(Abb. 62) und damit

$$\mu(M) = \lim_{k \rightarrow \infty} \int 1_k^- = \mu_L(M).$$

A5. Konvergenzsätze

Wir haben gesehen, dass $L^1(\mathbb{R}^n)$ vollständig ist, d.h. jede L^1 -Cauchyfolge (f_k) in $L(\mathbb{R}^n)$ konvergiert im L^1 -Sinn gegen eine integrierbare Funktion f , d.h. $\|f_k - f\|_1 \rightarrow 0$. Wir werden dafür $f_k \xrightarrow{L^1} f$ schreiben. Sind die Funktionen (f_k) stetig mit kompaktem Träger, also in $C_0(\mathbb{R}^n)$, so folgt nach Satz 2 in §A3 nach Übergang zu einer Teilfolge auch $f_k \xrightarrow{\cdot} f$, also punktweise Konvergenz fast überall. Diese Aussage wollen wir jetzt auch für beliebige L^1 -Cauchyfolgen in $L(\mathbb{R}^n)$ zeigen.

76. Satz 1. *Ist (f_k) eine Folge in $L(\mathbb{R}^n)$ mit $f_k \xrightarrow{L^1} f$, so folgt $f_{k_l} \xrightarrow{\cdot} f$ für eine Teilfolge (f_{k_l}) .*

Beweis. Jede L^1 -Funktion ist L^1 -Limes von $C_0(\mathbb{R}^n)$ -Funktionen. Es gibt also Folgen $(\phi_{kl})_{l \in \vec{N}}$ und $(\psi_l)_{l \in \vec{N}}$ in $C_0(\mathbb{R}^n)$ mit

$$\psi_l \xrightarrow{L^1} f, \quad \phi_{kl} \xrightarrow{L^1} f_k$$

für alle $k \in \vec{N}$. Nach Satz 2 gibt es nach Übergang zu Teilfolgen ϵ -Mengen S'_l, S''_{kl} zu $\epsilon := 1/2^l$ mit

$$(4) \quad \begin{aligned} |\psi_l - f| &\leq 1/2^l \text{ auf } \mathbb{C}S'_l, \\ |\phi_{kl} - f_k| &\leq 1/2^l \text{ auf } \mathbb{C}S''_{kl}. \end{aligned}$$

Für jedes $k \in \vec{N}$ wählen wir ein festes $l \geq k$, das der letzteren Bedingung genügt, und setzen

$$\phi_k := \phi_{kl}, \quad S''_k := S''_{kl}$$

für dieses l . Dann gilt jedenfalls $\mu_L(S''_k) < 1/2^k$ und

$$(5) \quad |\phi_k - f_k| \leq 1/2^k \text{ auf } \mathbb{C}S''_k.$$

Wir dürfen weiterhin ohne Einschränkung zusätzlich annehmen, dass etwa

$$\|\phi_k - f_k\|_1 < 1/2^{2k+1}, \quad \|\psi_k - f\|_1 < 1/2^{2k+1},$$

so dass $\|\phi_k - \psi_k\|_1 < 1/2^{2k}$ und somit nach Lemma 1, §A3 gilt:

$$(6) \quad |\phi_k - \psi_k| \leq 1/2^k \text{ auf } \mathbb{C}T_k,$$

wobei T_k eine Jordan-messbare Menge (endliche Vereinigung von Quadern) ist mit $\mu(T_k) < 1/2^k$. Aus (4), (5), (6) folgt

$$|f_k - f| \leq |f_k - \phi_k| + |\phi_k - \psi_k| + |\psi_k - f| \leq 3/2^k$$

außerhalb der Menge $S_k := S'_k \cup S''_k \cup T_k$, die von Quadern mit Gesamtmaß $\leq 3/2^k$ überdeckt wird ("3/2^k-Menge"). Setzen wir nun $R_j := \bigcup_{k>j} S_k$, so gilt $|f_k - f| < 3/2^k$ auf $\mathbb{C}R_j$ für alle $k > j$ gleichzeitig, und R_j wird immer noch von Quadern mit Gesamtmaß $\leq \sum_{k>j} 3/2^k = 3/2^j$

überdeckt. Dann ist $N = \bigcap_j R_j$ eine Nullmenge, da $N \subset R_j$ für alle j , und $f_k(x) \rightarrow f(x)$ für alle $x \in \mathbb{C}N$. \square

77. Korollar. *Ist (f_k) eine L^1 -Cauchyfolge in $L(\mathbb{R}^n)$ mit $f_k \dot{\rightarrow} f$, so ist $f \in L(\mathbb{R}^n)$ und $f_k \xrightarrow{L^1} f$.*

Beweis. Wegen der Vollständigkeit gibt es $\tilde{f} \in L(\mathbb{R}^n)$ mit $f_k \xrightarrow{L^1} \tilde{f}$, und nach Satz 1 folgt $f_{k_j} \dot{\rightarrow} \tilde{f}$, also $\tilde{f} \doteq f$ und die Behauptung folgt. \square

Nachdem wir nun gesehen haben, dass L^1 -Konvergenz und punktweise Konvergenz außerhalb einer Nullmenge für L^1 -Cauchyfolgen im wesentlichen dasselbe sind, können wir die beiden wichtigsten Konvergenzsätze für das Lebesgue-Integral zeigen: Den Satz von Levi über die monotone Konvergenz und den Satz von Lebesgue über die majorisierte Konvergenz.

78. Satz 2. (B. Levi) (Monotone Konvergenz) *Es sei (f_k) eine monotone Folge von Lebesgue-integrierbaren Funktionen mit beschränkter Integralfolge. Dann ist (f_k) fast überall konvergent, und $f := \lim f_k$ ist Lebesgue-integrierbar mit $f_k \xrightarrow{L^1} f$; insbesondere ist $\int f = \lim \int f_k$.*

Beweis. Ohne Einschränkung sei die Folge (f_k) monoton wachsend, also $f_k \leq f_{k+1}$. g auch die Folge $(\int f_k)$ monoton wachsend (vgl. Satz 5 in §A3) und nach Voraussetzung beschränkt, also an jeder Stelle eine Cauchyfolge in \mathbb{R} . Damit ist (f_k) eine L^1 -Cauchyfolge, denn für $k < l$ gilt wegen $f_k \leq f_l$:

$$\|f_l - f_k\|_1 = \int (f_l - f_k) = \int f_l - \int f_k < \epsilon,$$

wenn k genügend groß ist. Wegen der Vollständigkeit von $L^1(\mathbb{R}^n)$ gibt es $f \in L(\mathbb{R}^n)$ mit $f_k \xrightarrow{L^1} f$, und nach Satz 1 folgt $f_k \dot{\rightarrow} f$. Insbesondere gilt $\int f_k \rightarrow \int f$, denn

$$\left| \int f_k - \int f \right| \leq \int |f_k - f| = \|f_k - f\|_1 \rightarrow 0.$$

\square

79. Folgerung 1. *Ist (S_k) eine monoton wachsende Folge von Lebesgue-messbaren Mengen im \mathbb{R}^n , also $S_k \subset S_{k+1}$ für alle k , so dass die Folge der Maße $(\mu_L(S_k))$ beschränkt ist, so ist $S = \bigcup_k S_k$ Lebesgue-messbar mit $\mu_L(S) = \lim \mu_L(S_k)$.*

Beweis. Die Funktionen $f_k := 1_{S_k}$ erfüllen die Voraussetzungen von Satz 2 mit $f_k \rightarrow f := 1_S$. \square

80. Folgerung 2. Jede Riemann-integrierbare Funktion $f : \mathbb{R}^n \rightarrow \mathbb{R}$ ist auch Lebesgue-integrierbar, und Riemann- und Lebesgue-Integral stimmen überein.

Beweis. Nach Definition der Riemann-Integrierbarkeit hat f kompakten Träger, der in einem Quader Q enthalten ist, und es gibt monotone Folgen von Treppenfunktionen $t_k^- \nearrow, t_k^+ \searrow$ (zu immer feineren Unterteilungen von Q) mit

$$\begin{aligned} t_k^- &\leq f \leq t_k^+, \\ \|t_k^+ - t_k^-\|_1 &= \int (t_k^+ - t_k^-) \rightarrow 0. \end{aligned}$$

(Wir sahen schon in §A4, dass Treppenfunktionen Lebesgue-integrierbar sind mit dem "richtigen" Integralwert.) Nach Satz 1 folgt daraus $t_k^+ - t_k^- \rightarrow 0$ und daher $t_k^- \rightarrow f$ und $t_k^+ \rightarrow f$. Nach Satz 2 ist $f \in L(\mathbb{R}^n)$ mit Lebesgue-Integral $\int f = \lim \int t_k^-$, und dieser Wert ist gleich dem Riemann-Integral für f . \square

81. Satz 4. (H. Lebesgue) (Majorisierte Konvergenz) Es sei (f_k) eine Folge Lebesgue-integrierbarer Funktionen mit $f_k \rightarrow f$, und es gebe eine Funktion $F \in L(\mathbb{R}^n)$ mit $|f_k| \leq F$ für alle $k \in \mathbb{N}$. Dann ist $f \in L(\mathbb{R}^n)$ und $f_k \xrightarrow{L^1} f$; insbesondere folgt $\int f = \lim \int f_k$.

Beweis. Wir machen die Konvergenz monoton, um den Satz von Levi anwenden zu können. Dazu definieren wir $g_k : \mathbb{R}^n \rightarrow \mathbb{R}$,

$$g_k(x) = \sup\{f_l(x); l \geq k\} \leq F(x)$$

Offensichtlich ist die Folge $(g_k(x))$ monoton fallend, da das Supremum über eine immer kleiner werdende Menge gebildet wird. Der Limes dieser Folge ist der Limes Superior der Folge $(f_k(x))$; da diese Folge für fast alle x konvergiert, gilt

$$\lim_k g_k = \limsup f_k \doteq \lim f_k \doteq f.$$

Um den Satz von Levi anwenden zu können, müssen wir $g_k \in L(\mathbb{R}^n)$ nachweisen. Dazu benutzen wir abermals den Satz von Levi, indem wir g_k als Limes einer monoton wachsenden Folge darstellen, nämlich der Folge $(g_{km})_{m \geq k}$ mit

$$g_{km}(x) = \max\{f_l(x); m \geq l \geq k\}.$$

Das Maximum von zwei integrierbaren Funktionen a, b ist wieder integrierbar, da $\max(a, b) = \frac{1}{2}(a+b+|a-b|)$, also ist auch das Maximum von endlich vielen L^1 -Funktionen wieder integrierbar, also ist $g_{km} \in L(\mathbb{R}^n)$ für alle $k \leq m$. Natürlich gilt $g_{km} \nearrow g_k$ für $m \rightarrow \infty$, denn es wird

ja das Maximum über immer mehr Werte gebildet. Da $g_{km} \leq F$, ist die Integralfolge $(\int g_{km})_{m \geq k}$ durch $\int F$ nach oben beschränkt. Damit erfüllt die Folge $(g_{km})_{m \geq k}$ alle Voraussetzungen des Satzes von Levi, und es folgt $g_k \in \mathbf{L}(\mathbb{R}^n)$.

Nun können wir den Satz von Levi auf die monoton fallende Folge (g_k) anwenden. Die Integralfolge ist beschränkt, da $|g_k| \leq F$. Also ist $f \in \mathbf{L}(\mathbb{R}^n)$ und $g_k \xrightarrow{L^1} f$. Dies ist aber noch nicht ausreichend, denn wir wollen $f_k \xrightarrow{L^1} f$ zeigen. Dazu wiederholen wir das ganze Argument, wobei wir jetzt das Supremum durch das Infimum ersetzen. Wir definieren also $h_k : \mathbb{R}^n \rightarrow \mathbb{R}$,

$$h_k(x) = \inf\{f_l(x); l \geq k\} \geq -F(x).$$

Ganz analog wie vorher gilt $h_k \in \mathbf{L}(\mathbb{R}^n)$ und $h_k \nearrow f$ fast überall mit $\int h_k \leq \int F$ für alle k , also folgt nach Levi auch $h_k \xrightarrow{L^1} f$. Insbesondere folgt $g_k - h_k \xrightarrow{L^1} 0$. Da $h_k \leq f_l \leq g_k$ für alle $l \geq k$ und damit auch $h_k \leq f \leq g_k$, ist $|f_l - f| \leq |g_k - h_k|$ und damit

$$\|f_l - f\|_1 \leq \|g_k - h_k\|_1 \rightarrow 0.$$

□

A6. Integration über Teilmengen, Transformationssatz, Fubini

Wir nennen eine Funktion $g : \mathbb{R}^n \rightarrow \mathbb{R}$ *lokal integrierbar*, wenn $\phi \cdot g \in \mathbf{L}(\mathbb{R}^n)$ für alle $\phi \in C_0(\mathbb{R}^n)$. Der Raum aller lokal integrierbaren Funktionen heie $\mathbf{L}_{lok}(\mathbb{R}^n)$. Jede integrierbare Funktion f ist natrlich erst recht lokal integrierbar, denn ist $f = \dot{\lim} \phi_k$ fr eine L^1 -Cauchyfolge (ϕ_k) in $C_0(\mathbb{R}^n)$, so ist auch $(\phi \cdot \phi_k)$ eine L^1 -Cauchyfolge in $C_0(\mathbb{R}^n)$ fr jede Funktion $\phi \in C_0(\mathbb{R}^n)$, da

$$\|\phi \cdot \phi_k - \phi \cdot \phi_l\|_1 \leq \max|\phi| \cdot \|\phi_k - \phi_l\|_1,$$

und $\phi \cdot \phi_k \rightarrow \phi \cdot f$, also $\phi \cdot f \in \mathbf{L}(\mathbb{R}^n)$. Weitere Beispiele von lokal integrierbaren Funktionen sind charakteristische Funktionen von beliebigen offenen oder abgeschlossenen Mengen $M \subset \mathbb{R}^n$. Ist nmlich $\phi \in C_0(\mathbb{R}^n)$, so ist $\phi \cdot 1_M = \phi \cdot 1_{M'}$ fr eine *beschrnkte* offene oder abgeschlossene Menge M' ; wir schneiden dazu einfach M mit einem offenen oder abgeschlossenen Ball, der den Trger von ϕ enthlt. Die charakteristische Funktion $\chi := 1_{M'}$ lsst sich (nach §A4) durch eine L^1 -Cauchyfolge (1_k) in $C_0(\mathbb{R}^n)$ approximieren, und wie oben ist $(\phi \cdot 1_k)$ ebenfalls eine L^1 -Cauchyfolge in $C_0(\mathbb{R}^n)$, die gegen $\phi \cdot 1_M$ konvergiert, also ist $\phi \cdot 1_M \in \mathbf{L}(\mathbb{R}^n)$ und damit $1_M \in \mathbf{L}_{lok}(\mathbb{R}^n)$.

Allgemeiner heie eine Teilmenge $M \subset \mathbb{R}^n$ *lokal messbar*, wenn $1_M \in L_{lok}(\mathbb{R}^n)$. Ist M eine solche Menge, so heie eine Funktion $f : \mathbb{R}^n \rightarrow \mathbb{R}$ *integrierbar ber M* , wenn $f \cdot 1_M \in L(\mathbb{R}^n)$, und $L(M)$ sei der Raum der ber M integrierbaren Funktionen. Eine ber M integrierbare Funktion braucht natrlich nur auf M definiert zu sein; wir setzen sie einfach durch 0 auf $\mathbb{R}^n \setminus M$ fort.

Bemerkung. Auch $L(M)$ ist vollstndig bezglich der L^1 -Norm. Die in §A3 bewiesenen Eigenschaften von $L(\mathbb{R}^n)$ lassen sich auf $L(M)$ bertragen.

82. Satz 1. *Ist $f \in L(\mathbb{R}^n)$ und $g \in L_{lok}(\mathbb{R}^n)$ mit $|g| < C$ (also g beschrnkt), so gilt $f \cdot g \in L(\mathbb{R}^n)$.*

Beweis. Es gibt eine Folge (ϕ_k) in $C_0(\mathbb{R}^n)$ mit $\phi_k \dot{\rightarrow} f$ und $\phi_k \xrightarrow{L^1} f$. Dann ist $\phi_k \cdot g \in L(\mathbb{R}^n)$ (nach Definition der lokalen Integrierbarkeit) und $\phi_k \cdot g \dot{\rightarrow} f \cdot g$. Die Folge $(\phi_k \cdot g)$ ist eine L^1 -Cauchyfolge in $L(\mathbb{R}^n)$, denn

$$\|\phi_k \cdot g - \phi_l \cdot g\|_1 = \int (|\phi_k - \phi_l| \cdot |g|) \leq C \cdot \|\phi_k - \phi_l\|_1.$$

Nach Satz 1 in §A5 gibt es (nach bergang zu einer Teilfolge) $\tilde{f} \in L(\mathbb{R}^n)$ mit $\phi_k \cdot g \xrightarrow{L^1} \tilde{f}$ und $\phi_k \cdot g \dot{\rightarrow} \tilde{f}$, also $\tilde{f} \doteq f \cdot g$, und damit $f \cdot g \in L(\mathbb{R}^n)$. \square

83. Korollar. *Jede integrierbare Funktion $f \in L(\mathbb{R}^n)$ ist ber jede lokal messbare Teilmenge $M \subset \mathbb{R}^n$ integrierbar, da $f \cdot 1_M \in L(\mathbb{R}^n)$.*

84. Satz 2 (Transformationsatz) *Es seien U und V offene Teilmengen von \mathbb{R}^n und $\Phi : U \rightarrow V$ ein C^1 -Diffeomorphismus und $f \in L(V)$. Dann ist $(f \circ \Phi)|\det D\Phi| \in L(U)$ und*

$$\int_V f = \int_U ((f \circ \Phi)|\det D\Phi|).$$

Beweis. Es sei $g := f \cdot 1_V \in L(\mathbb{R}^n)$. Dann gibt es eine L^1 -Cauchyfolge (ϕ_k) in $C_0(\mathbb{R}^n)$ mit $\phi_k \xrightarrow{L^1} g$ und $\phi_k \dot{\rightarrow} g$. Wir drfen annehmen, dass ϕ_k Trger in V hat. Dazu ersetzen wir ϕ_k durch $\phi_k 1_l^-$ fr gengend groes l , wobei $1_l^- \nearrow 1_V$ die in §A4 angegebene stetige Approximation von 1_V ist: Da $1_l^- \phi_k \dot{\rightarrow} 1_U \phi_k$ und $|1_l^- \phi_k| \leq 1_U |\phi_k| \in L(\mathbb{R}^n)$, folgt $\|1_l^- \phi_k - 1_U \phi_k\|_1 \rightarrow 0$ nach Lebesgue (Satz 4 in §A5), und da ferner $\|1_U \phi_k - g\|_1 = \int_U |\phi_k - g| \leq \|\phi_k - g\|_1 \rightarrow 0$, folgt diese Behauptung.

Nach dem Transformationssatz für stetige Funktionen mit kompaktem Träger (Spezialfall von Satz 26.2) ist

$$\|((\phi_k - \phi_l) \circ \Phi) | \det D\Phi |\|_1 = \|\phi_k - \phi_l\|_1,$$

also ist $(\phi_k \circ \Phi) | \det D\Phi |_{k \in \bar{N}}$ eine L^1 -Cauchyfolge in $C_0(\mathbb{R}^n)$, und da diese Folge fast überall gegen $g \circ \Phi | \det D\Phi |$ konvergiert, ist die letztere Funktion integrierbar mit

$$\int (g \circ \Phi) | \det D\Phi | = \lim \int (\phi_k \circ \Phi) | \det D\Phi | = \lim \int \phi_k = \int g.$$

Dies war zu zeigen, denn $(g \circ \Phi) | \det D\Phi | = ((f \circ \Phi) | \det D\Phi |) \cdot 1_U$. \square

Jetzt seien $X \subset \mathbb{R}^p$ und $Y \subset \mathbb{R}^q$ lokal messbar, z.B. $X = \mathbb{R}^p$ und $Y = \mathbb{R}^q$. Der Satz von Fubini, den wir für Riemann-integrierbare Funktionen bereits kennengelernt haben, führt die Integration über $X \times Y \subset \mathbb{R}^{p+q}$ auf Integrationen über X und Y zurück. Für jede Funktion $f : X \times Y \rightarrow \mathbb{R}$ und jedes $y \in Y$ sei $f^y : X \rightarrow \mathbb{R}$ definiert durch,

$$f^y(x) = f(x, y).$$

85. Satz 3. (Fubini) *Es sei $f \in L(X \times Y)$. Dann gilt:*

- (a) $f^y \in L(X)$ für fast alle $y \in Y$,
- (b) Eine Funktion $F : Y \rightarrow \mathbb{R}$ mit $F(y) = \int_X f^y$ für fast alle $y \in Y$ ist integrierbar, d.h. $F \in L(Y)$ mit

$$\int_Y F = \int_{X \times Y} f,$$

$$\text{also } \int_{X \times Y} f = \int_Y \left(\int_X f(x, y) dx \right) dy.$$

Beweis. Auf den ersten Blick scheint der Beweis klar zu sein: f ist Limes einer L^1 -Cauchyfolge (ϕ_k) in $C_0(X \times Y)$, also $\phi_k \rightarrow f$ und $\phi_k \xrightarrow{L^1} f$. Da für stetige Funktionen mit kompaktem Träger (sogar für alle Riemann-integrierbaren Funktionen) der Satz von Fubini bereits gilt, folgt er durch einen Grenzübergang wohl auch für beliebige $f \in L(X \times Y)$. Allerdings gibt die Formulierung des Satzes schon etwas zu denken: Wir müssen ja zunächst einmal die Integrierbarkeit der Funktionen f^y und F aus dem Satz nachweisen! Die Sätze von Levi und Lebesgue bieten hier keine Hilfe; wir müssen schon auf die Definition der integrierbaren Funktionen als Limiten von L^1 -Cauchyfolgen zurückgreifen. Wir müssen also zunächst zeigen, dass die Folge (ϕ_k^y) für fast alle y eine L^1 -Cauchyfolge in $C_0(X)$ bilden (wobei wieder $\phi_k^y(x) := \phi_k(x, y)$ gesetzt wird).

Aus der Cauchyfolgen-Eigenschaft der Folge (ϕ_k) folgt

$$\int_Y \left(\int_X |\phi_k^y(x) - \phi_{k-1}^y(x)| dx \right) dy \rightarrow 0 \quad (*)$$

für $k \rightarrow \infty$. Wir wollen das innere Integral $\psi_k(y)$ nennen und definieren damit Funktionen $\psi_k \in C_0(Y)$ mit

$$\psi_k(y) = \int_X |\phi_k^y(x) - \phi_{k-1}^y(x)| dx = \|\phi_k^y - \phi_{k-1}^y\|_1.$$

Nach (*) ist $(\int \psi_k)$ eine Nullfolge; durch Übergang zu einer Teilfolge dürfen wir annehmen, dass

$$\int \psi_k = \|\psi_k\|_1 < 1/2^{2k}.$$

Mit dem vielbenutzten Lemma 1 in §A3 (S. 124) folgt $|\psi_k| < 1/2^k$ außerhalb einer kleinen Jordan-messbaren Menge $T_k \subset Y$ mit $\mu(T_k) < 1/2^k$. Außerhalb der Vereinigung $S_k = \bigcup_{l>k} T_l$ gilt daher

$$|\psi_l| < 1/2^l \quad \forall l > k,$$

und S_l ist eine ϵ -Menge für $\epsilon = 1/2^k$, also eine $1/2^k$ -Menge. Für alle $y \in Y \setminus S_k$ und für alle $m > l \geq k$ gilt daher

$$\|\phi_l^y - \phi_m^y\|_1 \leq \sum_{j=l+1}^m \psi_j < \sum_{j>l} 1/2^j = 1/2^l.$$

Somit ist (ϕ_l^y) eine L^1 -Cauchyfolge in $C_0(X)$ für alle $y \in Y \setminus S_k$ für beliebiges k , also für $y \in \bigcup_k (Y \setminus S_k) = Y \setminus \bigcap_k S_k$, und $N_1 := \bigcap_k S_k$ ist eine Nullmenge in Y (da $N_1 \subset S_k$ für alle k).

Für alle $y \in Y \setminus N_1$ gibt es also eine Funktion $\tilde{f}^y \in L(X)$ mit $\phi_l^y \xrightarrow{\cdot} \tilde{f}^y$ und $\phi_l^y \xrightarrow{L^1} \tilde{f}^y$. Wir möchten gerne $\tilde{f}^y = f^y$ schließen, aber das ist nicht ohne weiteres möglich, da ϕ_l ja nur außerhalb einer Nullmenge N in $X \times Y$ gegen f strebt; wir müssen wissen, ob der Schnitt von N mit den Mengen $X \times \{y\}$ wieder eine Nullmenge ist. Dies sagt uns das folgende Lemma, das ein Spezialfall des Satzes von Fubini für charakteristische Funktionen von Nullmengen ist:

86. Lemma. *Ist $N \subset \mathbb{R}^p \times \mathbb{R}^q$ eine Nullmenge, so ist $N^y := \{x \in \mathbb{R}^p; (x, y) \in \hat{N}\}$ für fast alle $y \in \mathbb{R}^q$ eine Nullmenge.*

Wir verschieben den Beweis dieses Lemmas an das Ende dieses Abschnittes und beenden zunächst den Beweis von Satz 2. Nach dem Lemma wissen wir jetzt, dass $\phi_l^y \rightarrow f^y$ für alle y außerhalb einer Nullmenge $N_2 \subset Y$. Also gilt $\tilde{f}^y \doteq f^y$ für alle $y \in Y \setminus N_3$, wobei $N_3 = N_1 \cup N_2$. Für diese y folgt also $f^y \in L(X)$ und $\phi_l^y \xrightarrow{L^1} f^y$. Insbesondere sehen wir für alle $y \in Y \setminus N_3$:

$$\Phi_l(y) := \int_X \phi_l^y \rightarrow \int_X f^y =: F(y).$$

Damit sind Funktionen $\Phi_l, F : Y \rightarrow \mathbb{R}$ definiert mit $\Phi_l \in C_0(Y)$ und $\Phi_l \rightarrow F$ (wobei F auf N_3 z.B. Null gesetzt wird). Die Φ_l bilden sogar eine L^1 -Cauchyfolge in $C_0(Y)$, denn für genügend große $l < m$ ist

$$\|\Phi_l - \Phi_m\|_1 = \int_Y |\Phi_l - \Phi_m| = \int_{X \times Y} |\phi_l - \phi_m| < \epsilon.$$

Somit ist $F \doteq \lim_l \Phi_l \in L(Y)$ und $\Phi_l \xrightarrow{L^1} F$, und insbesondere folgt $\int_Y \Phi_l \rightarrow \int_Y F$. Andererseits gilt nach Wahl der ϕ_l :

$$\int_Y \Phi_l = \int_{X \times Y} \phi_l \rightarrow \int_{X \times Y} f,$$

und somit folgt die Behauptung $\int_{X \times Y} f = \int_Y F$. \square

Beweis des Lemmas. Wir setzen $X := \mathbb{R}^p$ und $Y := \mathbb{R}^q$. Für jede Teilmenge $S \subset X \times Y$ und alle $y \in Y$ sei S^y der Schnitt von S mit $X \times \{y\}$, genauer:

$$S^y := \{x \in X; (x, y) \in S\}.$$

Unsere Nullmenge N liegt für jedes $\epsilon > 0$ in einer Lebesgue-messbaren Menge $S = \bigcup_{k=1}^{\infty} A_k$, wobei A_k Quader in $X \times Y$ von beliebig kleinem Gesamtmaß sind, etwa

$$\mu_L(S) \leq \sum_k \mu(A_k) \leq \epsilon^2.$$

Jeder dieser Quader A_k ist kartesisches Produkt $A_k = B_k \times C_k$ von Quadern $B_k \subset X$ und $C_k \subset Y$. Für festes $y \in Y$ ist S^y die Vereinigung von einigen dieser Quader B_k , und zwar gehört B_k genau dann zu S^y , wenn der zugehörige "Partner" C_k das y trifft. Insbesondere ist S^y als höchstens abzählbare Vereinigung von Quadern wieder Lebesgue-messbar (Folgerung 1 in §A5, S. 137). Wir definieren jetzt

$$Y_S := \{y \in Y : \mu_L(S^y) > \epsilon\}.$$

Wir wollen zeigen, dass Y_S Lebesgue-messbar ist mit $\mu_L(Y_S) \leq \epsilon$ falls $\mu_L(S) < \epsilon^2$. Wenn S selbst ein Quader $B \times C$ ist, ist dies klar: Dann

ist $Y_S = C$ falls $\mu(B) > \epsilon$ (andernfalls ist $Y_S = \emptyset$), und $\mu(C) < \epsilon$, da $\mu(B)\mu(C) \leq \epsilon^2$. Ebenso leicht folgt dies, wenn S endliche Vereinigung von Quadern ist, etwa $S = A_1 \cup \dots \cup A_N$ mit $\sum_i \mu(A_i) \leq \epsilon^2$ und $A_i = B_i \times C_i$. Ohne Einschränkung sei y genau in C_1, \dots, C_k enthalten, und unter den Quadern B_1, \dots, B_k gelte $\mu(B_i) > \epsilon$ genau für $i = 1, \dots, j$ (mit $j \leq k$), also $Y_S = C_1 \cup \dots \cup C_j$. Dann ist

$$\epsilon^2 \geq \sum_{i=1}^j \mu(B_i)\mu(C_i) > \epsilon \cdot \sum_{i=1}^j \mu(C_i) \geq \epsilon \cdot \mu(Y_S).$$

Ist schließlich S unendliche Vereinigung von Quadern A_1, A_2, \dots , so ist $S = \bigcup_N S^N$ mit $S^N = A_1 \cup \dots \cup A_N$, und $Y_S = \bigcup_N Y_{S^N}$. Also ist Y_S abzählbare Vereinigung von Quadern und damit Lebesgue-messbar, und da $\mu(S^N) \leq \epsilon^2$ für alle N , folgt $\mu_L(Y_S) = \lim_N \mu(Y_{S^N}) \leq \epsilon$.

Wir wählen nun eine monoton fallende Mengenfolge $(S_k)_{k \in \mathbb{N}}$ mit $S_k \supset N$, so dass jedes S_k eine $1/2^{2k}$ -Menge ist, also von Quadern mit Gesamtmaß $\leq 1/2^{2k}$ überdeckt wird. Definieren wir

$$Y_k = \{y \in Y; \mu((S_k)^y) > 1/2^k\},$$

so ist nach der vorstehenden Überlegung $\mu_L(Y_k) \leq 1/2^k$. Dann hat $Y^k := \bigcup_{l > k} Y_l$ ebenfalls Lebesgue-Maß $\mu_L(Y^k) \leq 1/2^k$, und für alle $y \in Y \setminus Y^k$ und für alle $l > k$ gilt $\mu_L((S_l)^y) \leq 1/2^l$. Für jedes $y \in \bigcup_k (Y \setminus Y^k) = Y \setminus N'$ mit $N' = \bigcap_k Y_k$ gilt also $\mu_L((S_l)^y) \leq 1/2^l$ für genügend große l . Da $N^y \subset (S_l)^y$ für alle l , ist N^y eine Nullmenge in X für alle $y \in Y \setminus N'$, und N' ist eine Nullmenge in Y . Dies war zu zeigen.

Literatur

- [Alt] H.W..Alt: Lineare Funktionalanalysis, Springer 1985
- [Aulbach] B.Aulbach: Analysis I/II. Vorlesungsskriptum 1992/93, Institut für Mathematik, Universität Augsburg
- [Barner-Flohr] M.Barner, F.Flohr: Analysis II. De Gruyter 1983
- [Forster] O. Forster: Analysis I,II,III, Vieweg 1977-1981
- [Heuser] H. Heuser: Gewöhnliche Differentialgleichungen, Teubner 1989
- [Milnor] J. Milnor: Morse Theory. Princeton 1962
- [Königsberger] K. Königsberger: Analysis II, Springer 1990
- [Spivak] M. Spivak: Calculus on Manifolds, Benjamin 1965
- [Wüst] R. Wüst: Höhere Mathematik für Physiker, Teil 2, De Gruyter 1995

Symbolverzeichnis

- $\|A\|$ Matrixnorm 21
- A^T Transposition 16
- $B_r(x)$ offener Ball vom Radius r um x
- C^1 (C^k) (k -mal) stetig differenzierbar 28
- $C_0(\mathbb{R}^n)$ Stetige Funktionen mit kompaktem Träger 80
- $\overline{C_0(\mathbb{R}^n)}$ Vervollständigung von $C_0(\mathbb{R}^n)$ 84
- CT Komplementmenge 84
- $D_i f$ i -te partielle Ableitung 13
- ∂T Rand 5
- $\partial_M T$ relativer Rand (bezüglich M) 5
- $\|f\|$ Supremumsnorm 8
- $\|f\|_1$ L^1 -Norm 83
- $f_k \xrightarrow{L^1} f$ L^1 -Konvergenz 92
- $f_k \xrightarrow{glm} f$ gleichmäßige Konvergenz 8
- $f_k \xrightarrow{\dot{\rightarrow}} f$ Konvergenz fast überall 92
- $f \xrightarrow{pw} f$ Punktweise Konvergenz 8
- $f_k \nearrow f$ monotone Konvergenz 79
- $f \doteq g$ fast überall gleich 83
- $f \circ g$ Verkettung $(f \circ g)(x) := f(g(x))$ 20
- $K_r(x)$ abgeschlossene Kugel vom Radius r um x 55
- $L(\mathbb{R}^n)$ Lebesgue-integrierbare Funktionen 83
- $L_{lok}(\mathbb{R}^n)$ lokal Lebesgue-integrierbare Funktionen 96
- $L(M)$ Lebesgue-integrierbare Funktionen über M 96
- $L^1(\mathbb{R}^n)$ (fast überall gleiche L^1 -Funktionen werden identifiziert) 84
- $\mu(T)$ Jordan-Maß 50, 52
- $\mu_L(M)$ Lebesgue-Maß 90
- $R(Q)$ Riemann-integrierbare Funktionen über Q 56
- $R(\mathbb{R}^n)$ Riemann-integrierbare Funktionen über \mathbb{R}^n 64
- $S\Delta T$ symmetrische Mengendifferenz 70
- \bar{T} Abschluß 5
- T_b Translation $x \mapsto x + b$ 65
- $T_x M$ Tangentialraum 36
- U^n, V^n offene Teilmengen des \mathbb{R}^n 5
- $\langle x, y \rangle$ Skalarprodukt 2
- $[x, y]$ Strecke von x nach y 25
- \approx_ϵ ungefähr gleich, Fehler $< \epsilon$ 70
- \perp senkrecht 2
- \forall "für alle" 6

- \exists “existiert” \exists
- \wedge “und” \wedge

INDEX

- abgeschlossen, 11–13, 134
- Ableitung, 19, 24, 25, 29, 39, 43, 46, 68
- Abstand, 7, 131
- Abstandsfunktion, 119
- Abstraktion, 119
- affine Abbildung, 69
- affine Menge, 52
- affiner Teilraum, 52
- analytisch, 64
- äquivalent, 9
- Arbeit, 23, 38
- Archimedes, 73, 91
- äußeres Maß, 77, 117, 134

- Ball, 11
- Banachraum, 18, 42, 121
- Banachscher Fixpunktsatz, 40, 44
- Bereichsadditivität, 80
- beschränkte Funktion, 78
- beschränkt, 76
- beschränkte Folge, 11
- Betrag, 7, 9
- Bogenlänge, 22, 116

- Cauchy-Schwarz-Ungleichung, 8, 22, 23, 33, 35
- Cauchyfolge, 10, 18, 41, 87, 118–120, 122, 123, 127
- Cavalieri-Prinzip, 90, 91, 93, 115
- Charakteristische Funktion, 77, 78, 142
- Codimension, 55

- definit, 65
- De l'Hôpital, 21
- Determinante, 45, 72, 96, 110, 111, 113
- Determinantenkriterium, 65
- Diffeomorphismus, 43, 47, 49, 52, 53, 68, 69, 86, 110, 111, 140
- Differentialgleichung, 42, 111
- differenzierbar, 19, 20, 25–27, 29–31, 33, 38, 39, 43, 48, 59, 65
- Dimension, 6, 11, 52, 54, 55, 62, 115, 122
- diskret, 56
- Divergenz, 104, 107, 109

- Divergenzsatz, 109, 114–116
- Dreiecksungleichung, 8, 9

- Eigenvektor, 60
- Eigenwert, 60
- Einheitssphäre, 35
- ϵ -Menge, 122, 127, 136, 142
- euklidischer Raum, 6
- Euler, 62
- explizit, 47, 48, 50, 51
- ϵ -Menge, 125

- fast überall, 122, 123, 125–127, 130, 134, 136, 137
- Fibonaccizahlen, 41
- Fläche, 53, 57, 115
- Flächenelement, 104
- Flächeninhalt, 73, 74
- Fluss, 111–113
- Folge, 10, 11, 118, 122, 138
- Form, 61, 65
- Fubini, 87, 88, 100, 139, 141, 142
- Funktion, 14

- Gammafunktion, 101
- Gaußverteilung, 100
- genügend groß, 14
- geometrische Reihe, 41
- gleichmäßig konvergent, 17, 117, 125
- gleichmäßig stetig, 16, 82
- Gleichung, 40, 44, 47
- goldener Schnitt, 41
- Grad, 61
- Gradient, 36–38, 56
- Graph, 19, 47, 95
- Grenzwert, 10

- Häufungspunkt, 11
- Hauptsatz, 22, 39, 42, 87, 96
- Hesseform, 63, 66
- Hessematrix, 66
- homogenes Polynom, 61
- Hyperfläche, 55, 109, 115

- implizit, 47
- impliziter Funktionensatz, 47
- indefinit, 66
- Indikatorfunktion, 77

- inneres Maß, 77, 78
 Integral, 22, 23, 78, 109, 117
 Integration, 74, 94, 109, 139, 141
 integrierbar, 76, 78, 80, 117, 122

 Jacobimatrix, 29, 30, 96
 Jim Knopf, 38

 Karte, 53, 110
 kartesische Koordinaten, 6
 Kegel, 91, 92
 Kettenregel, 33, 36, 50, 59, 96
 kofinal, 120, 121, 126
 kompakt, 13, 15, 16, 85, 86, 115
 komplex differenzierbar, 47
 konservativ, 38
 Kontraktion, 40, 41
 konvergent, 10, 120
 konvex, 40
 Koordinaten, 68
 Koordinatensystem, 53, 68
 Kraftfeld, 23, 38
 Kreuzprodukt, 115
 kritischer Punkt, 59, 60
 kubische Form, 61
 Kugel, 44, 56, 72, 73, 91, 93, 103, 114
 Kugelkoordinaten, 72
 Kurve, 19
 Kurvenintegral, 23, 115

 Lagrange-Multiplikator, 59
 Länge, 74
 Längenelement, 104
 Lebesgue-Integral, 117
 Lebesgue-integrierbar, 122
 Lebesgue-Maß, 131
 Lemniskate, 67
 Limes, 10
 lineare Approximation, 27, 32
 Linearform, 61
 Linearität, 80
 Lipschitz-stetig, 40
 Lissajou-Figuren, 68
 lokal integrierbar, 139
 lokale Extrema, 58, 65

 Mannigfaltigkeit, 52–55, 58, 59, 68, 85, 110, 115
 Maß, 73–75, 78, 84, 90, 102, 111, 114, 131, 134

 Matrixnorm, 34, 39
 Maximum, 15, 58
 messbar, 78, 84, 90, 92, 96, 112, 117, 131, 133, 134
 Metrik, 119
 metrischer Raum, 120, 121
 Minimum, 15, 58
 Mittelwertsatz, 20, 26, 39, 80
 Monotonie, 80

 Nebenbedingung, 59
 nicht-trivial, 31
 Niveaumenge, 52, 54, 55, 58
 Norm, 9, 18, 33, 119
 Normalenvektor, 104
 Normalraum, 55
 Nullmenge, 81, 84, 85, 91, 96, 122

 Oberfläche, 111, 114
 Oberflächenintegral, 109
 Oberintegral, 77
 Obersumme, 76
 offen, 12, 15, 133
 offene Abbildung, 69
 offene Überdeckung, 86
 Orthonormale Matrix, 103

 Parametrisierung, 58, 69, 110, 116
 partiell differenzierbar, 25, 30
 partielle Ableitung, 25
 partielle Integration, 89, 94
 Perpetuum Mobile, 38
 Polarkoordinaten, 70, 71, 99
 Polynom, 61
 Potential, 38
 Projektion, 48, 52, 68
 punktweise konvergent, 16, 17, 119, 123

 Quader, 74, 75, 82
 quadratische Ergänzung, 8, 65
 quadratische Form, 61, 63, 68
 Quadrik, 57

 radiale Funktion, 100
 Rand, 11, 13, 81, 109, 112, 115, 133
 Randpunkt, 11, 13
 rationale Funktion, 16, 34
 reelle Zahl, 120
 regulär, 19, 54, 59

- relativ abgeschlossen, 13
- relativ offen, 13
- Restglied, 32, 63, 65
- Richtungsableitung, 36
- Riemann-Integral, 78
- Rotation, 39, 104, 115, 116
- Rotationstorus, 56

- Sattelpunkt, 60, 66
- Satz von Gauß, 104
- Satz von Lebesgue, 138
- Satz von Levi, 137
- Satz von Stokes, 104, 115, 116
- Scherung, 90
- Schranksatz, 39
- senkrecht, 8, 56
- Singulärwertzerlegung, 104
- Sphäre, 56
- Spur, 107
- stetig, 14–17, 80, 82
- Strecke, 13, 22, 39, 40
- Substitutionsregel, 24, 95, 99
- Supremumsnorm, 17, 18
- symmetrische Matrix, 60

- Tangentenvektor, 19, 20, 53
- Tangentialraum, 53
- Tangentialvektor, 53
- Taylorformel, 63, 97
- Taylorpolynom, 64
- Taylorreihe, 64
- Teilfolge, 11
- Tetraeder, 92
- totales Differential, 29
- Träger, 118
- Transformationssatz, 95, 140
- Treppenfunktion, 124, 129, 138
- trivial, 31

- Umkehrsatz, 43
- Unterintegral, 77
- Urbild, 14

- Vektor, 6
- Vektorfeld, 23, 24, 38, 104, 109, 115
- Vektorprodukt, 115
- Verfeinerung, 74, 75, 78
- Vervollständigung, 119, 123, 130
- Vollständigkeit, 10, 18, 120, 121, 136
- Volumen, 73, 74, 91, 93, 96, 114

- Volumenelement, 104
- Würfel, 85

- Zeit, 19, 22, 23, 112
- Zerlegung, 74, 75
- zusammenhängend, 15
- Zwischenwertsatz, 15, 80
- Zylinderkoordinaten, 71