

Analysis II

Universität Mannheim

Dozent: H.-J. Bartels
Schrift: P. Höpner
Überarbeitung von D. Boldin 2009

Inhaltsverzeichnis

1	Differentialrechnung im \mathbb{R}^n	4
1.1	Der n-dimensionale Raum	4
1.2	Funktionen und Abbildungen	14
1.3	Kurven im \mathbb{R}^n	23
1.4	Differenzierbarkeit	30
1.5	Höhere Ableitungen, Taylor-Entwicklung und lokale Extremwerte	40
1.6	Implizite Funktionen, Umkehrabbildung und Extremwerte mit Nebenbedingungen	49
2	Integralrechnung von Funktionen mehrerer Veränderlicher	60
2.1	Jordanscher Inhalt und Lebesguesches Maß	60
2.2	Das Lebesguesche Integral	65
2.3	Grenzwertsätze	69
2.4	Der Satz von Fubini und das Cavalierische Prinzip	71
2.5	Transformation von Integralen	77
A	Integrationstheorie	80
A.1	Der Approximationssatz von Stone-Weierstraß	80
A.2	Das Integral für stetige Funktionen mit kompaktem Träger	90
A.3	Die Ausdehnung des Integrals auf halbstetige Funktionen	98
A.4	Der Daniell-Lebesgue-Prozess, Teil 1: Das Integral für halbstetige Funktionen	99
A.5	Der Daniell-Lebesgue-Prozess, Teil 2: Das äußere Integral	108
A.6	Die integrierbaren Funktionen	113
A.7	Integrierbarkeitskriterien	120
A.8	Nullmengen	123
A.9	Der Satz von Fubini	125
A.10	Die Transformationsformel	129
A.11	Messbarkeit	139

Vorbemerkung

In dem zweiten Teil der Analysis-Vorlesung geht es um zwei große Themen:

(i) Zunächst beschäftigen wir uns mit der Differentialrechnung von Funktionen mehrerer Veränderlicher. Nach einer Einführung in die Topologie des \mathbb{R}^n geht es um partielle Ableitungen, totale Differenzierbarkeit, Taylor-Formel, Extremwerte und um die Auflösung von Gleichungssystemen (implizite Funktionen).

(ii) Der zweite Teil ist der Integrationstheorie gewidmet, die wir im Spezialfall einer Dimension bereits in der Analysis I mit dem Riemannsches Ansatz behandelt haben. Hier geht es uns darum, einer möglichst großen Klasse mehrdimensionaler Funktionen ein Integral zuzuordnen (das Lebesgue-Integral). Wir zeigen den auch in der Wahrscheinlichkeits- und Maßtheorie viel verwendeten Satz von Fubini und im darauffolgenden Paragraphen die Transformationsformel.

Bei der Integrationstheorie werden im Skript zwei Zugänge dargestellt: Der erste folgt dem klassischen maßtheoretischen Zugang, während der zweite im Anhang dargestellte Zugang das Integral dadurch einführt, dass man ein lineares Funktional zunächst auf geeigneten elementaren Funktionen definiert und dann auf eine größere Klasse von Funktionen fortsetzt. Hier folgen wir in der Darstellung eng einem Skript von E. Freitag [5].

Wie in der Vorlesung Analysis I wird versucht, eine allzu große Abstraktion zu vermeiden, und die Theorie wird durch konkrete Beispiele erläutert.

Kapitel 1

Differentialrechnung im \mathbb{R}^n

1.1 Der n -dimensionale Raum

Für die Untersuchung von Funktionen mehrerer Veränderlicher werden einige topologische Begriffe im \mathbb{R}^n benötigt, wie z.B. offene Mengen, abgeschlossene Mengen, Rand, abgeschlossene Hülle usw. Dabei verzichten wir hier auf die Behandlung allgemeiner metrischer Räume; die Überlegungen dieses Paragraphen sind ganz auf den \mathbb{R}^n zugeschnitten, auch wenn eine Reihe von Schlußweisen auf normierte Vektorräume übertragbar sind. Wir wollen uns auf das Notwendige beschränken.

Betrachtet werden die n -dimensionalen \mathbb{R} - (bzw. \mathbb{C} -)Vektorräume

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

bzw.

$$\mathbb{C}^n = \{(z_1, \dots, z_n) : z_i \in \mathbb{C}\}.$$

Setzt man

$$x_{2k-1} := \operatorname{Re} z_k, \quad x_{2k} := \operatorname{Im} z_k \quad (k = 1, \dots, n),$$

so hat man eine Bijektion

$$\mathbb{C}^n \rightarrow \mathbb{R}^{2n}.$$

Alle Aussagen in diesem Paragraphen, die für endlichdimensionale reelle Räume bewiesen werden, sind dann auch für \mathbb{C}^n gültig.

\mathbb{R}^n besitzt eine \mathbb{R} -Vektorraumstruktur. Die Addition zweier Vektoren $x = (x_1, \dots, x_n)$ und $y = (y_1, \dots, y_n)$ ist dabei komponentenweise definiert:

$$x + y := (x_1 + y_1, \dots, x_n + y_n),$$

und entsprechend definiert man die skalare Multiplikation:

$$\lambda \cdot x := (\lambda x_1, \dots, \lambda x_n) \quad (\lambda \in \mathbb{R}, x \in \mathbb{R}^n).$$

Als **Betrag** des Vektors $x = (x_1, \dots, x_n)$ wird definiert:

$$|x| = \sqrt{x_1^2 + \dots + x_n^2}, \quad (1.1)$$

und es gilt für Vektoren $a, b \in \mathbb{R}^n$ die **Dreiecksungleichung**

$$|a + b| \leq |a| + |b| \quad (1.2)$$

Als Verallgemeinerung eindimensionaler Intervalle in $\mathbb{R} = \mathbb{R}^1$ können folgende Mengen angesehen werden (mit $a \in \mathbb{R}^n, r \in \mathbb{R}, r > 0$):

$$\begin{aligned} K(a, r) &:= \{x \in \mathbb{R}^n : |x - a| < r\}, \\ \overline{K}(a, r) &:= \{x \in \mathbb{R}^n : |x - a| \leq r\}. \end{aligned} \quad (1.3)$$

Man spricht in diesem Fall von einer **offenen** bzw. **abgeschlossenen Kugel mit Mittelpunkt a und Radius r** .

Eine andere Verallgemeinerung sind, mit $a, b \in \mathbb{R}^n$, die Mengen

$$\begin{aligned} Q(a, b) &:= \{x \in \mathbb{R}^n : a_k < x_k < b_k \text{ für } k = 1, \dots, n\}, \\ \overline{Q}(a, b) &:= \{x \in \mathbb{R}^n : a_k \leq x_k \leq b_k \text{ für } k = 1, \dots, n\}. \end{aligned} \quad (1.4)$$

Hier spricht man von einem **offenen** (bzw. **abgeschlossenen**) n -dimensionalen **Quader** mit Kantenlängen $b_k - a_k$. Sind die Kantenlängen alle gleich, so spricht man von einem offenen (bzw. abgeschlossenen) **Würfel**.

Sei $A \subseteq \mathbb{R}^n$ eine beliebige Menge. Dann wird definiert:

Definition 1.1. Ein Punkt $a \in \mathbb{R}^n$ heißt

- a) **innerer Punkt** von A , wenn es ein $r > 0$ gibt mit $K(a, r) \subseteq A$;
- b) **äußerer Punkt** von A , wenn es ein $r > 0$ gibt mit $K(a, r) \subseteq \mathbb{R}^n \setminus A$;
- c) **Randpunkt** von A , wenn a weder innerer noch äußerer Punkt ist, d.h. wenn die Kugel $K(a, r)$ für jedes $r > 0$ sowohl Punkte aus A wie auch Punkte aus $\mathbb{R}^n \setminus A$ enthält.

Die inneren Punkte und die Randpunkte von A heißen **Berührungspunkte** von A .

Bemerkung 1.2. Genau dann ist a ein Berührungspunkt von A , wenn $K(a, r)$ für jedes $r > 0$ Punkte von A enthält.

Anstelle von Zahlenfolgen im eindimensionalen Fall betrachten wir nun Folgen von Vektoren im \mathbb{R}^n :

Definition 1.3. Die Folge $x^{(m)}$ von Vektoren des \mathbb{R}^n konvergiert gegen den Grenzwert y genau dann, wenn es zu jedem $\varepsilon > 0$ ein N gibt mit

$$|x^{(m)} - y| \leq \varepsilon \quad \text{für alle } m \geq N.$$

Schreibweise: $y = \lim_{m \rightarrow \infty} x^{(m)}$.

Satz 1.4. Genau dann ist a Berührungspunkt einer Menge $A \subseteq \mathbb{R}^n$, wenn es eine Folge $x^{(m)} \in A$ gibt mit $\lim_{m \rightarrow \infty} x^{(m)} = a$.

Beweis. Sei zunächst a ein Berührungspunkt von A . Zu jedem $m \in \mathbb{N}$ existiert dann ein $x^{(m)} \in A \cap K(a, \frac{1}{m})$. Dann gilt

$$|x^{(m)} - a| < \frac{1}{m} \quad \text{für alle } m \in \mathbb{N}.$$

Daraus folgt $\lim_{m \rightarrow \infty} x^{(m)} = a$.

Für die umgekehrte Richtung sei $x^{(m)} \in A$ mit $\lim_{m \rightarrow \infty} x^{(m)} = a$. Zu jedem $\varepsilon > 0$ existiert dann also ein N mit $|x^{(m)} - a| \leq \frac{\varepsilon}{2}$ für $m \geq N$. Jeder dieser Punkte $x^{(m)}$ liegt daher in $A \cap K(a, \varepsilon)$, und daher ist a in der Tat ein Berührungspunkt von A . \square

Definition 1.5. Eine Teilmenge $A \subseteq \mathbb{R}^n$ heißt **abgeschlossen**, wenn A alle seine Randpunkte enthält, und **offen**, wenn sie keinen Randpunkt enthält, d.h. wenn A nur aus inneren Punkten besteht.

Beispiel 1.6.

1. \mathbb{R}^n und die leere Menge \emptyset sind sowohl offen wie auch abgeschlossen.
2. Sei $A = \{(x, y) : 1 < x^2 + y^2 \leq 2\} \subseteq \mathbb{R}^2$.

Die Menge der inneren Punkte von A ist

$$\{(x, y) : 1 < x^2 + y^2 < 2\}.$$

Die Menge der äußeren Punkte ist

$$\{(x, y) : x^2 + y^2 < 1\} \cup \{(x, y) : x^2 + y^2 > 2\}.$$

Die Menge der Randpunkte von A ist

$$\{(x, y) : x^2 + y^2 = 1\} \cup \{(x, y) : x^2 + y^2 = 2\}.$$

Offensichtlich ist die Menge A weder offen noch abgeschlossen.

3. Sei nun $A := \mathbb{Q} = \{\text{rationale Zahlen}\} \subseteq \mathbb{R} = \mathbb{R}^1$.

Die Menge der inneren Punkte von A ist leer. Die Menge der Randpunkte von A und damit die Menge der Berührungspunkte von A ist \mathbb{R} . Auch diese Menge A ist weder offen noch abgeschlossen.

4. Genau dann ist a innerer (bzw. äußerer) Punkt einer Menge A , wenn a äußerer (bzw. innerer) Punkt von $\mathbb{R}^n \setminus A$ ist. a ist Randpunkt von A genau dann, wenn a Randpunkt von $\mathbb{R}^n \setminus A$ ist. Weiter gilt:

A ist offen (bzw. abgeschlossen) $\iff \mathbb{R}^n \setminus A$ ist abgeschlossen (bzw. offen).

Satz 1.7. Die Menge A° der inneren Punkte von A ist die größte offene Menge, die in A enthalten ist. A° heißt der **offene Kern** von A .

Beweis. Für jedes $a \in A^\circ$ gibt es ein $r > 0$ mit $K(a, r) \subseteq A$ und damit $K(a, \frac{r}{2}) \subseteq A^\circ$, da wegen der Dreiecksungleichung für $b \in K(a, \frac{r}{2})$ stets $K(b, \frac{r}{2}) \subseteq A$ gilt. Also ist A° offen. Ist $B \subseteq A$ eine offene Menge, so existiert zu jedem $b \in B$ eine Kugel K mit $b \in K \subseteq B \subseteq A$ und damit $b \in A^\circ$. Also gilt $B \subseteq A^\circ$. \square

Satz 1.8. Die Menge \overline{A} aller Berührungspunkte von A ist die kleinste abgeschlossene Menge, in der A enthalten ist. \overline{A} heißt die **abgeschlossene Hülle** von A .

Beweis. Der Beweis ergibt sich mit (1.7) aus dem Beispiel 1.6-4. und sei dem Leser überlassen. \square

Satz 1.9.

- (i) Der Durchschnitt beliebig vieler und die Vereinigung endlich vieler abgeschlossener Mengen ist abgeschlossen.
- (ii) Der Durchschnitt endlich vieler und die Vereinigung beliebig vieler offener Mengen ist offen.

Beweis. Für beliebige Mengen $A_i \subseteq \mathbb{R}^n$, wobei i aus einer Indexmenge I ist, gelten folgende Gleichungen:

$$\mathbb{R}^n \setminus \bigcap_{i \in I} A_i = \bigcup_{i \in I} (\mathbb{R}^n \setminus A_i),$$

$$\mathbb{R}^n \setminus \bigcup_{i \in I} A_i = \bigcap_{i \in I} (\mathbb{R}^n \setminus A_i).$$

Wegen dem Beispiel 1.6-4 sind daher die Aussagen (i) und (ii) von Satz (1.9) gleichwertig. Also genügt es, (ii) zu beweisen. Hierzu seien B_i für $i \in I$ offene Mengen. Es gilt zunächst

$$a \in \bigcup_{i \in I} B_i \implies \exists j \in I : a \in B_j.$$

Da alle B_j offen sind, gilt $K(a, r) \subseteq B_j$ für passendes $a \in \mathbb{R}^n$, $r > 0$. Damit gilt $K(a, r) \subseteq \cup_{i \in I} B_i$, d.h. $\cup_{i \in I} B_i$ ist offen.

Nun sei $I = \{1, \dots, m\}$ endlich und sei $a \in \cap_{i \in I} B_i = \cap_{i=1}^m B_i$. Für alle i liegt dann a in B_i , und weil B_i offen ist, gibt es ein $r_i > 0$ mit $K(a, r_i) \subseteq B_i$. Setzt man $r = \min_i r_i$, so folgt $K(a, r) \subseteq \cap_{i=1}^m B_i$. \square

Wir wollen nun die **Konvergenz** von Punktfolgen im \mathbb{R}^n noch eingehender untersuchen. Es gilt der

Satz 1.10. Sei $x^{(m)} = (x_{m1}, \dots, x_{mn})$ eine Punktfolge im \mathbb{R}^n . Die Folge $\{x^{(m)}\}$ konvergiert gegen $y = (y_1, \dots, y_n)$ genau dann, wenn gilt: Jede der n Zahlenfolgen $\{x_{mk}\}$ ($k = 1, \dots, n$) konvergiert und zwar gegen y_k .

Mit anderen Worten: Koordinatenweise Konvergenz ist notwendig und hinreichend für die Konvergenz einer Punktfolge im \mathbb{R}^n .

Beweis. Für beliebige $a = (a_1, \dots, a_n) \in \mathbb{R}^n$ gilt

$$\max_k |a_k| \leq |a| = \sqrt{\sum_{k=1}^n a_k^2} \leq \sqrt{n} \cdot \max_k |a_k|.$$

Speziell für $a := x^{(m)} - y$ ergibt dann

$$\max_k |x_{mk} - y_k| \leq |x^{(m)} - y| \leq \sqrt{n} \cdot \max_k |x_{mk} - y_k|$$

die behauptete Äquivalenz. \square

Bemerkung 1.11. Mit Hilfe von Satz (1.10) können die Rechenregeln für Grenzwerte von Zahlenfolgen auf die analogen Rechenregeln für Vektorfolgen übertragen werden. Auch das Cauchysche Konvergenzkriterium für Zahlenfolgen lässt sich übertragen.

Satz 1.12 (Cauchysches Konvergenzkriterium). *Die Folge $x^{(m)} \in \mathbb{R}^n$ konvergiert genau dann, wenn es zu jedem $\varepsilon > 0$ ein N gibt mit $|x^{(p)} - x^{(q)}| \leq \varepsilon$ für alle $p, q \geq N$.*

Definition 1.13. *Eine Menge $M \subseteq \mathbb{R}^n$ heißt **beschränkt**, wenn es ein $r > 0$ gibt, so dass $|x| \leq r$ für alle $x \in M$ gilt, d.h. wenn $M \subseteq \overline{K}(0, r)$ gilt.*

Bemerkung 1.14. M ist genau dann beschränkt, wenn die Beträge der Koordinaten der Punkte von M beschränkt sind.

Analog zum eindimensionalen Fall gilt:

Satz 1.15 (Satz von Bolzano-Weierstraß). *Jede beschränkte Punktfolge enthält eine konvergente Teilfolge.*

Beweis. Sei $m \in \mathbb{N}$ und seien $x_m := (x_{m1}, \dots, x_{mn})$ die Folgenglieder der gegebenen beschränkten Folge. Nach der Bemerkung (1.14) sind die n Koordinatenfolgen (x_{mk}) für $k = 1, \dots, n$ ebenfalls beschränkt.

Da der Satz von Bolzano-Weierstraß im eindimensionalen Fall gilt, gibt es eine Teilfolge $\{x_m^{(1)}\} = \{(x_{m1}^{(1)}, \dots, x_{mn}^{(1)})\}$ von x_m , so dass die Zahlenfolge $\{x_{m1}^{(1)}\}$ konvergiert.

Dann ist $\{x_{m2}^{(1)}\}$ als Teilfolge der Zahlenfolge $\{x_{m2}\}$ ebenfalls beschränkt. Wir können daher wiederum eine Teilfolge $\{x_m^{(2)}\}$ von $\{x_m^{(1)}\}$ wählen, so dass die Folge $\{x_{m2}^{(2)}\}$ konvergiert. Setzt man nun dieses Verfahren fort, so erhalten wir eine Teilfolge

$$\{x_m^{(n)}\}$$

von $\{x_m\}$, so dass alle Koordinatenfolgen

$$x_{mk}^{(n)}$$

für alle $k = 1, \dots, n$ konvergieren. Wegen (1.10) konvergiert dann auch die Punktfolge $\{x_m^{(n)}\}$. □

Wir werden später auch den Begriff der **kompakten** Teilmenge des \mathbb{R}^n benötigen. Zur Einführung dieses wichtigen Begriffs verwenden wir folgende

Definition 1.16. Sei A irgendeine Teilmenge des \mathbb{R}^n . Eine Familie

$$\mathcal{U} = \{U_i : i \in I\}$$

heißt **Überdeckung** von A , wenn gilt:

$$A \subseteq \bigcup_{i \in I} U_i.$$

Dabei ist I eine beliebige (endliche oder unendliche) Indexmenge. \mathcal{U} heißt **offene Überdeckung** (von A), wenn alle $U_i \subseteq \mathbb{R}^n$ offene Mengen sind.

Beispiel 1.17.

(i) Die Menge der beiden Intervalle

$$\mathcal{U}_1 := \left\{ \left[0, \frac{1}{2} \right], \left[\frac{1}{2}, 1 \right] \right\}$$

ist eine Überdeckung des Intervalls $(0, 1)$.

(ii) Die Menge

$$\mathcal{U}_2 := \left\{ \left[\frac{1}{2n}, \frac{1}{n} \right] : n \in \mathbb{N} \right\}$$

ist Überdeckung von $(0, 1)$. In diesem Fall reichen endlich viele der Überdeckungsintervalle $\left[\frac{1}{2n}, \frac{1}{n} \right]$ **nicht** aus, um das ganze Intervall $(0, 1)$ zu überdecken.

Es gilt aber der folgende

Satz 1.18 (Überdeckungssatz von Heine-Borel). Sei $A \subseteq \mathbb{R}^n$ beschränkt und abgeschlossen. Ist

$$\mathcal{U} = \{U_i : i \in I\}$$

eine offene Überdeckung von A , so wird A bereits durch endlich viele der U_i überdeckt.

Beweis. Im Fall $A = \emptyset$ ist die Behauptung trivial. Sei also $A \neq \emptyset$. Ist A beschränkt, so ist A in einem abgeschlossenen Würfel $W(l)$ mit Kantenlänge l enthalten. Wir nehmen an, A könne nicht durch endlich viele der U_i überdeckt werden.

Wir halbieren zunächst die Kantenlänge von $W(l)$. Auf diese Weise wird $W(l)$ in 2^n viele abgeschlossene Teilwürfel $W_j^{(1)} \left(\frac{l}{2} \right)$ zerlegt. Dies ergibt eine

Zerlegung von A in 2^n Teilmengen $A \cap W_j^{(1)}\left(\frac{l}{2}\right)$, von denen möglicherweise einige leer sind. Mindestens eine dieser Teilmengen lässt sich nicht durch endlich viele der U_i überdecken. Wir nehmen an, $A \cap W_1^{(1)}\left(\frac{l}{2}\right)$ könne nicht durch endlich viele der U_i überdeckt werden.

Wir setzen dieses Verfahren fort, indem wir den Würfel $W_1^{(1)}\left(\frac{l}{2}\right)$ erneut in 2^n Teilwürfel $W_j^{(2)}\left(\frac{l}{4}\right)$ zerlegen. Dann kann wieder mindestens eine der Mengen $A \cap W_j^{(2)}\left(\frac{l}{4}\right)$, etwa $A \cap W_1^{(2)}\left(\frac{l}{4}\right)$, nicht durch endlich viele der U_i überdeckt werden. Fortgesetzte Halbierung ergibt dann eine Folge ineinandergeschachtelter Würfel

$$W^{(1)}\left(\frac{l}{2}\right) \supseteq W^{(2)}\left(\frac{l}{4}\right) \supseteq W^{(3)}\left(\frac{l}{8}\right) \supseteq \dots \supseteq W^{(n)}\left(\frac{l}{2^n}\right) \supseteq \dots,$$

so dass $A \cap W^{(n)}\left(\frac{l}{2^n}\right)$ nicht durch endlich viele der U_i überdeckt werden kann.

Diese „Würfelschachtelung“ (die in ihrem Wesen genau einer Intervallschachtelung entspricht) hat genau einen Punkt x gemeinsam. Wählt man $x_m \in A \cap W^{(m)}\left(\frac{l}{2^m}\right)$, so folgt $\lim_{m \rightarrow \infty} x_m = x$. A ist nach Voraussetzung abgeschlossen, und daher folgt $x \in A$.

Aus diesem Grund gibt es ein j mit $x \in U_j$, und weil U_j offen ist, existiert ein $\varepsilon > 0$ mit $K(x, \varepsilon) \subseteq U_j$. Somit gibt es ein $N > 0$ mit $U_j \supseteq W^{(m)}\left(\frac{l}{2^m}\right)$ für alle $m \geq N$, und wir haben einen Widerspruch. Damit ist Satz (1.18) bewiesen. \square

Definition 1.19. Eine Teilmenge $A \subseteq \mathbb{R}^n$ heißt **kompakt**, wenn es zu **jeder** offenen Überdeckung

$$\mathcal{U} = \{U_i : i \in I\}$$

endlich viele Indizes $i_1, \dots, i_k \in I$ gibt, so dass gilt

$$A \subseteq \bigcup_{\nu=1}^k U_{i_\nu}.$$

Bemerkung. Bei der Verwendung dieser Definition ist Vorsicht geboten. Die Definition sagt nämlich **nicht**, dass A kompakt sei, wenn A eine endliche offene Überdeckung besitzt. (Jede Teilmenge A des \mathbb{R}^n kann durch endlich viele offene Mengen überdeckt werden, z.B. durch die eine offene Menge \mathbb{R}^n .) Vielmehr wird verlangt, dass jede **beliebig vorgegebene** offene Überdeckung von A eine endliche Teilüberdeckung besitzt. Der Satz (1.18) besagt daher, dass jede beschränkte und abgeschlossene Teilmenge $A \subseteq \mathbb{R}^n$ kompakt ist.

Es gilt sogar der

Satz 1.20. Eine Teilmenge $A \subseteq \mathbb{R}^n$ ist genau dann kompakt, wenn A beschränkt und abgeschlossen ist.

Beweis. Es ist nur noch nötig zu zeigen, dass alle unbeschränkten und alle nicht abgeschlossenen Mengen $A \subseteq \mathbb{R}^n$ nicht kompakt sind.

Ist A nicht beschränkt, so bilden die Kugeln

$$K(0, m), \quad m \in \mathbb{N}$$

eine Überdeckung von A durch offene Mengen, von denen aber offenbar A nicht durch endlich viele überdeckt werden kann.

Ist A nicht abgeschlossen, so wählen wir einen Randpunkt

$$a \in \partial A, \quad a \notin A$$

(wobei ∂A den Rand von A bezeichnet). Die Mengen

$$\mathbb{R}^n \setminus \overline{K}\left(a, \frac{1}{m}\right), \quad m \in \mathbb{N}$$

bilden eine offene Überdeckung von A . Es ist aber nicht möglich, A durch endlich viele dieser Mengen zu überdecken, weil jede Kugel $\overline{K}\left(a, \frac{1}{m}\right)$ Punkte von A enthält. Dies beendet den Beweis von Satz (1.20). \square

Bemerkung. Die meisten Überlegungen in diesem Abschnitt hätte man, statt im \mathbb{R}^n , auch allgemeiner in normierten Vektorräumen durchführen können (auch in metrischen Räumen). Man vergleiche hierzu etwa [7]. Hier folgt daher nun ein kurzer Einschub über normierte Vektorräume und metrische Räume.

Sei V ein \mathbb{R} - oder \mathbb{C} -Vektorraum. Eine **Norm** auf V ist eine Abbildung $\|\cdot\| : V \rightarrow \mathbb{R}$ mit folgenden Eigenschaften:

- (i) $\|x\| \geq 0$ für alle $x \in V$, und aus $\|x\| = 0$ folgt $x = 0$.
- (ii) $\|c \cdot x\| = |c| \cdot \|x\|$ für alle $x \in V, c \in \mathbb{R}$ oder \mathbb{C} .
- (iii) $\|x + y\| \leq \|x\| + \|y\|$ für $x, y \in V$.

Man nennt dann das Paar $(V, \|\cdot\|)$ einen **normierten** Vektorraum.

Eine Folge $(x^{(m)}) \in V$ heißt **Cauchy-Folge**, wenn mit der Norm $\|\cdot\|$ die Eigenschaft aus dem Cauchyschen Konvergenzkriterium (1.12) erfüllt ist, und **konvergent mit dem Grenzwert** $y \in V$, wenn bezüglich $\|\cdot\|$ die Eigenschaft aus (1.3) gilt.

Jede konvergente Folge in einem normierten Vektorraum ist Cauchy-Folge (der Beweis verläuft wörtlich wie im Fall $V = \mathbb{R}$). Gilt die Umkehrung, so heißt der normierte Vektorraum **vollständig** (oder **Banachraum**). Beispiele vollständiger Vektorräume sind \mathbb{R}^n mit der Norm $\|\cdot\|$, die durch den Betrag aus (1.1) gegeben ist und auch die **euklidische Norm** heißt.

Jeder normierte Vektorraum $(V, \|\cdot\|)$ ist bezüglich der Abbildung

$$d: \begin{cases} V \times V \rightarrow \mathbb{R} \\ (x, y) \mapsto \|x - y\| \end{cases}$$

ein metrischer Raum, d.h. die Abbildung d ist eine **Metrik** auf V . Dieser Begriff ist dadurch definiert, dass gilt

- (i) $d(x, y) \geq 0$ für alle $x, y \in V$, und $d(x, y) = 0$ genau dann, wenn $x = y$ ist.
- (ii) $d(x, y) = d(y, x)$ für alle $x, y \in V$.
- (iii) $d(x, z) \leq d(x, y) + d(y, z)$ für alle $x, y, z \in V$.

In jedem metrischen Raum sind die Begriffe „Konvergenz“, „Cauchy-Folge“, „Vollständigkeit“, „offene Menge“ und „abgeschlossene Menge“ in analoger Weise erklärt wie in normierten Vektorräumen (wobei man $\|x - y\|$ durch $d(x, y)$ ersetzt). Die „kompakten“ Teilmengen eines metrischen Raums sind dann wieder die Mengen, für die (1.19) gilt.

Nun weiter im Spezialfall der normierten Vektorräume $V = \mathbb{R}^n$. Statt der euklidischen Norm

$$|x| = \sqrt{\sum_{i=1}^n x_i^2}$$

hätte man auch andere Normen auf dem \mathbb{R}^n verwenden können - etwa die Maximumsnorm

$$\|x\| := \max_{i=1, \dots, n} |x_i|$$

oder die Norm

$$\|x\| := \sum_{i=1}^n |x_i|.$$

Die Sätze dieses Paragraphen, wie etwa der Überdeckungssatz (1.18) oder das Cauchysche Konvergenzkriterium (1.12), gelten unverändert und werden in diesen Fällen analog bewiesen, was der Leser aus didaktischen Gründen noch einmal überprüfen sollte.

Bemerkung 1.21. Ist $p \geq 1$ eine reelle Zahl, so ist durch

$$\|x\|_p := \left(\sum_{\nu=1}^n |x_\nu|^p \right)^{1/p}$$

eine Norm auf dem \mathbb{R}^n gegeben (genannt die p -Norm). Die Dreiecksungleichung (iii) ist in diesem Fall die so genannte **Minkowski-Ungleichung**.

1.2 Funktionen und Abbildungen

Unter einer **Funktion von n Veränderlichen** versteht man eine Abbildung einer Teilmenge $D \subseteq \mathbb{R}^n$ nach \mathbb{R} bzw. \mathbb{C} . Ist der Bildbereich eine Teilmenge des \mathbb{R}^m , so sprechen wir von Abbildungen von D in den Raum \mathbb{R}^m . Die Schreibweise hierfür ist

$$f : D \rightarrow \mathbb{R} \quad \text{bzw.} \quad D \rightarrow \mathbb{C}$$

bzw.

$$f : D \rightarrow \mathbb{R}^m.$$

Ist $f : D \rightarrow \mathbb{R}$, so ist der **Graph** von f die Menge

$$\Gamma_f := \{(x, y) \in D \times \mathbb{R} : y = f(x)\} \subseteq \mathbb{R}^{n+1}.$$

Für $n = 2$ kann man sich den Graphen Γ_f von f als Fläche im dreidimensionalen Raum vorstellen.

Beispiel 1.22. Sei

$$D = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}$$

und sei

$$z := f(x, y) = \sqrt{1 - x^2 - y^2}.$$

Der Graph Γ_f ist dann die obere Hälfte der Einheitskugelfläche.

Es stellt sich die Frage, wie man eine solche Fläche in einem zweidimensionalen Bild veranschaulicht. Man kann dies wie folgt tun: Für konstante Werte von z zeichnet man im Definitionsbereich D die Linien $f(x, y) = z$ ein. Dies ergibt die sogenannten **Niveaulinien** oder **Höhenlinien**. In unserem Beispiel sind die Niveaulinien Kreise um den Ursprung der x - y -Ebene.

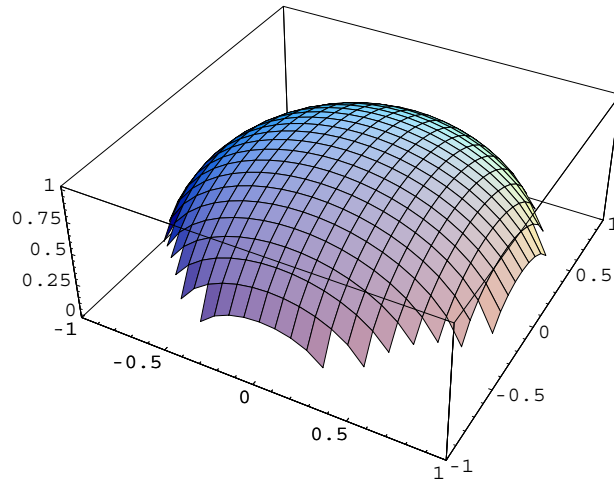


Abbildung 1.1: Der dreidimensionale Graph der Funktion $f(x, y) = \sqrt{x^2 + y^2}$ auf dem Definitionsbereich $D = \{(x, y) : x^2 + y^2 \leq 1\}$.

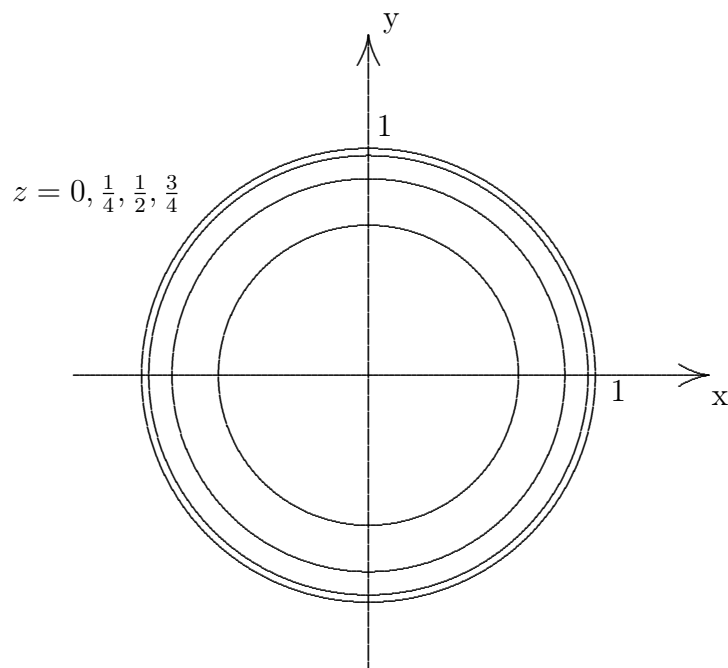


Abbildung 1.2: Nochmals die Funktion $f(x, y) = \sqrt{x^2 + y^2}$, hier dargestellt durch Niveaulinien.

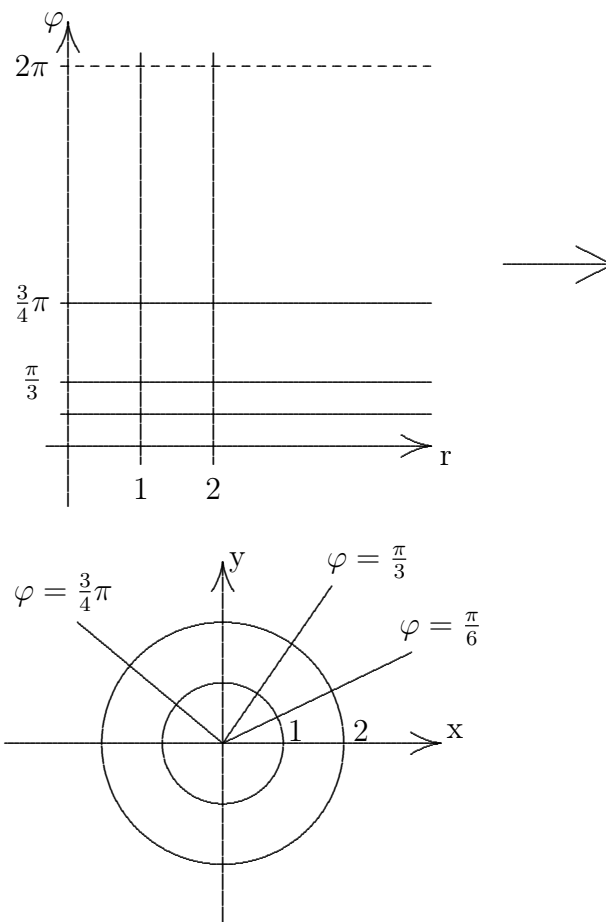


Abbildung 1.3: Der Übergang zwischen kartesischen Koordinaten und (zweidimensionalen) Polarkoordinaten.

Eine Abbildung $f : D \rightarrow \mathbb{R}^m$ wird durch m Koordinatenfunktionen f_1, \dots, f_m gegeben, wobei jedes f_m eine Abbildung von D nach \mathbb{R} ist. Im Fall $n = 2$, $m = 2$ hat man Abbildungen eines ebenen Bereiches in einen ebenen Bereich. Um solche Abbildungen besser zu verstehen, kann man ermitteln, in welche Kurven die Koordinatenlinien $x_i = \text{const}$ ($i = 1, 2$) übergehen.

Beispiel 1.23.

1. Wir betrachten zuerst die **zweidimensionalen Polarkoordinaten** $x = r \cdot \cos \varphi$, $y = r \cdot \sin \varphi$. Genauer: wir betrachten, mit dem Definitionsbereich

$$D = \{(r, \varphi) : 0 \leq r, 0 \leq \varphi < 2\pi\},$$

die Abbildung

$$\begin{cases} D \rightarrow \mathbb{R}^2 \\ (r, \varphi) \mapsto (x, y). \end{cases}$$

Für konstante Werte von r erhält man Kreise um den Nullpunkt, für konstante Werte von φ erhält man Halbgeraden, deren Startpunkt der (zweidimensionale) Koordinatenursprung ist.

2. Wir gehen nun zu den **dreidimensionalen Polarkoordinaten** über. Bei diesen ist

$$x = r \cos \varphi \sin \vartheta, \quad y = r \sin \varphi \sin \vartheta, \quad z = r \cos \vartheta$$

mit $r > 0$. Wir haben dann den Definitionsbereich

$$D = \{(\varphi, \vartheta) : 0 \leq \varphi < 2\pi, 0 \leq \vartheta \leq \pi\},$$

und unsere Abbildung f mit $f(\varphi, \vartheta) := (x, y, z)$ mit Definitionsbereich D hat als Bildbereich die Oberfläche der Kugel $\overline{K}(0, r) \subseteq \mathbb{R}^3$.

Lässt man nun ϑ konstant sein, so ergeben sich die Breitenkreise, und lässt man φ konstant, so ergeben sich die Meridiane.

Wann erhält man bei den Beispielen 1. und 2. injektive Abbildungen? Dazu sind in 1. und 2. die Definitionsbereiche noch etwas abzuändern:

$$D_1 := \{(r, \varphi) : 0 < r, 0 \leq \varphi < 2\pi\} \cup \{(0, 0)\}$$

bzw.

$$D_2 := \{(\varphi, \vartheta) : 0 \leq \varphi < 2\pi, 0 < \vartheta < \pi\} \cup \{(0, 0), (0, \pi)\}.$$

Dann können

$$(r, \varphi) \mapsto (x, y) \quad \text{bzw.}$$

$$(\varphi, \vartheta) \mapsto (x, y, z)$$

als Koordinatentransformationen aufgefasst werden, die eine Umkehrabbildung

$$(x, y) \mapsto (r, \varphi) \quad \text{bzw.}$$

$$(x, y, z) \mapsto (\varphi, \vartheta)$$

besitzen mit (x, y, z) als Punkt der Oberfläche der Kugel $K(0, r) \subseteq \mathbb{R}^3$ (bzw. $\subseteq \mathbb{R}^2$). Streng genommen redet man erst dann von Polarkoordinaten (r, φ) in der Ebene bzw. Polarkoordinaten (φ, ϑ) auf der Kugel.

Analog dem eindimensionalen Fall definiert man:

Definition 1.24. Sei $D \subseteq \mathbb{R}^m$ und sei $f : D \rightarrow \mathbb{R}^n$ eine Abbildung. Ist a ein Berührungspunkt von D , so sagen wir: Der Grenzwert

$$b = \lim_{x \rightarrow a} f(x)$$

existiert, falls es zu jedem $\varepsilon > 0$ ein $\delta > 0$ gibt mit $|f(x) - b| \leq \varepsilon$ für alle $x \in D \cap \overline{K}(a, \delta)$.

Die Schreibweise $\lim_{x \rightarrow a} f(x)$ setzt im Folgenden stillschweigend voraus, dass a ein Berührungspunkt des Definitionsbereiches von f ist.

Bemerkung 1.25. Ist $f(x) = (f_1(x), \dots, f_m(x))$ und $b = (b_1, \dots, b_m)$, so ist

$$\lim_{x \rightarrow a} f(x) = b$$

gleichbedeutend mit

$$\lim_{x \rightarrow a} f_k(x) = b_k \quad \text{für } k = 1, \dots, m.$$

Die aus der Analysis I bekannten Rechenregeln für Grenzwerte übertragen sich daher auf den höherdimensionalen Fall. Man beachte aber: Es genügt für den Grenzübergang $x \rightarrow a$ **nicht**, nacheinander die Komponenten x_i gegen a_i streben zu lassen. Dies zeigt das nächste Beispiel

Beispiel 1.26.

$$D = \mathbb{R}^2 \setminus \{(0, 0)\}, \quad f(x, y) = \frac{x^2}{x^2 + y^2}.$$

Es ist einerseits

$$\lim_{x \rightarrow 0} \left(\lim_{y \rightarrow 0} \frac{x^2}{x^2 + y^2} \right) = \lim_{x \rightarrow 0} 1 = 1,$$

aber für $y \neq 0$ gilt $\lim_{x \rightarrow 0} \frac{x^2}{x^2 + y^2} = 0$ und daher andererseits

$$\lim_{y \rightarrow 0} \lim_{x \rightarrow 0} \frac{x^2}{x^2 + y^2} = 0.$$

In diesem Beispiel ist $f(x, y)$ konstant auf jeder Geraden durch den Nullpunkt. So etwa ist $f(x, y) = \frac{1}{10}$ für $y = 3x$, oder $f(x, y) = \frac{9}{10}$ für $y = \frac{x}{3}$.

Zu jedem $\delta > 0$ existieren daher Punkte $x = (x_1, y_1)$ und $y = (x_2, y_2)$ aus $D \cap \overline{K}(0, \delta)$ mit $|f(x_1, y_1) - f(x_2, y_2)| > \frac{1}{2}$. Der Grenzwert

$$\lim_{(x,y) \rightarrow (0,0)} f(x, y)$$

existiert also nicht.

Wenn dagegen

$$\lim_{(x,y) \rightarrow (a,b)} f(x, y)$$

existiert, dann stimmen die iterierten Grenzwerte

$$\lim_{x \rightarrow a} \lim_{y \rightarrow b} f(x, y), \quad \lim_{y \rightarrow b} \lim_{x \rightarrow a} f(x, y)$$

überein (und existieren). Genauer gilt:

Satz 1.27. *Ist*

$$\lim_{(x,y) \rightarrow (a,b)} f(x, y) = c$$

und existiert $\lim_{y \rightarrow b} f(x, y)$, *so existiert auch*

$$\lim_{x \rightarrow a} (\lim_{y \rightarrow b} f(x, y))$$

und ist gleich c .

Beweis. Nach Voraussetzung existiert zu jedem $\varepsilon > 0$ ein $\delta > 0$ mit $|f(x, y) - c| \leq \varepsilon$ für $(x, y) \in \overline{K}((a, b), \delta)$. Wenn $\lim_{y \rightarrow b} f(x, y)$ existiert, gilt daher auch

$$|\lim_{y \rightarrow b} f(x, y) - c| \leq \varepsilon$$

für alle x mit $|x - a| \leq \delta$. □

Definition 1.28. *Eine Funktion* $f : D \rightarrow \mathbb{R}^m$ *heißt* **stetig** *im Punkt* $a \in D$, *wenn*

$$\lim_{x \rightarrow a} f(x) = f(a)$$

gilt, d.h. wenn es zu jedem $\varepsilon > 0$ *ein* $\delta > 0$ *gibt mit*

$$|f(x) - f(a)| \leq \varepsilon \quad \text{für alle } x \in D \cap \overline{K}(a, \delta).$$

f heißt **stetig**, *wenn f in jedem Punkt von* D *stetig ist.*

Bemerkung 1.29. Wird f durch Koordinatenfunktionen f_1, \dots, f_m gegeben, so ist f genau dann stetig, wenn alle f_i für $i = 1, \dots, m$ stetig sind.

Wie im Fall von Funktionen einer Veränderlicher gilt:

Korollar 1.30. Sind die Funktionen f und g an der Stelle a stetig, so sind auch die Funktionen $f \pm g$, $f \cdot g$ an der Stelle a stetig. Ebenso ist dann auch $\frac{f}{g}$ in a stetig, sofern $g(a) \neq 0$ gilt.

Insbesondere ist also ein Polynom

$$\sum_{(\nu_1, \dots, \nu_n)} a_{\nu_1 \dots \nu_n} x_1^{\nu_1} \dots x_n^{\nu_n}$$

überall stetig, eine rationale Funktion (also ein Quotient zweier Polynome) ist überall dort stetig, wo der Nenner nicht verschwindet.

Im Hinblick auf die Komposition von Abbildungen gilt:

Satz 1.31. Sei $D \subseteq \mathbb{R}^n$ und seien

$$f : D \rightarrow \mathbb{R}^m, \quad g : E \rightarrow \mathbb{R}^l$$

Abbildungen mit $f(D) \subseteq E \subseteq \mathbb{R}^m$ und sei

$$h = g \circ f : D \rightarrow \mathbb{R}^l$$

die Komposition von f und g . Ist

$$\lim_{x \rightarrow a} f(x) = b \quad \text{und} \quad \lim_{y \rightarrow b} g(y) = c,$$

so gilt

$$\lim_{x \rightarrow a} h(x) = c.$$

Insbesondere gilt: Ist f stetig an der Stelle a und g stetig an der Stelle $f(a)$, so ist $h = g \circ f$ stetig im Punkt a .

Beweis. Zu jedem $\varepsilon > 0$ existiert ein $\delta > 0$ mit $|g(y) - c| \leq \varepsilon$ für alle $y \in E \cap \overline{K}(b, \delta)$. Ebenso gibt es zu jedem $\delta > 0$ ein $\eta > 0$ mit $|f(x) - b| \leq \delta$ für alle $x \in D \cap \overline{K}(a, \eta)$.

Für diese x gilt $|h(x) - c| = |g(f(x)) - c| \leq \varepsilon$. Daher ist $\lim_{x \rightarrow a} h(x) = c$, und damit ist Satz (1.31) bewiesen. \square

Offene oder abgeschlossene Mengen werden häufig durch Ungleichungen analytisch beschrieben. Hier gilt folgender

Satz 1.32. Sei $D \subseteq \mathbb{R}^n$ und sei $f : D \rightarrow \mathbb{R}^m$ eine stetige Abbildung. Ist $A \subseteq \mathbb{R}^m$ eine offene (bzw. abgeschlossene) Menge, so ist

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

ein Durchschnitt von D mit einer offenen (bzw. abgeschlossenen) Teilmenge von \mathbb{R}^n . Ist D offen (bzw. abgeschlossen), z.B. $D = \mathbb{R}^n$, so ist $f^{-1}(A)$ selbst offen (bzw. abgeschlossen). Ist f eine auf ganz \mathbb{R}^n definierte reellwertige Funktion und ist $c \in \mathbb{R}$, so ist die Menge

$$\{x \in \mathbb{R}^n : f(x) < c\}$$

offen, und die Mengen

$$\{x \in \mathbb{R}^n : f(x) \leq c\}, \quad \{x \in \mathbb{R}^n : f(x) = c\}$$

sind abgeschlossen.

Beweis.

- (i) Sei $A \subseteq \mathbb{R}^m$ offen und sei $x \in f^{-1}(A)$. Dann existiert eine Kugel $K(f(x), \varepsilon)$, die ganz zu A gehört. Ist f stetig, so existiert eine Kugel $K(x, \delta)$ mit

$$f(D \cap K(x, \delta)) \subseteq K(f(x), \varepsilon),$$

und dies besagt

$$D \cap K(x, \delta) \subseteq f^{-1}(A).$$

Wir setzen nun

$$B := \bigcup_{x \in D} K(x, \delta)$$

als Vereinigung aller so konstruierten offenen Kugeln für alle $x \in D$. Dann ist B offen, und es gilt $f^{-1}(A) = D \cap B$.

- (ii) Sei nun $A \subseteq \mathbb{R}^m$ abgeschlossen. Dann ist $\mathbb{R}^m \setminus A$ offen, und aus (i) folgt $f^{-1}(\mathbb{R}^m \setminus A) = D \cap B$ mit einer offenen Menge B . Dann ist

$$f^{-1}(A) = D \cap (\mathbb{R}^n \setminus B).$$

Da nun $\mathbb{R}^n \setminus B$ abgeschlossen ist, folgt die Behauptung aus (i). Der Rest folgt durch die spezielle Wahl der Menge A

$$A = \{y \in \mathbb{R} : y < c\} \quad \text{oder} \quad A = \{y \in \mathbb{R} : y \leq c\} \quad \text{oder} \quad A = \{c\}.$$

□

Bemerkung 1.33. Ist A offen (bzw. abgeschlossen) und f stetig, so ist im Allgemeinen $f(A)$ nicht wieder offen (bzw. abgeschlossen).

Beispiel 1.34. Sei $m = 1$, $f(x) = \frac{1}{x^2+1}$. Der Definitionsbereich $D = \mathbb{R}$ von A ist offen und abgeschlossen. Allerdings ist $f(A)$ nicht offen, da $1 \in f(A)$ ist, ohne dass 1 ein innerer Punkt von $f(A)$ wäre (wegen $f(x) \leq 1$ für alle $x \in D$). Auch ist $f(A)$ nicht abgeschlossen, da der Punkt 0 nicht zu $f(A)$ gehört und dennoch Randpunkt von $f(A)$ ist.

Es gilt aber der

Satz 1.35. $D \subseteq \mathbb{R}^n$ sei kompakt und $f : D \rightarrow \mathbb{R}^m$ stetig. Dann ist $f(D)$ kompakt, d.h. beschränkt und abgeschlossen im \mathbb{R}^m .

Beweis. Sei $\mathcal{U} = \{U_i : i \in I\}$ eine offene Überdeckung von $f(D)$. Aus (1.32) folgt dann $f^{-1}(U_i) = D \cap V_i$ mit $V_i \subseteq \mathbb{R}^n$ offen. Dann bildet

$$\mathcal{V} := \{V_i : i \in I\}$$

eine offene Überdeckung von D . Da D kompakt ist, wird D bereits durch endlich viele der V_i überdeckt. Ist nun $x \in D \cap V_i$, so folgt $f(x) \in f(D \cap V_i) \subseteq U_i$. Also wird dann $f(D)$ durch endlich viele U_i überdeckt und ist damit kompakt. Damit ist Satz (1.35) bewiesen. \square

Ist in (1.35) speziell $m = 1$, d.h. ist f reellwertig, so ist damit $f(D)$ beschränkt. Ist nun $D \neq \emptyset$, so hat $f(D)$ eine obere Grenze c . Für $x > c$ gilt daher $x \notin f(D)$. Dennoch ist

$$[c - \varepsilon, c] \cap f(D) \neq \emptyset \quad \text{für alle } \varepsilon > 0,$$

d.h. c ist Randpunkt von $f(D)$. Da $f(D)$ abgeschlossen ist, folgt daraus $c \in f(D)$, und wir haben folgende Aussage:

Korollar 1.36. Jede auf einer nichtleeren kompakten Menge $D \subseteq \mathbb{R}^n$ definierte stetige reellwertige Funktion ist beschränkt und nimmt auf D ihr Maximum und Minimum an.

Auch der Begriff der gleichmäßigen Stetigkeit (der oft bereits in der Analysis I definiert wird) lässt sich auf den n -dimensionalen Fall übertragen:

Definition 1.37. Eine Funktion $f : D \rightarrow \mathbb{R}^m$ mit $D \subseteq \mathbb{R}^n$ heißt **gleichmäßig stetig** auf D , wenn es zu jedem $\varepsilon > 0$ ein $\delta > 0$ gibt mit

$$|f(x) - f(y)| \leq \varepsilon \quad \text{für alle } x, y \in D \text{ mit } |x - y| \leq \delta.$$

Anders als bei dem gewöhnlichen Stetigkeitsbegriff hängt δ nur von ε , nicht aber zusätzlich von x ab. Dies ist eine schärfere Bedingung als die der Stetigkeit von f auf D .

Beispiel 1.38. Die Funktion $f(x) = \frac{1}{x}$ ist stetig auf $(0,1)$, aber nicht gleichmäßig stetig. Hierzu setzen wir $\delta > 0$ beliebig (oBdA sei $\delta < 1$). Mit $x = \delta$ und $y = \frac{\delta}{2}$ gilt $|x - y| \leq \delta$, aber $|f(x) - f(y)| = \frac{1}{\delta} > \varepsilon = 1$.

Es gilt der

Satz 1.39. *Jede auf einer kompakten Menge $D \subseteq \mathbb{R}^n$ definierte stetige Funktion ist gleichmäßig stetig auf D .*

Beweis. Zu $\varepsilon > 0$ und $x \in D$ existiert ein $\delta = \delta(x) > 0$, so dass

$$|f(y) - f(x)| \leq \frac{\varepsilon}{2} \quad \text{falls } |y - x| < \delta, y \in D.$$

Die offenen Kugeln $K(x, \frac{\delta}{2})$ überdecken ganz D . Da D nach Voraussetzung kompakt ist, kann D bereits durch endlich viele Kugeln $K(x_i, \frac{\delta_i}{2})$ überdeckt werden.

Wir setzen nun $\delta := \frac{1}{2} \min_i(\delta_i)$. Sind dann $x, y \in D$ mit $|x - y| < \delta$ und $x \in K(x_i, \frac{\delta_i}{2})$, so folgt

$$|y - x_i| \leq |y - x| + |x - x_i| < \delta + \frac{\delta_i}{2} \leq \delta_i$$

und damit

$$|f(y) - f(x)| \leq |f(y) - f(x_i)| + |f(x_i) - f(x)| \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon,$$

so dass f tatsächlich gleichmäßig stetig ist. □

1.3 Kurven im \mathbb{R}^n

Wir betrachten nun Funktionen $f : D \rightarrow \mathbb{R}^n$ mit $D \subseteq \mathbb{R}^1$.

Definition 1.40. *Unter einer **stetigen Kurve im \mathbb{R}^n** versteht man eine stetige Abbildung $f : D \rightarrow \mathbb{R}^n$, wobei $D \subseteq \mathbb{R}^1$ ein (eigentliches oder uneigentliches) Intervall ist.*

Jede Funktion $f : D \rightarrow \mathbb{R}^n$ ist gegeben durch n Komponentenfunktionen $f_i : D \rightarrow \mathbb{R}$. Im Fall einer stetigen Funktion f sind alle f_i stetig.

Die Kurve f heißt **(stetig) differenzierbar**, wenn alle Funktionen f_i für $i = 1, \dots, n$ (stetig) differenzierbar sind. f heißt **k -mal differenzierbar**, wenn sämtliche Funktionen f_i jeweils k -mal differenzierbar sind.

Beispiel 1.41.

1. Sei $D = \mathbb{R}$ und $f(t) = a$ mit einem festen Punkt $a \in \mathbb{R}^n$. Dann besteht $f(D)$ nur aus einem einzigen Punkt. Manche Autoren lassen solche Fälle nicht zu und definieren den Begriff der „Kurve“ daher anders.
2. Es sei $r > 0$. Ein **Kreis vom Radius** r wird beschrieben durch die Kurve

$$f : \begin{cases} [0, 2\pi] \rightarrow \mathbb{R}^2 \\ t \mapsto (r \cos t, r \sin t) . \end{cases}$$

3. Sei $D = [-1, 1]$ und sei $f(x) = (x, \sqrt{1-x^2})$. Dann ist $f(D)$ die obere Hälfte des im Uhrzeigersinn durchlaufenen Einheitskreises.
4. Sei $a \in \mathbb{R}^n$, $v \in \mathbb{R}^n \setminus \{0\}$. Die Kurve

$$f : \begin{cases} \mathbb{R} \rightarrow \mathbb{R}^n \\ t \mapsto a + vt \end{cases}$$

beschreibt dann eine Gerade im \mathbb{R}^n durch den Punkt a mit Richtungsvektor v .

Ist die Kurve $f : D \rightarrow \mathbb{R}^n$ im Punkt $t_0 \in D$ differenzierbar, so existiert

$$f'(t_0) := \lim_{t \rightarrow t_0} \frac{1}{t - t_0} (f(t) - f(t_0)).$$

Man nennt dann $f'(t_0) = (f'_1(t_0), \dots, f'_n(t_0))$ den **Tangentialvektor** an die Kurve (im Punkt $f(t_0)$). Ist $f'(t_0) \neq 0$, so heißt

$$g(t) := f(t_0 + t \cdot f'(t_0))$$

die zum Parameterwert t_0 gehörige **Tangente** an die Kurve f .

Geometrische Interpretation: Der Tangentialvektor $f'(t_0)$ lässt sich als Limes von Sekanten auffassen. Es ist nämlich

$$f'(t_0) = \lim_{t \rightarrow t_0} \frac{f(t) - f(t_0)}{t - t_0}.$$

Bei einer physikalischen Interpretation entspricht $f'(t_0)$ dem Geschwindigkeitsvektor im Zeitpunkt t_0 der durch $f : D \rightarrow \mathbb{R}^n$ beschriebenen Bewegung eines Punktes im Raum (Grenzwert des Quotienten aus Orts- und Zeitdifferenz).

Ist $f'(t_0) = 0$, so ist die Tangente durch $g(t) = f(t_0)$ gegeben. Die Kurve braucht dann aber in der Nähe von $f(t_0)$ nicht „glatt“ (siehe Definition (1.43)) zu sein.

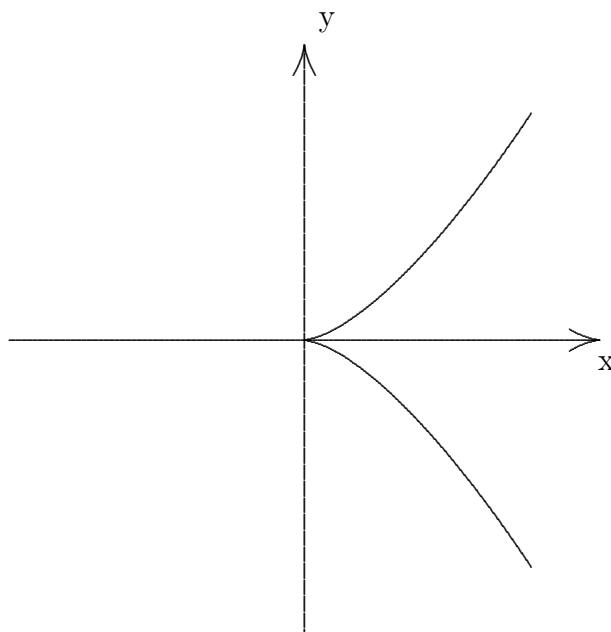


Abbildung 1.4: Die Neilsche Parabel.

Beispiel 1.42 (Neilsche Parabel). Sei

$$D = \mathbb{R}, \quad f(t) = (t^2, t^3).$$

Dann ist $f(\mathbb{R}) = \{(x, y) \in \mathbb{R}^2 : x \geq 0 \text{ und } y = \pm x^{3/2}\}$. Für $t = 0$ ist $f'(t) = (2t, 3t^2) = 0$. Wir haben hier eine Spitze (auch **Singularität** genannt) im Nullpunkt.

Definition 1.43. Die Kurve $f : D \rightarrow \mathbb{R}^n$ heißt **glatt** (oder **regulär**), wenn sie stetig differenzierbar ist und überall $f'(t) \neq 0$ gilt. Ein Parameterwert $t \in D$ mit $f'(t) = 0$ heißt **singulär** (oder **Singularität**).

Wir beschäftigen uns nun mit der Aufgabe, einer Kurve eine **Bogenlänge** zuzuordnen. Die Idee hierbei ist, den Graphen der Kurve durch Polygonzüge zu approximieren.

Hierzu sei $D = [a, b]$ ein abgeschlossenes Intervall und $f : [a, b] \rightarrow \mathbb{R}$ eine Kurve. Wir wählen eine Zerlegung

$$\{t_0, t_1, \dots, t_k\}$$

von $[a, b]$ mit $a = t_0 < t_1 < \dots < t_k = b$. Verbindet man nun für $1 \leq l \leq k$ jeweils $f(t_{l-1})$ geradlinig mit $f(t_l)$, so hat man einen Polygonzug \mathcal{Z} . Die

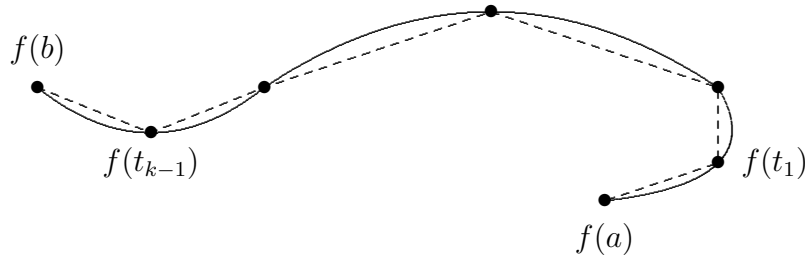


Abbildung 1.5: Polygonzug und stetige Kurve

Länge dieses Polygonzugs ist

$$\mathcal{L}_f(\mathcal{Z}) := \mathcal{L}_f(t_0, \dots, t_k) := \sum_{i=1}^k |f(t_i) - f(t_{i-1})|.$$

Die Länge der Kurve wird nun als Grenzwert der Längen der Polygonzüge bei immer feineren Unterteilungen definiert:

Definition 1.44. Eine Kurve $f : [a, b] \rightarrow \mathbb{R}^n$ heißt **rektifizierbar mit der Länge L** , wenn es zu jedem $\varepsilon > 0$ ein $\delta > 0$ gibt, so dass für jede Zerlegung

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

der Feinheit $\leq \delta$ (d.h. $\max_{i=1, \dots, k} (t_i - t_{i-1}) \leq \delta$) gilt:

$$|\mathcal{L}_f(\mathcal{Z}) - L| \leq \varepsilon.$$

Satz 1.45. Jede stetig differenzierbare Kurve $f : [a, b] \rightarrow \mathbb{R}^n$ ist rektifizierbar, und für ihre Länge L gilt

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

Bemerkung. Die stetige Differenzierbarkeit von f ist nicht notwendig für die Rektifizierbarkeit der Kurve f . Andererseits gibt es stetige Kurven, die nicht rektifizierbar sind.

Zum Beweis von Satz (1.45) benötigen wir zunächst den

Hilfssatz 1.46. Sei $f : [a, b] \rightarrow \mathbb{R}^n$ stetig differenzierbar. Dann existiert zu jedem $\varepsilon > 0$ ein $\delta > 0$, so dass

$$\left| \frac{f(t) - f(t^*)}{t - t^*} - f'(t) \right| \leq \varepsilon$$

für alle $t, t^* \in [a, b]$ mit $0 < |t - t^*| \leq \delta$ gilt.

Beweis.

- (i) Wir betrachten zunächst den Fall $n = 1$. Ist dann $f' : [a, b] \rightarrow \mathbb{R}$ stetig, so ist f' nach Satz (1.39) auch gleichmäßig stetig auf $[a, b]$. Zu jedem $\varepsilon > 0$ gibt es also ein $\delta > 0$ mit

$$|f'(t) - f'(t_0)| \leq \varepsilon \quad \text{für alle } s, t \in [a, b] \text{ mit } |s - t| \leq \delta.$$

Sind $t, t^* \in [a, b]$ mit $0 < |t - t^*| \leq \delta$, so existiert nach dem Mittelwertsatz der Differentialrechnung ein s zwischen t und t^* mit

$$\frac{f(t) - f(t^*)}{t - t^*} = f'(s).$$

Es ist dann

$$\left| \frac{f(t)}{f(t^*)} t - t^* - f'(t) \right| = |f'(s) - f'(t)| \leq \varepsilon,$$

und dies beweist (i).

- (ii) Ist $n \geq 1$ beliebig und $f = (f_1, \dots, f_n)$, so folgt

$$\left| \frac{f(t) - f(t^*)}{t - t^*} - f'(t) \right| \leq \sqrt{n} \cdot \max_{i=1, \dots, n} \left| \frac{f_i(t) - f_i(t^*)}{t - t^*} - f'_i(t) \right|.$$

Daraus folgt mit (i) die Behauptung des Hilfssatzes.

□

Wir wollen nun den Satz (1.45) beweisen.

Beweis. Gegeben sei ein $\varepsilon > 0$. Für genügend kleine $\delta_1 > 0$ und jede Zerlegung $a = t_0 < t_1 < \dots < t_k = b$ von $[a, b]$ mit Feinheit $\leq \delta_1$ gilt

$$\left| \int_a^b |f'(t)| dt - \sum_{i=1}^k |f'(t_i)| \cdot (t_i - t_{i-1}) \right| \leq \frac{\varepsilon}{2} \quad (1.5)$$

(Approximation von Integralen durch Riemannsche Summen).

Aufgrund von Hilfssatz (1.46) existiert ein $\delta > 0$, $\delta < \delta_1$ mit

$$\left| \frac{f(t_i) - f(t_{i-1})}{t_i - t_{i-1}} - f'(t_i) \right| \leq \frac{\varepsilon}{2(b-a)} \quad \text{für } i = 1, \dots, k.$$

Daher ist

$$|f(t_i) - f(t_{i-1})| - |f'(t_i)|(t_i - t_{i-1})| \leq \frac{t_i - t_{i-1}}{b - a} \cdot \frac{\varepsilon}{2}, \quad (1.6)$$

und wegen (1.5) und (1.6) ist

$$\left| \sum_{i=1}^k |f(t_i) - f(t_{i-1})| - \int_a^b |f'(t)| dt \right| \leq \varepsilon.$$

Dies beendet den Beweis von Satz (1.45). □

Beispiel 1.47.

1. Sei $s > 0$ und

$$f : \begin{cases} [0, s] \rightarrow \mathbb{R}^2 \\ t \mapsto f(t) = (\cos t, \sin t). \end{cases}$$

Dann ist $f'(t) = (-\sin t, \cos t)$ und $|f'(t)| = \sqrt{\sin^2 t + \cos^2 t} = 1$ für alle t . Für die Bogenlänge L gilt daher

$$L = \int_0^s \sqrt{\sin^2 t + \cos^2 t} dt = \int_0^s dt = s.$$

Insbesondere hat man für $s = 2\pi$ den Umfang des Einheitskreises.

2. Eine ebene Kurve der Form $y = f(x)$ ($a \leq x \leq b$) wird in der Parameterdarstellung

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

gegeben. Als Bogenlänge dieser ebenen Kurve ergibt sich

$$\int_a^b \sqrt{1 + (f'(t))^2} dt.$$

Manchmal ist es schwierig, solche Integrale explizit zu berechnen. Ist beispielsweise die Länge eines Bogens der Ellipse

$$y = \sqrt{1 - cx^2} \quad (c \neq 1, c > 0)$$

zu berechnen, so ergibt sich folgendes Integral:

$$\int \sqrt{1 + \left(\frac{-cx}{\sqrt{1 - cx^2}} \right)^2} dx =$$

$$\begin{aligned}
&= \int \sqrt{\frac{1 + c(c-1)x^2}{1 - cx^2}} dx = \\
&= \int \frac{1 + c(c-1)x^2}{\sqrt{(1 - cx^2)(1 + c(c-1)x^2)}} dx.
\end{aligned}$$

Integrale, bei denen der Integrand sich rational aus x und der Quadratwurzel aus einem Polynom dritten und vierten Grades zusammensetzt, nennt man wegen des hier dargestellten Zusammenhangs auch **elliptische Integrale**. Sie lassen sich, wenn das Polynom lauter verschiedene Nullstellen hat, im Allgemeinen **nicht** geschlossen auswerten, sondern führen auf neue transzendente Funktionen, die **elliptischen Funktionen**.

3. Ist eine ebene Kurve durch Angabe der Polarkoordinaten r, φ als Funktion von $r(t), \varphi(t)$ des Parameters t gegeben, so stellt sich f in kartesischen Koordinaten wie folgt dar:

$$f(t) = (r(t) \cos \varphi(t), r(t) \sin \varphi(t)).$$

Für die Bogenlänge gilt daher

$$\begin{aligned}
\int_a^b |f'(t)| dt &= \int_a^b \sqrt{(r' \cos \varphi - r \cdot \varphi' \sin \varphi)^2 + (r' \sin \varphi + r \cdot \varphi' \cdot \cos \varphi)^2} dt \\
&= \int_a^b \sqrt{r'(t)^2 + r^2(t) \varphi'(t)^2} dt.
\end{aligned}$$

Etwas formal beschreibt man dies in manchen Büchern als die Gleichung für das **Differential der Bogenlänge**

$$ds = \sqrt{dr^2 + r^2(d\varphi)^2},$$

wobei wir den Symbolen $ds, dr, d\varphi$ noch keine konkrete Bedeutung gegeben haben.

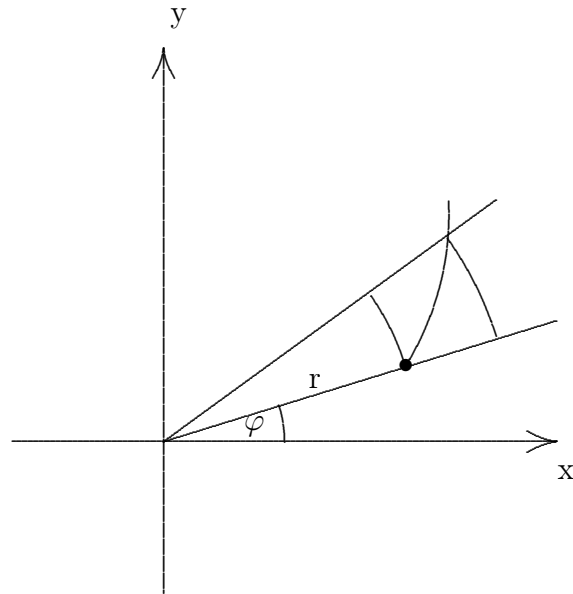


Abbildung 1.6: Zum Differential der Bogenlänge.

1.4 Differenzierbarkeit

Wir betrachten in diesem Abschnitt, umgekehrt zum Kapitel (1.3), Funktionen $f : D \rightarrow \mathbb{R}$ mit $D \subseteq \mathbb{R}^n$. Es ist dann also f eine Funktion von n Veränderlichen x_1, \dots, x_n .

Definition 1.48. Ist $y \in D$ ein innerer Punkt von D , so heißt f an der Stelle $y \in D$ **partiell nach x_k differenzierbar** (oder **partiell differenzierbar bezüglich der k -ten Koordinatenrichtung**), falls der Limes

$$\lim_{h \rightarrow 0} \frac{f(y + h \cdot e_k) - f(y)}{h} =: \frac{\partial f}{\partial x_k}(y)$$

existiert. Dabei ist $e_k \in \mathbb{R}^n$ der k -te Einheitsvektor:

$$e_k = (0, \dots, 0, 1, 0, \dots, 0)$$

mit der 1 an der k -ten Stelle. Für den Limes hat man sich auf diejenigen $h \in \mathbb{R}$ zu beschränken, für die $h \neq 0$ und $y + he_k \in D$ gilt. Man nennt $\frac{\partial f}{\partial x_k}(y)$ die **k -te partielle Ableitung** von f in y .

Es gibt für die partielle Ableitung auch die Schreibweisen

$$\frac{\partial f}{\partial x_k}(y) = f_{x_k}(y) = D_{x_k} f(y).$$

Man nennt f an der Stelle $y \in D$ **partiell differenzierbar**, falls jede der partiellen Ableitungen $\frac{\partial f}{\partial x_k}(y)$ für $k = 1, \dots, n$ existiert.

In gleicher Weise werden partielle Ableitungen für Abbildungen $f : D \rightarrow \mathbb{R}^n$ definiert. Ist $f = (f_1, \dots, f_n)$, so setzt man

$$\frac{\partial f}{\partial x_k}(x) := \left(\frac{\partial f_1}{\partial x_k}(x), \dots, \frac{\partial f_n}{\partial x_k}(x) \right).$$

Partielle Differenzierbarkeit ist für Funktionen (bzw. Abbildungen) von mehreren Veränderlichen eine recht schwache Eigenschaft - aus ihr folgt beispielsweise **nicht** die Stetigkeit der Funktion an der gegebenen Stelle.

Beispiel 1.49. Sei

$$f(x, y) = \begin{cases} \frac{xy}{x^2+y^2} & \text{für } (x, y) \neq (0, 0) \\ 0 & \text{für } x = y = 0. \end{cases}$$

Dann ist f überall, also auch im Nullpunkt, partiell differenzierbar. Aber f ist im Nullpunkt nicht stetig, denn es gilt

$$\lim_{t \rightarrow 0^+} f(t, t) = \frac{1}{2} \neq f(0, 0).$$

Dies ist allerdings nicht sehr verwunderlich: Partielle Differenzierbarkeit nach der Variablen x_k ist keine Aussage über das Verhalten von f in einer vollen Umgebung von y , sondern nur auf der Geraden $x_i = y_i$ mit $i \neq k$.

Bei Funktionen $f(x)$ einer Veränderlicher bedeutet die Differenzierbarkeit an einer Stelle x_0 , dass die durch $y = f(x)$ in der (x, y) -Ebene dargestellte Kurve im Punkt $(x_0, f(x_0))$ eine Tangente besitzt, dass sich also $f(x)$ bei x_0 gut linear approximieren lässt.

Analog geht man bei Funktionen mehrerer Veränderlicher vor:

Definition 1.50. Die Funktion f mit dem Definitionsbereich $D \subseteq \mathbb{R}^n$ heißt **total differenzierbar** (oder kurz **differenzierbar**) in dem inneren Punkt $x \in D$, wenn es eine lineare Funktion $L : \mathbb{R}^n \rightarrow \mathbb{R}$ gibt, etwa

$$L(z) = \sum_{i=1}^n a_i z_i \quad (z = (z_1, \dots, z_n) \in \mathbb{R}^n),$$

so dass

$$f(x) - f(y) = L(x - y) + R(x)$$

mit $\lim_{x \rightarrow y} \frac{R(x)}{|x-y|} = 0$ gilt.

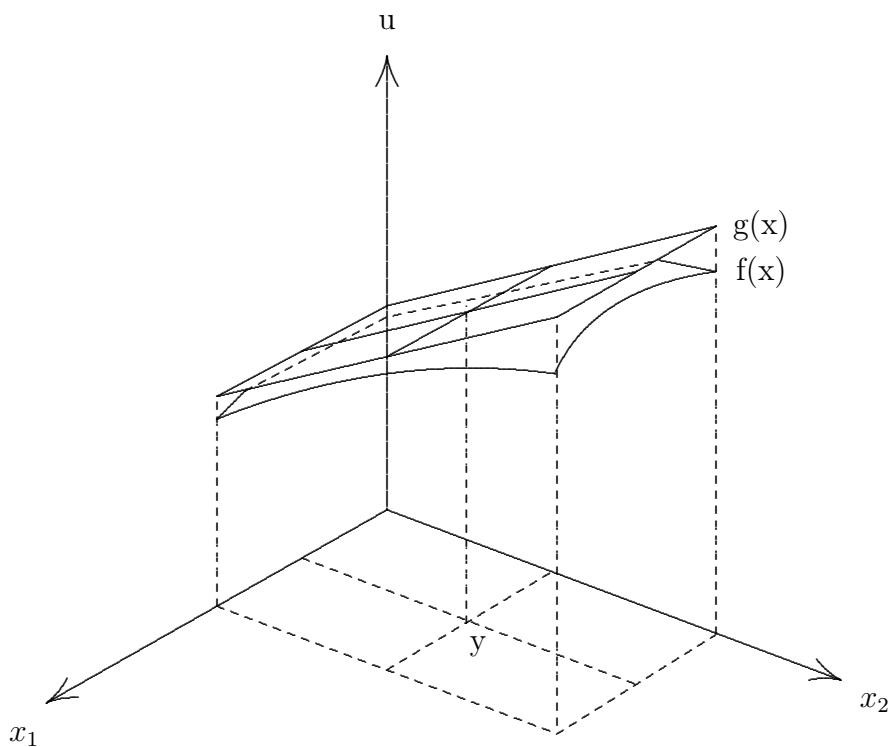


Abbildung 1.7: $f(x)$ ist Funktion von x_1 und x_2 , die zugehörige Tangentialebene im Punkt $(y, f(y))$ ist durch $g(x)$ gegeben.

Die geometrische Deutung ist folgende: Die Gleichung

$$u = f(x)$$

stellt im Raum mit den Koordinaten x_1, \dots, x_n, u eine Fläche dar (auch **Hyperfläche** genannt). Ist f an der Stelle y total differenzierbar, so berührt die durch

$$u = g(x) = f(y) + L(x - y)$$

gegebene Hyperebene die Fläche im Punkt

$$p = (y_1, \dots, y_n, f(y))$$

und heißt deshalb die **Tangentialebene** (oder **Tangentialhyperebene**) der Fläche im Punkt p .

Satz 1.51. *Ist f an der Stelle y total differenzierbar, dann ist f an dieser Stelle stetig und auch partiell differenzierbar.*

Beweis. Wir führen den Grenzübergang $x \rightarrow y$ durch. Das aus (1.50) bekannte Restglied $R(x)$ strebt dann gegen Null, ebenso gilt $L(x - y) \rightarrow 0$. Offenbar strebt dann $f(x)$ gegen $f(y)$, d.h. f ist an der Stelle y stetig.

Speziell gilt, mit $x = (y_1, \dots, y_{i-1}, x_i, y_{i+1}, \dots, y_n)$: Strebt x gegen y , so strebt x_i für $1 \leq i \leq n$ gegen y_i . Damit folgt

$$\begin{aligned} \frac{f(y_1, \dots, y_{i-1}, x_i, y_{i+1}, \dots, y_n) - f(y_1, \dots, y_n)}{x_i - y_i} &= \\ &= \frac{a_i(x_i - y_i) + R(x)}{x_i - y_i} = a_i \pm \frac{R(x)}{|x - y|}. \end{aligned}$$

Für $x_i \rightarrow y_i$ folgt

$$\frac{\partial f}{\partial x_i}(y) = a_i.$$

Die Koeffizienten der Linearform L sind also gerade die partiellen Ableitungen $\frac{\partial f}{\partial x_i}(y)$ von f . Für eine an der Stelle y total differenzierbare Funktion $f(x)$ gilt daher

$$f(x) - f(y) = L(x - y) + R(x) \quad (1.7)$$

mit

$$L(z) = \sum_{i=1}^n \frac{\partial f}{\partial x_i}(y) z_i \quad \text{und} \quad \lim_{x \rightarrow y} \frac{R(x)}{|x - y|} = 0.$$

□

Umgekehrt hat man

Satz 1.52. *Die Funktion f sei in einer Kugel um den Punkt y partiell differenzierbar, und die partiellen Ableitungen $\frac{\partial f}{\partial x_i}$ seien im Punkt y stetig. Dann ist f an der Stelle y total differenzierbar.*

Beweis. Es ist zunächst

$$f(x) - f(y) = \sum_{i=1}^n (f(y_1, \dots, y_{i-1}, x_i, \dots, x_n) - f(y_1, \dots, y_i, x_{i+1}, \dots, x_n)).$$

Wir wenden auf den i -ten Summanden den Mittelwertsatz an. Dies ergibt als i -ten Term

$$\frac{\partial f}{\partial x_i}(y_1, \dots, y_{i-1}, \xi_i, x_{i+1}, \dots, x_n)(x_i, y_i)$$

mit ξ_i zwischen x_i und y_i . Für das Restglied $R(x)$ gilt daher die Abschätzung

$$|R(x)| = \left| f(x) - f(y) - \sum_{i=1}^n \frac{\partial f}{\partial x_i}(y) \cdot (x_i - y_i) \right| \leq$$

$$\leq \sum_{i=1}^n \left| \frac{\partial f}{\partial x_i}(y_1, \dots, y_{i-1}, \xi_i, x_{i+1}, \dots, x_n) - \frac{\partial f}{\partial x_i}(y) \right| \cdot |y_i - x_i|.$$

Aus der Stetigkeit der partiellen Ableitungen in y folgt: Es gibt zu jedem $\varepsilon > 0$ ein $\delta > 0$ mit

$$\left| \frac{\partial f}{\partial x_i}(z) - \frac{\partial f}{\partial x_i}(y) \right| \leq \frac{\varepsilon}{n}$$

für $i = 1, \dots, n$ und alle z mit $|z - y| \leq \delta$. Ist $|x - y| \leq \delta$, so gilt dies insbesondere für $z = (y_1, \dots, y_{i-1}, \xi_i, x_{i+1}, \dots, x_n)$. Die obige Abschätzung für das Restglied liefert dann

$$|R(x)| \leq \frac{\varepsilon}{n} \sum_{i=1}^n |y_i - x_i| \leq \varepsilon \cdot \frac{1}{\sqrt{n}} \cdot |y - x|,$$

und daraus folgt $\lim_{x \rightarrow y} \frac{R(x)}{|y - x|} = 0$. Dies beendet den Beweis von Satz (1.52). \square

Definition 1.53. Sei $D \subseteq \mathbb{R}^n$ offen und $f : D \rightarrow \mathbb{R}$ eine Funktion. f heißt in D **stetig differenzierbar**, wenn f in D überall partiell differenzierbar ist und sämtliche partiellen Ableitungen von f dort stetig sind.

Satz (1.52) besagt nun genau: Ist f stetig differenzierbar auf D , so ist f auch total differenzierbar auf D . Insgesamt haben wir nun: Aus der stetigen Differenzierbarkeit folgt die totale Differenzierbarkeit, aus dieser weiter die Stetigkeit und die partielle Differenzierbarkeit. Die umgekehrten Richtungen sind im Allgemeinen falsch.

Definition 1.54. Sei $D \subseteq \mathbb{R}^n$ und y ein innerer Punkt von D . Eine Abbildung $D \rightarrow \mathbb{R}^m$ heißt **total differenzierbar** (oder auch kurz **differenzierbar**) in y , wenn es eine lineare Abbildung $L : \mathbb{R}^n \rightarrow \mathbb{R}^m$ gibt, so dass gilt

$$f(x) - f(y) = L(x - y) + R(x) \quad \text{und} \quad \lim_{x \rightarrow y} \frac{R(x)}{|y - x|} = 0$$

mit einer Abbildung $R : D \rightarrow \mathbb{R}^m$.

Bemerkung. Totale Differenzierbarkeit von f an der Stelle y bedeutet: In einer kleinen Kugel um y kann man f in erster Näherung durch die lineare Abbildung L ersetzen. Die approximierende lineare Abbildung wird mit

$$Df(y) \quad \text{oder} \quad df(y) \quad \text{oder} \quad f'(y)$$

bezeichnet.

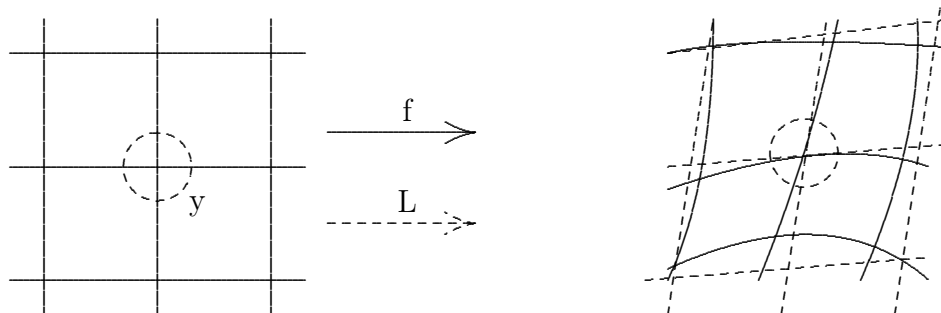


Abbildung 1.8: Ist f in y total differenzierbar, so kann man f in einer kleinen Kugel um den Punkt y in erster Näherung durch die lineare Abbildung L ersetzen.

Sind f_1, \dots, f_m die Koordinatenfunktionen von f , d.h. ist $f(x) = (f_1(x), \dots, f_m(x))$ für $x \in D$, so gilt:

Satz 1.55. Die Abbildung f ist genau dann im Punkt y total differenzierbar, wenn die Funktionen f_1, \dots, f_m im Punkt y total differenzierbar sind.

Beweis. Der Beweis ist analog zum Beweis von Satz (1.10) (Konvergenz von Punktfolgen im \mathbb{R}^n). Wie dort verwendet man die für alle $a \in \mathbb{R}^n$ gültigen Ungleichungen

$$\max_k |a_k| \leq |a| \leq \sqrt{n} \cdot \max_k |a_k|.$$

Insbesondere gelten die Sätze (1.51) und (1.52) analog für Abbildungen. \square

Die lineare Abbildung $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ wollen wir nun explizit beschreiben. Nach Wahl der Standardbasen wird L durch eine $m \times n$ -Matrix mit m Zeilen und n Spalten beschrieben, wenn man die Vektoren des \mathbb{R}^n bzw. \mathbb{R}^m als Spaltenvektoren schreibt. Ist $A^{(m,n)}$ die L zugeordnete Matrix, so gilt:

$$A^{(m,n)} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(y) & \cdots & \frac{\partial f_1}{\partial x_n}(y) \\ \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1}(y) & \cdots & \frac{\partial f_m}{\partial x_n}(y) \end{pmatrix} =: \left(\frac{\partial f_i}{\partial x_j}(y) \right).$$

Denn für die Koordinatenfunktionen f_i ($i = 1, \dots, m$) gilt

$$f_i(x) - f_i(y) = \sum_{j=1}^n \frac{\partial f_i}{\partial x_j}(y)(x_j - y_j) + R_i(y),$$

und aus dieser Gleichung folgt

$$L(z) = \left(\sum_{j=1}^n \frac{\partial f_i}{\partial x_j}(y) z_j \right) =$$

$$= \left(\frac{\partial f_i}{\partial x_j}(y) \right)_{1 \leq i \leq m, 1 \leq j \leq n} \begin{pmatrix} z_1 \\ \vdots \\ z_n \end{pmatrix}.$$

Die Matrix $\left(\frac{\partial f_i}{\partial x_j}(y) \right)$ heißt **Fundamentalmatrix** von f an der Stelle y .

Für die Übertragung der Kettenregel auf Abbildungen benötigen wir noch einige Vorbereitungen:

Hilfssatz 1.56. *Zu einer linearen Abbildung $A : \mathbb{R}^m \rightarrow \mathbb{R}^n$ existiert eine Konstante $C \in \mathbb{R}$, so dass*

$$|Az| \leq C \cdot |z| \quad \text{für alle } z \in \mathbb{R}^m \text{ gilt.}$$

Beweis. Sei A durch die Matrix (a_{ik}) gegeben und bezeichne $(Az)_i$ die i -te Koordinate von Az . Dann gilt

$$|(Az)_i| \leq \sum_{k=1}^m |a_{ik}| |z_k| \leq |z| \cdot \sum_{k=1}^m |a_{ik}|$$

und daher

$$|Az| \leq \sum_{i=1}^n |(Az)_i| \leq C \cdot |z|$$

mit

$$C := \sum_{i=1}^n \sum_{k=1}^m |a_{ik}|.$$

□

Hilfssatz 1.57. *Sei $D \subseteq \mathbb{R}^m$ und sei $f : D \rightarrow \mathbb{R}^n$ im Punkt y total differenzierbar. Dann gibt es Konstanten $\eta > 0$ und C , so dass*

$$|f(x) - f(y)| \leq C|x - y| \quad \text{für alle } x \in K(y, \eta)$$

gilt.

Beweis. Da f total differenzierbar in y ist, schreiben wir

$$f(x) - f(y) = f'(y)(x - y) + R(x).$$

Zu jedem $\varepsilon > 0$, speziell zu $\varepsilon = 1$, existiert ein $\eta > 0$ mit

$$|R(x)| \leq |x - y| \quad \text{für alle } x \text{ mit } |x - y| \leq \eta.$$

Wir verwenden nun (1.56): Zu der linearen Abbildung $f'(y)$ existiert eine Konstante C' mit

$$|f'(y) \cdot (x - y)| \leq C' \cdot |x - y| \quad \text{für beliebige } x.$$

Setzen wir nun $C := C' + 1$, so folgt

$$|f(x) - f(y)| \leq C \cdot |x - y|,$$

und dies beendet den Beweis von (1.57). □

Satz 1.58 (Kettenregel). *Seien $D \subseteq \mathbb{R}^l$, $E \subseteq \mathbb{R}^m$ und seien $g : D \rightarrow \mathbb{R}^m$, $f : E \rightarrow \mathbb{R}^n$ Abbildungen mit $g(D) \subseteq E$. Ist g total differenzierbar im Punkt $x \in D$ und f total differenzierbar im Punkt $g(x)$, so ist auch $h := f \circ g : D \rightarrow \mathbb{R}^n$ total differenzierbar im Punkt x , und es gilt*

$$h'(x) = f'(g(x)) \cdot g'(x),$$

in expliziter Schreibweise

$$\left(\frac{\partial h_i}{\partial x_k}(x) \right) = \left(\frac{\partial f_i}{\partial y_j}(g(x)) \right) \cdot \left(\frac{\partial g_j}{\partial x_k}(x) \right),$$

wobei der Punkt für die Matrixmultiplikation steht.

Beweis. Da f und g total differenzierbar sind, gilt

$$f(g(y)) - f(g(x)) = A(g(y) - g(x)) + R(g(y))$$

und

$$g(y) - g(x) = B(y - x) + S(y)$$

mit linearen Abbildungen $A : \mathbb{R}^m \rightarrow \mathbb{R}^n$, $B : \mathbb{R}^l \rightarrow \mathbb{R}^m$ und Restgliedern R und S , für die gilt:

$$\lim_{z \rightarrow g(x)} \frac{R(z)}{|z - g(x)|} = 0, \quad \lim_{y \rightarrow x} \frac{S(y)}{|y - x|} = 0.$$

Es folgt

$$h(y) - h(x) = A(g(y) - g(x)) + R(g(y)) = AB(y - x) + AS(y) + R(g(y)).$$

Das Matrixprodukt AB beschreibt gerade die in der Behauptung angegebene lineare Abbildung. Es ist zu zeigen, dass für das Restglied $T(y) := AS(y) + R(g(y))$ wieder

$$\lim_{y \rightarrow x} \frac{T(y)}{|y - x|} = 0$$

gilt.

Wir betrachten dazu die zwei Summanden von T einzeln: Zunächst soll $R(g(y))$ abgeschätzt werden.

Zu jedem $\varepsilon > 0$ gibt es ein $\delta > 0$ mit $|R(z)| \leq \varepsilon \cdot |z - g(x)|$ für alle $z \in K(g(x), \delta)$. Da g in x stetig ist, existiert zu diesem δ ein $\eta > 0$ mit $|g(y) - g(x)| \leq \delta$ für alle $y \in K(x, \eta)$. Nach Hilfssatz (1.57) gibt es Konstanten $C > 0$, $\zeta > 0$ mit

$$|g(y) - g(x)| \leq C|y - x| \quad \text{für alle } y \in K(x, \zeta).$$

Gilt nun $|y - x| \leq \min(\zeta, \eta)$, so folgt $|R(g(y))| \leq \varepsilon \cdot |g(y) - g(x)| \leq \varepsilon \cdot C \cdot |y - x|$. Daraus folgt

$$\lim_{y \rightarrow x} \frac{R(g(y))}{|y - x|} = 0.$$

Es ist noch das Glied $AS(y)$ zu behandeln. Zu jedem $\varepsilon > 0$ existiert zunächst ein $\beta > 0$ mit $|S(y)| \leq \varepsilon \cdot |y - x|$ für alle $y \in K(x, \beta)$. Aus Hilfssatz (1.56) folgt nun die Existenz einer Konstanten \tilde{C} mit $|Az| \leq \tilde{C} \cdot |z|$ für alle z . Es ist daher $|AS(y)| \leq \tilde{C} \cdot \varepsilon \cdot |y - x|$ für alle $y \in K(x, \beta)$, und wir haben damit

$$\lim_{y \rightarrow x} \frac{AS(y)}{|y - x|} = 0.$$

Damit ist Satz (1.58) bewiesen. □

Ist speziell $l = 1$ und ist D ein Intervall in \mathbb{R} , und ist $g : D \rightarrow \mathbb{R}^m$ eine differenzierbare Kurve, und ist außerdem f eine (reellwertige) Funktion von m Veränderlichen, so gilt

$$\frac{d}{dt} f(g(t)) = \sum_{k=1}^m \frac{\partial f}{\partial y_k}(g(t)) \cdot g'_k(t). \quad (1.8)$$

Die rechte Seite von Gleichung (1.8) ist dann das Skalarprodukt des Tangentialvektors

$$\frac{d}{dt} g(t) = (g'_1(t), \dots, g'_m(t))$$

mit dem Vektor

$$\left(\frac{\partial f}{\partial y_1}(g(t)), \dots, \frac{\partial f}{\partial y_m}(g(t)) \right).$$

Der letztgenannte Vektor heißt der **Gradient** der Funktion f und wird mit

$$\text{grad } f = \left(\frac{\partial f}{\partial y_1}, \dots, \frac{\partial f}{\partial y_m} \right)$$

bezeichnet. In kürzerer Form lässt sich (4.13) dann wie folgt schreiben:

$$\frac{d}{dt}f(g(t_0)) = g'(t_0) \cdot \text{grad}f(g(t_0)), \quad (1.9)$$

und dabei bezeichnet $x \cdot y$ das Skalarprodukt der Vektoren x und y im \mathbb{R}^m .

Ist $|g'(t_0)| \neq 0$, so gilt

$$\frac{\frac{d}{dt}f(g(t_0))}{|g'(t_0)|} = e \cdot \text{grad} f(g(t_0)), \quad (1.10)$$

wobei e den Vektor $\frac{g'(t_0)}{|g'(t_0)|}$ bezeichnet. Der Wert des Skalarprodukts (1.10) hängt dann nicht mehr von der Kurve selbst, sondern nur noch von der Richtung des Einheitsvektors e ab.

Allgemein nennt man

$$e \cdot \text{grad} f(x)$$

die **Richtungsableitung von f im Punkt x in der Richtung e** , wenn e ein beliebiger Vektor vom Betrag 1 ist.

Aus der Cauchy-Schwarzschen Ungleichung folgt

$$|e \cdot \text{grad} f| \leq |\text{grad} f|. \quad (1.11)$$

Zeigt e in Richtung des Gradienten, so gilt in (1.11) Gleichheit, und damit folgt: Der Gradient gibt die **Richtung des stärksten Wachsens (oder Fallens)** der Funktion f an, sein Betrag die Stärke des Wachsens (bzw. Fallens).

Der Gradient hat noch eine weitere geometrische Bedeutung:

Die „Flächen“ (bzw. im Fall einer Funktion von zwei Veränderlichen die Kurven), auf denen f einen festen Wert annimmt, nennt man **Niveauflächen** (bzw. Niveaulinien) der Funktion. Ist g eine Kurve, die ganz in einer Niveaufläche verläuft, so ist

$$\frac{d}{dt}f(g(t)) = 0,$$

weil $f(g(t))$ eine Konstante ist. Also gilt:

$$g'(t) \cdot \text{grad} f = 0.$$

Der Gradient steht also auf dem Tangentialvektor senkrecht, und dies gilt für alle ganz in den Niveauflächen verlaufenden Kurven. Das heißt, dass der Gradient, sofern er nicht der Nullvektor ist, auf der Niveaufläche durch den betreffenden Punkt senkrecht steht.

1.5 Höhere Ableitungen, Taylor-Entwicklung und lokale Extremwerte

Für den Differentialkalkül mehrerer Veränderlicher sind folgende Abkürzungen nützlich:

Für ein n -Tupel $i = (i_1, \dots, i_n) \in \mathbb{N}^n$ natürlicher Zahlen sei

$$|i| := i_1 + \dots + i_n$$

und

$$i! := i_1! \cdot i_2! \cdot \dots \cdot i_n!$$

Für $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ sei

$$x^i := x_1^{i_1} \cdot x_2^{i_2} \cdot \dots \cdot x_n^{i_n}.$$

Definition 1.59. Sei $D \subseteq \mathbb{R}^n$, $f : D \rightarrow \mathbb{R}$ und a ein innerer Punkt von D . f heißt **k -mal differenzierbar** im Punkt a , wenn f in einer ganzen Umgebung von a , d.h. in jedem Punkt einer Kugel $K(a, r)$ mit $r > 0$ $(k-1)$ -mal differenzierbar ist und wenn sämtliche $(k-1)$ -ten partiellen Ableitungen

$$\frac{\partial^{k-1} f}{\partial x_{i_{k-1}} \dots \partial x_{i_2} \partial x_{i_1}} = \frac{\partial}{\partial x_{i_{k-1}}} \dots \frac{\partial}{\partial x_{i_2}} \frac{\partial}{\partial x_{i_1}} f = f_{x_{i_1} x_{i_2} \dots x_{i_{k-1}}}$$

im Punkt a total differenzierbar sind. f heißt **k -mal stetig differenzierbar** im Punkt a , wenn f in einer Umgebung von a k -mal differenzierbar ist und alle k -ten partiellen Ableitungen von f im Punkt a stetig sind.

Bemerkung 1.60. Aus den Zahlen $1, \dots, n$ kann man n^k viele verschiedene k -Tupel i_1, \dots, i_k bilden, und man erhält so insgesamt n^k k -te partielle Ableitungen von f . In der Regel sind diese Ableitungen nicht alle voneinander verschieden.

Genauer gilt:

Satz 1.61. Ist f im Punkt a k -mal stetig differenzierbar, so ist

$$\frac{\partial^k f}{\partial x_{i_1} \dots \partial x_{i_k}}(a)$$

unabhängig von der Reihenfolge der x_{i_j} .

Jede Permutation eines k -Tupels (i_1, \dots, i_k) ergibt sich durch eine Folge von Vertauschungen benachbarter Paare des gegebenen k -Tupels. Daher können wir ohne Einschränkung $k = 2$ annehmen, d.h. es genügt zu zeigen

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}.$$

Hier gilt unter schwächeren Voraussetzungen als in (1.61):

Satz 1.62. Ist $f(x, y)$ in einem Kreis um den Punkt (x_0, y_0) partiell differenzierbar, und ist $\frac{\partial f}{\partial x}(x, y)$ in diesem Kreis nach y differenzierbar sowie $\frac{\partial^2 f}{\partial y \partial x}$ stetig im Punkt (x_0, y_0) , so existiert auch $\frac{\partial^2 f}{\partial x \partial y}$ in diesem Punkt und ist gleich $\frac{\partial^2 f}{\partial y \partial x}(x_0, y_0)$.

Beweis. Es ist zu zeigen, dass der Grenzwert

$$\lim_{x \rightarrow x_0} \frac{f_y(x, y_0) - f_y(x_0, y_0)}{x - x_0}$$

existiert und gleich $f_{xy}(x_0, y_0)$ ist. Unser Quotient ist der Grenzwert von

$$\frac{(f(x, y) - f(x, y_0)) - (f(x_0, y) - f(x_0, y_0))}{(x - x_0)(y - y_0)}$$

für $y \rightarrow y_0$. Indem man den Mittelwertsatz auf $g(x) := f(x, y) - f(x, y_0)$ anwendet, erhält man die Existenz eines x_1 zwischen x_0 und x , für das der Quotient den Wert

$$\frac{f_x(x_1, y) - f_x(x_1, y_0)}{y - y_0}$$

annimmt. Nochmalige Anwendung des Mittelwertsatzes liefert ein y_1 zwischen y_0 und y , für das der obige Quotient den Wert $f_{xy}(x_1, y_1)$ annimmt. Wir wählen nun zu einem gegebenen $\varepsilon > 0$ ein $\delta > 0$, so dass

$$|f_{xy}(x_1, y_1) - f_{xy}(x_0, y_0)| \leq \varepsilon$$

für $|x_1 - x_0| \leq \delta$ und $|y_1 - y_0| \leq \delta$ gilt. Dann folgt für $|x - x_0| \leq \delta$ und $|y - y_0| \leq \delta$

$$\left| \frac{(f(x, y) - f(x, y_0)) - (f(x_0, y) - f(x_0, y_0))}{(x - x_0)(y - y_0)} - f_{xy}(x_0, y_0) \right| \leq \varepsilon.$$

Wir führen nun den Grenzübergang $y \rightarrow y_0$ durch, wobei $|x - x_0| \leq \delta$ gelte. Dann ist

$$\left| \frac{f_y(x, y_0) - f_y(x_0, y_0)}{x - x_0} - f_{xy}(x_0, y_0) \right| \leq \varepsilon.$$

Also existiert dann $f_{yx}(x_0, y_0)$ und ist gleich $f_{xy}(x_0, y_0)$. Dies beendet den Beweis von Satz (1.62). \square

Bemerkung 1.63. Da die Reihenfolge der partiellen Ableitungen unerheblich ist, wenn f k -mal stetig differenzierbar ist, verwendet man für $i = (i_1, \dots, i_n)$ mit $|i| = k$ auch häufig die Schreibweise

$$\frac{\partial^k f}{\partial x_1^{i_1} \dots \partial x_n^{i_n}} = D^i f = D_1^{i_1} \dots D_n^{i_n} f,$$

wobei $D_k^{i_k} = D_k \dots D_k$ die i_k -fache Anwendung des Operators $D_k = \frac{\partial}{\partial x_k}$ bedeutet.

Wir interessieren uns nun für die **Taylor-Formel** im Fall einer Funktion f von n Veränderlichen. Das Ziel ist, $f(x+y)$ bei festem x und kleinem y durch ein Polynom in den Koordinaten von y anzunähern. Hierzu betrachten wir $g(t) := f(x + ty)$ (mit $t \in \mathbb{R}$ und $x + ty$ im Definitionsbereich von f) und wenden darauf die eindimensionale Taylor-Formel an.

Sei f in einer Umgebung von x , etwa in einer offenen Kugel um x , hinreichend oft differenzierbar. Die m -te Ableitung von $g(t)$ berechnet sich durch m -malige Anwendung der Kettenregel wie folgt:

$$g^{(m)}(t) = \sum \frac{\partial^m f}{\partial x_{i_m} \dots \partial x_{i_1}}(x + ty) y_{i_1} \dots y_{i_m}. \quad (1.12)$$

Dabei ist in (1.12) über **alle** m -gliedrigen Folgen (i_1, \dots, i_m) von Indizes i_k mit $i_k \in \{1, \dots, n\}$ zu summieren. Da es auf die Reihenfolge der Ableitungen nicht ankommt, kommen die meisten Glieder mehrfach vor, und man kann sie zusammenfassen. Es gilt:

$$g^{(m)}(t) = \sum_{|i|=m} \frac{m!}{i!} D^i f(x + ty) y^i, \quad (1.13)$$

und dabei summiert man über alle n -Tupel $(i_1, \dots, i_n) \in \mathbb{N}^n$ mit $|i| = m$. Dies kann man symbolisch auch wie folgt schreiben:

$$g^{(m)}(t) = \left(y_1 \frac{\partial}{\partial x_1} + \dots + y_n \frac{\partial}{\partial x_n} \right)^m f(x + ty). \quad (1.14)$$

Dies ist so zu verstehen: Man multipliziere die m -te Potenzformel aus, fasse in jedem Glied die y_i zu einem Potenzprodukt zusammen, ebenso die $\frac{\partial}{\partial x_i}$ zu einem Ausdruck $\frac{\partial^m}{\partial x_{i_1} \dots \partial x_{i_m}}$. Dann wende man diesen Ausdruck auf f an und setze dabei als Argument $x + ty$ ein.

Der **Beweis** der Formel (1.12) bzw. (1.13) erfolgt durch Induktion über m . Für $m = 1$ liefert die Anwendung der Kettenregel

$$\frac{dg}{dt}(t) = \frac{d}{dt} f(x_1 + ty_1, \dots, x_n + ty_n) = \sum_{i=1}^n \frac{\partial}{\partial x_i} f(x + ty) \cdot y_i.$$

Nun der Induktionsschluss von $m - 1$ nach m :

$$\frac{d^m}{dt^m} g(t) = \frac{d}{dt} \left(\sum_{i_1, \dots, i_{m-1}=1}^n \frac{\partial}{\partial x_{i_{m-1}}} \dots \frac{\partial}{\partial x_{i_1}} f(x + ty) y_{i_1} \dots y_{i_{m-1}} \right) =$$

$$\begin{aligned}
&= \sum_{j=1}^n \frac{\partial}{\partial x_j} \left(\sum_{i_1, \dots, i_{m-1}=1}^n \frac{\partial}{\partial x_{i_{m-1}}} \cdots \frac{\partial}{\partial x_{i_1}} f(x+ty) y_{i_1} \cdots y_{i_{m-1}} \right) \cdot y_j = \\
&= \sum_{i_1, \dots, i_m=1}^n \frac{\partial}{\partial x_{i_m}} \cdots \frac{\partial}{\partial x_{i_1}} f(x+ty) y_{i_1} \cdots y_{i_m}.
\end{aligned}$$

Kommt unter den Indizes (i_1, \dots, i_m) der Index 1 genau α_1 -mal vor, der Index 2 genau α_2 -mal usw. und ebenso der Index n jeweils α_n -mal, so gilt:

$$\begin{aligned}
&\frac{\partial}{\partial x_{i_m}} \cdots \frac{\partial}{\partial x_{i_1}} f(x+ty) y_{i_1} \cdots y_{i_m} = \\
&= D_1^{\alpha_1} \cdots D_n^{\alpha_n} f(x+ty) y_1^{\alpha_1} \cdots y_n^{\alpha_n}.
\end{aligned}$$

Es gibt genau $\frac{m!}{\alpha_1! \cdots \alpha_n!}$ viele m -Tupel (i_1, \dots, i_m) von Zahlen $1 \leq i_k \leq n$, bei denen die Zahl ν für $\nu = 1, \dots, n$ jeweils genau α_ν -mal vorkommt mit $\alpha_1 + \dots + \alpha_n = m$. Daher ist

$$\begin{aligned}
\frac{d^m g}{dt^m}(t) &= \sum_{i_1, \dots, i_m=1}^n \frac{\partial}{\partial x_{i_m}} \cdots \frac{\partial}{\partial x_{i_1}} f(x+ty) y_{i_1} \cdots y_{i_m} = \\
&= \sum_{|\alpha|=m} \frac{m!}{\alpha_1! \cdots \alpha_n!} D_1^{\alpha_1} \cdots D_n^{\alpha_n} f(x+ty) y_1^{\alpha_1} \cdots y_n^{\alpha_n} = \\
&= \sum_{|\alpha|=m} \frac{m!}{\alpha!} D^\alpha f(x+ty) y^\alpha.
\end{aligned}$$

Satz 1.64 (Taylor-Formel mit Restglied in Integralform). *Sei $U \subseteq \mathbb{R}^n$ offen, $x \in U$ und $y \in \mathbb{R}^n$ so dass $x+ty \in U$ für alle $t \in [0, 1]$ gilt. Die Funktion $f : U \rightarrow \mathbb{R}$ sei $(m+1)$ -mal stetig differenzierbar. Dann gilt:*

$$\begin{aligned}
f(x+y) &= \sum_{i=0}^m \frac{1}{i!} \left(y_1 \frac{\partial}{\partial x_1} + \cdots + y_n \frac{\partial}{\partial x_n} \right)^i f(x) + R_m = \\
&= \sum_{|\alpha| \leq m} \frac{D^\alpha f(x)}{\alpha!} y^\alpha + R_m
\end{aligned}$$

mit

$$R_m = \frac{1}{m!} \int_0^1 (1-t)^m \left(y_1 \frac{\partial}{\partial x_1} + \cdots + y_n \frac{\partial}{\partial x_n} \right)^{m+1} f(x+ty).$$

Beweis. Für den Beweis braucht man nur die Taylor-Formel für eindimensionale Funktionen g , die hinreichend oft differenzierbar sind, anzuwenden:

$$g(t) = \sum_{k=0}^m \frac{g^{(k)}(0)}{k!} t^k + R_m(t)$$

mit

$$R_m(t) = \frac{1}{m!} \int_0^t (t - \tau)^m g^{(m+1)}(\tau) d\tau$$

(vgl. Analysis I). □

Die Taylor-Formel wird angewendet, um lokale Extremwerte einer Funktion zu bestimmen. Dazu wird zunächst definiert:

Definition 1.65. Die Funktion $f : D \rightarrow \mathbb{R}$, $D \subseteq \mathbb{R}^n$ hat an der Stelle x ein

- a) **absolutes Maximum (bzw. Minimum)**, wenn $f(y) \leq f(x)$ (bzw. $f(y) \geq f(x)$) für alle y aus dem Definitionsbereich von f gilt;
- b) **relatives Maximum (bzw. Minimum)**, wenn es ein $\delta > 0$ gibt, so dass $f(y) \leq f(x)$ (bzw. $f(y) \geq f(x)$) für alle $y \in D$ mit $|y - x| \leq \delta$ gilt.

Hat f in x ein relatives Extremum, und ist x ein innerer Punkt des Definitionsbereichs D_f von f , so gilt: Die Funktion

$$g(t) := f(x + ty) \quad (y \in \mathbb{R}^n)$$

hat bei $t = 0$ ein relatives Extremum. Daher ist dann

$$g'(0) = \sum_{i=1}^n \frac{\partial f}{\partial x_i}(x) y_i = 0,$$

und dies gilt für alle Vektoren $y \in \mathbb{R}^n$. Daher ist

$$\frac{\partial f}{\partial x_i}(x) = \dots = \frac{\partial f}{\partial x_n}(x) = 0.$$

D.h.: Alle ersten partiellen Ableitungen von f an der Stelle x verschwinden.

Definition 1.66. Ein innerer Punkt x des Definitionsbereichs von f mit $\frac{\partial f}{\partial x_i}(x) = 0$ für $i = 1, \dots, n$ heißt **stationärer Punkt** von f .

Liegt in x ein Maximum vor, so muss weiter gelten

$$g''(x) = \sum_{i,k=1}^n \frac{\partial^2 f}{\partial x_i \partial x_k}(x) y_i y_k \leq 0$$

und entsprechend ≥ 0 im Fall eines Minimums. Also: Mit $a_{ik} := \frac{\partial^2 f}{\partial x_i \partial x_k}(x)$ folgt

$$\sum_{i,k=1}^n a_{ik} y_i y_k \leq 0 \quad \text{für alle } y. \quad (1.15)$$

Die quadratische Form aus (1.15) in den Variablen y_1, \dots, y_n heißt in diesem Fall **negativ semidefinit**. Man nennt diese quadratische Form **positiv semidefinit**, wenn $\sum a_{ik} y_i y_k \geq 0$ für alle y gilt. Analog nennt man die quadratische Form **positiv definit**, wenn sogar $\sum_{i,k=1}^n a_{ik} y_i y_k > 0$ für alle $y \neq 0$ gilt, und **negativ definit** im Fall des Zeichens „ $<$ “. Quadratische Formen, die weder positiv noch negativ semidefinit sind, heißen **indefinit**.

Satz 1.67. *Die Funktion f sei in einer Umgebung der Stelle $x \in \mathbb{R}^n$ zweimal stetig differenzierbar. Hat f an der Stelle x ein relatives Maximum (Minimum), so verschwinden dort alle ersten partiellen Ableitungen, und die quadratische Form*

$$Q(y_1, \dots, y_n) = \sum_{i,k=1}^n \frac{\partial^2 f}{\partial x_i \partial x_k}(x) y_i y_k$$

ist negativ (positiv) semidefinit. Verschwinden umgekehrt alle ersten Ableitungen und ist Q negativ (positiv) definit, so hat f bei x ein relatives Maximum (bzw. Minimum).

Beweis. Die erste Hälfte des Satzes ist bereits bewiesen. Zur Umkehrung genügt es **nicht**, festzustellen, dass die Funktion $g(t) := f(x + ty)$ für jedes y bei $t = 0$ ein relatives Minimum hat, denn hieraus folgt noch nicht, dass f ein relatives Extremum hat. Dies zeigt das folgende Beispiel:

$$f(x) := (x_2 - x_1^2)(x_2 - 2x_1^2)$$

an der Stelle $x = (0, 0)$. Es gilt nämlich $f(0) = 0$, aber es ist $f(x) < 0$ zwischen den beiden Parabeln $x_2 = x_1^2$ und $x_2 = 2x_1^2$ und $f(x) \geq 0$ sonst. Also hat f an der Stelle 0 kein relatives Extremum. Aber auf jeder Geraden durch den Nullpunkt liegt bei 0 ein relatives Minimum vor.

Zum Beweis der Umkehrung benötigen wir den

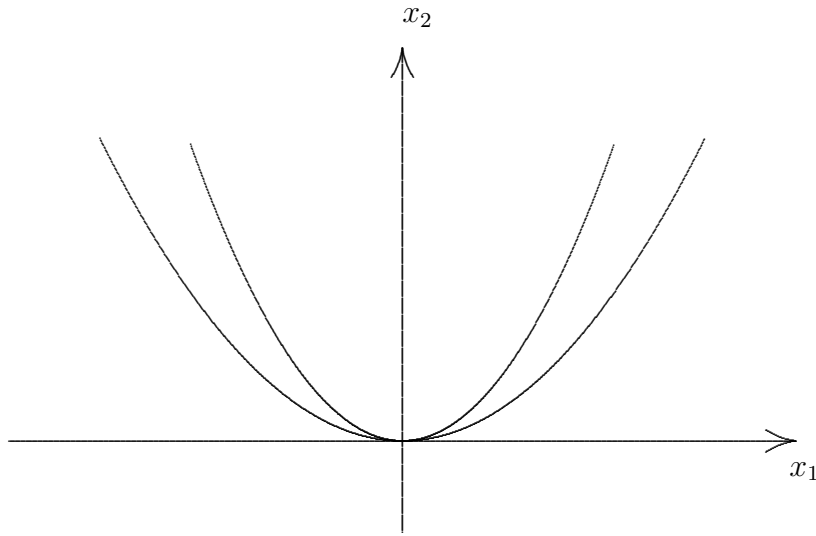


Abbildung 1.9: Ist $f(x) = (x_2 - x_1^2)(x_2 - 2x_1^2)$, so ist f zwischen diesen beiden Parabeln negativ und sonst ≥ 0 . Also liegt bei 0 kein relatives Extremum dieser Funktion vor, wohl aber auf jeder Geraden durch den Nullpunkt.

Hilfssatz 1.68. *Ist die quadratische Form $Q(y)$ positiv definit, so gibt es eine Konstante $c > 0$, so dass $Q(y) \geq c|y|^2$ für alle $y \in \mathbb{R}^n$ gilt.*

Beweis. Die Menge $M = \{y \in \mathbb{R}^n : |y| = 1\}$ ist beschränkt und abgeschlossen, also kompakt. Also nimmt Q auf M das Minimum an, d.h. es gibt ein $z \in M$ mit $Q(z) = \min_{y \in M} Q(y)$. Wir setzen $c := Q(z)$, und da Q positiv definit ist, gilt $c > 0$.

Ist nun $y \neq 0$, so gilt $\frac{y}{|y|} \in M$ und daher $c \leq Q\left(\frac{y}{|y|}\right) = \frac{1}{|y|^2}Q(y)$. Daher ist $Q(y) \geq c \cdot |y|^2$, und dies beendet den Beweis von (1.68). \square

Fortsetzung vom Beweis von Satz (1.67):

Die Taylor-Formel mit $m = 1$ liefert

$$\begin{aligned} f(x+y) - f(x) &= R_1 = \int_0^1 (1-t) \sum_{i,k} \frac{\partial^2 f}{\partial x_i \partial x_k}(x+ty) y_i y_k dt = \\ &= \frac{1}{2} Q(y) + \sum_{i,k} b_{ik} y_i y_k \end{aligned}$$

mit

$$b_{ik} = \int_0^1 (1-t) \left(\frac{\partial^2 f}{\partial x_i \partial x_k}(x+ty) - \frac{\partial^2 f}{\partial x_i \partial x_k}(x) \right) dt.$$

Ist Q negativ definit, so ist $-Q$ positiv definit, und es existiert ein $c > 0$ mit $-Q(y) \geq c|y|^2$ für alle $y \in \mathbb{R}^n$. Aus der Stetigkeit der zweiten Ableitungen folgt: Zu $\varepsilon = \frac{c}{4n^2}$ existiert ein $\delta > 0$ mit $|b_{ik}| \leq \varepsilon$ für alle y mit $|y| \leq \delta$ und $i, k = 1, \dots, n$. Für diese y gilt dann

$$f(x+y) - f(x) \leq -\frac{c}{2}|y|^2 + n^2 \cdot \frac{c}{4n^2} \cdot |y|^2 = -\frac{c}{4} \cdot |y|^2 \leq 0,$$

so dass f an der Stelle x ein relatives Maximum hat.

Für positiv definite Q erhält man entsprechend $f(x+y) - f(x) \geq \frac{c}{4}|y|^2 \geq 0$, so dass dann f bei x ein relatives Minimum hat. Damit ist Satz (1.67) bewiesen. \square

Um das Kriterium von Satz (1.67) anwenden zu können, muss man entscheiden können, wann quadratische Formen Q positiv oder negativ definit sind.

Relativ einfach sind die Verhältnisse noch bei $n = 2$. Dann nämlich schreibt sich die quadratische Form Q als

$$Q(x, y) = ax^2 + 2bxy + cy^2$$

mit $a, b, c \in \mathbb{R}$. Ist $a = 0$, so ist Q weder positiv noch negativ definit. Ist $a \neq 0$, so kann man schreiben

$$Q(x, y) = \frac{1}{a} \left((ax + by)^2 + (ac - b^2)y^2 \right).$$

Also ist Q definit (d.h. positiv oder negativ definit) genau dann, wenn $ac - b^2 > 0$ gilt. Genau dann ist Q positiv definit, wenn sowohl $a > 0$ wie $ac - b^2 > 0$ gilt. Ebenso ist Q genau dann negativ definit, wenn $a < 0$ und außerdem $ac - b^2 > 0$ ist.

Allgemeiner kann man folgenden Satz beweisen:

Satz 1.69. Die symmetrische $n \times n$ -Matrix $B = (b_{ij})$ ist genau dann positiv definit, wenn die Hauptuntermatrizen

$$B_m = \begin{pmatrix} b_{11} & \cdots & b_{1m} \\ \vdots & & \vdots \\ b_{m1} & \cdots & b_{mm} \end{pmatrix}$$

der Matrix $B = B_n$ für $m = 1, 2, \dots, n$ positive Determinanten haben:

$$\det B_1 = b_{11} > 0, \det B_2 = b_{11}b_{22} - b_{12}^2 > 0, \dots, \det B_n > 0. \quad (1.16)$$

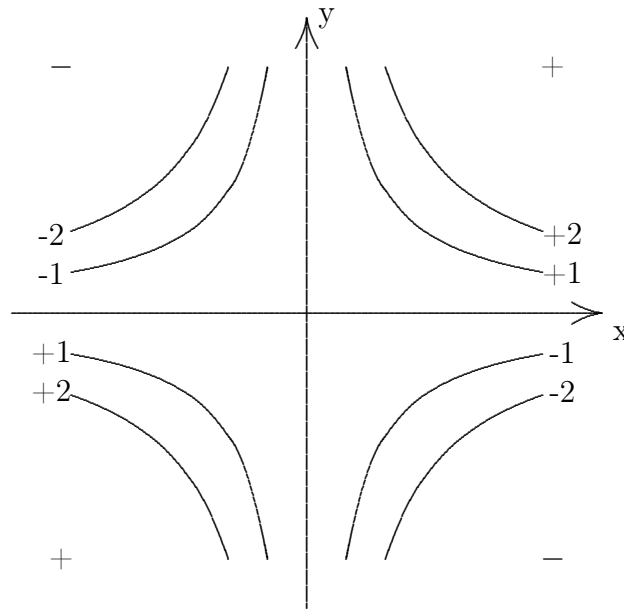


Abbildung 1.10: Höhenlinien der Funktion $z = xy$. Alle ersten Ableitungen im Nullpunkt verschwinden, aber die quadratische Form ist indefinit - und so liegt hier im Nullpunkt kein Extremum vor.

Bemerkung. Dieser Satz wird hier nicht bewiesen. Die Notwendigkeit der Bedingungen (1.16) ist relativ leicht einzusehen. Bedeutend schwieriger ist es einzusehen, dass die Bedingung (1.16) hinreichend ist. Hierzu vergleiche man [2].

Bemerkung 1.70. Ist die quadratische Form Q indefinit, so liegt kein Extremum vor.

Beispiel 1.71. Sei $f(x, y) = xy$. Die beiden ersten Ableitungen verschwinden im Nullpunkt (denn es ist $\frac{\partial f}{\partial x} = y$ und $\frac{\partial f}{\partial y} = x$). Die quadratische Form der zweiten Ableitungen ist indefinit: Es gilt $\frac{\partial^2 f}{\partial x^2} = \frac{\partial^2 f}{\partial y^2} = 0$ und $\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x} = 1$, so dass sich für B die Matrix $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ ergibt.

Bemerkung 1.72. Im Fall der semidefiniten Formen liefert Satz (1.67) keine Entscheidung, ob ein Extremwert vorliegt oder nicht.

Beispiel 1.73. Wir betrachten $f(x, y) = x^2 \pm y^4$. Im Falle des Pluszeichens liegt bei 0 ein Minimum, im Fall des Minuszeichens aber ein Sattelpunkt vor. Es ist hier

$$\frac{\partial f}{\partial x}(0) = \frac{\partial f}{\partial y}(0) = 0$$

und

$$\begin{pmatrix} \frac{\partial^2 f}{\partial x^2}(0) & \frac{\partial^2 f}{\partial x \partial y}(0) \\ \frac{\partial^2 f}{\partial y \partial x}(0) & \frac{\partial^2 f}{\partial y^2}(0) \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}.$$

Die zugehörige quadratische Form ist positiv semidefinit.

1.6 Implizite Funktionen, Umkehrabbildung und Extremwerte mit Nebenbedingungen

Es geht in diesem Abschnitt um die Auflösung von Gleichungen bzw. Gleichungssystemen. Ist zum Beispiel $D \subseteq \mathbb{R}^2$ und $F : D \rightarrow \mathbb{R}$, so sagt man, durch

$$F(x, y) = 0$$

sei auf dem Intervall $I \subseteq \mathbb{R}$ eine **implizite Funktion**

$$f : I \rightarrow K \subseteq \mathbb{R}$$

definiert, wenn es zu jedem $x \in I$ genau ein $y = f(x) \in K$ gibt mit $(x, y) \in D$ und $F(x, y) = 0$.

Beispiel 1.74. Sei $D = \mathbb{R}^2$ und $F(x, y) = 3x + 2y - 1 = 0$. Man kann diese Gleichung nach y auflösen und erhält damit für y eine explizite Darstellung:

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

Dass eine solche explizite Darstellung nicht immer möglich ist, zeigt der folgende Fall:

Beispiel 1.75. Sei $D = \mathbb{R}^2$ und

$$F(x, y) = e^y + y^3 + x^3 + x^2 - 1.$$

Dann gibt es wieder zu jedem $x \in \mathbb{R}$ genau ein $y \in \mathbb{R}$ mit $F(x, y) = 0$. Die Funktion $h(y) := e^y + y^3$ ist nämlich streng monoton wachsend mit Wertebereich \mathbb{R} . Aber man kann die Gleichung $F(x, y) = 0$ in diesem Fall nicht durch algebraische Umformungen nach y auflösen.

Wir nehmen nun an, durch $F(x, y) = 0$ sei wie in den Beispielen (1.74) und (1.75) eine implizite Funktion $y = f(x)$ gegeben, und F und f seien differenzierbar. Nach der Kettenregel folgt dann aus der Gleichung $F(x, f(x)) = 0$ die Gleichung

$$\frac{\partial F}{\partial x}(x, f(x)) + \frac{\partial F}{\partial y}(x, f(x))f'(x) = 0.$$

Ist jetzt $\frac{\partial F}{\partial y}(x, f(x)) \neq 0$, so kann man bei gegebenem F hieraus $f'(x)$ berechnen. Ist dagegen $\frac{\partial F}{\partial y}(x, f(x)) = 0$, so folgt: $\frac{\partial F}{\partial x}(x, f(x)) = 0$, und $(x, f(x))$ ist stationärer Punkt von F .

Die Frage nach der Auflösung der Gleichung $F(x, y) = 0$ in der Umgebung solcher Punkte ist meist schwieriger zu behandeln. Wir werden daher im Folgenden stets $\frac{\partial F}{\partial y}(x, y) \neq 0$ voraussetzen. Aber selbst das garantiert nicht die Lösbarkeit der Gleichung $F(x, y) = 0$.

Beispiel 1.76. Sei $D = \mathbb{R}^2$ und $F(x, y) = x^2 + e^y$. Wegen $x^2 \geq 0$ und $e^y > 0$ ist stets $x^2 + e^y > 0$, außerdem gilt $\frac{\partial F}{\partial y} = e^y \neq 0$. Hier liegt für die Gleichung $F(x, y) = 0$ überhaupt keine Lösung vor.

Später wird gezeigt: Existiert wenigstens ein (\bar{x}, \bar{y}) mit $\frac{\partial F}{\partial y}(\bar{x}, \bar{y}) \neq 0$ und $F(\bar{x}, \bar{y}) = 0$, dann lässt sich $F(x, y) = 0$ in einer hinreichend kleinen Umgebung von (\bar{x}, \bar{y}) eindeutig auflösen.

Für die allgemeine Theorie behandeln wir gleich allgemeiner Systeme von mehreren Gleichungen:

$$F_i(x_1, \dots, x_m; y_1, \dots, y_n) = 0 \quad \text{für } i = 1, \dots, n. \quad (1.17)$$

Dies sind n Gleichungen mit n Unbekannten y_1, \dots, y_n . Wir setzen alle F_i als stetig differenzierbar voraus. Wie überträgt sich die Bedingung $\frac{\partial F}{\partial y} \neq 0$? Um diese Frage zu beantworten, nehmen wir Folgendes an:

Annahme: Es gibt n differenzierbare Funktionen $y_i = f_i(x_1, \dots, x_m)$ ($i = 1, \dots, n$), die Lösungen von (1.17) sind. Dann folgt durch Ableiten nach x_k :

$$\frac{\partial F_i}{\partial x_k} + \sum_{j=1}^n \frac{\partial F_i}{\partial y_j} \frac{\partial f_j}{\partial x_k} = 0, \quad (1.18)$$

d.h. mit den Abkürzungen

$$\frac{\partial F}{\partial x} := \left(\frac{\partial F_i}{\partial x_k} \right), \quad \frac{\partial F}{\partial y} := \left(\frac{\partial F_i}{\partial y_j} \right), \quad \frac{\partial f}{\partial x} := \left(\frac{\partial f_j}{\partial x_k} \right)$$

gilt die Matrixgleichung

$$\frac{\partial F}{\partial x} + \frac{\partial F}{\partial y} \cdot \frac{\partial f}{\partial x} = 0.$$

Auf beiden Seiten des Pluszeichens steht je eine $n \times m$ -Matrix. Das lineare Gleichungssystem ist eindeutig nach den $\frac{\partial f_j}{\partial x_k}$ auflösbar, wenn $\det \frac{\partial F}{\partial y} \neq 0$ gilt. Man nennt

$$\det \left(\frac{\partial F_i}{\partial y_j} \right) = \det \frac{\partial F}{\partial y}$$

auch die **Funktionaldeterminante** der Funktionen F_1, \dots, F_n nach den Variablen y_1, \dots, y_n und verwendet dafür auch die Bezeichnung

$$\frac{\partial(F_1, \dots, F_n)}{\partial(y_1, \dots, y_n)}.$$

Satz 1.77. Die Funktionen $F_i(x_1, \dots, x_m; y_1, \dots, y_n)$ seien für $i = 1, \dots, n$ stetig differenzierbar auf einer offenen Teilmenge des \mathbb{R}^{n+m} und $y_i = f_i(x_1, \dots, x_m)$ seien stetige Lösungen des Gleichungssystems (1.17). Ist die Funktionaldeterminante

$$\frac{\partial(F_1, \dots, F_n)}{\partial(y_1, \dots, y_n)}$$

an der Stelle

$$\bar{x} = (\bar{x}_1, \dots, \bar{x}_m), \bar{y} = (\bar{y}_1, \dots, \bar{y}_n), \bar{y}_j = f_j(\bar{x})$$

von Null verschieden, so sind die Funktionen f_j an der Stelle \bar{x} differenzierbar, und ihre partiellen Ableitungen ergeben sich durch Auflösen des linearen Gleichungssystems (1.18). Sind die F_i jeweils k -mal stetig differenzierbar, so auch die f_j .

Beweis. Sei $x := (x_1, \dots, x_m)$ und $f = (f_1, \dots, f_n)$ und $y = f(x)$. Wir nehmen an, x liege so nahe bei \bar{x} , dass y nahe genug bei $\bar{y} = f(\bar{x})$ liegt: genauer, so nahe, dass die Verbindungsstrecke von (\bar{x}, \bar{y}) nach (x, y) ganz im Definitionsbereich von F liegt. Auf

$$F(x, y) = F(\bar{x} + (x - \bar{x}), \bar{y} + (y - \bar{y}))$$

kann man die Taylor-Formel anwenden mit $m = 0$ (vgl. Satz 1.64). Damit ergibt sich

$$F(x, y) = F(\bar{x}, \bar{y}) + \int_0^1 \left(\frac{\partial F}{\partial x}(\bar{x}, \bar{y}) + t((x, y) - (\bar{x}, \bar{y})) \right) \cdot (x - \bar{x}) +$$

$$+ \frac{\partial F}{\partial y} \left((\bar{x}, \bar{y}) + t((x, y) - (\bar{x}, \bar{y})) \right) dt,$$

und es folgt:

$$\begin{aligned} 0 &= \int_0^1 \frac{\partial F}{\partial x} ((\bar{x}, \bar{y}) + t(x, y)) dt \cdot (x - x') + \\ &+ \int_0^1 \frac{\partial f}{\partial y} ((\bar{x}, \bar{y}) + t(x, y)) dt \cdot (y - \bar{y}) = \\ &= A(x - x') + B(y - y') \end{aligned} \quad (1.19)$$

mit Matrizen A und B , die sich von $\frac{\partial F}{\partial x}(\bar{x}, \bar{y}) =: \bar{A}$ und $\bar{B} := \frac{\partial F}{\partial y}(\bar{x}, \bar{y})$ wegen der Stetigkeit der Ableitungen nur wenig unterscheiden, wenn x, y nahe genug bei \bar{x}, \bar{y} gewählt werden. Genauer ist

$$\lim_{(x,y) \rightarrow (\bar{x}, \bar{y})} A = \frac{\partial F}{\partial x}(\bar{x}, \bar{y}) = \bar{A}, \quad \lim_{(x,y) \rightarrow (\bar{x}, \bar{y})} B = \bar{B}.$$

Wegen $\det \bar{B} = \det \frac{\partial F}{\partial y}(\bar{x}, \bar{y}) \neq 0$ ist $\det B \neq 0$ für alle (x, y) , die nahe genug bei (\bar{x}, \bar{y}) liegen.

Da die Koeffizienten von B^{-1} rationale Funktionen in den Koeffizienten von B sind mit Nenner $\det B$, konvergiert auch B^{-1} gegen \bar{B}^{-1} für $(x, y) \rightarrow (\bar{x}, \bar{y})$. Die Auflösung von (1.19) liefert dann:

$$y - \bar{y} = f(x) - f(\bar{x}) = -B^{-1}A(x - \bar{x}) = -\bar{B}^{-1}\bar{A}(x - \bar{x}) - (B^{-1}A - \bar{B}^{-1}\bar{A})(x - \bar{x}).$$

Hieraus kann man wie folgt sehen, dass $f(x)$ differenzierbar ist: Sei $\varepsilon > 0$. Zu zeigen ist

$$|(B^{-1}A - \bar{B}^{-1}\bar{A})(x - \bar{x})| \leq \varepsilon |x - \bar{x}|,$$

falls x genügend nahe bei \bar{x} gewählt wird. Dazu wähle man x so nahe bei \bar{x} , dass sämtliche Matrixelemente von $B^{-1}A - \bar{B}^{-1}\bar{A}$ kleiner als $\frac{\varepsilon}{mn}$ sind.

Die Behauptung über die k -malige Differenzierbarkeit ergibt sich durch Induktion über k unter Beachtung der Tatsache, dass sich die Ableitungen der f_j durch die der F_i ausdrücken lassen mittels Auflösung der Gleichungen

$$\frac{\partial F_i}{\partial x_k} + \sum_{j=1}^n \frac{\partial F_i}{\partial y_j} \frac{\partial f_j}{\partial x_k} = 0.$$

Es muss noch die Auflösbarkeit des Gleichungssystems aus (1.17)

$$F_i(x_1, \dots, x_m; y_1, \dots, y_n) = 0 \quad \text{für } i = 1, \dots, n$$

nachgewiesen werden. □

Hierzu dient folgender

Satz 1.78. *Sind die Funktionen $F_i(x_1, \dots, x_m; y_1, \dots, y_n)$ in einem offenen Bereich des \mathbb{R}^{m+n} r -mal stetig differenzierbar und ist $\bar{x} = (\bar{x}_1, \dots, \bar{x}_m)$, $\bar{y} = (\bar{y}_1, \dots, \bar{y}_n)$ Lösung des Gleichungssystems (1.17) mit $\frac{\partial(F_1, \dots, F_n)}{\partial y_1, \dots, y_n}(\bar{x}, \bar{y}) \neq 0$, so existieren reelle Zahlen $a > 0$, $b > 0$ mit folgender Eigenschaft: Das Gleichungssystem*

$$F_i(x_1, \dots, x_m, y_1, \dots, y_n) = 0$$

hat für jedes x mit $|x - \bar{x}| \leq a$ genau eine Lösung $y = f(x)$ mit $|y - \bar{y}| \leq b$. Die Abbildung $f : K(\bar{x}, a) \rightarrow \mathbb{R}^n$ ist r -mal stetig differenzierbar.

Beweis. Man fasse die F_i zu einem Spaltenvektor F zusammen. Dann ist (1.17) gleichbedeutend mit $F(x, y) = 0$, und mit $\bar{A} := \frac{\partial F}{\partial y}(\bar{x}, \bar{y})$ ist dies äquivalent zu

$$y = y - \bar{A}^{-1}F(x, y) =: G(x, y). \quad (1.20)$$

Es ist $G(\bar{x}, \bar{y}) = \bar{y}$, und G ist stetig differenzierbar mit $\frac{\partial G}{\partial y}(\bar{x}, \bar{y}) = E_m - \bar{A}^{-1} \cdot \bar{A}$ (wobei E_m die $m \times m$ -Einheitsmatrix bezeichnet). Es ist daher

$$\lim_{(x, y) \rightarrow (\bar{x}, \bar{y})} \frac{\partial G_i}{\partial y_j}(x, y) = 0.$$

Mit der Taylor-Formel folgt hieraus

$$|G(x, y') - G(x, y)| \leq \frac{1}{2}|y' - y| \quad (1.21)$$

für alle (x, y') , (x, y) , die nahe genug bei (\bar{x}, \bar{y}) sind; etwa für $|x - \bar{x}| \leq a$, $|y - \bar{y}| \leq b$ sowie $|y' - \bar{y}| \leq b$.

Mit (1.21) lässt sich zunächst zeigen, dass es zu jedem x mit $|x - \bar{x}| \leq a$ höchstens ein y mit $|y - \bar{y}| \leq b$ gibt, so dass (1.20) und damit (1.17) gilt.

Eindeutigkeit: Aus

$$y = G(x, y), \quad |y - \bar{y}| \leq b, \quad |x - \bar{x}| \leq a \quad \text{und} \quad y' = G(x, y'), \quad |y' - \bar{y}| \leq b$$

folgt wegen (1.21)

$$|y - y'| = |G(x, y) - G(x, y')| \leq \frac{1}{2}|y' - y|,$$

also $|y - y'| = 0$ und damit $y = y'$.

Es bleibt zu zeigen, dass $G(x, y) = y$ eine stetige Lösung $y = f(x)$ besitzt. Dazu wenden wir ein sog. **Iterationsverfahren** an:

Setze

$$g_0(x) := \bar{y} \quad (1.22)$$

und

$$g_{k+1}(x) := G(x, g_k(x)) \quad (1.23)$$

für $k \geq 0$. Die gesuchte Lösung wird dann durch den Grenzübergang

$$f(x) = \lim_{k \rightarrow \infty} g_k(x)$$

gefunden.

Dazu ist zunächst zu zeigen, dass (1.23) wirklich g_{k+1} definiert, d.h. dass die Werte $g_k(x)$ in dem Bereich liegen, in dem die Ungleichung (1.21) gilt.

Für g_0 ist das klar, für $k > 0$ führen wir einen Induktionsbeweis durch: Zu zeigen ist

$$|x - \bar{x}| \leq a, |y - \bar{y}| \leq b \implies |G(x, y) - \bar{y}| \leq b,$$

falls a, b passend gewählt werden.

Es ist

$$\begin{aligned} |G(x, y) - \bar{y}| &= |G(x, y) - G(\bar{x}, \bar{y})| \leq \\ &\leq |G(x, y) - G(x, \bar{y})| + |G(x, \bar{y}) - G(\bar{x}, \bar{y})|. \end{aligned}$$

Der erste Summand auf der rechten Seite ist $\leq \frac{1}{2}|y - \bar{y}|$ nach (1.22). Für den zweiten Summanden ergibt sich aus der Taylor-Formel eine Abschätzung

$$\leq C \cdot |x - \bar{x}|$$

mit einer Konstanten C , die aus oberen Schranken für die partiellen Ableitungen $\frac{\partial G_i}{\partial x_k}$ erhalten wird. Insgesamt ist daher

$$|G(x, y) - \bar{y}| \leq \frac{1}{2}|y - \bar{y}| + C|x - \bar{x}| \leq \frac{b}{2} + Ca$$

und damit $|G(x, y) - \bar{y}| \leq b$ gilt, falls $Ca \leq \frac{b}{2}$ ist. Trifft dies noch nicht von alleine zu, dann verkleinere man a durch den Wert $\frac{b}{2C}$.

Also: Ist $|x - \bar{x}| \leq a$ und $|y - \bar{y}| \leq b$, so gilt dann $|G(x, y) - \bar{y}| \leq b$. Damit ist gezeigt, dass die Rekursion (1.23) wohldefiniert ist.

Aus (1.21) folgt:

$$|g_{k+1}(x) - g_k(x)| = |G(x, g_k(x)) - G(x, g_{k-1}(x))| \leq \frac{1}{2}|g_k(x) - g_{k-1}(x)|,$$

also

$$|g_{k+1}(x) - g_k(x)| \leq \frac{1}{2^k}|g_1(x) - g_0(x)| \leq \frac{b}{2^k}.$$

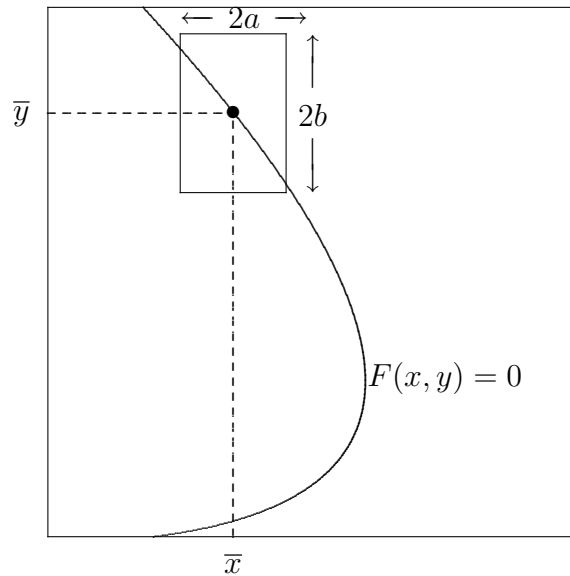


Abbildung 1.11: Satz (1.78) mit $m = n = 1$.

Nach dem Majorantenkriterium konvergiert daher

$$\sum_{k=0}^{\infty} (g_{k+1}(x) - g_k(x))$$

absolut und gleichmäßig für $|x - x'| \leq a$. Damit konvergiert auch die Folge $g_k(x)$ gleichmäßig auf $|x - \bar{x}| \leq a$ gegen eine Grenzfunktion f :

$$f(x) = \lim_{k \rightarrow \infty} g_k(x).$$

Als gleichmäßiger Limes stetiger Funktionen ist f wieder stetig, und es gilt wegen der Stetigkeit von G nach (1.23):

$$f(x) = G(x, f(x)).$$

Damit ist die Lösbarkeit von (1.17) durch eine stetige Funktion f gezeigt. Die Differenzierbarkeit der Lösung f folgt aus Satz (1.77), und damit ist Satz (1.78) bewiesen. \square

Im Spezialfall $m = n$ und

$$F_i(x_1, \dots, x_n; y_1, \dots, y_n) = h_i(y_1, \dots, y_n) - x_i$$

hat man eine Abbildung $h : D_h \rightarrow \mathbb{R}^n$. Die Auflösbarkeit der Gleichung $F(x, y) = 0$ bedeutet in diesem Fall, dass man in der Nähe eines Bildpunktes $\bar{x} = h(\bar{y})$ eine Umkehrabbildung zu h sucht. Satz (1.78) liefert dann:

Satz 1.79. Sind die Funktionen $h_i(y_1, \dots, y_n)$ für $i = 1, \dots, n$ in einer offenen Teilmenge $D \subseteq \mathbb{R}^n$ stetig differenzierbar und ist dort die Funktionaldeterminante $\frac{\partial(h_1, \dots, h_n)}{\partial(y_1, \dots, y_n)} \neq 0$, so ist das Bild $h(U)$ jeder offenen Teilmenge $U \subseteq D$ offen in \mathbb{R}^n , und zu jedem $\bar{y} \in D$ gibt es eine offene Menge $V \subseteq D$, $\bar{y} \in V$, so dass $h : V \rightarrow h(V)$ eine stetig differenzierbare Umkehrabbildung besitzt.

Beweis. Sei $U \subseteq D$ offen und $\bar{y} \in U$. Nach (1.78) existieren dann Kugeln $K(h(\bar{y}), a)$ und $K(\bar{y}, b) \subseteq U$, so dass die Gleichung $h(y) = x$ für $x \in K(h(\bar{y}), a)$ durch genau ein $y \in K(\bar{y}, b)$ lösbar ist. Es ist dann $h(\bar{y})$ ein innerer Punkt von $h(U)$, und $h(U)$ ist offen. Ebenfalls offen ist

$$V := \{y \in K(\bar{y}, b) : h(y) \in K(h(\bar{y}), a)\}$$

als Urbild einer offenen Menge unter einer stetigen Abbildung; vergleiche Satz (1.32).

Nun ist $h : V \rightarrow h(V)$ bijektiv, die Umkehrabbildung ist nach Satz (1.78) also wieder stetig differenzierbar. Dies beendet den Beweis von Satz (1.79). \square

Bemerkung 1.80. Unter den Voraussetzungen von (6.13) gibt es im Allgemeinen keine Umkehrabbildung von h auf ganz D .

Das zeigt das

Beispiel 1.81. Sei

$$h : \begin{cases} \mathbb{R}_+ \times \mathbb{R} \rightarrow \mathbb{R}^2 \\ (r, \varphi) \mapsto (r \cos \varphi, r \sin \varphi) = (x, y) \end{cases}$$

(dies ist die Abbildung, die den Polarkoordinaten die kartesischen Koordinaten zuordnet.)

Die Funktionalmatrix ist

$$\begin{pmatrix} \frac{\partial h_1}{\partial r} & \frac{\partial h_1}{\partial \varphi} \\ \frac{\partial h_2}{\partial r} & \frac{\partial h_2}{\partial \varphi} \end{pmatrix} = \begin{pmatrix} \cos \varphi & -r \sin \varphi \\ \sin \varphi & r \cos \varphi \end{pmatrix}.$$

Die Funktionaldeterminante ist r , d.h. die Funktionalmatrix ist im Fall $r > 0$ stets invertierbar. Also ist die Abbildung h lokal umkehrbar für $r > 0$.

Eine Umkehrabbildung gibt es beispielsweise dann, wenn man φ auf ein offenes Intervall der Länge 2π beschränkt, nicht aber für den gesamten Bereich $r > 0$. Die explizite Gestalt dieser Umkehrabbildung kann man leicht angeben (siehe z.B. Forster, [4], Kap. I, 6.8, Beispiel 8.4).

Als Anwendung der Sätze über implizite Funktionen betrachten wir noch die Bestimmung von **lokalen Extremwerten unter Nebenbedingungen**. Das entscheidende Hilfsmittel hierbei sind die sog. **Lagrangeschen Multiplikatoren**. Es gilt der

Satz 1.82. Sei $U \subseteq \mathbb{R}^n$ offen und seien $f : U \rightarrow \mathbb{R}$ und $f_i : U \rightarrow \mathbb{R}$ ($i = 1, \dots, r$) stetig differenzierbare Funktionen. Sei

$$M := \{x \in U : f_1(x) = \dots = f_r(x) = 0\}.$$

Ist $a \in M$ ein Punkt, an dem die Funktion f einen relativen Extremwert annimmt, so ist entweder der Rang der Funktionalmatrix

$$\left(\frac{\partial f_i}{\partial x_k}(a) \right)$$

kleiner als r oder: es gibt Konstanten λ_i ($i = 1, \dots, r$) mit der Eigenschaft, dass für die Funktion

$$F = f - \sum_{i=1}^r \lambda_i f_i$$

die Gleichungen $\frac{\partial F}{\partial x_k}(a) = 0$ gelten ($k = 1, \dots, n$).

Beweis. Ist $a \in M$ eine Stelle, an der f ein relatives Extremum annimmt und an der die Funktionalmatrix $\left(\frac{\partial f_i}{\partial x_k}(a) \right)$ den höchstmöglichen Rang r hat, so können wir uns die Veränderlichen so nummeriert denken, dass die mit x_1, \dots, x_r gebildete Determinante

$$\det \left(\frac{\partial f_i}{\partial x_k}(a) \right) \neq 0$$

ist. Die Gleichungen $f_i = 0$ können wir in der Nähe von a durch Funktionen

$$x_i = g_i(x_{r+1}, \dots, x_n)$$

auflösen und haben also in a einen lokalen Extremwert der Funktion

$$f(g_1, \dots, g_r, x_{r+1}, \dots, x_n).$$

Damit gilt dort

$$\sum_{j=1}^r \frac{\partial f}{\partial x_j} \frac{\partial g_j}{\partial x_k} + \frac{\partial f}{\partial x_k} = 0 \quad \text{für } k = r+1, \dots, n. \quad (1.24)$$

Wegen $f_i(g_1, \dots, g_r, x_{r+1}, \dots, x_n) = 0$ ist andererseits

$$\sum_{j=1}^r \frac{\partial f_i}{\partial x_j} \frac{\partial g_j}{\partial x_k} + \frac{\partial f_i}{\partial x_k} = 0 \quad \text{für } i = 1, \dots, r, \quad k = r+1, \dots, n. \quad (1.25)$$

Weil in der Nähe von a stets $\det \begin{pmatrix} f_i \\ x_j \end{pmatrix} \neq 0$ gilt, gibt es $\lambda_1, \dots, \lambda_r \in \mathbb{R}$ mit

$$\frac{\partial f}{\partial x_j}(a) = \sum_{i=1}^r \lambda_i \frac{\partial f_i}{\partial x_j}(a) \quad \text{für } j = 1, \dots, r. \quad (1.26)$$

Multipliziert man (1.25) mit λ_i und summiert über i , so ergibt sich

$$\sum_{j=1}^r \left(\sum_{i=1}^r \lambda_i \frac{\partial f_i}{\partial x_j} \right) \frac{\partial g_j}{\partial x_k} + \sum_{i=1}^r \frac{\partial A_i f_i}{\partial x_k} = 0,$$

und andererseits gilt nach (1.24)

$$\sum_{j=1}^r \frac{\partial f}{\partial x_j} \frac{\partial g_j}{\partial x_k} + \frac{\partial f}{\partial x_k} = 0 \quad \text{für } k = r+1, \dots, n.$$

Wir subtrahieren unsere letzte Gleichung von der vorletzten. Dies ergibt zusammen mit (1.26)

$$\frac{\partial}{\partial x_k} \left(f - \sum_{i=1}^r \lambda_i f_i \right) = 0 \quad \text{für } k = r+1, \dots, n.$$

Für $k = 1, \dots, r$ gilt diese Gleichung nach Definition der λ_i ebenso. Dies beendet den Beweis von Satz (1.82). \square

Beispiel 1.83. Sei $A = (a_{ik}) \in M_n(\mathbb{R})$ eine symmetrische $n \times n$ -Matrix („symmetrisch“ bedeutet $a_{ki} = a_{ik}$). Sei $f : \mathbb{R}^n \rightarrow \mathbb{R}$ die zugehörige quadratische Form $f(x_1, \dots, x_n) = \sum_{i,k=1}^n a_{ik} x_i x_k$. Was sind die Extrema der Form f auf der Kugeloberfläche $x_1^2 + \dots + x_n^2 = 1$?

Da $M = S^{n-1} := \{x \in \mathbb{R}^n : |x| = 1\}$ kompakt ist und f stetig ist, nimmt f auf M das absolute Maximum und Minimum an. Die Funktionalmatrix ist wie folgt gegeben:

$$\frac{\partial(x_1^2 + \dots + x_n^2)}{\partial x_k} = 2x_k$$

und hat daher die Gestalt $(2x_1, \dots, 2x_n)$. Ihr Rang ist auf M stets 1. Mit $F(x_1, \dots, x_n) := \sum_{i,k=1}^n a_{ik} x_i x_k - \lambda (\sum_{i=1}^n x_i^2 - 1)$ erhält man nach Satz (1.82) die folgende Bedingung:

$$\sum_{k=1}^n a_{ik} x_k - \lambda x_i = 0 \quad \text{für } i = 1, \dots, n. \quad (1.27)$$

In Matrixform lautet dieses homogene lineare Gleichungssystem

$$(A - \lambda E_n) \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = 0.$$

Eine von $(0,0,\dots,0)$ verschiedene Lösung existiert genau dann, wenn die Determinante

$$\det(A - \lambda E_n) = \begin{vmatrix} a_{11} - \lambda & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} - \lambda & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \vdots & a_{nn} - \lambda \end{vmatrix}$$

gleich Null ist. Dies ist eine Polynomgleichung vom Grad n für λ . Sie hat höchstens n verschiedene reelle Lösungen.

Also: Ein $a \in M$, für das f extremal wird, ist ein Eigenvektor der Matrix A , und der Lagrangesche Multiplikator λ ist der zugehörige Eigenwert.

Wegen

$$f(a) = (a_1, \dots, a_n) \cdot A \cdot \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} = \langle a, Aa \rangle = \lambda \cdot |a| = \lambda$$

wird das Maximum von f auf M bei einem Eigenvektor zum größten Eigenwert angenommen, das Minimum bei einem Eigenvektor zum kleinsten Eigenwert.

Da die Existenz des Maximums (bzw. Minimums) schon oben nachgewiesen wurde, ergibt sich so auch ein Beweis dafür, dass jede reelle symmetrische Matrix $A = (a_{ik})$ mindestens einen reellen Eigenwert hat. In der linearen Algebra zeigt man mittels Induktion, dass alle Eigenwerte reell sind.

Kapitel 2

Integralrechnung von Funktionen mehrerer Veränderlicher

2.1 Jordanscher Inhalt und Lebesguesches Maß

Wir beginnen unsere Betrachtungen damit, dass wir versuchen, möglichst vielen Teilmengen $A \subseteq \mathbb{R}^n$ in anschaulicher Weise ein sogenanntes **Maß** zuzuordnen (das Lebesgue-Maß). Dies dient uns dann im nächsten Paragraphen als Grundlage für die Integration von Funktionen $f : A \rightarrow \mathbb{R}^m$, wobei $A \subseteq \mathbb{R}^n$ eine messbare Menge ist.

Wir definieren zunächst den elementareren Begriff des **Inhalts**. Sei stets $\mathfrak{M} \subseteq \mathcal{P}(\mathbb{R}^n)$, d.h. \mathfrak{M} sei eine Menge von Teilmengen des \mathbb{R}^n . Sei $\mathbb{R}_+ = \{x \in \mathbb{R} : x \geq 0\}$.

Definition 2.1. Eine Mengenfunktion $m : \mathfrak{M} \rightarrow \mathbb{R}_+ \cup \{\infty\}$ mit den Eigenschaften

$$(M1) \quad A, B \in \mathfrak{M} \Rightarrow A \cup B, A \cap B, A \setminus B = \{x \in A : x \notin B\} \in \mathfrak{M}$$

$$(M2) \quad A \cap B = \emptyset \Rightarrow m(A \cup B) = m(A) + m(B)$$

heißt **Inhalt** (auf \mathfrak{M}).

Folgerung 2.2.

1. Sind die Mengen A_1, \dots, A_k Elemente von \mathfrak{M} mit $A_i \cap A_j = \emptyset$ für $i \neq j$, so gilt

$$m \left(\bigcup_{i=1}^k A_i \right) = \sum_{i=1}^k m(A_i).$$

2. Für beliebige $A, B \in \mathfrak{M}$ gilt $m(A \cup B) + m(A \cap B) = m(A) + m(B)$.
3. $m(A \cup B) \leq m(A) + m(B)$.
4. $m(A) \leq m(B)$, falls $A \subseteq B$.

Definition 2.3. Ein Inhalt m auf \mathfrak{M} heißt **Maß**, wenn anstelle von (M2) folgende schärfere Bedingung gilt:

(M3) Ist A_1, A_2, \dots eine Folge paarweise disjunkter Mengen aus \mathfrak{M} , so ist $\cup_{i=1}^{\infty} A_i \in \mathfrak{M}$ und es gilt $m(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} m(A_i)$.

Bemerkung 2.4. (M3) ist äquivalent zu

(M3') Ist B_1, B_2, \dots eine aufsteigende Folge von Mengen aus \mathfrak{M} , d.h. $B_i \subseteq B_{i+1}$ für $i \in \mathbb{N}$, so ist $\cup_{i=1}^{\infty} B_i \in \mathfrak{M}$ und es gilt:

$$m(\cup_{i=1}^{\infty} B_i) = \lim_{n \rightarrow \infty} m(B_n).$$

Folgerung 2.5 (aus M3). Sei A_1, A_2, \dots eine Folge beliebiger Mengen aus \mathfrak{M} . Dann gilt

$$\cup_{i=1}^{\infty} A_i \in \mathfrak{M} \quad \text{und} \quad m(\cup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} m(A_i).$$

Folgerung 2.6 (Zu (M3') duale Aussage). Sei A_1, A_2, \dots eine absteigende Folge (d.h. $A_1 \supseteq A_2 \supseteq A_3 \supseteq \dots$) von Mengen aus \mathfrak{M} , mit $m(A_k) < \infty$ für alle k . Dann gehört $\cap_{k=1}^{\infty} A_k$ zu \mathfrak{M} , und es gilt

$$m\left(\bigcap_{k=1}^{\infty} A_k\right) = \lim_{k \rightarrow \infty} m(A_k).$$

Bemerkung 2.7. Die Bedingung $m(A_k) < \infty$ ist wesentlich, wie etwa das Beispiel $A_k = \{x \in \mathbb{R} : x \geq k\} \subseteq \mathbb{R}^1$ zeigt.

Wir werden nun an das Lebesgue-Maß eine zusätzliche Forderung (M4) stellen, und es wird sich zeigen, dass diese bei geeigneter Menge \mathfrak{M} ein Maß eindeutig definiert.

Gegeben seien reelle Zahlen $a_i < b_i$ für $i = 1, \dots, n$. Unter einem **offenen Quader** verstehen wir eine Menge

$$Q_1 = \{x : a_i < x_i < b_i \text{ für } i = 1, \dots, n\}$$

und unter einem **abgeschlossenen Quader** eine Menge der Form $Q_2 = \overline{Q_1}$. Jede Menge Q mit $Q_1 \subseteq Q \subseteq Q_2$ heie **Quader**, insbesondere nennen wir

$$Q = Q(a_i, b_i) = \{x : a_i \leq x_i < b_i \quad \text{fur } i = 1, \dots, n\}$$

einen **halboffenen Quader**.

Forderung:

(M4) Ist $Q = Q(a_i, b_i)$ ein halboffener Quader, so ist $Q \in \mathfrak{M}$ und es gilt $m(Q) = \prod_{i=1}^n (b_i - a_i)$.

Problem: Existieren m und \mathfrak{M} , so dass (M1) bis (M4) gelten?

Wir setzen zunchst

$$\mathfrak{M}_0 := \{\text{Vereinigungen endlich vieler halboffener Quader}\}.$$

Diese Menge \mathfrak{M}_0 erfllt fur $A, B \in \mathfrak{M}_0$ offensichtlich $A \cup B \in \mathfrak{M}_0$. Ebenso ist fur $A, B \in \mathfrak{M}_0$ auch (mit Quadern Q_i, R_j)

$$A \cap B = (\cup_{i=1}^r Q_i) \cap (\cup_{j=1}^s R_j) = \cup_{i,j} (Q_i \cap R_j) \in \mathfrak{M}_0,$$

da $Q_i \cap R_j$ ein Quader oder die leere Menge ist. Es bleibt zu zeigen, dass mit $A, B \in \mathfrak{M}_0$ auch $A \setminus B \in \mathfrak{M}_0$ gilt.

Hilfssatz 2.8. *Jede endliche Vereinigung $A = \cup_{i=1}^r Q_i$ von halboffenen Quadern ist darstellbar als endliche Vereinigung von paarweise punktfremden Quadern.*

Mit dem Satz gilt somit: Zu $A = \cup_{i=1}^r Q_i$ und $B = \cup_{j=1}^s R_j$ gibt es endlich viele paarweise punktfremde Quader $\{S_k\}$ ($k \in I$) derart, dass $A = \cup_{k \in K} S_k$ und $B = \cup_{n \in L} S_n$ gilt mit gewissen Teilmengen K, L der Indexmenge I . Daraus ergibt sich

$$A \setminus B = \bigcup_{k \in K \setminus L} S_k,$$

und somit liegt $A \setminus B$ in \mathfrak{M}_0 , was zu zeigen war.

Definition 2.9.

- Fur einen halboffenen Quader $Q = Q(a_k, b_k)$ sei $m(Q) = \prod_{k=1}^n (b_k - a_k)$.
- Ist $A = \cup_{i=1}^r Q_i$ eine Vereinigung punktfremder Quader Q_i , so sei $m(A) = \sum_{i=1}^r m(Q_i)$.

Hier ist zu zeigen, dass die Definition (b) nicht von der Wahl der Zerlegung von A in punktfremde Quader abhängt. D.h.: ist $A = \cup_{i=1}^r Q_i = \cup_{j=1}^s R_j$ mit paarweise disjunkten Quadern Q_i bzw. R_j , so ist zu zeigen

$$\sum_{i=1}^r m(Q_i) = \sum_{j=1}^s m(R_j).$$

Hat man dies gezeigt, so hat man: Die Einschränkung von m auf \mathfrak{M}_0 erfüllt (M1), (M2) und (M4). Die Eigenschaft (M3) ist nicht erfüllt, da nicht jede Vereinigung abzählbar vieler Quader in \mathfrak{M}_0 liegt.

Daher definieren wir nun

$$\mathfrak{M}_1 := \{\text{Vereinigungen abzählbar vieler halboffener Quader}\}.$$

Es ist leicht zu sehen, dass \mathfrak{M}_1 zugleich die Menge aller Vereinigungen aufsteigender Folgen aus \mathfrak{M}_0 ist. Für $A, B \in \mathfrak{M}_1$ gilt stets $A \cup B \in \mathfrak{M}_1$ wie auch $A \cap B \in \mathfrak{M}_1$. Hingegen gehört $A \setminus B$ im allgemeinen nicht zu \mathfrak{M}_1 (Beispiel siehe unten), d.h. \mathfrak{M}_1 erfüllt (M1) nicht vollständig.

Hilfssatz 2.10. *Jede offene Menge gehört zu \mathfrak{M}_1 .*

Beweis. klar, Übungsaufgabe! □

Insbesondere gehören alle **offenen** Kreisscheiben zu \mathfrak{M}_1 . Ist aber beispielsweise $A = K(0, 1)$ und $B = K(0, 2)$, so liegt $B \setminus A$ nicht in \mathfrak{M}_1 , da der Kreisrand überabzählbar ist. Wir erweitern nun m auf \mathfrak{M}_1 :

Definition 2.11. *Sei A_1, A_2, \dots eine aufsteigende Folge von Mengen aus \mathfrak{M}_0 und $A = \cup_{i=1}^{\infty} A_i$. Dann sei*

$$m(A) := \lim_{i \rightarrow \infty} m(A_i).$$

Dieser Grenzwert existiert oder ist $+\infty$, da $(m(A_i))$ eine monoton steigende Folge ist. Wieder ist zu zeigen, dass die Definition (2.11) sinnvoll ist, also nicht von der Wahl der gegen A aufsteigenden Folge abhängt.

Wenn man dies gezeigt hat, lässt sich weiter zeigen, dass die Mengenfunktion m auf \mathfrak{M}_1 die Bedingungen (M2) und (M3) erfüllt. Weiter gelten, obwohl (M1) nicht vollständig erfüllt ist, die zu Beginn als Folgerungen festgehaltenen Regeln

1. A_1, \dots, A_k paarweise disjunkt $\Rightarrow m(\cup_{i=1}^k A_i) = \sum_{i=1}^k m(A_i)$.
2. $m(A \cup B) + m(A \cap B) = m(A) + m(B)$.

$$3. m(A \cup B) \leq m(A) + m(B).$$

$$4. m(A) \leq m(B), \text{ falls } A \subseteq B.$$

Bemerkung. Obwohl $m|\mathfrak{M}_1$ die meisten geforderten Eigenschaften hat, ist (M1) nicht vollständig erfüllt, und vielen anschaulichen Mengen ist auf diese Weise noch kein Inhalt zugeordnet — beispielsweise allen abgeschlossenen Kreisscheiben. Idee: man versucht, solche Mengen durch Mengen aus \mathfrak{M}_1 zu approximieren.

Definition 2.12. Für $A \subseteq \mathbb{R}^n$ heißt

$$\bar{m}(A) := \inf\{m(A') : A' \in \mathfrak{M}_1 \text{ und } A \subseteq A'\}$$

äußeres Lebesguesches Maß und

$$\bar{i}(A) = \inf\{m(A') : A' \in \mathfrak{M}_0 \text{ und } A \subseteq A'\}$$

äußerer Jordanscher Inhalt. In beiden Fällen lassen wir beliebige Teilmengen A des \mathbb{R}^n zu.

Allerdings kann man für die Elemente von \mathfrak{M} nicht alle Teilmengen des \mathbb{R}^n zulassen, da man in diesem Fall zeigen kann, dass (M2) nicht gilt.

Definition 2.13. A heißt **messbar** im Sinne von Lebesgue (oder kurz **L-messbar** bzw. **messbar**), d.h. $A \in \mathfrak{M}$, wenn es zu jedem $\varepsilon > 0$ Mengen A' und A'' aus \mathfrak{M}_1 gibt mit

$$A \subseteq A', \quad A' \setminus A \subseteq A'', \quad m(A'') < \varepsilon.$$

Ist A messbar, so heißt $m(A) := \bar{m}(A)$ das **Lebesgue-Maß** (L -Maß) von A .

Definition 2.14. A heißt **messbar im Sinne von Jordan** (kurz **J-messbar**), wenn es zu jedem $\varepsilon > 0$ Mengen A' und A'' aus \mathfrak{M}_0 gibt mit $A \subseteq A'$, $A' \setminus A \subseteq A''$ und $m(A'') < \varepsilon$. Ist A J-messbar, so heißt $i(A) := \bar{i}(A)$ der **Jordansche Inhalt** von A .

Bemerkung 2.15. Eine Menge $A \subseteq \mathbb{R}^n$ ist genau dann J-messbar, wenn sie sich von innen und außen durch endliche Quadersummen approximieren läßt, d.h. wenn zu $\varepsilon > 0$ Mengen $B, C \in \mathfrak{M}_0$ existieren mit $B \subseteq A \subseteq C \Rightarrow m(C \setminus B) \leq \varepsilon$. (Denn man kann $C = A'$, $B = A' \setminus A''$ wählen.)

Bemerkung 2.16. Es existieren keine hierzu analogen Aussagen (Approximation mit abzählbaren Quadersummen) für L-messbare Mengen. Grund: $A' \setminus A''$ liegt im allgemeinen nicht in \mathfrak{M}_1 .

Man kann nun zeigen: Das durch (2.13) definierte Lebesgue-Maß $m|\mathfrak{M}$ erfüllt alle Eigenschaften (M1)-(M4). Das Ziel dieses Paragraphen ist dann erreicht. Analog gilt: Der oben definierte Jordansche Inhalt i erfüllt (M1), (M2) und (M4).

Weiter gilt

Behauptung 2.17. *Jede kompakte Menge ist messbar und hat endliches Maß.*

Behauptung 2.18. *Jede Menge A mit $\overline{m}(A) = 0$ ist messbar. Eine solche Menge heißt Nullmenge.*

Jede Teilmenge einer Nullmenge ist Nullmenge, und jede abzählbare Vereinigung von Nullmengen ist wieder eine Nullmenge. Nullmengen sind insbesondere alle einpunktigen Mengen und damit auch alle abzählbaren Mengen. Daraus folgt

Bemerkung 2.19. Die Punkte des Einheitsintervalls (und damit auch die einer Kreislinie) sind nicht abzählbar. Sonst hätte das Einheitsintervall im \mathbb{R}^1 das Maß 0, was einen Widerspruch zu (M4) bildet.

Was sind die Mengen aus \mathfrak{M} mit endlichem Maß?

Satz 2.20. *Die Menge A gehört zu \mathfrak{M} und hat endliches Maß genau dann, wenn es zu jedem $\varepsilon > 0$ eine Menge A_0 aus \mathfrak{M}_0 gibt mit*

$$\overline{m}((A_0 \setminus A) \cup (A \setminus A_0)) \leq \varepsilon.$$

2.2 Das Lebesguesche Integral

Wir gehen ähnlich wie bei der Definition des Riemanschen Integrals im \mathbb{R}^1 vor. Stets sei $A \subseteq \mathbb{R}^n$ und $f : A \rightarrow \mathbb{R}^m$ (wobei alle Zielmengen \mathbb{R} , \mathbb{C} , \mathbb{R}^k , \mathbb{C}^k zugelassen sind).

Die **Schwankung** $s(f, A)$ von f auf A sei definiert durch

$$s(f, A) := \sup_{x, y \in A} |f(y) - f(x)|.$$

Ein Mengensystem $\{A_i\} = \mathcal{Z}$ heißt **Zerlegung** von A , wenn A die *disjunkte* Vereinigung $A = \cup_{i=1}^{\infty} A_i$ ist mit $A_i \in \mathfrak{M}$ und $m(A_i) < \infty$. Gilt jeweils $s(f, A_i) < \infty$, so heißt

$$S(f, \mathcal{Z}) = \sum_{i=1}^{\infty} s(f, A_i) m(A_i)$$

die **Schwankungssumme** von f bezüglich \mathcal{Z} .

Hilfssatz 2.21. Sei A messbar. Folgende Aussagen sind äquivalent:

- (i) Zu jedem $\varepsilon > 0$ existiert eine Zerlegung \mathcal{Z} von A mit $S(f, \mathcal{Z}) \leq \varepsilon$.
- (ii) Für jede offene Teilmenge $C \subseteq \mathbb{R}^m$ ist $f^{-1}(C) = \{x \in \mathbb{R}^n : f(x) \in C\}$ messbar.

Definition 2.22. Eine Funktion f heißt auf A **messbar**, wenn A messbar ist und die äquivalenten Bedingungen (i), (ii) von (2.21) erfüllt sind.

Sei nun f auf A messbar und $\mathcal{Z} = \{A_i\}$ eine Zerlegung von A , so dass jedes A_i ein endliches Maß hat. Wähle nun jeweils ein $x_i \in A_i$ und setze

$$R(f, \mathcal{Z}) = \sum_{i=1}^{\infty} f(x_i)m(A_i).$$

Problem: Diese Summe konvergiert nicht für jedes f und beliebige Zerlegungen \mathcal{Z} von A .

Hilfssatz 2.23. Sei f auf A messbar und $\mathcal{Z} = \{A_i\}$ eine Zerlegung von A mit $S(f, \mathcal{Z}) \leq \varepsilon$. In jedem A_i seien zwei Punkte x_i, x'_i gewählt. Wenn von den beiden Reihen

$$\sum_{i=1}^{\infty} f(x_i)m(A_i), \quad \sum_{i=1}^{\infty} f(x'_i)m(A_i)$$

eine absolut konvergiert, so auch die andere, und es gilt

$$\left| \sum_{i=1}^{\infty} f(x_i)m(A_i) - \sum_{i=1}^{\infty} f(x'_i)m(A_i) \right| \leq \varepsilon.$$

Definition 2.24. Eine auf A messbare Funktion f heißt **über A integrierbar**, wenn für jede Zerlegung $\mathcal{Z} = \{A_i\}$ von A mit $S(f, \mathcal{Z}) < \infty$ die Reihe $R(f, \mathcal{Z}) = \sum_{i=1}^{\infty} f(x_i)m(A_i)$ mit $x_i \in A_i$ absolut konvergiert.

Definition 2.25. Die Zerlegung $\mathcal{Z}' = \{B_j\}$ von A heißt eine **Verfeinerung** der Zerlegung $\mathcal{Z} = \{A_i\}$ von A (oder **feiner als \mathcal{Z}**), wenn jedes B_j in einem der A_i enthalten ist.

Hilfssatz 2.26. Ist die Zerlegung $\mathcal{Z}' = \{B_j\}$ von A feiner als $\mathcal{Z} = \{A_j\}$, dann gilt $S(f, \mathcal{Z}') \leq S(f, \mathcal{Z})$.

Hilfssatz 2.27. Sei f auf A messbar und seien $\mathcal{Z} = \{A_i\}$, $\mathcal{Z}' = \{B_j\}$ zwei Zerlegungen von A mit $S(f, \mathcal{Z}) < \infty$ und $S(f, \mathcal{Z}') < \infty$. Konvergiert von zwei Näherungssummen $R(f, \mathcal{Z})$ und $R(f, \mathcal{Z}')$ die eine absolut, so auch die andere, und es gilt

$$|R(f, \mathcal{Z}) - R(f, \mathcal{Z}')| \leq S(f, \mathcal{Z}) + S(f, \mathcal{Z}').$$

Definition 2.28. Sei f über A integrierbar und $\mathcal{Z}_1, \mathcal{Z}_2, \dots$ eine Folge von Zerlegungen von A mit $\lim_{i \rightarrow \infty} S(f, \mathcal{Z}_i) = 0$. Dann heißt

$$\int_A f(x) dx = \int_A f(x_1, \dots, x_n) dx_1 \dots dx_n := \lim_{i \rightarrow \infty} R(f_i, \mathcal{Z}_i)$$

das **Lebesguesche Integral** von f über A .

Hierbei folgt aus (2.27), dass der Limes existiert und unabhängig von der Wahl der Zerlegungsfolge ist.

Integrationsregeln:

- (i) Seien f, g über A integrierbar und a, b Konstanten. Dann ist auch $af + bg$ über A integrierbar, und es gilt

$$\int_A af(x) + bg(x) dx = a \int_A f(x) dx + b \int_A g(x) dx.$$

- (ii) Seien A, B messbare Mengen mit $A \cap B = \emptyset$. Genau dann ist f über $A \cup B$ integrierbar, wenn f über A und über B integrierbar ist. Dann gilt

$$\int_{A \cup B} f(x) dx = \int_A f(x) dx + \int_B f(x) dx.$$

- (iii) Sind f und g über A integrierbar und $f \leq g$, so gilt

$$\int_A f(x) dx \leq \int_A g(x) dx.$$

Dies folgt aus der entsprechenden Ungleichung für Näherungssummen.

- (iv) Ist f auf A messbar (bzw. integrierbar), so ist auch $|f|$ auf A messbar (bzw. integrierbar).

Zum Beweis verwendet man, dass $s(|f|, B) \leq s(f, B)$ für beliebige $B \subseteq A$ gilt und dass die Definition der Integrierbarkeit auf absoluter Konvergenz der Näherungssummen beruht.

Bemerkung. Aus der Messbarkeit von $|f|$ folgt i.a. nicht die Messbarkeit von f .

(v) Ist f über A integrierbar, so gilt

$$\left| \int_A f(x) dx \right| \leq \int_A |f(x)| dx.$$

Dies folgt für reellwertige Funktionen aus (iii) und lässt sich im allgemeinen Fall mit Hilfe der Dreiecksungleichung zeigen.

(vi) Sei f auf A messbar und g über A integrierbar. Gilt $|f(x)| \leq g(x)$ auf A , so ist f über A integrierbar.

(vii) Sei A messbar mit $m(A) < \infty$. Dann ist jede konstante Funktion $f(x) \equiv c$ über A integrierbar und es gilt $\int_A f(x) dx = c \cdot m(A)$. (Wähle die Zerlegung $\mathcal{Z} = \{A\}$; dann ist $S(f, \mathcal{Z}) = 0$.)

Satz 2.29. *Jede auf einer messbaren Menge A stetige Funktion ist messbar.*

Zum Beweis verwende man (2.21) und die Tatsache, dass stetige Urbilder offener Mengen offen sind.

Satz 2.30. *Ist A messbar mit $m(A) < \infty$ und f auf A messbar und beschränkt, so ist f über A integrierbar.*

Satz 2.31. *Ist A messbar mit $m(A) < \infty$ und ist f stetig und beschränkt auf A , so ist f über A integrierbar.*

Speziell gilt: Ist A kompakt, so ist A auch messbar mit $m(A) < \infty$, und jede stetige Funktion f auf A ist beschränkt. Daraus folgt

Satz 2.32. *Eine auf einer kompakten Menge A stetige Funktion ist über A integrierbar.*

Bemerkung 2.33. Ist f über $[a, b]$ Riemann-integrierbar (im eigentlichen Sinn), so ist f über $[a, b]$ auch Lebesgue-integrierbar und es gilt

$$\int_{[a,b]} f(x) dx = \int_a^b f(x) dx.$$

Aber: es ist möglich, dass das uneigentliche Riemann-Integral existiert, ohne dass das entsprechende Lebesgue-Integral existiert. Ein Beispiel ist

$$\int_1^\infty \frac{\sin x}{x} dx.$$

2.3 Grenzwertsätze

Wir interessieren uns für die Frage, in welchen Fällen die Limesbildung mit dem Lebesgue-Integral vertauschbar ist, ohne dabei gleichmäßige Konvergenz vorauszusetzen.

Satz 2.34. *Ist f_1, f_2, \dots eine Folge von über A messbaren Funktionen $f_i : A \rightarrow \mathbb{R}^m$ und existiert für alle $x \in A$ der Grenzwert $\lim_{n \rightarrow \infty} f_n(x) =: f(x)$, so ist f über A messbar.*

Zum Beweis ist zu zeigen: Ist $C \subseteq \mathbb{R}^n$ offen, so ist $f^{-1}(C)$ messbar. Urbilder offener Mengen kann man nach oben durch messbare Mengen abschätzen, Urbilder abgeschlossener Mengen sind entsprechend nach unten durch messbare Mengen abschätzbar. Das wird benutzt, um durch einen Kunstgriff $f^{-1}(C)$ selbst durch Vereinigungen bzw. Durchschnitte messbarer Mengen darzustellen. Man setzt

$$C_n := \{x \in C : \exists r > 1/n \text{ mit } K(x, r) \subseteq C\}$$

und zeigt als erste Feststellung, dass alle C_n offen sind. Als zweite Feststellung ergibt sich sodann $C = \bigcup_{n=1}^{\infty} C_n$, als dritte Feststellung dann $\overline{C_n} =: D_n \subseteq C_{n+1}$. Damit folgt $f^{-1}(C) = \bigcup_{n=1}^{\infty} f^{-1}(C_n) \subseteq \bigcup_{n=1}^{\infty} f^{-1}(D_n) \subseteq \bigcup_{n=1}^{\infty} f_n^{-1}(C_{n+1}) = f^{-1}(C)$.

Wir führen nun folgende Sprechweise ein: Eine Eigenschaft $E(x)$ gilt für **fast alle** x aus A , bzw. **fast überall** in A , wenn es eine Nullmenge $B \subseteq A$ gibt, so dass $E(x)$ für alle $x \in A \setminus B$ gilt. Ein Beispiel: Seien f, g Funktionen $A \rightarrow \mathbb{R}^m$ und sei $E(x)$ die Eigenschaft $f(x) = g(x)$. Gilt $E(x)$ für fast alle $x \in A$, so folgt

$$\int_A f(x) dx = \int_A g(x) dx,$$

falls eines der beiden Integrale existiert. Dies folgt aus

Behauptung 2.35. *Ist B eine Nullmenge und f eine beliebige Funktion auf B , so ist f über B integrierbar und $\int_B f(x) dx = 0$.*

Hier nun (in diesem Exzerpt ohne Beweis) zwei wichtige und oft gebrauchte Grenzwertsätze:

Satz 2.36 (B. Levi, Satz von der monotonen Konvergenz). *Sei A messbar, f_1, f_2, \dots auf A definierte reelle Funktionen mit $0 \leq f_1 \leq f_2 \leq \dots$. Die f_n seien über A integrierbar mit $\int_A f_n(x) dx < s$. Dann ist $\{f_n(x)\}$ für fast*

alle $x \in A$ beschränkt, die Grenzfunktion $f(x) = \lim_{n \rightarrow \infty} f_n(x)$ ist über A integrierbar, und es gilt:

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

Bemerkung 2.37. Die Voraussetzung $f_1(x) \geq 0$ ist überflüssig, da man sonst f_i durch $f_i - f_1$ ersetzen kann. Ebenso ist der Satz richtig für monoton fallende Funktionenfolgen $f_i \leq f_{i-1}$, denn man kann f_i durch $-f_i$ ersetzen.

Satz 2.38 (H. Lebesgue, Satz von der majorisierten Konvergenz). *Die Funktionen f_i ($i \in \mathbb{N}$) und g seien über A integrierbar mit $|f_i(x)| \leq g(x)$ für alle $x \in A$. Für fast alle $x \in A$ konvergiere $(f_n(x))$ gegen den Grenzwert $f(x)$. Dann ist f über A integrierbar, und es gilt*

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

Anwendungen der Sätze von Levi und Lebesgue:

Satz 2.39. *Sei $f(x, y)$ definiert für $x \in A \subseteq \mathbb{R}^n$, $y \in B \subseteq \mathbb{R}^m$, und es existiere $\lim_{y \rightarrow y_0} f(x, y) = h(x)$ für alle $x \in A$. $f(x, y)$ sei für jedes feste y integrierbar über A , und für alle $x \in A$ und alle $y \in B$ gelte $|f(x, y)| \leq g(x)$ mit einer über A integrierbaren Funktion $g(x)$. Dann ist $h(x)$ über A integrierbar und es gilt*

$$\int_A h(x) dx = \lim_{y \rightarrow y_0} \int_A f(x, y) dx.$$

Beispiel 2.40. Wir betrachten die Gammafunktion

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt, \quad 0 < x < \infty.$$

Behauptung 2.41. Γ ist stetig und sogar differenzierbar.

Beweis. Dazu wählen wir $0 < a < b < \infty$. Für $x \in [a, b]$ ist dann

$$|t^{x-1} e^{-t}| \leq e^{-t} \cdot (t^{a-1} + t^{b-1}) =: g(t).$$

Für beliebige $t \in [0, \infty)$ ist der Integrand stetig in x , und aus (2.39) folgt daher die Stetigkeit von $\Gamma(x)$ auf $[a, b]$ und damit in $(0, \infty)$. Zum Beweis der Differenzierbarkeit:

$$\left| \frac{t^{y-1} - t^{x-1}}{y - x} \cdot e^{-t} \right| \leq (t^{a-1} + t^{b-1}) e^{-t} \cdot \log t =: g(t).$$

□

Satz 2.42. Ist $A \in \mathfrak{M}$ und $A = \cup_{i=1}^{\infty} A_i$ die disjunkte Vereinigung von $A_i \in \mathfrak{M}$, so ist f genau dann über A integrierbar, wenn f für jedes n über A_n integrierbar ist und $\sum_{n=1}^{\infty} \int_{A_n} |f(x)| dx$ konvergiert. Es gilt dann

$$\int_A f(x) dx = \sum_{n=1}^{\infty} \int_{A_n} f(x) dx.$$

Wir interessieren uns nun für den Zusammenhang des Lebesgue-Integrals mit dem uneigentlichen (Riemann-)Integral. Sei $f : [a, \infty) \rightarrow \mathbb{R}$ über $[a, b]$ eigentlich integrierbar für jedes b . Dann gilt: f ist über $[a, \infty)$ L-integrierbar genau dann, wenn $\int_a^{\infty} f(x) dx$ absolut konvergiert. In diesem Fall ist

$$\int_{[a, \infty)} f(x) dx = \int_a^{\infty} f(x) dx.$$

Die analoge Aussage gilt für nach links unbeschränkte Intervalle, ebenso für beschränkte Intervalle und unbeschränkte Funktionen f .

Aus Satz (2.42) folgt nun: Das Lebesgue-Integral $\int_{[a, \infty)} f(x) dx$ existiert und ist gleich $\int_a^{\infty} f(x) dx$. Daher auch die bei der Gammafunktion verwendete Schreibweise $\int_0^{\infty} f(x) dx$.

Bemerkung 2.43. Falls $\int_a^{\infty} f(x) dx$ existiert im Riemannschen Sinne, nicht aber $\int_a^{\infty} |f(x)| dx$, so existiert das Lebesgue-Integral $\int_{[a, \infty)} f(x) dx$ nicht. Aber: Für endliche Intervalle liefert das Lebesgue-Integral mehr als das eigentliche Riemann-Integral (vgl. (2.33)).

2.4 Der Satz von Fubini und das Cavalierische Prinzip

Sei zunächst

$$Q = \{x : a_i \leq x_i \leq b_i \text{ für } i = 1, 2\} \subseteq \mathbb{R}^2$$

ein Quader im \mathbb{R}^2 . Dann läßt sich Q als Produkt $I_1 \times I_2$ von reellen Intervallen I_1, I_2 schreiben. Das Maß von Q ist dann das Produkt der Maße dieser beiden Intervalle.

Allgemeiner: Sei $Q \subseteq \mathbb{R}^n = \mathbb{R}^{r+s}$ ein Quader und sei

$$x = (y_1, \dots, y_r, z_1, \dots, z_s) = (y, z)$$

mit $y \in \mathbb{R}^r$ und $z \in \mathbb{R}^s$. Dann ist

$$Q = \{(y, z) : y \in Q', z \in Q''\}$$

mit Quadern $Q' \subseteq \mathbb{R}^r$, $Q'' \subseteq \mathbb{R}^s$. Wir haben

$$m(Q) = m_r(Q') \cdot m_s(Q''), \quad (2.1)$$

wobei m_r das Lebesgue-Maß im \mathbb{R}^r und m_s das Lebesgue-Maß im \mathbb{R}^s bezeichnet.

Definition 2.44. Sei $A \subseteq \mathbb{R}^n$ eine beliebige Punktmenge und $x \in \mathbb{R}^n$. Dann heißt

$$\chi_A(x) := \begin{cases} 1 & \text{falls } x \in A \\ 0 & \text{falls } x \notin A \end{cases}$$

die **charakteristische Funktion** oder **Indikatorfunktion** von A . Genau dann ist A messbar mit endlichem Maß, wenn χ_A integrierbar ist. Dann gilt

$$m(A) = \int \chi_A(x) dx.$$

Behauptung 2.45. (2.1) ist gleichbedeutend mit

$$\int \chi_A(x) dx = \int \left(\int \chi_A(y, z) dy \right) dz. \quad (2.2)$$

Zum Beweis verwendet man, dass $\int \chi_A(y, z) dy$ für fast alle $z \in \mathbb{R}^s$ definiert und integrierbar ist, wobei Integration die rechte Seite von (2.1) ergibt. Allgemeiner gilt:

Satz 2.46. Sei $f(x)$ integrierbar über \mathbb{R}^n . Dann existiert das Integral $\int f(y, z) dy$ für fast alle $z \in \mathbb{R}^s$, und es gilt

$$\int f(x) dx = \int \left(\int f(y, z) dy \right) dz.$$

Der **Beweis** erfolgt in mehreren Schritten. Zunächst ist zu zeigen, dass (2.46) für charakteristische Funktionen von Quadern im \mathbb{R}^n gilt. Anschließend zeigt man die Richtigkeit für Funktionen des Typs

$$f(x) = \sum_{i=1}^k a_i \chi_{Q_i}(x),$$

wobei diese Funktionen auch **Elementarfunktionen** genannt werden. Schließlich approximiert man dann beliebige Funktionen durch Elementarfunktionen. Hierzu ein

Hilfssatz 2.47. Sei f_1, f_2, \dots eine Folge von Elementarfunktionen, und die Reihe $\sum_{i=1}^{\infty} \int |f_i(x)| dx$ konvergiere. Dann gilt

- (a) Für fast alle $x \in \mathbb{R}^n$ konvergiert die Reihe $f(x) = \sum_{i=1}^{\infty} f_i(x)$ absolut.
 (b) Es gibt eine Nullmenge $B \subseteq \mathbb{R}^s$ und für jedes $z \notin B$ eine Nullmenge $C(z) \subseteq \mathbb{R}^n$ derart, dass

$$f(y, z) = \sum_{i=1}^{\infty} f_i(y, z) \quad \text{für } z \notin B, y \notin C(z)$$

absolut konvergiert und $\int f(x) dx = \int (\int f(y, z) dy) dz$ gilt.

Dieser Hilfssatz wird mittels (2.36) bewiesen. Aus ihm kann man zunächst folgern, dass der Satz von Fubini für Funktionen gilt, die außerhalb einer Nullmenge $A \subseteq \mathbb{R}^n$ verschwinden. Der letzte Schritt ist dann der Beweis folgender Aussage: Zu jeder über \mathbb{R}^n integrierbaren Funktion g existiert eine Folge von Elementarfunktionen f_i mit den Voraussetzungen von (2.47) und mit der Eigenschaft, dass $f = \sum_{i=1}^{\infty} f_i$ fast überall mit g übereinstimmt. Dann folgt: (2.46) gilt nach (2.47) für f und damit auch für $g - f$, also auch für g .

Hilfssatz 2.48. Sei f integrierbar und $\varepsilon > 0$. Dann existiert eine Elementarfunktion $h(x)$ mit

$$\int |f(x) - h(x)| dx \leq \varepsilon.$$

Zum Beweis verwendet man (2.20). Dies sichert die Existenz einer Folge $h_i(x)$ mit $\int |f(x) - h_i(x)| dx \rightarrow 0$ für $i \rightarrow \infty$. Insbesondere gilt, da man die Rollen von \mathbb{R}^r und \mathbb{R}^s vertauschen kann,

$$\int \int f(y, z) dy dz = \int \int f(y, z) dz dy.$$

Bemerkung 2.49.

- a) Aus der Existenz der iterierten Integrale folgt i.a. noch nicht die Integrierbarkeit von f , d.h. die Existenz von $\int f(x) dx$.
 b) Es gibt Beispiele nicht integrierbarer Funktionen, bei denen die iterierten Integrale existieren, aber voneinander verschieden sind!

Man kann nun durch mehrfache Anwendung des Satzes von Fubini die Integration im \mathbb{R}^n auf n einfache Integrationen zurückführen. Hierzu benutzt man z.B. im \mathbb{R}^2 die Schreibweise

$$\int f(x)dx = \int \int f(x_1, x_2)dx_1dx_2.$$

Es ist auch nicht erforderlich, dass f über ganz \mathbb{R}^n definiert ist, sondern man kann f als eine Funktion auf irgendeiner messbaren Teilmenge $A \subseteq \mathbb{R}^n$ betrachten. Das Integral schreibt sich dann als

$$\int_A f(x)dx,$$

oder, mit der Funktion

$$g(x) = \begin{cases} f(x) & \text{falls } x \in A \\ 0 & \text{falls } x \notin A \end{cases}$$

gilt $\int g(x)dx = \int_A f(x)dx$. Auf g kann man den Satz von Fubini anwenden und erhält

$$\int_A f(x)dx = \int g(x)dx = \int \int g(y, z)dydz.$$

Für $z \in \mathbb{R}^s$ betrachten wir nun die **Schnittmenge**

$$A_z = \{y \in \mathbb{R}^r : (y, z) \in A\}.$$

Dann ist leicht zu zeigen, dass A_z für fast alle z messbar als Teilmenge des \mathbb{R}^r ist. Ist $m(A) < \infty$, so zeigt man die Messbarkeit von $A_z = \{y : \chi_{A_z}(y) > 0\}$. Ist $m(A) = \infty$, so stellt man A_z als Vereinigung einer aufsteigenden Folge von Schnittmengen $(A_i)_z$ dar und folgert daraus, dass A_z für fast alle z messbar ist. Damit folgt

$$\int g(y, z)dy = \int_{A_z} g(y, z)dy = \int_{A_z} f(y, z)dy$$

und

$$\int_A f(x)dx = \int \int_{A_z} f(y, z)dydz = \int_B \int_{A_z} f(y, z)dydz. \quad (2.3)$$

Dabei ist B irgendeine messbare Teilmenge des \mathbb{R}^s mit

$$\{z : (y, z) \in A \text{ für mindestens ein } y\} \subseteq B.$$

Beispiel 2.50.

1. Sei

$$A = \{(y, z) : a(y) \leq z \leq b(y), c \leq y \leq d\} \subseteq \mathbb{R}^2$$

mit integrierbaren Funktionen a, b und sei f über A integrierbar. Dann gilt

$$\int_A \int f(y, z) dy dz = \int_c^d \int_{a(y)}^{b(y)} f(y, z) dz dy. \quad (2.4)$$

Ist speziell $f = \chi_A$, so erhält man

$$\int_A f(x) dx = m(A) = \int_c^d (b(y) - a(y)) dy, \quad (2.5)$$

und dies ist der Flächeninhalt von A (siehe auch die Vorlesung Analysis I).

2. Sei $A \subseteq \mathbb{R}^3$ messbar und $A_z = \{(x, y) : (x, y, z) \in A\}$. Dann gilt

$$\int_A f(x, y, z) dx dy dz = \int \left(\int_{A_z} f(x, y, z) dx dy \right) dz$$

und speziell für $f = \chi_A$

$$m_3(A) = \int m_2 A_z dz. \quad (2.6)$$

Daraus ergibt sich das

Cavalierische Prinzip: Sind A und B zwei messbare Mengen im \mathbb{R}^3 und gilt $m_2 A_z = m_2 B_z$ für alle z , so haben A und B das gleiche Maß.

Ist beispielsweise A der Rotationskörper

$$\{(x, y, z) : c \leq z \leq d, x^2 + y^2 \leq f(z)^2\},$$

so gilt $m_2 A_z = \pi(f(z))^2$ und wegen (2.6) daher

$$m_3(A) = \pi \int_c^d f(z)^2 dz. \quad (2.7)$$

Beispiel 2.51. Sei

$$A = \{(x, y, z) \in \mathbb{R}^3 : (x, y) \in B, a(x, y) \leq z \leq b(x, y)\}.$$

Der Satz von Fubini liefert

$$\int_A f(x, y, z) dx dy dz = \int_B \left(\int_{a(x,y)}^{b(x,y)} f(x, y, z) dz \right) dx dy. \quad (2.8)$$

Ist nun $f = \chi_A$, so gilt

$$m(A) = \int_B (b(x, y) - a(x, y)) dx dy. \quad (2.9)$$

Beispiel 2.52. Zu berechnen ist das iterierte Integral

$$\int e^{-(x_1^2 + \dots + x_n^2)} dx_1 \dots dx_n.$$

Man setzt zunächst

$$A = \{(x_0, x_1, \dots, x_n) : 0 < x_0 < e^{-(x_1^2 + \dots + x_n^2)} \subseteq \mathbb{R}^{n+1}\}$$

und

$$A_N := \{(x_0, \dots, x_n) \in A : |x_i| < N \text{ für } i = 1, \dots, n\}.$$

Alle A_N sind offen und beschränkt mit dem Maß

$$m(A_N) = \left(\int_{-N}^N e^{-x^2} dx \right)^n$$

(vgl. (2.8) und Satz von Fubini). Das uneigentliche Riemann-Integral

$$\int_{-\infty}^{\infty} e^{-x^2} dx =: F$$

konvergiert. Nun ist $m(A_N) \leq F^n$ und $m(A) = F^n$.

Andererseits kann man $m(A)$ wie folgt berechnen:

$$\int \left(\int \chi_A(x_0, x_1, \dots, x_n) dx_1 \dots dx_n \right) dx_0$$

wobei der Integrand für $0 < x_0 < 1$ das Volumen des Körpers ist, der durch $x_1^2 + \dots + x_n^2 < -\log x_0$ gegeben ist. D.h.

$$\int \chi_A(x_0, x_1, \dots, x_n) dx_1 \dots dx_n = I_n \cdot (-\log x_0)^{n/2}$$

mit I_n als Volumen der n -dimensionalen Einheitskugel. Es folgt

$$F^n = m(A) = I_n \cdot \int_0^1 (-\log x)^{n/2} dx,$$

und mit der Substitution $x = e^{-y}$ ergibt sich die rechte Seite zu

$$I_n \cdot \int_0^\infty y^{n/2} \cdot e^{-y} dy = I_n \cdot \Gamma\left(\frac{n}{2} + 1\right)$$

(mit der Gammafunktion $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$).

Ist nun n gerade, also $n = 2m$, so liefert $\Gamma(m + 1) = m!$ eine Beziehung zwischen I_{2m} und F . Etwa für $m = 1$ ist $I_2 = \pi$ und damit

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

Es folgt

$$I_{2m} = \frac{\pi^m}{m!}.$$

Ist andererseits $n = 2m + 1$, so ergibt sich mit $m = 0$

$$\Gamma(3/2) = \frac{1}{2}\sqrt{\pi}$$

und

$$\Gamma(1/2) = \int_0^\infty e^{-t} \frac{dt}{t} = \sqrt{\pi}.$$

Per Induktion erhält man

$$I_{2m+1} = \frac{2^{2m+1} \cdot \pi^m}{(m+1)(m+2)\dots(2m+1)}.$$

2.5 Transformation von Integralen

Wir erinnern zunächst an die Substitutionsregel im \mathbb{R}^1

$$\int_{g(a)}^{g(b)} f(y) dy = \int_a^b f(g(x)) \cdot g'(x) dx$$

etwa unter den Voraussetzungen, dass f stetig und g stetig differenzierbar ist. Falls weiter $g'(x) \neq 0$ auf dem Intervall $[a, b] = A$ gilt, ist

$$\int_{g(A)} f(y)dy = \int_A f(g(x)) \cdot |g'(x)|dx. \quad (2.10)$$

Wenn g' keine Nullstelle hat, ist g umkehrbar eindeutig. Wir wollen nun diese Überlegungen auf den allgemeinen Fall des \mathbb{R}^n übertragen.

Dazu verlangen wir die Ein-Eindeutigkeit von g und das Nichtverschwinden der Funktionaldeterminante der Abbildung $g : A \rightarrow \mathbb{R}^n$ ($A \subseteq \mathbb{R}^n$). Für f genügt als Voraussetzung die Integrierbarkeit, und so erhält man auch für den eindimensionalen Fall eine erhebliche Verallgemeinerung.

Satz 2.53. *Sei $A \subseteq \mathbb{R}^n$ offen, $g : A \rightarrow g(A) \subseteq \mathbb{R}^n$ eine umkehrbar eindeutige, stetig differenzierbare Abbildung, deren Funktionaldeterminante nirgends verschwindet, und f eine auf $g(A)$ definierte Funktion. Wenn eines der beiden Integrale*

$$\int_{g(A)} f(y)dy \quad \text{und} \quad \int_A f(g(x)) \left| \det \left(\frac{\partial g_i}{\partial x_j} \right) \right| dx$$

existiert, so auch das andere und beide Integrale sind gleich.

Zum **Beweis** wendet man für stetiges f Induktion nach n an und setzt dabei zunächst weitere Eigenschaften von g und A voraus (zunächst ist A ein offener Quader, g stetig differenzierbar auf ganz \bar{A}), die dann später schrittweise abgebaut werden. Das entscheidende Hilfsmittel in diesem Beweis ist der Satz von Fubini, ferner die eindimensionale Substitutionsregel. Um den Fall zu behandeln, dass f nicht stetig ist, wird noch benötigt:

Lemma 2.54. *Sei $g : A \rightarrow g(A)$ ($A, g(A) \subseteq \mathbb{R}^n$) eine umkehrbar eindeutige, stetig differenzierbare Abbildung, deren Funktionaldeterminante nirgendwo verschwindet.*

(i) *Ist B eine offene Teilmenge von A , so gilt*

$$m(g(B)) = \int_B \left| \det \left(\frac{\partial g_i}{\partial x_j} \right) \right| dx.$$

(ii) *Ist $C \subseteq A$ eine Nullmenge, so ist auch $g(C)$ eine Nullmenge.*

Damit kann nun gezeigt werden, dass (2.53) richtig ist für den Fall $f = \chi_Q$, $Q \subseteq A$ offen, und insbesondere für Quader Q . Der Satz gilt damit für Elementarfunktionen. Der letzte Schritt ist dann die Approximation der Funktion f durch eine Folge von Elementarfunktionen.

Zwei Spezialfälle von Satz (2.53):

1. Sei g eine affine Transformation, d.h. $g(x) = Cx + b$ mit einer quadratischen Matrix C . Dann ist $\det C$ die Funktionaldeterminante. Ist dann $f(y) = \chi_A(y)$, so gilt

$$m(g(A)) = |\det C| \cdot m(A).$$

Ist nun speziell g eine orthogonale Transformation, d.h. $|\det C| = 1$, so erhalten wir:

Korollar 2.55. *Das Lebesgue-Maß ist bewegungsinvariant.*

2. Wir betrachten die Abbildung g , die im \mathbb{R}^2 den Polarkoordinaten r, φ die kartesischen Koordinaten $x, y \in \mathbb{R}^1$ zuordnet. Dann gilt

$$\frac{\partial(x, y)}{\partial(r, \varphi)} = r.$$

Sei $A \subseteq \mathbb{R}^2$ und $g : A \rightarrow B$. Dann folgt aus (2.53)

$$\int \int_B f(x, y) \, dx dy = \int \int_A f(r \cos \varphi, r \sin \varphi) r \, dr \, d\varphi.$$

Die Eineindeutigkeit erzeugt man häufig dadurch, dass man eine Nullmenge C wählt und B durch $B \setminus C$ ersetzt. Etwa

$$B = \{(x, y) : x^2 + y^2 < R^2\}, \quad C = \{(x, y) : y = 0, 0 \leq x < R\},$$

$$A = \{(r, \varphi) : 0 < r < R, 0 < \varphi < 2\pi\}.$$

In diesem Fall folgt

$$\int \int_B f(x, y) \, dx dy = \int_0^{2\pi} \int_0^R f(r, \cos \varphi, r \sin \varphi) r \, dr \, d\varphi.$$

Anhang A

Integrationstheorie

Bei der mehrdimensionalen Integrationstheorie wird der Approximationssatz von Stone-Weierstraß benötigt. Dieser Satz ist Gegenstand des folgenden Paragraphen. In der Darstellung folgen wir eng dem Skript von E. Freitag[5].

A.1 Der Approximationssatz von Stone-Weierstraß

Der Satz von Stone-Weierstraß beschäftigt sich mit der gleichmäßigen Approximation einer stetigen Funktion $f : [a, b] \rightarrow \mathbb{R}$ durch eine vorgegebene Klasse von Funktionen h . Wir treffen die Einschränkung, dass es sich in jedem Fall um Klassen **stetiger** Funktionen handelt.

Satz A.1 (Approximationssatz von Stone-Weierstraß, erste Variante). *Es sei $X \subseteq \mathbb{R}^n$ eine kompakte Teilmenge, die mindestens zwei Punkte enthält. Weiterhin sei W eine Menge von stetigen reellwertigen Funktionen auf X mit den Eigenschaften*

1) $f, g \in W \implies f + g \in W$ und $cf \in W$ für $c \in \mathbb{R}$.

2) $f \in W \implies |f| \in W$.

3) (Punktentrennungseigenschaft):

Seien a, b zwei verschiedene Punkte von X . Dann existiert ein $f \in W$ mit $f(a) = 0$ aber $f(b) \neq 0$.

Unter diesen Voraussetzungen gilt: Ist $f \in C(X)$ eine beliebige stetige Funktion auf X und $\varepsilon > 0$, so existiert eine Funktion

$$h \in W \text{ mit } \|f - h\| \leq \varepsilon$$

(wobei $\|\cdot\|$ die Supremumsnorm bezeichnet). Zusatz: Insbesondere existiert dann eine Folge (h_n) von Funktionen aus W , die gleichmäßig gegen f konvergiert. (Man setze $\varepsilon := \frac{1}{n}$ und bezeichne die zugehörige Funktion mit h_n .)

Beweis: Zunächst einige Bezeichnungen:

Sei $f : X \rightarrow \mathbb{R}$ eine Funktion. Dann sei

$$f^+ : X \rightarrow \mathbb{R}, \quad f^+(x) := \max(f(x), 0).$$

Offenbar gilt:

$$f^+ = \frac{1}{2}(|f| + f).$$

Insbesondere gilt

$$f \in W \implies f^+ \in W \quad (\text{wegen 1) und 2}).$$

Sei $g : X \rightarrow \mathbb{R}$ eine weitere Funktion. Man definiert

$$f \vee g : X \rightarrow \mathbb{R}, \quad (f \vee g)(x) := \max(f(x), g(x)),$$

$$f \wedge g : X \rightarrow \mathbb{R}, \quad (f \wedge g)(x) := \min(f(x), g(x)).$$

Offenbar gilt

$$f \vee g = f + (g - f)^+,$$

$$f \wedge g = f - (g - f)^+.$$

Hieraus kann man sofort folgern:

$$f, g \in W \implies f \vee g \in W, \quad f \wedge g \in W.$$

Das gleiche gilt dann auch für endlich viele Funktionen f_1, \dots, f_n .

Nun soll (A.1) in mehreren Schritten bewiesen werden. Wir formulieren erst die Schritte.

1. *Schritt.* Seien $a, b \in X$ verschiedene Punkte und A, B zwei reelle Zahlen. Dann existiert $f \in W$ mit

$$f(a) = A, \quad f(b) = B.$$

2. *Schritt.* Sei $f : X \rightarrow \mathbb{R}$ stetig, $a, b \in X$ und $\varepsilon > 0$. Dann gibt es eine offene Umgebung V_b von b und eine Funktion $h_{a,b} \in W$ mit

$$h_{a,b}(a) = f(a) - \frac{\varepsilon}{2} \quad \text{und} \quad h_{a,b}(x) < f(x) \quad \text{für alle } x \in V_b \cap X.$$

3. *Schritt.* Sei $f : X \rightarrow \mathbb{R}$ stetig, $a \in X$ und $\varepsilon > 0$. Dann gibt es eine offene Umgebung U_a von a und eine Funktion $h_a \in W$ mit

$$f(x) > h_a(x) \text{ für alle } x \in U_a$$

und

$$h_a(x) > f(x) - \varepsilon \text{ für alle } x \in U_a \cap X.$$

4. *Schritt:* Beweis des Theorems.

Beweis des 1. Schrittes. Nach 3) existieren Funktionen $g, h \in W$ mit

$$g(a) \neq 0, \quad g(b) = 0 \text{ und } h(a) = 0, \quad h(b) \neq 0.$$

Man setze

$$f := \frac{A}{g(a)}g + \frac{B}{h(b)}h.$$

Insbesondere existiert zu jedem $a \in X$ eine Funktion $f \in W$ mit vorgegebenem Funktionswert in a .

Beweis des 2. Schrittes. Nach dem 1. Schritt existiert eine Funktion $h_{a,b} \in W$ mit

$$h_{a,b}(a) = f(a) - \frac{\varepsilon}{2} \quad \text{und} \quad h_{a,b}(b) = f(b) - \frac{\varepsilon}{2}.$$

Die Menge

$$V_b := \mathbb{R}^n \setminus \{x \in X : h_{a,b}(x) \geq f(x)\}$$

ist eine offene Umgebung von b (denn b ist offenbar in V_b enthalten, und $\{x \in X : h_{a,b}(x) \geq f(x)\}$ ist eine abgeschlossene Menge, weil sie das Urbild von $[0, \infty)$ unter der stetigen Abbildung $f - h_{a,b}$ ist).

Beweis des 3. Schrittes. Sei $f \in C(X, \mathbb{R})$, $a \in X$ und $\varepsilon > 0$. Zu jedem $b \in X$ existieren $h_{a,b} \in W$ und eine offene Umgebung V_b mit den Eigenschaften, die im zweiten Schritt formuliert wurden. Es gilt

$$X \subseteq \bigcup_{b \in X} V_b.$$

X ist kompakt, kann also durch endlich viele der V_b überdeckt werden:

$$X \subseteq V_{b_1} \cup \dots \cup V_{b_n} \quad (b_1, \dots, b_n \text{ geeignet}).$$

Wir definieren nun

$$h_a := h_{a,b_1} \wedge \dots \wedge h_{a,b_n}.$$

Dies ist wieder eine Funktion aus W . Nach Konstruktion von $h_{a,b}$ gilt ferner

$$h_a(x) \leq h_{a,b_j}(x) < f(x) \text{ für } x \in X \cap V_{b_j}.$$

Da die V_{b_j} ganz X überdecken, gilt

$$h_a(x) < f(x) \text{ für alle } x \in X.$$

Das ist eine der Behauptungen des dritten Schrittes. Es bleibt die Umgebung U_a zu konstruieren. Wegen

$$h_a(a) = f(a) - \frac{\varepsilon}{2} > f(a) - \varepsilon$$

ist

$$U_a := \mathbb{R}^n \setminus \{x \in X : h_a(x) \leq f(x) - \varepsilon\}$$

wieder eine offene Umgebung von a .

4. Schritt: Beweis des Theorems. Sei $f \in C(X)$. Zu jedem $a \in X$ wurde eine Funktion $h_a \in W$ konstruiert mit

$$f(x) > h_a(x) \text{ für alle } x \in X,$$

aber

$$h_a(x) > f(x) - \varepsilon \text{ für alle } x \in X \cap U_a,$$

wobei U_a eine offene Umgebung von a ist. Es gilt

$$X \subseteq \bigcup_{a \in X} U_a$$

und somit aufgrund der Kompaktheit von X bereits

$$X \subseteq U_{a_1} \cup \dots \cup U_{a_n}.$$

Man definiere nun

$$h := h_{a_1} \vee \dots \vee h_{a_n}.$$

Offenbar liegt h in W und es gilt

$$f(x) > h(x) > f(x) - \varepsilon \text{ für alle } x \in X,$$

insbesondere also

$$|f(x) - h(x)| < \varepsilon \text{ für alle } x,$$

was genau bedeutet $\|f - h\| < \varepsilon$. □

Beispiel A.2. Wir nehmen an, X sei ein eindimensionales kompaktes Intervall:

$$X = [a, b], \quad a < b$$

und W die Menge aller stetigen Funktionen $f : [a, b] \rightarrow \mathbb{R}$, für die es Stützstellen

$$a = a_0 < a_1 < \dots < a_n = b$$

gibt, so dass f für $0 \leq \nu < n$ linear auf $[a_\nu, a_{\nu+1}]$ ist.

Die Menge W erfüllt die in (A.1) geforderten Voraussetzungen. Daher kann man jede stetige Funktion durch stückweise lineare Funktionen gleichmäßig approximieren.

Satz A.3 (Approximationssatz von Stone-Weierstraß, zweite Variante). *Sei X eine nichtleere kompakte Teilmenge des \mathbb{R}^n und $W \subseteq C(X, \mathbb{R})$ eine Menge von stetigen reellwertigen Funktionen auf X mit den Eigenschaften*

- 1) $f, g \in W \Rightarrow f + g \in W$ und $cf \in W$ für $c \in \mathbb{R}$.
- 2) $f, g \in W \Rightarrow f \cdot g \in W$ und die konstanten Funktionen liegen in W .
- 3) W besitzt die Punkttrennungseigenschaft aus (A.1).

Unter diesen Voraussetzungen gilt: Ist $f \in C(X)$ eine beliebige stetige Funktion auf X und $\varepsilon > 0$, so existiert eine Funktion

$$h \in W \text{ mit } \|f - h\| \leq \varepsilon.$$

Der wesentliche Unterschied zu (A.1) besteht darin, dass man die Bedingung

$$f \in W \Rightarrow |f| \in W$$

durch die Bedingung

$$f, g \in W \Rightarrow f \cdot g \in W$$

ersetzt hat.

Beweis von A.3: Sei \overline{W} der Abschluss von W in $C(X)$, d.h. \overline{W} besteht aus allen stetigen Funktionen auf X , die sich gleichmäßig durch Funktionen aus W approximieren lassen. Also:

$$f \in \overline{W} \iff \text{zu jedem } \varepsilon > 0 \text{ existiert } h \in W \text{ mit } \|f - h\| < \varepsilon.$$

Zunächst machen wir uns klar, dass \overline{W} wieder die Eigenschaften 1)-3) von A.3 erfüllt. Es ist zu zeigen:

$$f, g \in \overline{W} \Rightarrow f + g, f \cdot g, cf \in \overline{W}.$$

- (a) Wir zeigen $f + g \in \overline{W}$. Sei $\varepsilon > 0$ vorgegeben und f_1, g_1 so gewählt, dass

$$\|f - f_1\| < \varepsilon \text{ und } \|g - g_1\| < \varepsilon$$

gilt. Es folgt

$$\|f + g - (f_1 + g_1)\| < \|f - f_1\| + \|g - g_1\| < 2\varepsilon.$$

- (b) Es soll $f \cdot g \in \overline{W}$ gezeigt werden. Sei wieder $\varepsilon > 0$ vorgegeben. Wir setzen

$$\delta := \min\left(1, \frac{\varepsilon}{1 + \|f\| + \|g\|}\right)$$

und bestimmen $f_1, g_1 \in W$ so, dass

$$\|f - f_1\| < \delta \text{ und } \|g - g_1\| < \delta$$

gilt. Es folgt

$$\begin{aligned} \|fg - f_1g_1\| &= \|-(f - f_1)(g - g_1) + f(g - g_1) + g(f - f_1)\| \\ &\leq \|f - f_1\| \cdot \|g - g_1\| + \|f\| \cdot \|g - g_1\| + \|g\| \cdot \|f - f_1\| \\ &\leq \delta^2 + \delta\|f\| + \delta\|g\| = \delta(\delta + \|f\| + \|g\|) \\ &\leq \delta(1 + \|f\| + \|g\|) \leq \varepsilon. \end{aligned}$$

Hierbei wurde neben der Dreiecksungleichung $\|f + g\| \leq \|f\| + \|g\|$ auch noch die Ungleichung $\|f \cdot g\| \leq \|f\| \cdot \|g\|$ benutzt, welche leicht zu verifizieren ist.

Damit besitzt also auch \overline{W} die Eigenschaften 1)-3) aus (A.3). Die Punkt-trennungseigenschaft ist dabei trivial, denn es gilt ja $W \subseteq \overline{W}$.

Jetzt zeigen wir sogar

$$f \in \overline{W} \implies |f| \in \overline{W}.$$

Wenn wir dies gezeigt haben, so sind wir fertig, denn nach der ersten Variante (A.1) kann dann jede stetige Funktion $f \in C(X)$ gleichmäßig durch Funktionen aus \overline{W} approximiert werden, d.h. ist $\varepsilon > 0$, so existiert $g \in \overline{W}$ mit

$$\|f - g\| < \varepsilon.$$

Andererseits existiert nach Definition von \overline{W} eine Funktion

$$h \in W \text{ mit } \|f - h\| = \|(f - g) + (g - h)\| < 2\varepsilon.$$

Wir haben also gezeigt, dass $\overline{\overline{W}} = \overline{W}$ gilt, was offenbar in beliebigen metrischen Räumen richtig ist. \square

Halten wir noch einmal fest:

Die zweite Variante A.3 des Approximationssatzes von Stone-Weierstraß ist zurückgeführt auf die erste, wenn man zeigen kann, dass für $f \in W$ die Funktion $|f|$ gleichmäßig durch Funktionen aus W approximiert werden kann. Wir werden sogar zeigen:

Die Funktion $|f|$ lässt sich gleichmäßig durch endliche Linearkombinationen von $1, f, f^2, \dots$ approximieren.

Der Schlüssel zum Beweis dieser Behauptung ist der

Hilfssatz A.4. *Auf dem Intervall $[-1, 1]$ existiert eine Folge von Polynomen*

$$p_n : [-1, 1] \rightarrow \mathbb{R},$$

die gleichmäßig gegen die Funktion $|x|$ konvergiert.

Wir werden diese Folge (p_n) explizit angeben. Der Gedankengang ist der folgende: Man kann

$$|x| = \sqrt{1 - (1 - x^2)} \text{ für } -1 \leq x \leq 1$$

schreiben. Setzt man

$$y = 1 - x^2,$$

so variiert y zwischen 0 und 1.

Es genügt also, eine Folge von Polynomen $q_n(y)$ zu konstruieren, die gleichmäßig gegen die Funktion

$$\sqrt{1 - y}$$

konvergiert. Der zugrunde gelegte Definitionsbereich ist hierbei

$$D := \{y : 0 \leq y \leq 1\}.$$

(Anschließend setze man dann $p_n(x) := q_n(1 - x^2)$. Dies sind dann Polynome in x).

Zur Approximation von $\sqrt{1 - y}$ verwendet man die **Taylorreihe**. (Diese haben wir in der [1], Abschnitt 5.2, im Zusammenhang mit der Taylorschen Formel behandelt. Dort haben wir uns allerdings um die Randpunkte des Konvergenzintervalls nicht gekümmert, weshalb wir das Ganze hier noch einmal aufrollen wollen.) Man ermittelt für die Taylorreihe sofort

$$\sum_{\nu=0}^{\infty} \binom{\frac{1}{2}}{\nu} (-1)^\nu y^\nu,$$

wobei der verallgemeinerte Binomialkoeffizient $\binom{\alpha}{\nu}$ definiert ist durch

$$\binom{\alpha}{\nu} = \frac{\alpha(\alpha-1)\dots(\alpha-\nu+1)}{\nu!}$$

Der Konvergenzradius dieser Reihe ist offenbar 1. Es ist aber dennoch nicht von vornherein klar, dass $\sqrt{1-y}$ durch die Reihe dargestellt wird. Jedenfalls wird durch die Reihe eine Funktion

$$f : (-1, 1) \rightarrow \mathbb{R}$$

dargestellt, die, wie man leicht nachrechnet, der Differentialgleichung

$$(1-y)f'(y) = -\alpha f(y)$$

(mit $\alpha = \frac{1}{2}$) genügt. (Gliedweise Differentiation ist im Innern des Konvergenzintervalls erlaubt!)

Hieraus folgert man leicht, dass

$$\left(\frac{f(y)}{(1-y)^\alpha} \right)' = 0$$

gilt. Also ist

$$f(y) = c \cdot (1-y)^\alpha.$$

Durch Spezialisierung ($y = 0$) ermittelt man $c = 1$.

Wir haben also gezeigt: Die Folge der Polynome

$$q_n(y) := \sum_{\nu=0}^n \binom{\frac{1}{2}}{\nu} (-1)^\nu y^\nu$$

konvergiert in $(-1, 1)$ gegen $\sqrt{1-y}$ und zwar ist die Konvergenz gleichmäßig in jedem Teilintervall $[-1 + \varepsilon, 1 - \varepsilon]$ (mit $1 > \varepsilon > 0$, ε beliebig). Mehr ist aus der allgemeinen Theorie der Potenzreihen nicht zu erhalten.

Wir wollten jedoch die Funktion $\sqrt{1-y}$, welche ja in $y = 1$ noch stetig ist, sogar gleichmäßig in $[0, 1]$ approximieren. (Der Punkt $y = 1$ entspricht gerade der Knickstelle $x = 0$ der Funktion $|x|$ und ist damit der Angelpunkt!)

Wir müssen also noch zeigen:

- 1) Die Folge $q_n(y)$ konvergiert auch für $y = 1$ und zwar gegen $\sqrt{1-1} = 0$.

2) Die Konvergenz in $[0, 1]$ ist gleichmäßig.

Beweis: Eine Abzählung der Vorzeichen im Binomialkoeffizienten $\binom{\alpha}{\nu}$ für $\nu \geq 1$ ergibt, dass

$$(-1)^\nu \binom{\frac{1}{2}}{\nu} = - \left| \binom{\frac{1}{2}}{\nu} \right|$$

gilt. Hieraus folgt

$$q_n(x) = 1 - \sum_{\nu=1}^n \left| \binom{\frac{1}{2}}{\nu} \right| y^\nu.$$

Damit ist klar, dass für $0 \leq y \leq 1$ gilt:

$$q_1(y) \geq q_2(y) \geq \dots$$

Außerdem ist

$$\sqrt{1-y} = \lim_{n \rightarrow \infty} q_n(y) \geq 0 \text{ für } 0 \leq y < 1.$$

Hieraus folgt $q_n(y) \geq 0$ für alle n und $0 \leq y < 1$.

Die Funktionen q_n sind stetig in $y = 1$, also folgt sogar

$$q_n(y) \geq 0 \text{ für alle } n \text{ und } 0 \leq y \leq 1.$$

Die Folge $(q_n(1))$ ist also monoton fallend und nach unten (durch 0) beschränkt. Sie konvergiert daher. Das bedeutet nichts anderes, als dass die Reihe

$$\sum_{\nu=0}^{\infty} \left| \binom{\frac{1}{2}}{\nu} \right|$$

konvergiert. Für alle $y \in [0, 1]$ gilt

$$\left| \binom{\frac{1}{2}}{\nu} (-1)^\nu y^\nu \right| \leq \left| \binom{\frac{1}{2}}{\nu} \right|.$$

Nach dem Majorantenkriterium konvergiert also die Folge $(q_n(y))$ daher schon gleichmäßig für $0 \leq y \leq 1$. Die Grenzfunktion

$$\lim_{n \rightarrow \infty} q_n(y)$$

ist daher stetig in $[0, 1]$.¹ Sie stimmt mit $\sqrt{1-y}$ in $[0, 1)$ überein. Da auch letztere Funktion in $y = 1$ stetig ist, müssen die beiden Funktionen auch dort übereinstimmen. Damit ist (A.4) bewiesen. \square

¹Die Stetigkeit folgt auch aus dem Abelschen Grenzwertsatz.

Nun ist der Approximationssatz in der zweiten Variante leicht zu beweisen. Sei etwa $f \in W$. Wir approximieren $|f|$. Sei zunächst

$$g := \frac{f}{\|f\| + 1}.$$

Dann gilt offenbar

$$|g(x)| \leq 1 \text{ für alle } x \in X.$$

Nach (A.4) existiert ein Polynom

$$p_n(x) = a_0 + a_1x + \dots + a_nx^n$$

mit der Eigenschaft

$$|p_n(x) - |x|| < \varepsilon \text{ für alle } x \in [-1, 1].$$

Es folgt insbesondere

$$|p_n(g(x)) - |g(x)|| < \varepsilon \text{ für alle } x \in X.$$

Die Funktion

$$p_n(g(x)) = a_0 + a_1g(x) + \dots + a_ng(x)^n$$

ist eine Funktion aus W , damit ist gezeigt

$$|g| \in \overline{W} \implies |f| = (\|f\| + 1)|g| \in \overline{W}.$$

Folgerung A.5 (Weierstraßscher Approximationssatz). *Sei*

$$f : [a, b] \rightarrow \mathbb{R} \quad (a < b)$$

stetig. Es existiert eine Folge von Polynomen, die gleichmäßig gegen f konvergiert.

Beweis: Setze $X := [a, b]$ und $W := \{\text{Polynome auf } [a, b]\}$. Dann sind die Voraussetzungen von (A.3) erfüllt. \square

Es ist interessant festzustellen, dass obiger Beweis des Satzes von Stone-Weierstraß (A.3) sich zusammensetzt aus allgemeinen Betrachtungen und einem extremen Spezialfall ($f(x) = |x|$), in dem einmal eine solche Approximation explizit konstruiert werden musste.

Im Vorgriff auf das folgende Kapitel bemerken wir hier folgende Anwendung des Weierstraßschen Approximationssatzes: Es sei f eine stetige Funktion auf

dem Rechteck

$$[a, b] \times [c, d] \quad (a < b, \quad c < d).$$

Dann gilt

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

Beweis: Für Polynome

$$f(x, y) = \sum_{0 \leq i, k \leq n} a_{ik} x^i y^k$$

kann man dies leicht nachrechnen. Aufgrund des Approximationssatzes lässt sich aber jede stetige Funktion gleichmäßig durch Polynome approximieren. \square

Die nächsten Kapitel folgen im Wesentlichen dem Skript [5] von E. Freitag.

A.2 Das Integral für stetige Funktionen mit kompaktem Träger

Gegeben sei ein abgeschlossener Quader Q im \mathbb{R}^n , welcher durch die n -Tupel

$$a = (a_1, \dots, a_n), \quad b = (b_1, \dots, b_n), \quad a_\nu \leq b_\nu \text{ für } \nu = 1, \dots, n$$

definiert sei:

$$Q = \{x \in \mathbb{R}^n : a_\nu \leq x_\nu \leq b_\nu\}.$$

Es sei eine stetige Funktion

$$f : Q \rightarrow \mathbb{R}$$

gegeben. Wir versuchen, das (mehrfache) Integral von f

$$\int_Q f(x) dx_1 \dots dx_n$$

zu definieren. Dies soll induktiv geschehen. Zunächst kann man bei festem x_2, \dots, x_n das Integral

$$\int_{a_1}^{b_1} f(x) dx_1$$

definieren, da f als Funktion von x_1 stetig und somit integrierbar ist.

Lässt man jetzt x_2, \dots, x_n wieder variieren, so kann man dieses Integral als Funktion von x_2, \dots, x_n auffassen. Definitionsbereich ist der Quader im \mathbb{R}^{n-1} , der durch

$$a_\nu \leq x_\nu \leq b_\nu \quad \text{für } 2 \leq \nu \leq n$$

definiert ist.

Jetzt will man über x_2 integrieren, d.h.

$$\int_{a_2}^{b_2} \left[\int_{a_1}^{b_1} f(x) dx_1 \right] dx_2$$

betrachten, usw. und schließlich soll dann das mehrfache Integral von f über Q durch die Formel

$$\int_Q f(x) dx_1 \dots dx_n = \int_{a_n}^{b_n} \dots \int_{a_1}^{b_1} f(x) dx_1 \dots dx_n$$

definiert werden. Allerdings muss man wissen, dass alle diese Integrationen durchführbar sind.

Hilfssatz A.6. *Die Funktion*

$$f : Q \rightarrow \mathbb{R}, \quad Q = \{x \in \mathbb{R}^n : a_\nu \leq x_\nu \leq b_\nu\} \quad (a_\nu \leq b_\nu)$$

sei stetig. Dann ist auch die Funktion

$$F : Q' \rightarrow \mathbb{R}, \quad Q' = \{(x_2, \dots, x_n) \in \mathbb{R}^{n-1} : a_\nu \leq x_\nu \leq b_\nu\}$$

$$F(x_2, \dots, x_n) = \int_{a_1}^{b_1} f(x_1, x_2, \dots, x_n) dx_1$$

stetig.

Beweis: Wesentliches Hilfsmittel zum Beweis ist der Satz von der gleichmäßigen Stetigkeit (2.14). Versuchen wir zunächst einmal, die Stetigkeit direkt zu beweisen, etwa im Punkt (x_2^0, \dots, x_n^0) . Man muss zeigen

$$|F(x_2, \dots, x_n) - F(x_2^0, \dots, x_n^0)| < \varepsilon \quad \text{für } |x_\nu - x_\nu^0| < \delta(\varepsilon) \quad (2 \leq \nu \leq n).$$

Dabei sei $\varepsilon > 0$ beliebig vorgegeben, also

$$\left| \int_{a_1}^{b_1} [f(x_1, x_2, \dots, x_n) - f(x_1, x_2^0, \dots, x_n^0)] dx_1 \right| < \varepsilon.$$

Nun gilt wegen der Stetigkeit von f tatsächlich

$$|f(x_1, x_2, \dots, x_n) - f(x_1, x_2^0, \dots, x_n^0)| < \varepsilon \text{ für } |x_\nu - x_\nu^0| < \delta(\varepsilon)$$

für $2 \leq \nu \leq n$, und man kann abschätzen:

$$|F(x_2, \dots, x_n) - F(x_2^0, \dots, x_n^0)| < \int_{a_1}^{b_1} \varepsilon dx_1 = (b_1 - a_1)\varepsilon$$

für $|x_\nu - x_\nu^0| < \delta(\varepsilon)$ ($2 \leq \nu \leq n$).

Das Auftreten der Konstanten $b_1 - a_1$ störe natürlich in keiner Weise, aber der Beweis funktioniert nur dann, wenn $\delta(\varepsilon)$ unabhängig von x_1 gewählt werden kann.

Dies folgt aber aus dem Satz von der gleichmäßigen Stetigkeit. □

Damit ist folgende Definition möglich:

Definition A.7. *Das Integral einer stetigen Funktion*

$$f : Q \rightarrow \mathbb{R}, \quad Q = \{x : a_\nu \leq x_\nu \leq b_\nu\} \quad (a_\nu \leq b_\nu)$$

auf einem kompakten Quader Q wird induktiv definiert durch

$$\int_Q f(x) dx_1, \dots, dx_n = \int_{Q'} \left[\int_{a_1}^{b_1} f(x_1, x_2, \dots, x_n) dx_1 \right] dx_2 \dots dx_n.$$

($Q' = \{(x_2, \dots, x_n) \in \mathbb{R}^{n-1}; \quad a_\nu \leq x_\nu \leq b_\nu\}$ für $2 \leq \nu \leq n$)

$$\int_{[a,b]} f(x) dx = \int_a^b f(x) dx \quad (a < b) \text{ im Falle } n = 1.$$

Schreibweise:

$$\int_Q f(x) dx = \int_{a_n}^{b_n} \dots \int_{a_1}^{b_1} f(x) dx_1 \dots dx_n.$$

Der Aufbau der Integrationstheorie ist vergleichsweise kompliziert. Für pragmatisch denkende Anwender ist vielleicht der Hinweis nützlich, dass das Integral für stetige Funktionen mit kompaktem Träger einer Veränderlicher im Sinne des Regelintegrals bereits ausreicht, um Volumenberechnungen durchzuführen. Man gehe folgendermaßen vor: Sei $A \subseteq \mathbb{R}^n$ eine einigermaßen anständige, sagen wir kompakte Teilmenge, welche man sich als massiven n -dimensionalen Körper vorstelle. Durchschneidet man diesen Körper mit der

„Ebene“ $x_n = t$, so erhält man einen $(n-1)$ -dimensionalen Körper $A(t)$. Man will die Volumentheorie induktiv aufbauen, nimmt also an, dass man weiß, wie man das $(n-1)$ -dimensionale Volumen von $A(t)$ berechnen kann. Die Funktion $h(t) = \text{vol}(A(t))$ verschwindet außerhalb eines geeigneten Intervalls $[-C, C]$. In einigermaßen anständiger Situation darf man erwarten, dass sie eine Regelfunktion² ist, so dass man

$$\text{vol}(A) := \int_{-C}^C h(t) dt$$

definieren kann. Aber dieser naive Weg gibt keine direkte Einsicht, warum beispielsweise das Volumen gegenüber Drehungen des Körpers invariant bleibt. Dennoch: Obige Vorgehensweise reicht aus für alle praktischen Belange der Volumenbestimmung und der eine oder andere Hörer mag mit dieser Erkenntnis zufrieden sein und die Mühe des Aufbaus einer leistungsstarken Integrationstheorie als überflüssig erachten. Dies ist legitim, wenn er sich in der Mathematik von der Analysis wegorientiert und beispielsweise eine algebraische Linie bevorzugt. Tieferer Einstieg in Gebiete wie Analysis, Wahrscheinlichkeitstheorie etc. erfordert jedoch eine leistungsstarke Integrationstheorie mit guten Grenzwertsätzen (Vertauschbarkeit von Integration mit Limesbildungen wie beispielsweise Differentiation).

Für die Integrationstheorie ist fundamental, dass man die Reihenfolge der Integrationen vertauschen darf, dass also beispielsweise im Falle $n = 2$

$$\int_{a_2}^{b_2} \int_{a_1}^{b_1} f(x_1, x_2) dx_1 dx_2 = \int_{a_1}^{b_1} \int_{a_2}^{b_2} f(x_1, x_2) dx_2 dx_1$$

gilt. Dies kann man an vielen Beispielen nachprüfen, muss aber streng bewiesen werden. Wir führen dies nochmals durch:

Gegeben sei eine Umordnung (Permutation)

$$\sigma = \{\sigma_1, \sigma_2, \dots, \sigma_n\} \quad \text{der Zahlen } \{1, \dots, n\},$$

d.h. jede Zahl 1 bis n kommt unter den σ_ν genau einmal vor.

²Eine auf einem kompakten Intervall definierte beschränkte Funktion $f: [a, b] \rightarrow \mathbb{R}$ heißt Regelfunktion, wenn eine Folge von Treppenfunktionen h_n auf dem Intervall $[a, b]$ existiert, die in der Supremumsnorm gegen f konvergiert: $\|h_n - f\| \rightarrow 0$ für $n \rightarrow \infty$. Man kann zeigen (vgl. Aufgabe 4 des Übungsblatts 4 zur Analysis II): Eine beschränkte Funktion f ist genau dann Regelfunktion, wenn an jeder Stelle des betrachteten Intervalls sowohl der rechtsseitige als auch der linksseitige Grenzwert existiert.

Satz A.8. Gegeben sei eine stetige Funktion $f : Q \rightarrow \mathbb{R}$, Q ein kompakter Quader im \mathbb{R}^n und σ eine Permutation der Variablen. Dann gilt

$$\int_{a_n}^{b_n} \dots \int_{a_1}^{b_1} f(x) dx_1 \dots dx_n = \int_{a_{\sigma_n}}^{b_{\sigma_n}} \dots \int_{a_{\sigma_1}}^{b_{\sigma_1}} f(x) dx_{\sigma_1} \dots dx_{\sigma_n}.$$

Es kommt also auf die Reihenfolge der Integrationen nicht an.

Beweis: Die Behauptung ist trivial für Funktionen von dem speziellen Typ

$$f(x) = f_1(x_1) \cdot f_2(x_2) \dots f_n(x_n),$$

wobei

$$f_\nu : [a_\nu, b_\nu] \rightarrow \mathbb{R}$$

stetige Funktionen einer Variablen sind, denn dann zerfällt das Integral

$$\int_Q f(x) dx = \left(\int_{a_1}^{b_1} f_1(x_1) dx_1 \right) \dots \left(\int_{a_n}^{b_n} f_n(x_n) dx_n \right)$$

in ein Produkt von Integralen.

Wir bezeichnen mit A die Menge aller stetigen Funktionen $f : Q \rightarrow \mathbb{R}$, die sich als endliche Summe von Funktionen obigen speziellen Typs schreiben lassen, also

$$f(x) = \sum_{\nu=1}^m f_{1\nu}(x_1) \dots f_{n\nu}(x_n).$$

Es ist klar, dass die Behauptung auch für alle Funktionen dieser Klasse gilt. Satz A.8 wird dann mit Hilfe des *Approximationssatzes von Stone-Weierstraß* bewiesen (siehe A.1).

Zunächst behaupten wir:

Jede stetige Funktion $f : Q \rightarrow \mathbb{R}$ ist gleichmäßiger Grenzwert einer Folge von Funktionen aus A . Dazu weisen wir die Voraussetzungen des Approximationssatzes nach. Der einzige nicht völlig triviale Punkt ist die Punktentrennung. Seien also

$$x^0 = (x_1^0, \dots, x_n^0), \quad y^0 = (y_1^0, \dots, y_n^0)$$

zwei verschiedene Punkte aus Q , also etwa

$$x_\nu^0 \neq y_\nu^0.$$

Die Funktion f mit

$$f(x) = x_\nu - x_\nu^0$$

trennt die beiden Punkte $(f(x^0) = 0, f(y^0) \neq 0)$ und liegt in A . Da man jede stetige Funktion $f : Q \rightarrow \mathbb{R}$ gleichmäßig approximieren kann durch Funktionen, für die Satz (A.8) schon bewiesen ist, muss man nur noch wissen, dass das mehrfache Integral bezüglich gleichmäßiger Konvergenz stabil ist. \square

Hilfssatz A.9. Sei $Q \subseteq \mathbb{R}^n$ ein kompakter Quader und

$$f : Q \rightarrow \mathbb{R}$$

eine stetige Funktion. Es gilt

$$\left| \int_Q f(x) dx_1 \dots dx_n \right| \leq \|f\| (b_1 - a_1) \dots (b_n - a_n),$$

wobei $\|f\|$ die Maximumsnorm bezeichnet.

Beweis: Aus den Rechenregeln für eine Veränderliche folgt unmittelbar

$$\int_Q f(x) dx_1 \dots dx_n \leq \int_Q g(x) dx_1 \dots dx_n,$$

falls

$$f(x) \leq g(x) \quad \text{für alle } x \in Q$$

gilt. Insbesondere gilt

$$\begin{aligned} \left| \int_Q f(x) dx_1 \dots dx_n \right| &\leq \int_Q |f(x)| dx_1 \dots dx_n \leq \\ &\int_Q \|f\| dx_1 \dots dx_n = \|f\| (b_1 - a_1) \dots (b_n - a_n). \end{aligned}$$

\square

Folgerung A.10. Die Folge von stetigen Funktionen

$$f_k : Q \rightarrow \mathbb{R}, \quad k = 1, 2, 3, \dots,$$

konvergiere gleichmäßig gegen f . Dann gilt

$$\int_Q f(x) dx_1 \dots dx_n = \lim_{k \rightarrow \infty} \int_Q f_k(x) dx_1 \dots dx_n.$$

Beweis:

$$\left| \int_Q f(x) dx_1 \dots dx_n - \int_Q f_k(x) dx_1 \dots dx_n \right| \leq \|f - f_k\| (b_1 - a_1) \dots (b_n - a_n) \rightarrow 0 \quad \text{für } k \rightarrow \infty.$$

□

Man sollte sich gut vor Augen halten:

Die Stabilität des Integrals für stetige Funktionen auf kompakten Quadern ist eine Trivialität, welche auf der banalen Abschätzung A.9 beruht.

Stetige Funktionen mit kompaktem Träger

Unter dem **Träger** einer Funktion

$$f : X \rightarrow \mathbb{R}, \quad X \subseteq \mathbb{R}^n$$

versteht man den Abschluss der Menge aller Punkte, in denen f nicht verschwindet:

$$\text{Träger } f = \overline{\{x \in X; f(x) \neq 0\}}.$$

Ein Punkt $x \in X$ gehört genau dann zum Träger von f , wenn $f(x) \neq 0$ ist, oder wenn es eine Folge von Punkten $x_n \in X$ gibt mit

$$x_n \longrightarrow x \quad \text{für } n \rightarrow \infty \quad \text{und } f(x_n) \neq 0 \text{ für alle } n.$$

Bezeichnung:

Klasse der stetigen Funktionen mit kompaktem Träger auf X :

$$C_c(X) := \{f : X \rightarrow \mathbb{R} : f \text{ stetig, Träger}(f) \text{ ist } \mathbf{kompakt}\}$$

Offenbar gilt

$$f, g \in C_c(X) \Rightarrow f + g, f \cdot g \text{ und } cf \in C_c(X) \quad (c \in \mathbb{R}).$$

Die konstanten Funktionen liegen nur dann in $C_c(X)$, wenn X selbst kompakt ist.

Bemerkung. Eine stetige Funktion

$$f : \mathbb{R}^n \rightarrow \mathbb{R}$$

gehört dann und nur dann zur Klasse $C_C(\mathbb{R}^n)$, wenn eine Zahl $R > 0$ existiert, so dass

$$f(x) = 0 \quad \text{für } \|x\| > R$$

gilt.

Beweis:

1. Wenn der Träger von f kompakt ist, so existiert eine Zahl $R > 0$, so dass

$$\text{Träger}(f) \subseteq K_R(0).$$

(Jedes Kompaktum ist beschränkt.)

2. Wenn ein solches R existiert, so ist der Träger von f beschränkt, außerdem ist er nach Konstruktion abgeschlossen und daher nach dem Überdeckungssatz von Heine-Borel kompakt.

Im Fall $n > 1$ erhält man stetige Funktionen mit kompaktem Träger durch

$$f(x) = f_1(x_1) \dots f_n(x_n),$$

wobei die $f_i(x_i)$ solche in einer Veränderlichen sind. □

Das Integral für stetige Funktionen mit kompaktem Träger

Es sei

$$f : \mathbb{R}^n \rightarrow \mathbb{R}$$

eine stetige Funktion mit kompaktem Träger. Wir wählen $R > 0$, so dass gilt:

$$f(x) \neq 0 \implies |x_\nu| \leq R \quad \text{für } \nu = 1, \dots, n.$$

Dann definieren wir

$$\int_{\mathbb{R}^n} f(x) dx_1 \dots dx_n := \int_Q f(x) dx_1 \dots dx_n,$$

wobei Q den Quader

$$-R \leq x_\nu \leq R \quad \text{für } \nu = 1, \dots, n$$

bezeichne. Es ist klar, dass diese Definition von der Wahl von R nicht abhängt. R muss nur so groß sein, dass der Träger von f in Q liegt.

Das so definierte Integral für stetige Funktionen mit kompaktem Träger ist der Baustein des Lebesgue'schen Integrals, das nun mit Hilfe des Daniell-Lebesgue-Prozesses gewonnen werden soll.

A.3 Die Ausdehnung des Integrals auf halbstetige Funktionen

In A.2 haben wir die Klasse $C_c(\mathbb{R}^n)$ der stetigen Funktionen mit kompaktem Träger eingeführt und darauf ein Integral definiert:

$$f \mapsto \int_{\mathbb{R}^n} f(x) dx_1 \dots dx_n.$$

Wir stellen die Eigenschaften dieses Integrals, die wir im Folgenden benutzen, kurz zusammen.

Dabei benutzen wir die abkürzende Schreibweise

$$I(f) := \int_{\mathbb{R}^n} f(x) dx := \int_{\mathbb{R}^n} f(x) dx_1 \dots dx_n.$$

1) *Das Integral ist ein lineares Funktional, d.h.*

$$I(f + g) = I(f) + I(g), \quad I(cf) = cI(f).$$

2) *Das Integral ist „positiv“, d.h.*

$$I(f) \geq 0, \quad \text{falls } f(x) \geq 0 \text{ für alle } x \in \mathbb{R}^n.$$

Im Folgenden besteht das Problem, das Integral auf eine möglichst große Klasse von Funktionen auszudehnen.

Zunächst haben wir, das Integral für Funktionen einer Veränderlichen schon benutzend, das Integral für stetige Funktionen mit kompaktem Träger gewonnen. Allgemeinere Funktionen f versuchen wir jetzt, durch stetige Funktionen zu approximieren. *Dabei müssen wir allerdings das gelobte Land der gleichmäßigen Konvergenz verlassen, denn sonst kommen wir aus dem Bereich der stetigen Funktionen nicht heraus.*

Der Typ der Konvergenz, den wir betrachten wollen, ist die *monotone* Konvergenz.

A.4 Der Daniell-Lebesgue-Prozess, Teil 1: Das Integral für halbstetige Funktionen

Mit Hilfe der Integrationstheorie mehrerer Variablen will man u.a. mehrdimensionale Volumina von Bereichen im \mathbb{R}^n berechnen.

Sei also

$$f : D \rightarrow \mathbb{R}, \quad D \subseteq \mathbb{R}^n$$

eine Funktion, die der Einfachheit halber nirgends negativ sei,

$$f(x) \geq 0 \quad \text{für alle } x \in D \quad (\text{Schreibweise } f \geq 0).$$

Gesucht ist ein Maß für das Volumen des Bereiches im $(n+1)$ -dimensionalen Raum

$$\{(x, t) \in \mathbb{R}^{n+1}; \quad x \in D, \quad 0 \leq t \leq f(x)\}$$

Dieses Volumen soll gerade

$$\int_D f(x) dx_1 \dots dx_n$$

sein.

Man könnte natürlich wie im Falle $n = 1$ versuchen, das Integral durch Approximationen mit Hilfe von „Treppen“ zu definieren. Das ist möglich, aber schwierig. Im Fall $n = 1$ sind die Definitionsbereiche D in der Regel einfache Intervalle. Diese sind sehr einfach aufzuteilen in „kleine“ Intervalle, auf denen man dann die Treppen aufbaut.

Im Fall $n > 1$ kann schon der Definitionsbereich D relativ kompliziert sein. Man müsste ihn durch eine Quaderaufteilung pflastern (approximativ) und darauf die „Treppen“ aufbauen. Dieser Aufbau ist möglich. Stattdessen haben wir einen anderen Weg eingeschlagen. In sehr naheliegender Weise konnten wir direkt das Integral für stetige Funktionen definieren. Dazu machten wir einen Rückgriff auf das Regelintegral in einer Veränderlichen, das ja sehr leicht (relativ zum Fall $n \geq 1$) zu gewinnen war. Jetzt werden wir das Integral für allgemeinere Funktionentypen f dadurch gewinnen, dass wir f durch stetige Funktionen mit kompaktem Träger zu approximieren suchen.

Dabei müssen wir notgedrungen auf die gleichmäßige Konvergenz verzichten, sonst kommen wir nicht aus dem Bereich der stetigen Funktionen heraus.

Warum ist das Integral für allgemeinere Funktionstypen interessant?

Will man das n -dimensionale Volumen eines Bereiches $D \subseteq \mathbb{R}^n$ definieren und berechnen, so kann man folgendermaßen vorgehen:

Man betrachte die Funktion

$$f(x) = 1 \quad \text{für } x \in D.$$

Die Anschauung zeigt dann, dass die Definition

$$\text{Volumen}(D) = \int_D 1 dx_1 \dots dx_n$$

sinnvoll ist.

(Die Länge des Intervalls $[a, b]$ kann als Integral über die Funktion 1 interpretiert werden, also eindimensionales Volumen von $[a, b]$):

$$\int_a^b dx = b - a,$$

oder: Die Fläche der Kreisscheibe

$$E = \{(x, y) \in \mathbb{R} \times \mathbb{R} : x^2 + y^2 \leq 1\}$$

kann interpretiert werden als Volumen des dreidimensionalen Zylinders

$$Z = \{(x, y, t) \in \mathbb{R}^3 : x^2 + y^2 \leq 1, 0 \leq t \leq 1\},$$

das wäre das Integral

$$\text{Volumen von } Z = \int_{\substack{x^2+y^2 \leq 1 \\ 0 \leq t \leq 1}} dx dy dt.$$

Im folgenden werden wir nur Integrale über den \mathbb{R}^n berechnen, wir nehmen also an, dass die Funktion f auf dem ganzen \mathbb{R}^n definiert ist und streben von vorneherein das uneigentliche Integral

$$\int_{\mathbb{R}^n} f(x) dx_1 \dots dx_n$$

an.

Dies ist keine Einschränkung der Allgemeinheit: wenn eine Funktion

$$f : D \rightarrow \mathbb{R}, \quad D \subseteq \mathbb{R}^n$$

gegeben ist, so betrachten wir einfach $\tilde{f} : \mathbb{R}^n \rightarrow \mathbb{R}$,

$$\tilde{f}(x) = \begin{cases} f(x) & \text{für } x \in D, \\ 0 & \text{für } x \notin D, \end{cases}$$

und definieren

$$\int_D f(x) dx_1 \dots dx_n := \int_{\mathbb{R}^n} \tilde{f}(x) dx_1 \dots dx_n,$$

vorausgesetzt, dass die rechte Seite schon definiert ist. Für die Volumenmessung $\text{Volumen}(D)$ bedeutet dies folgendes: Man betrachte die sogenannte **charakteristische Funktion (Indikatorfunktion)** von D ,

$$\chi_D : \mathbb{R}^n \rightarrow \mathbb{R}, \quad \chi_D(x) = \begin{cases} 1 & \text{für } x \in D, \\ 0 & \text{für } x \notin D, \end{cases}$$

und definiert

$$\text{Volumen}(D) = \int_{\mathbb{R}^n} \chi_D(x) dx_1 \dots dx_n.$$

Damit ist gezeigt, dass das Konzept des Volumens eines Bereiches $D \subseteq \mathbb{R}^n$ sich unter den Begriffsapparat des Integrals

$$\int_{\mathbb{R}^n} f(x) dx_1 \dots dx_n, \quad f : \mathbb{R}^n \rightarrow \mathbb{R}$$

unterordnet, aber man beachte, dass die Funktion χ_D nicht stetig ist (die Randpunkte von D sind Unstetigkeitspunkte). Das Integral für stetige Funktionen mit kompaktem Träger reicht also keineswegs zur Berechnung von Volumina aus.

Wichtigstes Hilfsmittel für den ersten Schritt im Daniell-Lebesgue-Prozess ist

Satz A.11 (Dini). *Es sei X ein kompakter metrischer Raum und*

$$f_n : X \rightarrow \mathbb{R}, \quad n = 1, 2, 3, \dots$$

eine Folge von stetigen Funktionen, die monoton gegen Null fällt, d.h.

- a) $f_1(x) \geq f_2(x) \geq \dots$,
- b) $\lim_{n \rightarrow \infty} f_n(x) = 0$ für jedes $x \in X$.

Die Folge (f_n) konvergiert dann gleichmäßig gegen 0.

Beweis: Sei $\varepsilon > 0$ vorgegeben. Zu jedem $x \in X$ existiert eine natürliche Zahl $N(\varepsilon, x)$ mit der Eigenschaft

$$|f_n(x)| < \varepsilon \quad \text{für } n \geq N(\varepsilon, x).$$

Zu jedem x_0 existiert dann eine offene Umgebung $U(x_0)$ mit der Eigenschaft

$$|f_{N(\varepsilon, x_0)}(x)| < \varepsilon \quad \text{für } x \in U(x_0).$$

Endlich viele dieser Umgebungen überdecken X (Kompaktheit). Wir definieren

$$N = N(\varepsilon) = \max\{N(\varepsilon, x_0); x_0 \in \text{obiger endlicher Menge}\}.$$

Es gilt dann

$$|f_N(x)| < \varepsilon \quad \text{für alle } x.$$

Wegen der Monotonie der Folge $F_n(x)$ gilt sogar

$$|f_n(x)| < \varepsilon \quad \text{für alle } x \text{ und } n \geq N.$$

Das heißt gerade, dass f_n gleichmäßig gegen Null konvergiert. □

Der Satz von Dini gibt uns die Möglichkeit, das Integral für stetige Funktionen mit kompaktem Träger auf eine große Klasse von Funktionen auszudehnen und zwar auf solche Funktionen f , die sich *monoton* durch eine Folge von Funktionen aus $C_c(\mathbb{R}^n)$ *approximieren* lassen.

Bei der technischen Durchführung der Integrationstheorie hat es sich als zweckmäßig erwiesen, auch Funktionen zuzulassen, die die Werte $\pm\infty$ annehmen dürfen, d.h. wir erweitern die reelle Zahlengerade \mathbb{R} durch Hinzufügen von weiteren Elementen, für die wir ∞ und $-\infty$ schreiben:

$$\overline{\mathbb{R}} := \mathbb{R} \cup \{\infty\} \cup \{-\infty\} \quad (\text{erweiterte Zahlengerade})$$

und vereinbaren die folgenden Rechenregeln:

- a) $\infty > x$ für alle $x \in \mathbb{R} \cup \{-\infty\}$.
- b) $-\infty < x$ für alle $x \in \mathbb{R} \cup \{\infty\}$.

Diese Erweiterung hat folgenden Vorteil: Jede Teilmenge $M \subseteq \overline{\mathbb{R}}$, sofern sie nicht leer ist, besitzt nun eine obere (natürlich auch untere) Grenze, die wir mit $\text{Sup}M$ bezeichnen wollen.³

Das Supremum $\text{Sup} M$ einer nichtleeren Teilmenge $M \subseteq \overline{\mathbb{R}}$ ist die *kleinste* obere Schranke von M . Im selben Sinne verstehen wir $\text{Inf}M$.

³Wir haben dieselbe Konvention bereits im Zusammenhang mit der Berechnung des Konvergenzradius einer Potenzreihe verwendet: siehe Analysis I.

Ist insbesondere $M \subseteq \mathbb{R}$ eine in \mathbb{R} nach oben beschränkte nicht leere Teilmenge, so ist $\text{Sup } M = \sup M$ das gewöhnliche Supremum. Ist hingegen M durch kein Element aus \mathbb{R} nach oben beschränkt, so ist $\text{Sup } M = \infty$.

Ist a_1, a_2, a_3, \dots eine Folge von Elementen aus \mathbb{R} , so verstehen wir unter dem Supremum dieser Folge einfach das Supremum der Menge der Folgenglieder:

$$\text{Sup}(a_n) = \text{Sup}\{a_1, a_2, \dots\}$$

Die erweiterte Zahlengerade hat allerdings den Nachteil, dass es unmöglich ist, die algebraischen Rechenregeln $(+, \cdot)$ auf \mathbb{R} so zu erweitern, dass die üblichen Gesetze erfüllt sind (Kommutativ-, Assoziativ- und Distributivgesetz). Zu keinen Widersprüchen führt die Konvention

$$\begin{aligned} \infty + x &= \infty && \text{für alle } x \in \mathbb{R} \cup \{\infty\}, \\ -\infty + x &= -\infty && \text{für alle } x \in \mathbb{R} \cup \{-\infty\}, \\ x \cdot \infty &= \infty && \text{für alle } x \in \mathbb{R}, x > 0, \\ x \cdot (-\infty) &= -\infty && \text{für alle } x \in \mathbb{R}, x > 0, \\ \infty \cdot \infty &= \infty, \\ \infty \cdot (-\infty) &= -\infty, \\ (-\infty) \cdot (-\infty) &= \infty, \end{aligned}$$

wovon sich der Leser (am besten dort wo es verwendet wird) überzeugen mag. **Nicht definiert werden jedoch:** $+\infty - \infty$ und $-\infty + \infty$.

Wir betrachten nun Funktionen, die man monoton durch stetige Funktionen mit kompaktem Träger approximieren kann.

Definition A.12. *Eine Funktion*

$$f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$$

gehört der Klasse B^+ an (Bairesche Klasse), wenn es eine Folge von stetigen Funktionen mit kompaktem Träger

$$f_\nu : \mathbb{R}^n \rightarrow \mathbb{R}$$

mit den Eigenschaften

$$a) f_1(x) \leq f_2(x) \leq f_3(x) \leq \dots$$

$$b) f(x) = \text{Sup}_\nu f_\nu(x) \quad \text{für alle } x \in \mathbb{R}^n$$

gibt.

Im Folgenden schreiben wir einfach $f_n \uparrow f$, wenn die Eigenschaften a) und b) erfüllt sind.

(Die Bezeichnung $f_n \downarrow f$ versteht sich von selbst.)

Es ist klar, dass die Funktionen aus B^+ den Wert ∞ annehmen können, aber auf keinen Fall $-\infty$. Man kann Funktionen aus der Klasse B^+ addieren, ohne diese Klasse zu verlassen, aber wenn f in B^+ enthalten ist, braucht $-f$ noch lange nicht in B^+ enthalten zu sein. Die Menge B^+ ist also kein Vektorraum. Immerhin gilt noch

$$f \in B^+, C \geq 0 \implies Cf \in B^+.$$

Bemerkung A.13. Es seien $f \in B^+$ und $(f_\nu), (g_\nu)$ zwei Folgen von Funktionen aus C_c (stetige Funktionen mit kompaktem Träger) mit der Eigenschaft

$$f_\nu \uparrow f, \quad g_\nu \uparrow f.$$

Dann gilt

$$\text{Sup}_\nu \left(\int_{\mathbb{R}^n} f_\nu(x) dx \right) = \text{Sup}_\nu \left(\int_{\mathbb{R}^n} g_\nu(x) dx \right).$$

Diese Bemerkung - wir werden sie gleich beweisen - gibt dann Anlass zu

Definition A.14. Es sei $f \in B^+$. Das Integral von f wird durch die Formel

$$\int_{\mathbb{R}^n} f(x) dx = \text{Sup}_\nu \left(\int_{\mathbb{R}^n} f_\nu(x) dx \right)$$

definiert, wobei $f_\nu \in C_c(\mathbb{R}^n)$ irgend eine Folge von Funktionen (stetig mit kompaktem Träger) ist, die f monoton wachsend approximiert, d.h. $f_\nu \uparrow f$.

Die Bemerkung A.13 besagt gerade, dass diese Definition unabhängig von der Wahl der Folge (f_ν) ist. Außerdem beachte man, dass eine Funktion $f \in C_c$ auch in B^+ liegt, man kann sie durch die konstante Folge $f, f, f \dots$ monoton wachsend approximieren.

Also: Es gilt $C_c \subseteq B^+$ und das durch A.14 definierte Integral stimmt auf C_c mit dem früher definierten Integral (A.2) überein.

Wir erinnern an die Bezeichnungen für die Funktionen

$$\begin{aligned} & f \vee g \text{ und } f \wedge g, \text{ die durch} \\ (f \vee g)(x) &= \max(f(x), g(x)), \\ (f \wedge g)(x) &= \min(f(x), g(x)), \end{aligned}$$

definiert sind.

Es gilt: $f, g \in C_c \implies f \vee g$ und $f \wedge g \in C_c$.

Wir zeigen zunächst folgendes:

Sei

$$f_\nu \uparrow f, \quad f_\nu \text{ stetig mit kompaktem Träger,}$$

und sei g irgendeine stetige Funktion mit kompaktem Träger mit der Eigenschaft

$$f \geq g.$$

Dann ist

$$\text{Sup} \int_{\mathbb{R}^n} f_\nu(x) dx \geq \int_{\mathbb{R}^n} g(x) dx.$$

Beweis von A.13: Die Folge $g - (f_\nu \wedge g)$ fällt offenbar monoton gegen Null. Nach dem *Satz von Dini* konvergiert sie daher gleichmäßig gegen Null, und da das Integral stabil gegenüber gleichmäßiger Konvergenz ist (A.10), gilt

$$\lim_{\nu \rightarrow \infty} \int_{\mathbb{R}^n} (f_\nu \wedge g) dx = \int_{\mathbb{R}^n} g(x) dx.$$

Nun ist

$$f_\nu \wedge g \leq f_\nu \quad \text{für alle } \nu,$$

es folgt daher

$$\int_{\mathbb{R}^n} g(x) dx \leq \text{Sup}_\nu \int_{\mathbb{R}^n} f_\nu(x) dx.$$

Es sei nun

$$g_\nu \uparrow f, \quad g_\nu \in C_c.$$

Nach dem, was oben bewiesen wurde, gilt

$$\int_{\mathbb{R}^n} g_\mu(x) dx \leq \text{Sup}_\nu \int_{\mathbb{R}^n} f_\nu(x) dx \quad \text{für jedes } \mu.$$

Insbesondere gilt

$$\text{Sup}_\nu \int_{\mathbb{R}^n} g_\nu(x) dx \leq \text{Sup}_\nu \int_{\mathbb{R}^n} f_\nu(x) dx.$$

Die umgekehrte Ungleichung gilt genauso, da man die Rollen von f_k und g_k vertauschen kann. \square

Dieser Beweis zeigt in Wirklichkeit noch etwas mehr, nämlich

Hilfssatz A.15. Seien $f \leq g$ zwei Funktionen aus B^+ , dann gilt

$$\int_{\mathbb{R}^n} f(x)dx \leq \int_{\mathbb{R}^n} g(x)dx.$$

Rechenregeln für die Bairesche Klasse

Wir nennen manchmal die Funktionen aus B^+ auch (unter-) halbstetig. Das hat seine Berechtigung in folgender Eigenschaft der Funktionen aus B^+ , die für stetige Funktionen wohlbekannt ist:

Bemerkung A.16. Seien $f \in B^+$, C eine reelle Zahl und $a \in \mathbb{R}^n$ ein Punkt mit $f(a) > C$. Dann gilt

$$f(x) > C \quad \text{in einer vollen Umgebung von } a.$$

Beweis: Sei

$$f_\nu \uparrow f, f_\nu \in C_c.$$

Nach Definition des Supremums existiert ein Index l mit

$$f(a) \geq f_l(a) > C.$$

Für die stetige Funktion f_l stimmt aber die Behauptung und damit erst recht für f . Allgemein nennt man Funktionen mit der in Bemerkung A.16 genannten Eigenschaft **unterhalbstetig** (oder **halbstetig nach unten**). Man kann umgekehrt zeigen, dass jede unterhalbstetige Funktion $f : \mathbb{R}^n \rightarrow \mathbb{R}$ mit der Eigenschaft $f(x) \geq 0$ außerhalb eines Kompaktums zu B^+ gehört. \square

Als nächstes zeigen wir, dass das Integral für Funktionen aus B^+ stabil gegenüber monotoner Approximation ist.

Hilfssatz A.17. Sei

$$f_1 \leq f_2 \leq f_3 \leq \dots$$

eine monotone Folge aus B^+ . Dann gilt

a) $f = \text{Sup} f_\nu \in B^+$,

b) $\int_{\mathbb{R}^n} f(x)dx = \text{Sup} \int_{\mathbb{R}^n} f_\nu(x)dx.$

Beweis: Sei

$$f_k \uparrow f, \quad f_k \in B^+ \quad (\text{Bairesche Klasse}).$$

Es existiert also

$$f_{1k} \leq f_{2k} \leq f_{3k} \dots \quad \text{mit } f_k = \text{Sup}_i f_{ik}.$$

Wir bilden die Funktionenfolge

$$g_r = \bigvee_{i+k \leq r} f_{ik} \quad (\text{also } g_r(x) = \max_{i+k \leq r} (f_{ik}(x))).$$

Offenbar gilt $g_r \in C_c$.

Außerdem gelten die Ungleichungen

$$g_1 \leq g_2 \leq g_3 \leq \dots \leq f.$$

Hieraus folgt

$$g = \text{Sup } g_k \leq f.$$

Beachtet man außerdem

$$f_{ik} \leq g_{i+k} \leq g \quad \text{für alle } i, k,$$

so folgt

$$f_k \leq g \quad \text{für alle } k = 1, 2, \dots$$

und daher

$$f = \text{Sup } f_k \leq g,$$

insbesondere also $f = g$.

Damit ist gezeigt:

$$g_k \uparrow f, \quad \text{also } f \in B^+.$$

Außerdem gilt

$$\int_{\mathbb{R}^n} f(x) dx = \text{Sup}_k \int_{\mathbb{R}^n} g_k(x) dx.$$

Aus den Ungleichungen

$$g_k \leq f_k \quad \text{für alle } k$$

folgt (A.15)

$$\int_{\mathbb{R}^n} g(x) dx \leq \int_{\mathbb{R}^n} f_k(x) dx \leq \int_{\mathbb{R}^n} f(x) dx$$

und hiermit

$$\text{Sup} \int_{\mathbb{R}^n} f_k(x) dx = \int_{\mathbb{R}^n} f(x) dx.$$

□

Hilfssatz A.18. Seien $f, g \in B^+$ und a, b nicht negative Zahlen, $a \geq 0$, $b \geq 0$. Dann gilt $af + bg \in B^+$ und

$$\int_{\mathbb{R}^n} (af(x) + bg(x))dx = a \int_{\mathbb{R}^n} f(x)dx + b \int_{\mathbb{R}^n} g(x)dx.$$

Beweis: Seien

$$f_k \uparrow f, g_k \uparrow g, \quad f_k, g_k \in C_c.$$

Offenbar gilt

$$af_k + bg_k \uparrow af + bg \quad (\text{beachte: } a > 0, b > 0!)$$

und die Behauptung ist evident. □

Hier sieht man auch den wesentlichen Nachteil des Integrals für B^+ : Leider gilt im Allgemeinen nicht

$$f \in B^+ \implies -f \in B^+.$$

A.5 Der Daniell-Lebesgue-Prozess, Teil 2: Das äußere Integral

Wir ordnen nun **jeder** Funktion $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ ein gewisses äußeres Integral zu, das wir mit

$$\int^- f(x)dx_1 \dots dx_n = \int^- f(x)dx$$

bezeichnet werden. Zur Motivation nehmen wir einige Eigenschaften vorweg.

- 1) Wenn f in der Klasse B^+ liegt, dann stimmt das äußere Integral mit dem früher definierten Integral überein.
- 2) Das äußere Integral ist im Allgemeinen nicht additiv, es gilt z.B. im Allgemeinen nicht

$$\int^- f(x)dx = - \int^- (-f(x))dx.$$

Man nennt manchmal auch

$$\int_- f(x)dx := - \int^- (-f(x))dx$$

das *innere Integral* von $f(x)$.

Eine Funktion $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ heißt (im Lebesgue'schen Sinn) integrierbar, wenn äußeres und inneres Integral von f übereinstimmen und wenn diese einen endlichen Wert ($\neq \infty, -\infty$) annehmen.

Die Klasse $\mathcal{L}^1(\mathbb{R}^n)$ der Lebesgue-integrierbaren Funktionen hat dann alle die Eigenschaften, die man von einem „vernünftigen“ Integral erwartet. Diese werden dann in A.5 formuliert und bewiesen.

Definition A.19. Das *äußere Integral* (oder *Oberintegral*) einer Funktion

$$f : \mathbb{R}^n \rightarrow \mathbb{R}$$

wird durch die Formel

$$\int^- f(x) dx = \text{Inf}_g \left\{ \int_{\mathbb{R}^n} g(x) dx \right\}$$

definiert, wobei $g : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ alle Funktionen der Klasse B^+ mit der Eigenschaft

$$g(x) \geq f(x) \quad \text{für alle } x \in \mathbb{R}^n$$

durchläuft.

Fast unmittelbar aus der Definition kann man einige Eigenschaften des äußeren Integrals ableiten, die für das Folgende wichtig sind.

Zunächst wollen wir noch klarstellen, dass das Oberintegral überhaupt wohldefiniert ist. Da wir die Werte ∞ und $-\infty$ zulassen, muss dazu nur gezeigt werden, dass die Menge der Funktionen

$$g \in B^+ \text{ mit } g \geq f \quad (\text{d.h. } g(x) \geq f(x) \text{ für alle } x \in \mathbb{R}^n)$$

nicht leer ist. Dazu beachten wir einfach

Hilfssatz A.20. Die Funktion

$$f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}, \quad f(x) = \infty \quad \text{für alle } x \in \mathbb{R}^n$$

gehört der Klasse B^+ an.

Wir deuten den Beweis nur im Falle $n = 1$ an. Es ist dann sehr einfach, ihn auf den Fall $n > 1$ zu verallgemeinern. Die Funktion f kann approximiert werden durch die Folge von „Dreiecken“

$$f_k(x) = \begin{cases} -|x| + k & \text{für } |x| \leq k, \\ 0 & \text{für } |x| \geq k. \end{cases}$$

Damit ist also das äußere Integral wohldefiniert. Das äußere Integral ist erfreulicherweise ordnungstreu.

Hilfssatz A.21. *Es seien zwei Funktionen*

$$f, h : \mathbb{R}^n \rightarrow \overline{\mathbb{R}} \text{ mit } f \leq h$$

gegeben. Dann gilt für das äußere Integral

$$\int^- f(x) dx \leq \int^- h(x) dx.$$

Beweis: Ist $g \in B^+$ eine Funktion der Baireschen Klasse mit der Eigenschaft $g \geq h$, so gilt erst recht $g \geq f$. Bei der Definition des Oberintegrals h werden also weniger Funktionen zur Konkurrenz zugelassen als bei f . Dieses wird daher höchstens größer.

Für Funktionen aus der Baireschen Klasse B^+ bringt das äußere Integral nichts Neues. \square

Hilfssatz A.22. *Für jede Funktion $f \in B^+$ gilt*

$$\int_{\mathbb{R}^n} f(x) dx = \int^- f(x) dx = - \int^- (-f(x)) dx.$$

Beweis: 1. *Teil:* Die Gleichung

$$\int_{\mathbb{R}^n} f(x) dx = \int^- f(x) dx$$

ist klar, denn dann ist sogar

$$\int_{\mathbb{R}^n} f(x) dx = \min_{g \in B^+, g \geq f} \left\{ \int_{\mathbb{R}^n} g(x) dx \right\},$$

da unter den Funktionen $g \in B^+, g \geq f$ die Funktion f selbst vorkommt.

2. *Teil:* Die Ungleichung

$$- \int^- (-f(x)) dx \geq \int_{\mathbb{R}^n} f(x) dx :$$

(Beachte: wenn $f \in B^+$ enthalten ist, so braucht dies nicht für $-f$ zuzutreffen.)

Sei (f_ν) eine Folge von stetigen Funktionen mit kompaktem Träger, die f monoton wachsend approximiert

$$f_\nu \in C_c, \quad f_\nu \uparrow f.$$

Dann gilt

$$-f_\nu \geq -f,$$

also

$$\int_{\mathbb{R}^n} (-f_\nu)(x) dx \geq \int^- (-f)(x) dx$$

oder

$$-\int^- (-f(x)) dx \geq \int_{\mathbb{R}^n} f_\nu(x) dx.$$

Da dies für alle ν gilt, folgt die behauptete Ungleichung.

3. Teil: Die Ungleichung

$$\int^- (-f(x)) dx \geq - \int_{\mathbb{R}^n} f(x) dx.$$

Nach Definition des Oberintegrals als ein Infimum bedeutet dies nichts anderes als

$$\int_{\mathbb{R}^n} g(x) dx \geq - \int_{\mathbb{R}^n} f(x) dx,$$

wobei g alle Funktionen

$$g \in B^+ \quad \text{mit} \quad -f \leq g$$

durchläuft.

Dies wiederum ist äquivalent zu

$$\int_{\mathbb{R}^n} g(x) dx + \int_{\mathbb{R}^n} f(x) dx = \int_{\mathbb{R}^n} (g(x) + f(x)) dx \geq 0,$$

was aber wegen $g + f \geq 0$ trivial ist. □

Die Gleichung

$$\int^- f(x) dx = - \int^- (-f(x)) dx$$

ist für beliebige Funktionen f nicht richtig, wie komplizierte Gegenbeispiele zeigen, auf die wir hier nicht eingehen wollen.

Bezeichnung A.23 (Inneres Integral oder Unterintegral).

$$\int_- f(x) dx := - \int^- (-f(x)) dx.$$

Allgemein gilt noch die folgende Ungleichung

Hilfssatz A.24. Für jede Funktion $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ gilt

$$\int_{-} f(x)dx \leq \int_{-}^{-} f(x)dx.$$

Beweis: Diese Ungleichung ist äquivalent zu

$$\int_{-} f(x)dx + \int_{-}^{-} (-f(x))dx \geq 0,$$

wenn man die beiden Fälle

$$\int_{-} f(x)dx = \pm\infty \quad \text{und gleichzeitig} \quad \int_{-}^{-} (-f(x))dx = \mp\infty$$

ausschließt. In diesen Ausnahmefällen gilt sogar $\int_{-}^{-} = \int_{-}$. Nach Definition des Oberintegrals als ein Infimum ist diese Behauptung äquivalent zu

$$\int_{\mathbb{R}^n} g(x)dx + \int_{\mathbb{R}^n} h(x)dx \geq 0$$

für alle halbstetigen Funktionen

$$g, h \in B^{+}, \quad \text{mit } g \geq f, h \geq -f.$$

Hieraus folgt $g + h \geq 0$ und somit ist alles klar, denn für halbstetige Funktionen aus B^{+} gilt ja

$$\int_{\mathbb{R}^n} g(x)dx + \int_{\mathbb{R}^n} h(x)dx = \int_{\mathbb{R}^n} [g(x) + h(x)]dx.$$

□

Definition A.25. Eine Funktion $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ heißt integrierbar⁴, wenn

$$\int_{-} f(x)dx = \int_{-} f(x)dx$$

gilt und wenn dieser Wert endlich ist (also $\neq \infty, -\infty$).

⁴im Sinne von Lebesgue

Bezeichnung: Wenn eine Funktion f integrierbar ist, so setzen wir

$$\int_{\mathbb{R}^n} f(x)dx = \int^- f(x)dx = \int_- f(x)dx.$$

Diese Bezeichnung ist gerechtfertigt (Hilfssatz A.21).

Eine offensichtliche, den Begriff des Supremums umgehende Umformulierung des Begriffs der Integrierbarkeit ist:

Bemerkung A.26. Eine Funktion $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ ist genau dann integrierbar, wenn es zu jedem $\varepsilon > 0$ zwei Funktionen g, h der Baireschen Klasse mit folgenden beiden Eigenschaften gibt:

a) $g(x) \geq f(x), \quad h(x) \geq -f(x),$

b) $\int_{\mathbb{R}^n} (g + h)dx \leq \varepsilon.$

Aus dieser Umformulierung leiten wir eine wichtige Konsequenz ab.

Satz A.27. Sei f eine integrierbare Funktion. Dann sind auch die Funktionen f^+ und f^- integrierbar.

Wir erinnern an die Bezeichnungen

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

Beweis: Wenn f eine Funktion der Baireschen Klasse ist, so trifft dies auch für die beiden Funktionen f^+ und $-f^-$ zu (weil der entsprechende Sachverhalt für stetige Funktionen mit kompaktem Träger gilt und weil die Bildungen ordnungstreu sind). Mit den Bezeichnungen von A.26 gilt $g^+(x) \geq f^+(x)$ und $-h^-(x) \geq -(-f)^-(x) = -f^+(x)$. Außerdem gilt $g^+(x) - h^-(x) \leq g(x) + h(x)$. Diese Ungleichung ist klar, wenn $g(x) \geq 0$, denn dann ist $g(x) = g^+(x)$ und außerdem gilt stets $-h^-(x) \leq h(x)$. Im Falle $g(x) < 0$ gilt $f(x) < 0$ und $g(x)$ und $h^-(x)$ sind beide 0. Andererseits ist $g(x) + h(x) \geq 0$ für alle x (auch in den Unendlichkeitsstellen von f). \square

A.6 Die integrierbaren Funktionen

In diesem und im nächsten Paragraphen wird sich zeigen, dass das Lebesgue-Integral alle Eigenschaften hat, die man von einem „vernünftigen“ Integral erwartet.

Im Folgenden werden wir häufig nur noch Funktionen $f : \mathbb{R}^n \rightarrow \mathbb{R}$ betrachten, die also nur endliche Werte annehmen. Die Werte $+\infty$ und $-\infty$ haben nur noch für die technische Durchführung der Theorie Bedeutung. Außerdem werden wir noch zeigen (vgl. A.7), dass eine *integrierbare* Funktion $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ in den Unendlichkeitsstellen (das sind Stellen $x \in \mathbb{R}^n$ mit $f(x) = +\infty$ oder $-\infty$) beliebig abgeändert werden kann, ohne dass der Wert des Integrals verändert wird. Die Unendlichkeitsstellen einer integrierbaren Funktion haben kein positives Volumen.

Bezeichnung. Menge der integrierbaren Funktionen ohne Unendlichkeitsstellen:

$$\mathcal{L}^1 = \mathcal{L}^1(\mathbb{R}^n) = \{f : \mathbb{R}^n \rightarrow \mathbb{R} : f \text{ integrierbar}\}.$$

Satz A.28. *Die integrierbaren Funktionen aus \mathcal{L}^1 bilden einen Vektorraum und das Integral ist ein lineares Funktional, d.h. also:*

$$f, g \in \mathcal{L}^1 \implies f + g \in \mathcal{L}^1 \text{ und } cf \in \mathcal{L}^1 \text{ für } c \in \mathbb{R},$$

außerdem

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

und

$$\int_{\mathbb{R}^n} cf(x) dx = c \int_{\mathbb{R}^n} f(x) dx.$$

Der Beweis der Additivität beruht auf einer Ungleichung für das äußere Integral:

Hilfssatz A.29. *Es seien*

$$f, g, h : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$$

drei Funktionen mit der Eigenschaft

$$f(x) + g(x) = h(x),$$

falls die Summe $f(x) + g(x)$ wohldefiniert ist. (Wenn also $f(x) = \infty, g(x) = -\infty$ oder $f(x) = -\infty, g(x) = \infty$ gilt, so wird nichts gefordert, denn dann ist die Summe $f(x) + g(x)$ nicht definiert. Es ist dann gleichgültig, was $h(x)$ für einen Wert annimmt.)

Dann gilt

$$\int^- f(x) dx + \int^- g(x) dx \geq \int^- h(x) dx,$$

falls die Summe auf der linken Seite definiert ist. Für das Unterintegral gilt eine entsprechende Ungleichung in der anderen Richtung.

Im Spezialfall $g = -f$, $h = 0$ ist dies nichts anderes als Hilfssatz A.22. Der Beweis von A.29 erfolgt in Analogie zu dem von A.22.

Beweis: Man muss zeigen, dass für alle halbstetigen Funktionen

$$f^*, g^* \in B^+ \quad \text{mit } f^* \geq f, g^* \geq g$$

gilt

$$\int_{\mathbb{R}^n} f^*(x) dx + \int_{\mathbb{R}^n} g^*(x) dx \geq \int_{\mathbb{R}^n} h(x) dx.$$

Dies ist aber trivial, denn es gilt $f^* + g^* \geq h$ (auch in den Stellen, in denen $f(x) + g(x)$ nicht definiert ist!)

Der Beweis von A.28 ist eine triviale Folgerung aus A.29. Es gilt sogar mehr als wir in A.28 formuliert haben. Die in A.28 formulierte Additivität gilt auch für integrierbare Funktionen f, g mit eventuellen Unendlichkeitsstellen. Definiert man die Funktion h durch die Bedingungen $h(x) = f(x) + g(x)$ in allen Stellen x , in denen die Summe definiert ist und definiert man $h(x)$ an allen anderen Stellen beliebig - etwa $= 0$, so folgt aus A.29, dass auch h integrierbar ist und dass das Integral von h gleich der Summe der Integrale von f und g ist. In der Regel kommen wir mit der glatteren Variante A.28 aus. In Beweisen werden wir jedoch gelegentlich auf diese durch A.29 gegebene Verschärfung zurückgreifen.

Noch einfacher sieht man, dass

$$f \in \mathcal{L}^1 \Rightarrow cf \in \mathcal{L}^1 \text{ für } c \in \mathbb{R}$$

und

$$\int_{\mathbb{R}^n} cf(x) dx = c \int_{\mathbb{R}^n} f(x) dx \text{ für positive } c > 0.$$

Dies folgt unmittelbar aus der Definition des äußeren Integrals, wenn man beachtet, dass die entsprechende Gleichung für halbstetige Funktionen aus B^+ gilt. Der Übergang zu negativen c bei integrierbaren Funktionen f erfolgt über die Rechenregel

$$\int_{\mathbb{R}^n} (-f(x)) dx = - \int_{\mathbb{R}^n} f(x) dx.$$

□

Satz A.30. *Ist $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ eine integrierbare Funktion, so ist auch ihr Betrag $|f|$ integrierbar. Ist $g : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ eine weitere integrierbare Funktion, so sind auch die Maxima und Minima $f \vee g$ und $f \wedge g$ integrierbar.*

Zum Beweis der ersten Aussage verwende man die Formel $f(x) = f^+(x) - f^-(x)$. Wir wissen, dass die Funktionen f^\pm integrierbar sind (A.27). Die Behauptung folgt aus der Linearität des Integrals, wobei man die verschärfte Fassung A.20 benötigt. Die beiden restlichen Aussagen beweist man analog.

Die Stärke des Lebesgue-Integrals liegt in seiner Stabilität gegenüber Grenzprozessen.

Satz A.31 (Beppo Levi). *Gegeben sei eine monoton wachsende Folge*

$$f_1 \leq f_2 \leq f_3 \leq \dots$$

von integrierbaren Funktionen aus $\mathcal{L}^1(\mathbb{R}^n)$. Die Folge der Integrale

$$\int_{\mathbb{R}^n} f_\nu(x) dx$$

sei beschränkt. Die durch

$$f(x) = \sup_{\nu \in \mathbb{N}} f_\nu(x)$$

definierte Funktion $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ ist integrierbar und es gilt

$$\int_{\mathbb{R}^n} f(x) dx = \lim_{\nu \rightarrow \infty} \int_{\mathbb{R}^n} f_\nu(x) dx.$$

Beweis: Wir betrachten die Funktionen

$$h_\nu = f_\nu - f_{\nu-1} \geq 0 \quad (f_0 = 0).$$

Es gilt

$$\sum_{\nu=1}^k h_\nu = f_k.$$

Da jede monotone und beschränkte Folge konvergiert, existieren die Grenzwerte

$$R = \sum_{\nu=1}^{\infty} \int_{\mathbb{R}^n} h_\nu(x) dx = \lim_{k \rightarrow \infty} \int_{\mathbb{R}^n} f_k(x) dx.$$

1) Es gilt

$$R \leq \int_{-} f(x) dx.$$

Beweis: Man beachte die Ungleichung (A.30)

$$\int_{\mathbb{R}^n} f_k(x) dx \leq \int_{-} f(x) dx \quad \text{für } k = 1, 2, \dots$$

2) Es gilt

$$\int_{-} f(x) dx \leq R$$

Der Beweis folgt offenbar aus der Ungleichung

$$\int_{-} f(x) dx \leq \sum_{\nu=1}^{\infty} \int_{-} h_{\nu}(x) dx,$$

welche es nun zu beweisen gilt.

Nach Definition des äußeren Integrals können wir halbstetige Funktionen

$$\bar{h}_{\nu} \in B^{+} \quad \text{mit } \bar{h}_{\nu} \geq h_{\nu}$$

und

$$\int_{\mathbb{R}^n} \bar{h}_{\nu} dx \leq \int_{\mathbb{R}^n} h_{\nu}(x) dx + \frac{\varepsilon}{2^{\nu}}$$

finden ($\varepsilon > 0$ beliebig vorgegeben).

Wir setzen

$$\bar{f} = \sup_k \sum_{\nu=1}^k \bar{h}_{\nu} \geq f.$$

Da das Integral für halbstetige Funktionen stabil gegenüber monotoner Konvergenz ist, folgt

$$\begin{aligned} \int_{-} f(x) dx &\leq \int_{\mathbb{R}^n} \bar{f}(x) dx = \\ &\sum_{\nu=1}^{\infty} \int_{\mathbb{R}^n} \bar{h}_{\nu}(x) dx \leq \sum_{\nu=1}^{\infty} \left\{ \int_{\mathbb{R}^n} h_{\nu}(x) dx + \frac{\varepsilon}{2^{\nu}} \right\}. \end{aligned}$$

Beachtet man

$$\sum_{\nu=1}^{\infty} \frac{\varepsilon}{2^{\nu}} = \varepsilon \quad (\text{geometrische Reihe}),$$

so folgt

$$\int_{-} f(x) dx \leq \sum_{\nu=1}^{\infty} \int_{\mathbb{R}^n} h_{\nu}(x) dx + \varepsilon.$$

Da dies für alle $\varepsilon > 0$ gilt, ist die Behauptung bewiesen. □

Satz A.32 (Lebesgue'scher Grenzwertsatz). *Gegeben sei eine Folge*

$$f_k : \mathbb{R}^n \rightarrow \mathbb{R}, \quad k = 0, 1, 2, \dots,$$

*von integrierbaren Funktionen, die **punktweise** gegen eine Funktion $f : \mathbb{R}^n \rightarrow \mathbb{R}$ konvergiert.*

Es existiere eine Funktion $h : \mathbb{R}^n \rightarrow \mathbb{R}$ mit den Eigenschaften

a) $|f_k| \leq h$ für $k = 1, 2, 3, \dots$,

b) $\int^- h(x) dx < \infty$.

Dann ist auch f integrierbar und es gilt

$$\int_{\mathbb{R}^n} f(x) dx = \lim_{k \rightarrow \infty} \int_{\mathbb{R}^n} f_k(x) dx.$$

Beweis: Aus den Ungleichungen $|f_k| \leq h$ folgt $f \leq h$. Daher ist das äußere Integral von $|f|$ endlich:

$$\int^- |f(x)| dx < \infty.$$

Insbesondere sind daher äußeres und inneres Integral von f endlich ($\neq \pm\infty$). Aus $f \leq |f|$ und $-f \leq |f|$ folgt nämlich

$$\int^- f(x) dx < \infty \quad \text{und} \quad \int^- (-f(x)) dx < \infty,$$

also

$$-\infty < \int_- f(x) dx \leq \int^- f(x) dx < \infty.$$

Wir wollen den Lebesgue'schen Grenzwertsatz auf den Satz von Beppo Levi zurückführen und bilden hierzu

$$\begin{aligned} \mathfrak{g}_k(x) &= \text{Inf}\{f_\nu(x), \nu \geq k\}, \\ \bar{\mathfrak{g}}_k(x) &= \text{Sup}\{f_\nu(x), \nu \geq k\}. \end{aligned}$$

Dann gilt offenbar

$$\begin{aligned} \mathfrak{g}_1 &\leq \mathfrak{g}_2 \leq \mathfrak{g}_3 \leq \dots, \\ \bar{\mathfrak{g}}_1 &\geq \bar{\mathfrak{g}}_2 \geq \bar{\mathfrak{g}}_3 \geq \dots \end{aligned}$$

Aus $f = \lim_{k \rightarrow \infty} f_k$ folgert man leicht

$$\mathfrak{g}_k \uparrow f, \quad \bar{\mathfrak{g}}_k \downarrow f.$$

Als nächstes wird gezeigt, dass die \mathbf{g}_k und \bar{g}_k integrierbare Funktionen sind. Dazu wird der Satz von Beppo Levi ausgenutzt. Bildet man nämlich

$$\underline{G}_{kj} = f_k \wedge f_{k+1} \wedge \dots \wedge f_{k+j}$$

und

$$\bar{G}_{kj} = f_k \vee f_{k+1} \vee \dots \vee f_{k+j}$$

so sind \underline{G}_{kj} und \bar{G}_{kj} integrierbar (A.21) und es gelten die Ungleichungen

$$-h \leq \underline{G}_{kj} \leq f_k \leq \bar{G}_{kj} \leq h.$$

Ferner gilt offensichtlich

$$\underline{G}_{kj} \downarrow \mathbf{g}_k \quad (j \rightarrow \infty)$$

und

$$\bar{G}_{kj} \uparrow \bar{g}_k \quad (j \rightarrow \infty)$$

Daher sind nach dem Satz von Beppo Levi \mathbf{g}_k und \bar{g}_k integrierbar.

Ferner gelten die Ungleichungen

$$h \leq \mathbf{g}_k \leq f_k \leq \bar{g}_k \leq h,$$

also

$$-\int_{\mathbb{R}^n} h(x) dx \leq \int_{\mathbb{R}^n} \mathbf{g}_k(x) dx \leq \int_{\mathbb{R}^n} f_k(x) dx \leq \int_{\mathbb{R}^n} \bar{g}_k(x) dx \leq \int_{\mathbb{R}^n} h(x) dx.$$

Da \mathbf{g}_k und \bar{g}_k monoton gegen f konvergieren, ist f integrierbar und aus der letzten Ungleichung folgt

$$\begin{aligned} \lim_{k \rightarrow \infty} \int_{\mathbb{R}^n} \mathbf{g}_k(x) dx &= \lim_{k \rightarrow \infty} \int_{\mathbb{R}^n} f_k(x) dx \\ &= \lim_{k \rightarrow \infty} \int_{\mathbb{R}^n} \bar{g}_k(x) dx = \int_{\mathbb{R}^n} f(x) dx. \end{aligned}$$

□

Bezeichnung: Sei f eine Funktion, deren Definitionsbereich die Menge $D \subseteq \mathbb{R}^n$ umfasse. Wir definieren

$$\chi_D \cdot f : \mathbb{R}^n \longrightarrow \mathbb{R}$$

durch

$$\chi_D \cdot f(x) = \begin{cases} f(x) & \text{für } x \in D, \\ 0 & \text{für } x \notin D. \end{cases}$$

Die Funktion f heißt über D integrierbar, wenn $\chi_D \cdot f$ integrierbar ist, und man definiert

$$\int_D f(x) dx := \int_{\mathbb{R}^n} \chi_D \cdot f(x) dx.$$

A.7 Integrierbarkeitskriterien

Eine Teilmenge $A \subseteq \mathbb{R}^n$ heißt **endlich messbar**, wenn die charakteristische Funktion

$$\chi_A = \begin{cases} 1 & \text{für } x \in A \\ 0 & \text{sonst} \end{cases}$$

integrierbar ist und man nennt

$$v(A) := \int_{\mathbb{R}^n} \chi_A(x) dx$$

das (Euklidische) Volumen von A .

Satz A.33. *Die charakteristische Funktion*

$$\chi_U : \mathbb{R}^n \rightarrow \mathbb{R}$$

einer offenen Teilmenge $U \subseteq \mathbb{R}^n$ gehört der Klasse B^+ an.

Beweis: 1. *Schritt:* U ist ein offener Quader

$$U = \{x \in \mathbb{R}^n : a_\nu < x_\nu < b_\nu\} \quad (a_\nu \leq b_\nu).$$

Im Falle $n = 1$ approximiert man die charakteristische Funktion in naheliegender Weise von unten durch Trapeze. Im Fall $n > 1$ verfährt man ähnlich.

2. *Schritt.* U ist beliebig. Zunächst zeigen wir, dass U abzählbare Vereinigung von offenen Quadern ist:

$$U = U_1 \cup U_2 \cup U_3 \cup \dots, \quad U_\nu \text{ offene Quader.}$$

Man betrachte hierzu die Menge aller offenen Quader

$$Q \subseteq U; \quad Q = \{x : a_\nu < x_\nu < b_\nu, \quad 1 \leq \nu \leq n\},$$

wobei die Zahlen $a_1, \dots, a_n, b_1, \dots, b_n$ rational sind. Es ist klar, dass die Menge dieser Quader U überdeckt. Außerdem ist diese Menge abzählbar, weil die rationalen Zahlen und damit auch die n -Tupel von rationalen Zahlen abzählbar sind.

Man hat jetzt eine monotone Approximation von χ_U :

$$\chi_{U_1}, \quad \chi_{U_1 \cup U_2}, \quad \chi_{U_1 \cup U_2 \cup U_3}, \dots$$

□

Satz A.34. *Jede stetige und beschränkte Funktion $f : D \rightarrow \mathbb{R}$ auf einem offenen und beschränkten Teil $D \subseteq \mathbb{R}^n$ ist integrierbar.*

Beweis: Man kann annehmen, dass f nirgends negativ ist. Die Funktion

$$f \cdot \chi_D(x) = \begin{cases} f(x) & \text{für } x \in D, \\ 0 & \text{für } x \notin D, \end{cases}$$

gehört dann sogar der Klasse $B^+(\mathbb{R}^n)$ an. Sie wird approximiert durch die Folge

$$f_k \cdot f \cdot \chi_D,$$

wobei $f_k \in C_c$ eine Folge ist, die χ_D monoton approximiert. □

Folgerung A.35.

- a) *Jede beschränkte offene Teilmenge des \mathbb{R}^n ist endlich messbar.*
- b) *Jede kompakte Teilmenge des \mathbb{R}^n ist endlich messbar.*

Beweis: Es ist nur noch b) zu beweisen.

Sei jetzt $K \subseteq \mathbb{R}^n$ kompakt. Wir werden eine Folge von offenen beschränkten Mengen

$$U_1 \supseteq U_2 \supseteq U_3 \supseteq \cdots \supseteq K$$

konstruieren, so dass

$$\bigcap_{\nu=1}^{\infty} U_\nu = K$$

gilt. Dann können wir wieder den Grenzwertsatz (auf die Folge χ_{U_ν}) anwenden. Wir setzen

$$U_\nu = \left\{ x \in \mathbb{R}^n : \|x - y\| < \frac{1}{\nu} \text{ für mindestens ein } y \in K \right\}.$$

Es sei wieder dem Leser überlassen, die gewünschten Eigenschaften zu beweisen. □

Das Lebesgue'sche Integral ist eine Verallgemeinerung des Regin- tegrals

Sei

$$f : [a, b] \rightarrow \mathbb{R}$$

eine Regelfunktion. Dann ist f auch Lebesgue-integrierbar und das Regel- und Lebesgue-Integral stimmen überein. Dies ist für Treppenfunktionen einfach zu zeigen und folgt dann allgemein aus den Grenzwertsätzen.

Allgemeiner gilt:

Sei

$$f : D \rightarrow \mathbb{R}, \quad D \subseteq \mathbb{R} \quad \text{ein Intervall,}$$

eine Funktion auf einem nicht notwendigerweise abgeschlossenen Intervall. Die Einschränkung von f auf jedes geschlossene Intervall sei eine Regelfunktion.

Die Funktion f ist genau dann Lebesgue-integrierbar, wenn $|f|$ uneigentlich integrierbar ist im Sinne der [1], Abschnitt 4.3.

Das uneigentliche Integral und das Lebesgue-Integral stimmen dann überein. Der Beweis ergibt sich leicht aus dem Lebesgueschen Grenzwertsatz.

Die wichtigsten Integrierbarkeitskriterien liegen in den Grenzwertsätzen. Es gibt aber auch eine einfache direkte Charakterisierung der Integrierbarkeit, welche unabhängig von ihren Anwendungen von eigenem theoretischem Interesse ist.

Satz A.36. *Eine Funktion*

$$f : \mathbb{R}^n \rightarrow \mathbb{R}$$

ist genau dann integrierbar, wenn es zu jedem $\varepsilon > 0$ eine Funktion $g \in C_c$ mit

$$\int |f(x) - g(x)| dx < \varepsilon$$

gibt.

(Dieser Satz wird noch in A.8 kommentiert.)

Beweis: Es ist unmittelbar klar, dass die angegebene Bedingung notwendig für die Integrierbarkeit von f ist, denn man findet zunächst eine Bairesche Funktion, deren Integral sich beliebig wenig von dem von f unterscheidet und anschließend eine stetige Funktion mit kompaktem Träger, deren Integral beliebig nahe bei dem der Baireschen Funktion liegt. Wir müssen umgekehrt zeigen, dass diese Bedingung hinreichend ist.

Nach Voraussetzung findet man zu jedem $\varepsilon > 0$ eine stetige Funktion g mit kompaktem Träger und eine Bairesche Funktion h mit den Eigenschaften

$$|f(x) - g(x)| \leq h(x), \quad \int^- h(x) dx < \varepsilon.$$

Die Behauptung folgt nun aus der Charakterisierung A.26 (mit Hilfe der Baireschen Funktionen $h + g$ und $h - g$). \square

A.8 Nullmengen

Definition A.37.

(a) Eine Funktion $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ heißt Nullfunktion, wenn

$$\int^- |f(x)| dx = 0$$

gilt.

(b) Eine Teilmenge $A \subseteq \mathbb{R}^n$ heißt Nullmenge, wenn die charakteristische Funktion χ_A

$$\chi_A(x) = \begin{cases} 1 & \text{für } x \in A \\ 0 & \text{für } x \notin A \end{cases}$$

eine Nullfunktion ist.

Die Ungleichung $\int_- \leq \int^-$ zeigt, dass jede Nullfunktion in der Tat integrierbar ist. Außerdem ist mit f auch h eine Nullfunktion, wenn $|h| \leq |f|$ gilt.

Satz A.38. Eine Funktion $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ ist dann und nur dann Nullfunktion, wenn die Menge $\{x \in \mathbb{R}^n : f(x) \neq 0\}$ eine Nullmenge ist.

Beweis: Sei f eine Nullfunktion. Man wende den Satz von Beppo Levi auf die Folge

$$h_\nu(x) = \begin{cases} \nu|f(x)|, & \text{falls } f(x) \neq \pm\infty \\ \nu, & \text{falls } f(x) = \infty \end{cases}$$

an und erhält

$$\int_{\mathbb{R}^n} h(x) dx = 0,$$

wobei

$$h(x) = \begin{cases} \infty & \text{für } f(x) \neq 0, \\ 0 & \text{für } f(x) = 0. \end{cases}$$

Aus der Ungleichung

$$\chi_A \leq h \text{ mit } A = \{x \in \mathbb{R}^n : f(x) \neq 0\}$$

folgt dann, dass A eine Nullmenge ist.

Umkehrung. Sei $A = \{x \in \mathbb{R}^n : f(x) \neq 0\}$ eine Nullmenge. Der erste Teil dieses Beweises zeigt dann, dass h eine Nullfunktion ist und wegen $|f| \leq h$ ist dann auch f eine Nullfunktion. \square

Bemerkung A.39. Wenn eine Funktion f integrierbar ist, dann ist die Menge der Punkte

$$\{x \in \mathbb{R}^n : f(x) = \infty \text{ oder } -\infty\}$$

eine Nullmenge.

Beweis: Es sei A die Menge der Unstetigkeitsstellen von f und χ_A die charakteristische Funktion, offenbar gilt

$$\chi_A(x) = f(x) + (-f(x)),$$

wann immer diese Summe definiert ist. Nicht definiert sind die Ausdrücke $\infty + (-\infty)$.

Nach A.29 gilt daher

$$\int \chi_A(x) dx \leq \int_{\mathbb{R}^n} f(x) dx - \int_{\mathbb{R}^n} f(x) dx = 0.$$

\square

Satz A.40. Sei f eine integrierbare Funktion und g eine Funktion, welche nur auf einer Nullmenge von f verschieden ist. Dann ist auch g integrierbar und die Integrale von f und g stimmen überein.

Folgerung A.41. Man darf eine integrierbare Funktion in ihren Unendlichkeitsstellen beliebig abändern - etwa zu 0, ohne ihre Integrierbarkeit zu verlieren und den Wert des Integrals zu ändern.

Damit haben die Unendlichkeitsstellen ihre Bedeutung verloren!

Beweis: Man kann annehmen, dass die Funktion g in allen Punkten, in denen sie sich von f unterscheidet, verschwindet. Dann gilt aber $f = g + h$ mit einer Nullfunktion und man kann A.29 verwenden (A.28 genügt nicht, da wir noch Unendlichkeitsstellen zulassen). \square

Was haben wir nun gewonnen? Es wurde gezeigt, dass man integrierbare Funktionen in Nullmengen beliebig abändern kann ohne ihre Integrierbarkeit und den Wert des Integrals zu verändern.

Außerdem wissen wir, dass es auf die Unendlichkeitsstellen einer integrierbaren Funktion beim Integrieren nicht ankommt, denn diese bilden eine Nullmenge.

Satz A.42.

- (a) Jede Teilmenge einer Nullmenge ist eine Nullmenge.
- (b) Jeder Punkt ist eine Nullmenge.
- (c) Sind A_1, A_2, A_3, \dots Nullmengen, so ist auch $\bigcup_{\nu=1}^{\infty} A_{\nu}$ eine Nullmenge.

Beweis: a) und b) sind klar. c) folgert man leicht aus dem Satz von Beppo Levi. □

Insbesondere ist \mathbb{Q} in \mathbb{R} eine Nullmenge, womit noch einmal sehr drastisch gezeigt ist, dass es irrationale Zahlen geben muss.

Weiteres Beispiel einer Nullmenge: Sei U ein Untervektorraum des \mathbb{R}^n mit $\dim U < n$. Dann ist

$$L = x + U \text{ für jedes } x \in \mathbb{R}^n$$

eine Nullmenge.

Übungsaufgabe. Man beweise dies wenigstens für einen achsenparallelen Unterraum, d.h.

$$U = \{(x_1, \dots, x_n) : x_k = \dots = x_n = 0\} \quad (1 \leq k \leq n).$$

(Den allgemeinen Fall kann man auf diesen speziellen mittels der Transformationsformel (A.10) zurückführen.)

A.9 Der Satz von Fubini

In diesem Kapitel soll Satz A.8 über die Vertauschbarkeit der Integrationsreihenfolge auf integrierbare Funktionen verallgemeinert werden. Gegeben sei eine integrierbare Funktion

$$(x, y) \mapsto f(x, y) = f(x_1, \dots, x_n, y_1, \dots, y_m)$$

in $n + m$ Veränderlichen. Wir setzen $z = (x, y)$ und bezeichnen die Volumenelemente im

$$\mathbb{R}^{n+m} \text{ mit } dz, \text{ im } \mathbb{R}^n \text{ mit } dx \text{ und im } \mathbb{R}^m \text{ mit } dy.$$

Es soll eine Formel der Art

$$\int_{\mathbb{R}^{n+m}} f(x, y) dz = \int_{\mathbb{R}^m} \left[\int_{\mathbb{R}^n} f(x, y) dx \right] dy = \int_{\mathbb{R}^n} \left[\int_{\mathbb{R}^m} f(x, y) dy \right] dx$$

bewiesen werden.

Das Problem ist zunächst, dass wir nicht wissen, ob f bei festgehaltenem y bzw. x integrierbare Funktion von x bzw. y ist.

Daher behelfen wir uns zunächst mit äußerem und innerem Integral, also beispielsweise mit

$$x \mapsto \int_{\mathbb{R}^m}^- f(x, y) dy \quad (\text{dies ist eine Funktion von } x).$$

Von dieser Funktion kann man sagen, dass sie integrierbar ist.

Satz A.43 (Fubini). *Gegeben sei eine integrierbare Funktion*

$$f : \mathbb{R}^{n+m} \rightarrow \mathbb{R}, \quad (x, y) \mapsto f(x, y).$$

Dann sind die beiden Funktionen

$$x \mapsto \int_{\mathbb{R}^m}^- f(x, y) dy, \quad y \mapsto \int_{\mathbb{R}^n}^- f(x, y) dx$$

ebenfalls integrierbar (im \mathbb{R}^n bzw. \mathbb{R}^m) und es gilt die Formel

$$\begin{aligned} \int_{\mathbb{R}^{n+m}} f(x, y) dz &= \int_{\mathbb{R}^n} \int_{\mathbb{R}^m}^- f(x, y) dy dx \\ &= \int_{\mathbb{R}^m} \int_{\mathbb{R}^n}^- f(x, y) dx dy. \end{aligned}$$

Dieselbe Formel bleibt richtig, wenn man das äußere Integral durch das innere Integral ersetzt.

Es drängt sich natürlich die Frage auf, ob die Funktion $f(x, y)$ bei festem y (bzw. x) nicht sogar integrierbar ist. Aus A.43 folgt, dass dies bis auf eine Ausnahmemenge vom Maß 0, auf die es bei der Integration nicht ankommt, der Fall ist. Man muss nur beachten, dass aus A.43 beispielsweise

$$\int_{\mathbb{R}^n}^- \left[\int_{\mathbb{R}^m} f(x, y) dy - \int_{\mathbb{R}^m}^- f(x, y) dy \right] dx$$

folgt. Der Integrand ist also eine Nullfunktion. Man kann also die Fubini-Formel auch etwas unpräzise aber doch legitim in der Form

$$\int_{\mathbb{R}^{n+m}} f(x, y) dz = \int_{\mathbb{R}^m} \left[\int_{\mathbb{R}^n} f(x, y) dx \right] dy = \int_{\mathbb{R}^n} \left[\int_{\mathbb{R}^m} f(x, y) dy \right] dx$$

schreiben. Für die Anwendungen ist dieser Zusatz jedoch irrelevant. Man sieht die Integrierbarkeit meist direkt.

Beweis:

1. *Schritt:* $f \in C_c$: Die Behauptung folgt aus A.8.

2. *Schritt.* $f \in B^+(\mathbb{R}^{n+m})$. Es existiert eine Folge

$$f_k \uparrow f, \quad f_k \in C_c.$$

Bei festgehaltenem x gilt dann

$$\int_{\mathbb{R}^m} f(x, y) dy \uparrow \int_{\mathbb{R}^m} f_k(x, y) dy.$$

Nach A.6 ist die Folge dieser Integrale stetig, d.h. das Integral

$$\int_{\mathbb{R}^m} f(x, y) dy$$

liegt in B^+ (als Funktion von x) und außerdem gilt

$$\int_{\mathbb{R}^n} \left[\int_{\mathbb{R}^m} f(x, y) dy \right] dx = \lim_{k \rightarrow \infty} \int_{\mathbb{R}^n} \left[\int_{\mathbb{R}^m} f_k(x, y) dy \right] dx.$$

Da die Behauptung für die stetige Funktion f mit kompaktem Träger f_k richtig ist, gilt sie auch für f .

3. *Schritt.* Sei $h : \mathbb{R}^{n+m} \rightarrow \mathbb{R}$ eine Funktion, $h \in B^+$ und $f \leq h$. Dann gilt

$$\begin{aligned} \int_{\mathbb{R}^{n+m}} h(x, y) dz &\stackrel{2.Schritt}{=} \int_{\mathbb{R}^m} \int_{\mathbb{R}^n} h(x, y) dx dy \\ &= \int_{\mathbb{R}^m}^- \left[\int_{\mathbb{R}^n}^- h(x, y) dx \right] dy \geq \int_{\mathbb{R}^m}^- \left[\int_{\mathbb{R}^n}^- f(x, y) dx \right] dy. \end{aligned}$$

Hieraus folgt mit Hilfe der Definition des Integrals

$$\int_{\mathbb{R}^{n+m}} f(x, y) dz = \int_{h \in B^+, f \leq h} \left\{ \int_{\mathbb{R}^{n+m}} h(x, y) dz \right\}$$

$$\geq \int_{\mathbb{R}^m} \left[\int^- f(x, y) dx \right] dy.$$

Diese Ungleichung kann man auch für $-f$ anstelle von f anwenden. Beachtet man

$$\int^- (-f) dx = - \int^- f dx,$$

so folgt

$$\int_{\mathbb{R}^m} \left[\int^- f(x, y) dx \right] dy \geq \int_{\mathbb{R}^{n+m}} f(x, y) dz.$$

Aus der Ungleichung

$$\int^- \leq \int^-$$

folgt dann

$$\int^- \left[\int^- f(x, y) dx \right] dy \geq \int^- \left[\int^- f(x, y) dx \right] dy \geq \int^- \left[\int^- f(x, y) dx \right] dy.$$

Andererseits haben wir eben die Ungleichungen

$$\int^- \int^- \geq \int^- \int^-$$

erhalten, also folgt

$$\begin{aligned} \int_{\mathbb{R}^{n+m}} f(x, y) dz &= \int^- \left[\int^- f(x, y) dx \right] dy \\ &= \int^- \left[\int^- f(x, y) dx \right] dy. \end{aligned}$$

Da äußeres und inneres Integral von $\int^- f(x, y) dx$ somit übereinstimmen, ist diese Funktion integrierbar, und es gilt die Formel

$$\int_{\mathbb{R}^{n+m}} f(x, y) dz = \int^- \left[\int^- f(x, y) dx \right] dy.$$

Entsprechend werden auch die anderen Formeln bewiesen. □

Eine Anwendung:

Satz A.44. Sei

$$f : \mathbb{R}^n \rightarrow \mathbb{R}, \quad f \geq 0,$$

eine integrierbare Funktion, die keine negativen Werte annimmt.

Sei

$$M = \{(x, y) \in \mathbb{R}^n \times \mathbb{R} : 0 \leq y \leq f(x)\}.$$

Dann ist M endlich messbar und es gilt

$$\text{vol}(M) = \int_{\mathbb{R}^n} f(x) dx.$$

Beweis: Das Volumen ist definiert durch die Formel

$$\text{vol}(M) = \int_{\mathbb{R}^{n+1}} \chi_M(x, y) dz.$$

Man integriere zunächst bei festem x über y und wende den Satz von Fubini an. \square

Beispiel: Hat man im \mathbb{R}^2 einen durch Ungleichungen beschriebenen Bereich

$$A = \{(y, z) : a(y) \leq z \leq b(y), c \leq y \leq d\}$$

mit integrierbaren Funktionen a und b , so gilt für eine über A integrierbare Funktion f

$$\int \int_A f(y, z) dy dz = \int_c^d \int_{a(y)}^{b(y)} f(y, z) dz dy.$$

Setzen wir speziell $f(x) = \chi_A(x)$, so bekommen wir für das Volumen von A die Formel

$$v(A) = \int_c^d (b(y) - a(y)) dy.$$

A.10 Die Transformationsformel

Wir erinnern uns daran, dass eine Funktion

$$f : D \rightarrow \mathbb{R}, \quad D \subseteq \mathbb{R}^n,$$

über D integrierbar heißt, wenn die Funktion

$$\tilde{f} : \mathbb{R}^n \rightarrow \mathbb{R}, \quad \tilde{f}(x) = \begin{cases} f(x) & \text{für } x \in D, \\ 0 & \text{für } x \notin D, \end{cases}$$

integrierbar ist, und wir setzen dann

$$\int_D f(x)dx = \int_{\mathbb{R}^n} \tilde{f}(x)dx.$$

Seien $A, B \subseteq \mathbb{R}^n$ offene Mengen. Unter einem **Diffeomorphismus**

$$\varphi : A \longrightarrow B$$

verstehen wir eine bijektive (=umkehrbare) stetig differenzierbare Abbildung mit nirgends verschwindender Funktionaldeterminante

$$j(\varphi, x) = \det J(\varphi; x) \neq 0 \quad \text{für alle } x \in A.$$

Nach dem Satz für implizite Funktionen ist dann auch φ^{-1} differenzierbar und es gilt

$$J(\varphi; x)^{-1} = J(\varphi^{-1}, \varphi(x)).$$

Satz A.45. *Es sei $u : A \rightarrow B$ ein Diffeomorphismus zwischen offenen Mengen $A, B \subseteq \mathbb{R}^n$. Ist $f : B \rightarrow \mathbb{R}$ eine integrierbare Funktion, so gilt*

$$\int_B f(y)dy = \int_A f(u(x))|j(u, x)|dx.$$

(Insbesondere wird also behauptet, dass $f(u(x))|j(u, x)|$ über A integrierbar ist.)

Beweis der Transformationsformel:

1. *Schritt.* Es sei $f : \mathbb{R}^n \rightarrow \mathbb{R}$ eine stetige Funktion mit kompaktem Träger. Dann gelten die Transformationsformeln

$$a) \quad \int_{\mathbb{R}^n} f(x_1, \dots, x_{i+1}, x_i + x_j, x_{i+1}, \dots, x_n)dx = \int_{\mathbb{R}^n} f(x)dx$$

$$b) \quad \int_{\mathbb{R}^n} f(a_1x_1, \dots, a_nx_n)dx = |a_1 \dots a_n| \int_{\mathbb{R}^n} f(x)dx$$

$$c) \quad \int_{\mathbb{R}^n} f(x_1, \dots, x_n)dx_{\sigma(1)} \dots dx_{\sigma(n)} = \int_{\mathbb{R}^n} f(x)dx_1 \dots dx_n$$

$$d) \quad \int_{\mathbb{R}^n} f(x + b)dx = \int_{\mathbb{R}^n} f(x)dx$$

Dabei sei $b = (b_1, \dots, b_n)$ ein festes n -Tupel.

Beweis: Man benutzt die Transformationsformel im Fall $n = 1$, sowie die Tatsache, dass das Integral für stetige Funktionen mit kompaktem Träger iterativ definiert ist, beispielsweise

$$\int_{\mathbb{R}^2} f(x_1 + x_2) dx = \int_{\mathbb{R}} \left[\int_{\mathbb{R}} f(x_1 + x_2, x_2) dx_1 \right] dx_2.$$

Im inneren Integral macht man die Substitution $t = x_1 + x_2$ und erhält für das innere Integral

$$\int_{\mathbb{R}} f(x_1, x_2) dx_1.$$

2. Schritt. Sei $f : \mathbb{R}^n \rightarrow \mathbb{R}$ stetig mit kompaktem Träger und

$$A = (a_{\nu\mu}) \quad 1 \leq \nu, \mu \leq n$$

eine feste Matrix mit von Null verschiedener Determinante, $b \in \mathbb{R}^n$ ein fester Vektor. Dann ist

$$\int_{\mathbb{R}^n} f(y) dy = \int_{\mathbb{R}^n} f(Ax + b) |\det A| dx.$$

Beweis: Man kann $b = 0$ annehmen (1. Schritt d)). Im Spezialfall, dass die Abbildung A zu den drei Typen

- a) Scherung
- b) $(x_1, \dots, x_n) \rightarrow (a_1 x_1, \dots, a_n x_n)$
- c) Permutation der Variablen

gehört, haben wir das im 1. Schritt erkannt.

Man muss jetzt nur aus der linearen Algebra wissen, dass sich jede lineare Abbildung A , $\det A \neq 0$, als Hintereinanderausführung von Transformationen des Typs a) - c) schreiben lässt. Dies ist nichts anderes als die Tatsache, dass sich jede Matrix A , $\det A \neq 0$, durch elementare Umformungen in die Einheitsmatrix überführen lässt.

Außerdem muss man benutzen, dass die Determinante multiplikativ ist, $\det(A_1 \dots, A_n) = \det A_1 \cdots \det A_n$.

3. *Schritt.* Sei $f : \mathbb{R}^n \rightarrow \mathbb{R}$ beliebig und $A = (a_{\nu,\mu})_{1 \leq \nu,\mu \leq n}$ eine Matrix mit $\det A \neq 0$, $b \in \mathbb{R}^n$. Dann ist

$$\int^- f(y)dy = \int^- f(Ax + b)|\det A|dx.$$

Beweis. Ist $f \in B^+$ und $f_k \uparrow f$ eine monotone Approximation durch Funktionen $f_k \in C_c$, so sind die Funktionen

$$g_k(x) = f_k(Ax + b)$$

ebenfalls stetig mit kompaktem Träger und approximieren monoton $g(x) = f(Ax + b)$. Damit ist die Behauptung für Funktionen $f \in B^+$ zurückgeführt auf den 2. Schritt. Ist f beliebig, so folgt die Behauptung unmittelbar aus der Definition des äußeren Integrals.

Die bisherigen Überlegungen zeigen:

1) Ist f integrierbar, so gilt

$$\int_{\mathbb{R}^n} f(y)dy = \int_{\mathbb{R}^n} f(Ax + b)|\det A|dx.$$

2) Jede Teilmenge $W_0 \subseteq \mathbb{R}^n$ der Form

$$W_0 = a + W, \quad W \subseteq \mathbb{R}^n \text{ ein Untervektorraum, } \dim W < n,$$

ist eine Nullmenge.

(Mit Hilfe einer linearen Transformation macht man W achsenparallel.)

4. *Schritt:* (A, B, u wie in A.45.) Die Funktion $f : \mathbb{R}^n \rightarrow \mathbb{R}$ sei stetig. Sei $a \in A$ fest, $b = u(a)$. Es gelte $f(b) > 0$. Außerdem sei $q > 1$ beliebig, aber fest gewählt. Dann existiert eine Umgebung $a \in U \subseteq A$,

$$U = U_{\mathbb{R}}(a) = \{x : \|x - a\| < R\} \quad (\text{hierbei sei } \|\cdot\| \text{ die Maximumsnorm}),$$

so dass für jeden Würfel

$$W = U_r(x_0) \subseteq U$$

gilt

$$q^{n+4} \int_{u(W)} f(y)dy \geq \int_W f(ux)|j(u, x)|dx.$$

(Es wird nicht gefordert, dass W den gleichen Mittelpunkt wie U hat!)

Beweis. Die Ungleichung wird auf den linearen Fall (3. Schritt) zurückgeführt, indem man $u(x)$ linear approximiert.

$$u(x) = u(A) + J(u, a)(x - a) + r(x).$$

Wir ersetzen u durch

$$u_0(x) = u(a) + J(u; a)(x - a)$$

und erhalten nach dem 3. Schritt

$$\int_{u_0(W)} f(y) dy = \int_W f(u_0x) |j(u, a)| dx.$$

Dabei sei r so klein gewählt, dass der abgeschlossene Würfel W noch in A enthalten ist, dann ist $f(u_0x)$ in W beschränkt und damit wegen A.34 über W integrierbar.

Wir denken uns R immer so klein gewählt, dass

$$f(u_0x) > 0 \quad \text{für } x \in U$$

gilt. (Beachte $u_0(a) = u(a) = b$ und $f(b) > 0$ nach Voraussetzung.) In einer solchen Umgebung kann man

$$\frac{f(ux)}{f(u_0x)}$$

betrachten. Diese Funktion ist stetig und konvergiert gegen 1, wenn x nach a strebt. Daher gilt

$$q > \frac{f(ux)}{f(u_0x)} \quad \text{für } x \in U, \quad r \text{ genügend klein.}$$

Aus demselben Grund gilt

$$q |j(u, a)| \geq |j(u, x)| \quad \text{für } x \in U,$$

wenn man R genügend klein wählt.

Also gilt

$$q^2 \int_{u_0(W)} f(y) dy \geq \int_W f(ux) |j(u, x)| dx.$$

In dieser Ungleichung müsste $u(W)$ anstelle von $u_0(W)$ stehen. Wir werden daher

$$u(W) \quad \text{und} \quad u_0(W)$$

vergleichen.

Behauptung. Wählt man R genügend klein, so gilt $u_0(W^*) \subseteq u(W)$. Dabei sei

$$W^* = \{x : \|x - x^0\| < rq^{-1}\}$$

der Würfel, der aus W durch Schrumpfung um den Faktor q^{-1} entsteht.

Beweis. Dies bedeutet nichts anderes als

$$v(W^*) \subseteq W \quad \text{mit } v = u^{-1}u_0,$$

also

$$\|x - x^0\| < rq^{-1} \Rightarrow \|v(x) - v(x^0)\| < r.$$

Wir schätzen $v(x) - v(x^0)$ nach dem verallgemeinerten Mittelwertsatz der Differentialrechnung ab:

$$\|v(x) - v(x^0)\| \leq \|x - x^0\| \sum_{\mu=1}^n |\partial_{\mu} v_{\nu}(\xi^{(\nu)})|.$$

Dabei ist $\xi^{(\nu)}$ ein Punkt auf der Verbindungsstrecke zwischen x und x^0 . Hieraus folgt

$$\|v(x) - v(x^0)\| \leq \|x - x^0\| < M$$

mit

$$M = \sup_{\xi \in U} \sum_{\mu=1}^n |\partial_{\mu} v_{\nu}(\xi)|.$$

Wir wollen ja U so bestimmen, dass gilt

$$\|v(x) - v(x^0)\| < r \quad \text{falls } \|x - x^0\| < rq^{-1}.$$

Dazu benötigt man ersichtlich die Ungleichung $M \leq q$, d.h.

$$\sum_{\mu=1}^n |\partial_{\mu} v_{\nu}(\xi)| < q \quad \text{für alle } \xi \in U.$$

Nun beachte man, dass nach der Kettenregel für $v = u^{-1} \cdot u_0$ gilt:

$$j(v, x) = \text{Einheitsmatrix},$$

d.h. obige Ungleichung ist im Punkt $\xi = a$ erfüllt (wegen $1 < q$). Sie gilt dann aus Stetigkeitsgründen auch in einer vollen Umgebung von a .

Damit erhalten wir nun die Ungleichung

$$q^2 \int_{u(W)} f(y) dy \geq \int_{W^*} f(u(x)) |j(u, x)| dx.$$

Die Abbildung u_0 ist damit eliminiert, wir müssen allerdings noch die Integrale

$$\int_{W^*} g(x) dx \quad \text{und} \quad \int_W g(x) dx \quad \text{mit} \quad g(x) = f(ux) |j(u, x)|$$

vergleichen.

Wählt man R genügend klein, so gilt

$$g(a) \cdot q > g(x) > g(a)q^{-1} \quad \text{für} \quad x \in U \quad (\text{beachte } q > 1 \text{ und } g(a) > 0).$$

Hieraus folgt

$$\int_{W^*} g(x) dx \geq g(a)q^{-1} = \text{vol}(W_{rq^{-1}}) = g(a)q^{-1} \cdot q^{-n}(2r)^n;$$

andererseits ist

$$\int_W g(x) dx \leq g(a) \cdot q \cdot \text{vol}(W) = g(a)q(2r)^n.$$

Aus den beiden Ungleichungen folgt

$$q^{n+2} \int_{W^*} g(x) dx \geq \int_W g(x) dx.$$

Wir erhalten die gewünschte Ungleichung

$$q^{n+4} \int_{u(W)} f(y) dy \geq \int_W f(ux) |j(u, x)| dx$$

für $W \subseteq U = U_R(A)$, R genügend klein.

5. Schritt, Konstruktion einer geeigneten Würfelüberdeckung.

Jede offene Teilmenge $U \subseteq \mathbb{R}^n$ lässt sich als abzählbare Vereinigung von abgeschlossenen Würfeln schreiben

$$U = \overline{W}_1 \cup \overline{W}_2 \cup \overline{W}_3 \cup \dots,$$

wobei die Würfel W_ν offen und paarweise disjunkt sind

$$\overline{W}_\nu = \{x : \|x - a\| \leq \varepsilon, \quad W_\nu = \{x : \|x - a\| < \varepsilon\}.$$

Beweis. Wir betrachten die Menge aller Würfel

$$W = \left\{ x : \frac{a_\nu u}{2^r} < x_\nu < \frac{a_\nu u + 1}{2^r}, \quad \nu = 1, \dots, n \right\},$$

wobei r alle natürlichen und a_ν alle ganzen Zahlen durchläuft.

Die Menge dieser Würfel ist abzählbar. (Die Menge $\mathbb{N} \times \mathbb{Z}^n$ ist abzählbar, weil allgemein das kartesische Produkt von abzählbaren Mengen abzählbar ist.) Man überlegt sich nun (dies im einzelnen durchzuführen sei dem Leser überlassen):

a) Sind W und W' zwei der beschriebenen Würfelmenge, so gilt

$$W \cap W' \neq \emptyset \Rightarrow W \subseteq W' \text{ oder } W' \subseteq W.$$

b) Jeder Punkt $a \in \mathbb{R}^n$ ist enthalten in einem der Würfel W , wobei noch r beliebig groß gewählt werden kann (und daher die Kantenlänge beliebig klein).

Aus a) und b) konstruiert man nun leicht eine Würfelaufteilung der gewünschten Art.

6. Schritt. Die Funktion

$$f : \mathbb{R}^n \longrightarrow \mathbb{R}, \quad f \geq 0,$$

sei stetig mit kompaktem Träger in B . Es gilt

$$\int_B f(y) dy \geq \int_A f(u(x)) |j(u, x)| dx.$$

Beweis indirekt: Die Ungleichung sei falsch. Dann kann jedenfalls f nicht identisch Null sein, aus Stetigkeitsgründen sind beide Integrale nicht Null und man kann eine Zahl $q > 1$ finden, so dass sogar

$$q^{n+4} \int_B f(y) dy < \int_A f(u(x)) |j(u, x)| dx$$

gilt.

Wir betrachten nun die Menge

$$A_0 = \{x \in \mathbb{R}^n : f(u(x)) \neq 0\} \subseteq A.$$

Diese Menge ist offen, da f stetig ist. Nach dem 5. Schritt existiert eine Überdeckung

$$A_0 = \overline{W}_1 \cup \overline{W}_2 \dots,$$

wobei die Würfel W_ν offen und paarweise disjunkt sind.

Behauptung. Für mindestens einen dieser Würfel, nennen wir ihn $W^{(1)}$, gilt die Ungleichung

$$q^{n+4} \int_{u(W^{(1)})} f(y) dy < \int_{W^{(1)}} f(u(x)) |j(u, x)| dx.$$

Beweis. Würde für alle Würfel W die Ungleichung „ \geq “ gelten, so könnte man mit Hilfe des Lebesgueschen Grenzwertsatzes sogar

$$\begin{aligned} q^{n+4} \int_B f(y) dy &\geq \sum_{\nu=1} q^{n+4} \int_{u(W_\nu)} f(y) dy \\ &\geq \sum_{\nu=1} \int_{W_\nu} f(u(x)) |j(u, x)| dx \\ &= \int_A f(u(x)) |j(u, x)| dx \end{aligned}$$

schließen. (Man beachte, dass das Integral über den Rand eines Würfels Null ergibt.)

Die Existenz eines Würfels $W^{(1)}$ mit

$$q^{n+4} \int_{u(W^{(1)})} f(y) dy < \int_{W^{(1)}} f(u(x)) |j(u, x)| dx$$

ist damit gesichert. Den Würfel $W^{(1)}$ kann man durch Halbieren der Kantenlänge in 2^n Würfel aufteilen. Mit der gleichen Schlussweise folgt die Existenz eines Würfels $W^{(2)}$ mit der obigen Ungleichung.

So fortfahrend erhält man eine Folge von Würfeln

$$W^{(1)} \supseteq W^{(2)} \supseteq W^{(3)} \supseteq \dots$$

mit den Eigenschaften

- a) Kantenlänge von $W^{(k)} \rightarrow 0$ für $k \rightarrow \infty$.
- b) $q^{n+4} \int_{u(W^{(k)})} f(y) dy < \int_{W^{(k)}} f(u(x)) |j(u, x)| dx$.

Nach dem verallgemeinerten Intervallschachtelungsprinzip existiert ein Punkt

$$a \in \overline{W^{(1)}} \cap \overline{W^{(2)}} \cap \overline{W^{(3)}} \dots$$

Zu diesem a betrachten wir die im 4. Schritt konstruierte Umgebung U . Wählt man k genügend groß, so gilt

$$W^{(k)} \subseteq U$$

und wir haben einen Widerspruch zwischen den Ungleichungen des 4. und 6. Schritts erhalten.

7. Schritt, Beweis von Theorem A.45 für stetige Funktionen mit kompaktem Träger f .

Man kann annehmen, dass $f \geq 0$ gilt. Dies liegt an der Möglichkeit des Aufspaltens

$$f = f^+ - f^-$$

mit

$$f^+ = f \vee 0, \quad f^- = (-f)^+.$$

Nach dem 6. Schritt gilt die Ungleichung

$$\int_B f(y) dy \geq \int_A f(u(x)) |j(u, x)| dx.$$

Wendet man diese Ungleichung an auf

$$\begin{aligned} & A \text{ anstelle von } B \\ & B \text{ anstelle von } A \\ & u^{-1} \text{ anstelle von } u \\ & f(u(x)) |j(u, x)| \text{ anstelle von } f, \end{aligned}$$

so resultiert die umgekehrte Gleichung.

8. Schritt, Beweis von A.45. Sei

$$f : \mathbb{R}^n \rightarrow \mathbb{R}, \quad f \geq 0$$

eine halbstetige Funktion. Dann ist auch $f \chi_B$ halbstetig. Man wähle

$$f_k \uparrow f$$

und

$$g_k \uparrow \chi_B; \quad g_k \geq 0,$$

wobei f_k und g_k stetig mit kompaktem Träger sei. Dann gilt

$$f_k \cdot g_k \uparrow f \cdot \chi_B.$$

Die Gleichung

$$\int_B f(y) dy = \int_A f(u(x)) |j(u, x)| dx$$

folgt nun aus dem 8. Schritt. Jetzt folgt die Transformationsformel unmittelbar für das äußere Integral und dann erst recht für das Integral. \square

A.11 Messbarkeit

Man kann jede Funktion $f : \mathbb{R}^n \rightarrow \mathbb{R}$ als Limes einer Folge von beschränkten Funktionen mit kompaktem Träger schreiben, nämlich

$$f = \lim f_k$$

mit

$$f_k(x) = \begin{cases} f_k(x), & \text{falls } \|x\| \leq k \text{ und } |f(x)| \leq k, \\ 0, & \text{sonst.} \end{cases}$$

Definition A.46. Eine Funktion $f : \mathbb{R}^n \rightarrow \mathbb{R}$ heißt messbar, falls die Funktionen f_k , $k = 1, 2, \dots$ alle integrierbar sind.

Es gilt offenbar

$$f_k = (\varphi_k \wedge f) \vee (-\varphi_k),$$

wenn φ_k die Funktion

$$\varphi_k(x) = \begin{cases} k & \text{für } \|x\| \leq k \\ 0 & \text{sonst} \end{cases}$$

bezeichnet. Da φ_k integrierbar ist, folgt

Bemerkung A.47. Jede integrierbare Funktion ist messbar.

Diese und die folgende Bemerkung folgen unmittelbar aus dem Lebesgueschen Grenzwertsatz.

Bemerkung A.48. Sei $f : \mathbb{R}^n \rightarrow \mathbb{R}$ eine messbare Funktion. Es existiere eine integrierbare Funktion h mit der Eigenschaft

$$|f(x)| \leq |h(x)| \quad \text{für alle } x.$$

Dann ist auch f integrierbar.

Die messbaren Funktionen haben alle wünschenswerten Stabilitätseigenschaften.

Satz A.49.

- 1) Summe und Produkt von messbaren Funktionen sind messbar, konstante Funktionen sind messbar.

- 2) Seien f, g messbare Funktionen. Dann sind auch $f \wedge g, f \vee g$, insbesondere $|f|$ messbar.
- 3) Sei f_n eine punktweise konvergente Folge messbarer Funktionen. Dann ist auch ihr Grenzwert messbar.
- 4) Man darf eine messbare Funktion auf einer Nullmenge abändern, ohne ihre Messbarkeit zu verlieren.

Beweis: Es genügt zu zeigen, dass das Produkt messbarer Funktionen messbar ist, da alle anderen Stabilitätseigenschaften aus entsprechenden Eigenschaften integrierbarer Funktionen folgen.

Wegen der Formel

$$4(f \cdot g) = (f + g)^2 - (f - g)^2$$

braucht man nur zu beweisen, dass mit f auch f^2 messbar ist. Da man f durch f_k ersetzen darf, genügt es zu zeigen:

Das Quadrat einer beschränkten integrierbaren Funktion ist integrierbar. Dies folgt aus der Charakterisierung A.36 unter Verwendung der Formel $f^2 - h^2 = (f + h)(f - h)$. \square

Als Anwendung des Begriffs der messbaren Funktion dient:

Ergänzung zum Satz von Fubini. Sei $f : \mathbb{R}^{n+m} \rightarrow \mathbb{R}$ eine messbare Funktion, so dass

$$\int^- \int^- |f(x, y)| dy dx$$

endlich ist. Dann ist f integrierbar und es gilt infolgedessen der Satz von Fubini A.43.

Der Beweis ist einfach und wird übergangen.

Als weitere Anwendung des Begriffs der messbaren Funktion führen wir die \mathcal{L}^p - und die L^p -Räume ein. Dabei kann X ein beliebiger lokal kompakter metrischer Raum sein, welcher abzählbar im Unendlichen ist und auf welchem ein Radonsches Maß ausgezeichnet ist. Für uns ist der Fall $X = \mathbb{R}^n$ mit dem Standardmaß ausreichend.

Den Fall $p = 1$ haben wir bereits behandelt (A.2). Auf Beweise gehen wir hier nicht ein. Man findet sie in der Standardliteratur über Maßtheorie. Hier geben wir nur einen Ausblick.

$\mathcal{L}^p(X)$, $p > 0$, bestehe aus allen messbaren Funktionen f , so dass $|f|^p$ integrierbar ist. Wir definieren

$$\|f\|_p = \sqrt[p]{\int_{\mathbb{R}^n} |f|_p dv} \quad \text{für } f \in \mathcal{L}^p(X).$$

Es gilt

- a) $\|cf\| = |c| \|f\|_p$,
- b) $\|f\|_p \geq 0$,
- c) Mit f und g ist auch $f + g$ in $\mathcal{L}^p(X)$ enthalten und es gilt
- d) $\|f + g\|_p \leq \|f\|_p + \|g\|_p$.

Identifiziert man zwei Funktionen, wenn sie sich nur auf einer Nullmenge unterscheiden, so erhält man den Raum

$$L^p(X) = \{[f], f \in \mathcal{L}^p(X)\}, \quad [f] = \{g \in \mathcal{L}^p(X), \|f - g\|_p = 0\}.$$

Die Definitionen

$$[f] + [g] = [f + g], \quad C[f] = [Cf], \quad \|[f]\|_p = \|f\|_p$$

hängen nicht von der Wahl der Repräsentanten ab.

In $L^p(X)$ gilt

$$\|[f]\|_p = 0 \implies [f] = [0],$$

d.h. $L^p(X)$ ist ein normierter Vektorraum. Aus dem Lebesgueschen Grenzwertsatz lässt sich (in nicht offensichtlicher Weise) folgern:

Satz A.50. Die Räume $(L^p(X), \|\cdot\|_p)$ sind für $p > 0$ **Banachräume**. Der Unterraum der Funktionen $[f]$, $f \in C_c(X)$, ist dicht in $L^p(X)$.

Übrigens: Wenn f und g stetige Funktionen mit kompaktem Träger sind, so gilt $[f] = [g] \implies f = g$. Wir können also (leicht unpräzise aber suggestiv) $C_c(X) \subseteq L^p(X)$ schreiben.

Der Fall $p = 2$ ist besonders wichtig. Man kann zeigen, dass durch

$$\langle [f], [g] \rangle := \int_X f(x) \overline{g(x)} dx$$

ein Skalarprodukt auf $L^2(X)$ definiert wird und erhält somit:

Satz A.51. Der Raum $L^2(X)$ ist ein **Hilbertraum** (wenn er mit obigem Skalarprodukt versehen wird).

(Dabei versteht man unter einem **Hilbertraum** einen vollständigen, mit einem Skalarprodukt versehenen \mathbb{R} - oder \mathbb{C} -Vektorraum. Die Forderung der Vollständigkeit bezieht sich stets auf die durch $\|x\| := \sqrt{\langle x, x \rangle}$ definierte Norm.)

Literaturverzeichnis

- [1] H.-J. Bartels Analysis I, Vorlesungsskriptum
- [2] C. Carathéodory: Variationsrechnung und partielle Differentialgleichungen erster Ordnung, Berlin 1935
- [3] J. Dieudonné: Foundations of Modern Analysis, Academic Press 1969
- [4] O. Forster: Analysis 2, Differentialrechnung im \mathbb{R}^n , Gewöhnliche Differentialgleichungen, Vieweg 1977 (oder Neuauflagen)
- [5] E. Freitag: Vorlesungen über Analysis, Teil II, Vorlesungsskriptum, Heidelberg 2001
- [6] H. Grauert/W. Fischer: Differential- und Integralrechnung II, Springer 1973
- [7] H. Heuser: Lehrbuch der Analysis, Teil 2, Teubner Stuttgart 1992
- [8] M. Kneser: Differential- und Integralrechnung II, Vorlesungsskriptum, Göttingen 1977
- [9] S. Lang: Undergraduate Analysis, Springer 1983