

# Angewandte Mathematik

Dr. Johannes Riesterer

11. November 2020



©Johannes Riesterer

## Vorwort

Kann jeder Mathematik lernen? Als Antwort auf diese Frage möchte ich auf den interessanten Lebenslauf einer der bedeutendsten Mathematikerinnen aller Zeiten eingehen (Auszug aus Wikipedia):

Emmy Noether war eine deutsche Mathematikerin, die grundlegende Beiträge zur abstrakten Algebra und zur theoretischen Physik lieferte. Insbesondere hat Noether die Theorie der Ringe, Körper und Algebren revolutioniert. Das nach ihr benannte Noether-Theorem gibt die Verbindung zwischen Symmetrien von physikalischen Naturgesetzen und Erhaltungsgrößen an.

Sie zeigte in mathematischer Richtung keine besondere Frühreife, sondern hatte in ihrer Jugend Interesse an Musik und Tanzen. Sie besuchte die Städtische Höhere Töchterschule – das heutige Marie-Therese-Gymnasium – in der Schillerstraße in Erlangen. Mathematik wurde dort nicht intensiv gelehrt. Im April 1900 legte sie die Staatsprüfung zur Lehrerin der englischen und französischen Sprache an Mädchenschulen in Ansbach ab. 1903 holte sie in Nürnberg die externe Abiturprüfung am Königlichen Realgymnasium – dem heutigen Willstätter-Gymnasium – nach.

# Inhaltsverzeichnis

<b>1 Vorwissen und Konventionen</b>	<b>6</b>
<b>2 Mehrdimensionale Differentialrechnung</b>	<b>7</b>
2.1 Richtungsableitung und Gradient reellwertiger Funktionen . . . . .	9
2.2 Extrema . . . . .	17
2.3 Gradient einer mehrdimensionalen Funktion . . . . .	23
<b>3 Mehrdimensionale Integralrechnung</b>	<b>25</b>
3.1 Lebesgue Maß . . . . .	25
3.2 Lebesgue Integral . . . . .	28
3.2.1 Die $L^1$ -Halbnorm . . . . .	30
3.2.2 Integration über den $\mathbb{R}^n$ . . . . .	31
3.2.3 Transformationssatz . . . . .	36
3.2.4 Parameterabhängige Integrale . . . . .	37
3.2.5 Faltung . . . . .	39
<b>Tabellenverzeichnis</b>	<b>40</b>
<b>Abbildungsverzeichnis</b>	<b>41</b>



# 1 Vorwissen und Konventionen

## Differenzierbarkeit reeller Funktionen

Eine reelle Funktion  $f : (a, b) \rightarrow \mathbb{R}$  heißt differenzierbar in  $x \in (a, b)$ , falls der Grenzwert  $\lim_{h \rightarrow 0} \frac{f(x+h)-f(x)}{h}$  existiert. In diesem Fall heißt dieser Grenzwert die Ableitung (Steigung) von  $f$  in  $x$  und wird mit  $f'(x)$  bezeichnet.



Abbildung 1: Quelle: Wikipedia: [https://de.wikipedia.org/wiki/Datei:Diferencial\\_quotient\\_of\\_a\\_function.svg](https://de.wikipedia.org/wiki/Datei:Diferencial_quotient_of_a_function.svg)

## Mittelwertsatz einer Veränderlichen

Sei  $f : [a, b] \rightarrow \mathbb{R}$  stetig und differenzierbar für alle  $x \in (a, b)$ . Dann gibt es  $\xi \in (a, b)$  mit  $f'(\xi) = \frac{f(b)-f(a)}{b-a}$ .

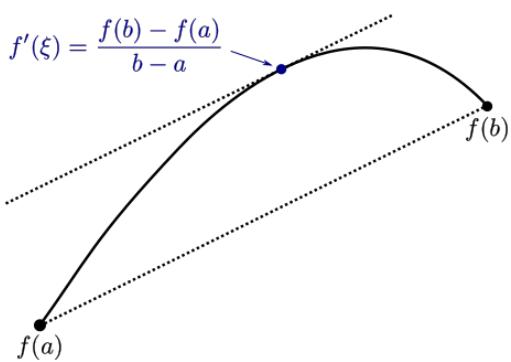


Abbildung 2: Quelle: Wikipedia: <https://commons.wikimedia.org/wiki/File:Mittelwertsatz3.svg>

### Taylorapproximation einer Veränderlichen

Jede reelle Funktion  $f$ , deren  $p+1$ -ten Ableitungen existieren und stetig sind lässt sich mit Hilfe des Taylorpolynoms

$$f(x) = f(a) + f'(a)(x-a) + \frac{1}{2!}f''(a)(x-a)^2 + \cdots + \frac{1}{p!}f^{(p)}(a)(x-a)^p + R_{p+1}(x, a)$$

und dem Restglied  $R_p(x, a) := \frac{1}{(p+1)!}f^{(p+1)}(\xi)(x-a)^{p+1}$  mit einem  $\xi \in (x, a)$  darstellen.

### Cauchy-Schwarzsche Gleichung

Für zwei Vektoren  $v, w \in \mathbb{R}^n$  gilt:

$$\frac{\langle v, w \rangle}{\|v\| \cdot \|w\|} = \cos(\varphi)$$

wobei  $\varphi$  der Innenwinkel zwischen  $v$  und  $w$  ist.

### Äquivalenz von Normen

Die Normen  $\|v\| := \sqrt{\sum_{i=1}^n v_i^2}$  und  $\|v\|_\infty := \max\{|v_1|, \dots, |v_n|\}$  sind Äquivalent. Sie lassen sich mit Konstanten  $k_1\|v\| < \|v\|_\infty k_2\|v\|$  gegeneinander abschätzen.

### Symmetrische Matrizen

Für eine symmetrische Matrix  $A \in \mathbb{R}^{n \times n}$  ist Äquivalent:

- $A$  hat positive Eigenwerte.
- $v^T A v > 0$  für alle  $v \neq 0$ .
- alle Unterdeterminanten sind positiv. Speziell für  $n = 2$  und  $A = \begin{pmatrix} a & b \\ b & d \end{pmatrix}$  bedeutet dies  $a > 0$  und  $ad - b^2 > 0$ .

**Definition 1** (Konventionen). In diesem Abschnitt ist  $U \subset \mathbb{R}^n$  stets eine offene

Teilmenge des  $\mathbb{R}^n$ .  $e_i := \begin{pmatrix} 0 \\ \vdots \\ 1(i\text{-te Zeile}) \\ \vdots \\ 0 \end{pmatrix}$  bezeichnet den  $i$ -ten Basisvektor des  $\mathbb{R}^n$ .

## 2 Mehrdimensionale Differentialrechnung

**Definition 2** (Konvergenz). Eine Folge  $(a_n)$  in  $\mathbb{R}^n$  heißt konvergent gegen den Grenzwert  $a \in \mathbb{R}^n$ , wenn gilt:

$$\forall \varepsilon > 0 \exists N \in \mathbb{N} \forall n > N : d(a, a_n) < \varepsilon$$

In Worten: Es gibt für jedes beliebige (noch so kleine)  $\varepsilon > 0$  einen Index  $N$  derart, dass für alle Indizes  $n > N$ , also alle weiteren Folgenglieder, gilt: Der Abstand  $d(a, a_n)$  ist kleiner als  $\varepsilon$ .



Abbildung 3: Quelle: Wikipedia: [https://commons.wikimedia.org/wiki/File:Epsilonschlauch\\_klein.svg](https://commons.wikimedia.org/wiki/File:Epsilonschlauch_klein.svg)

**Definition 3** (Grenzwert). Sei  $f : X \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$  eine Funktion und  $a \in X$ . Wir nennen  $L_a \in \mathbb{R}^m$  Grenzwert von  $f$  bezüglich der Annäherung von  $x$  an  $a$ , falls für jede konvergente Folge  $x_n \rightarrow a$  die Folge  $f(x_n)$  nach  $L_a$  konvergiert. In diesem Fall bezeichnen wir

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

Dies ist gleichbedeutend damit, dass für jedes  $\varepsilon > 0$  ein  $\delta > 0$  existiert, so dass  $d(f(x), L_a) < \varepsilon$  gilt für jedes  $x$  mit  $d(x, a) < \delta$ .



Abbildung 4: Quelle: Wikipedia: [https://de.wikipedia.org/wiki/Datei:Limes\\_Definition\\_Vektorgrafik.svg](https://de.wikipedia.org/wiki/Datei:Limes_Definition_Vektorgrafik.svg)



Abbildung 5: Quelle: Wikipedia: [https://commons.wikimedia.org/wiki/File:Upper\\_semi.svg](https://commons.wikimedia.org/wiki/File:Upper_semi.svg)

**Definition 4** (Stetige Funktion). Eine reellwertige Funktion  $f : U \rightarrow \mathbb{R}$  heißt stetig, wenn für alle  $y \in U$  der Grenzwert  $\lim_{x \rightarrow y} f(x) = L_y$  existiert.

## 2.1 Richtungsableitung und Gradient reellwertiger Funktionen

Ableitungen beschreiben bildlich gesprochen das Verhalten einer Funktion bezüglich beliebig kleiner Änderungen der Eingabewerte.

**Definition 5** (Richtungsableitung). Sei  $f : U \rightarrow \mathbb{R}$  eine Funktion. Für einen Vektor  $h \in \mathbb{R}^n$ ,  $t \in \mathbb{R}$  und einen Punkt  $a \in U$  heißt der Grenzwert (falls er existiert)

$$\partial_h f(a) := \lim_{t \rightarrow 0} \frac{f(a + th) - f(a)}{t}$$

Richtungsableitung von  $f$  am Punkt  $a$  in Richtung  $h$ . Sie misst die Änderung der Funktion in Richtung  $h$ .

Speziell nennen wir für die Standard Basisvektoren  $e_i$

$$\frac{\partial f(a)}{\partial x_i} := \partial_{e_i} f(a) := \lim_{t \rightarrow 0} \frac{f(a + te_i) - f(a)}{t}$$

die partielle Ableitung von  $f$  in  $a$  nach  $x_i$ .

**Definition 6** (Partielle Differenzierbarkeit). Eine Funktion  $f : U \rightarrow \mathbb{R}$  heißt partiell differenzierbar im Punkt  $a \in U$ , falls alle partiellen Ableitungen

$$\frac{\partial f(a)}{\partial x_1}, \dots, \frac{\partial f(a)}{\partial x_n}$$

existieren.

**Definition 7** (Differenzierbarkeit). Eine Funktion  $f : U \rightarrow \mathbb{R}$  heißt differenzierbar im Punkt  $a \in U$ , falls alle partiellen Ableitungen

$$\frac{\partial f(a)}{\partial x_1}, \dots, \frac{\partial f(a)}{\partial x_n}$$

existieren und stetig sind. Man nennt in diesem Fall die  $1 \times n$ -Matrix

$$df(a) := \left( \frac{\partial f(a)}{\partial x_1}, \dots, \frac{\partial f(a)}{\partial x_n} \right)$$

das Differential von  $f$  im Punkt  $a$ . Der Vektor

$$\nabla f(a) := \begin{pmatrix} \frac{\partial f(a)}{\partial x_1} \\ \vdots \\ \frac{\partial f(a)}{\partial x_n} \end{pmatrix}$$

wird als Gradient bezeichnet. Es ist  $df(a) \cdot h = \langle \nabla f(a), h \rangle$ .

★★★ Der Unterschied zwischen Differenzierbarkeit und partieller Differenzierbarkeit ist also, dass die partiellen Ableitungen zusätzlich zur Existenz auch stetig sein müssen.



Abbildung 6: Quelle: Wikipedia: <https://commons.wikimedia.org/wiki/File:Gradient2.svg>

**Bemerkung 1.** Für das Differential einer differenzierbaren Funktion  $f : U \rightarrow \mathbb{R}$  gilt für alle  $a \in U$ :

- $df(a)(h) := df(a) \cdot h$  ist eine lineare Abbildung von  $\mathbb{R}^n$  nach  $\mathbb{R}$ .
- $df(a) \cdot h = \partial_h f(a)$ .
- $d(f \cdot g) = g(a)d(f) + f(a)dg$
- $d(f + g) = df + dg$

*Beweis.* • Multiplikation mit einer Matrix ist eine lineare Abbildung.

- Für die Basisvektoren ist per Definition  $df(a) \cdot e_i = \partial_{e_i} f(a)$ . Da jeder Vektor  $h$  eine Linearkombination der Basisvektoren ist und  $df$  linear ist, folgt die Behauptung.
- Folgt direkt aus der entsprechenden Eigenschaft reeller Funktionen.
- Folgt direkt aus der entsprechenden Eigenschaft reeller Funktionen.

□

**Satz 1** (Steilste Anstiegsrichtung). Sei  $f : U \rightarrow \mathbb{R}$  differenzierbare Funktion,  $a \in U$  und  $v := \operatorname{argmax}_{h \in S^n} \partial_h f(a)$ . Dann gilt

$$\|\nabla f(a)\| v = \nabla f(a).$$

*Beweis.* Mit der Cauchy-Schwarz Gleichung folgt für beliebiges  $h$

$$\partial_h f(a) = df(a)h = \langle \nabla f(a), h \rangle = \|\nabla f(a)\| \cdot \|h\| \cdot \cos(\varphi)$$

wobei  $\varphi$  den Innenwinkel zwischen  $\nabla f(a)$  und  $h$  bezeichnet. Für  $\|h\| = 1$  wird somit  $\partial_h f(a)$  maximal, wenn  $\varphi = 0$  und somit  $h = \frac{\nabla f(a)}{\|\nabla f(a)\|}$  ist. □

**Satz 2** (Lokale Linearisierung). Ist  $f : U \rightarrow \mathbb{R}$  differenzierbar, so gibt es ein Restglied  $R(h)$  mit  $\lim_{h \rightarrow 0} \frac{R(h)}{\|h\|} = 0$  so dass für alle  $a \in U$  und  $h \in \mathbb{R}$

$$f(a + h) = f(a) + df(a) \cdot h + R(h)$$

gilt.

★★★ Eine differenzierbare Funktion kann auf hinreichend kleinen Umgebungen beliebig genau durch eine lineare Funktion approximiert werden.

♦♦ Der Beweis beruht im Wesentlichen auf dem Mittelwertsatz einer Veränderlichen.

*Beweis.* Wir wählen einen offenen, achsenparallelen Quader  $Q \subset U$ , so dass er vollständig in  $U$  enthalten und  $a \in Q$  ist. Jeder Punkt  $a + h \in Q$  lässt sich damit durch einen achsenparallelen Streckenzug durch die Punkte

$$\begin{aligned} a_0 &:= a \\ a_i &:= a_{i-1} + h_i e_i; \quad i = 1, \dots, n \end{aligned}$$

mit  $a$  verbinden.



Abbildung 7: Kantenzug mit achsenparallelen Kanten

Damit ist  $f(a+h) - f(a) = \sum_{i=1}^n (f(a_i) - f(a_{i-1}))$  und mit  $\varphi_i(t) := f(a_i + t\mathbf{e}_i)$  gilt  $f(a_i) - f(a_{i-1}) = \varphi_i(h_i) - \varphi_i(0)$ . Wegen dem Mittelwertsatz einer Veränderlichen gibt es  $\tau_i$  mit

$$\varphi_i(h_i) - \varphi_i(0) = h_i \varphi'_i(\tau_i).$$

Da  $\varphi'_i(t) = \frac{\partial f(a_{i-1} + t\mathbf{e}_i)}{\partial x_i}$  folgt mit  $\xi_i := a_i + \tau_i \mathbf{e}_i$

$$f(a+h) - f(a) - df(a) \cdot h = \sum_{i=1}^n \left( \frac{\partial f(\xi_i)}{\partial x_i} - \frac{\partial f(a)}{\partial x_i} \right) h_i$$

und damit

$$|f(a+h) - f(a) - df(a) \cdot h| \leq \|h\|_\infty \sum_{i=1}^n \left| \frac{\partial f(\xi_i)}{\partial x_i} - \frac{\partial f(a)}{\partial x_i} \right|.$$

Für  $h \rightarrow 0$  gilt  $\xi_i \rightarrow a$  und da die partiellen Ableitungen stetig sind nach Voraussetzung und alle Normen äquivalent sind folgt

$$\lim_{h \rightarrow 0} \frac{f(a+h) - f(a) - df(a) \cdot h}{\|h\|} = 0$$

und damit die Behauptung.  $\square$

**Bemerkung 2.** Umformuliert bedeutet Satz 2.3, dass für das Differential einer differenzierbaren Funktion

$$\lim_{h \rightarrow 0} \frac{f(a+h) - f(a) - df(a)h}{\|h\|} = 0 \quad (1)$$

gilt. Ist  $L$  eine weitere lineare Abbildung mit  $\lim_{h \rightarrow 0} \frac{f(a+h) - f(a) - L(a)h}{\|h\|}$ , so ist  $L = df$ . Das Differential ist somit eindeutig durch die Eigenschaft der lokalen Linearisierung bestimmt.

*Beweis.* Für  $v \in \mathbb{R}^n$  mit  $\|v\| = 1$  gilt

$$(L(a) - df(a))(v) = \lim_{t \rightarrow 0} (L(a) - df(a)) \left( \frac{tv}{\|tv\|} \right) = \lim_{t \rightarrow 0} \frac{(L(a) - df(a))(tv)}{\|tv\|} = 0$$

Da jeder Vektor als Linearkombination von Einheitsbasisvektoren dargestellt werden kann, folgt die Behauptung.  $\square$

**Definition 8** (Differenzierbarer Weg). Seien  $a, b \in \mathbb{R}$ . Ein Weg ist eine Abbildung

$$\gamma : [a, b] \rightarrow \mathbb{R}^n$$

$$\gamma(t) := \begin{pmatrix} \gamma_1(t) \\ \vdots \\ \gamma_n(t) \end{pmatrix}$$

mit reellen, stetigen Funktionen  $\gamma_i : [a, b] \rightarrow \mathbb{R}$  (damit ist auch  $\gamma$  stetig). Der Weg heißt differenzierbar, falls alle Ableitungen  $\gamma'_i(t)$  existieren. In diesem Fall definieren wir

$$\gamma'(t) := \begin{pmatrix} \gamma'_1(t) \\ \vdots \\ \gamma'_n(t) \end{pmatrix}$$



Abbildung 8: Quelle: Wikipedia: <https://de.wikipedia.org/wiki/EbeneKurve.png>

Im Folgenden gilt:  $I := [a, b]$  mit  $a, b \in \mathbb{R}$  und  $U \subset \mathbb{R}^n$ .

**Satz 3** ((Baby) Kettenregel). Sei  $\gamma : I \rightarrow U$  ein differenzierbarer Weg und  $f : U \rightarrow \mathbb{R}$  eine differenziertere Funktion. Dann ist  $f \circ \gamma : I \rightarrow \mathbb{R}$  differenzierbar und hat die Ableitung

$$\frac{d(f \circ \gamma)}{dt}(t) = df(\gamma(t))\gamma'(t) = \sum_{i=1}^n \frac{\partial f(\gamma(t))}{\partial x_i} \gamma'_i(t)$$

★★★ Das Differential einer differenzierbare Funktion kann als eine Abbildung von Tangenten interpretiert werden.

*Beweis.* Wegen der Differenzierbarkeit des Weges und der Funktion gilt für hinreichend kleine  $k \in \mathbb{R}$  und  $h \in \mathbb{R}^n$

$$\gamma(t+k) = \gamma(t) + k\gamma'(t) + r_1(k)|k|, \text{ mit } \lim_{k \rightarrow 0} r_1(k) = 0$$

$$f(\gamma(t)+h) = f(\gamma(t)) + df(\gamma(t)) \cdot \gamma'(t)h + r_2(h)||h||, \text{ mit } \lim_{h \rightarrow 0} r_2(h) = 0$$

Mit  $h := \gamma(t+k) - \gamma(t)$  folgt

$$f(\gamma(t+k)) = f(\gamma(t)) + df(\gamma(t)) \cdot \gamma'(t)k + R(k)$$

mit dem Restglied

$$R(k) := df(\gamma(t))r_1(k)|k| + r_2(\gamma(t+k) - \gamma(t))||\gamma'(t)k + r_1(k)|k||$$

Da  $\lim_{k \rightarrow 0} R(k) = 0$  folgt die Behauptung.  $\square$

**Satz 4** (Mittelwertsatz). *Sei  $f : U \rightarrow \mathbb{R}$  eine differenziertere Funktion und  $a, b \in U$ , so dass die Verbindungsstrecke von  $a$  nach  $b$  vollständig in  $U$  liegt. Dann gibts es einen Punkt  $\xi \in [a, b]$  mit*

$$f(b) - f(a) = df(\xi)(b - a).$$



Abbildung 9: Quelle: Wikipedia

*Beweis.* Setze  $\gamma(t) := a + t(b - a)$ ,  $t \in [0, 1]$  (Verbindungsgerade zwischen  $a$  und  $b$ ) und  $F := f \circ \gamma : [0, 1] \rightarrow \mathbb{R}$ . Damit ist  $F(1) - F(0) = f(1) - f(0)$  und nach der Kettenregel ist  $F$  differenzierbar. Mit dem Mittelwertsatz einer Veränderlichen gibt es  $\tau \in (0, 1)$  mit  $F(1) - F(0) = F'(\tau) = df(\gamma(\tau))(b - a)$ . Somit ist  $\xi := \gamma(\tau)$  der gesuchte Punkt.  $\square$

**Satz 5** (Satz von Schwarz). *Wenn Für eine Funktion  $f : U \rightarrow \mathbb{R}$  die Ableitungen  $\frac{\partial}{\partial x_i} f(a)$ ,  $\frac{\partial}{\partial x_j} f(a)$  und  $\frac{\partial}{\partial x_i} \frac{\partial}{\partial x_j} f(a)$  existieren und letztere stetig ist, dann existiert auch  $\frac{\partial}{\partial x_j} \frac{\partial}{\partial x_i} f(a)$  und es gilt*

$$\frac{\partial}{\partial x_i} \frac{\partial}{\partial x_j} f(a) = \frac{\partial}{\partial x_j} \frac{\partial}{\partial x_i} f(a)$$

★★★ Bei wiederholten Ableiten spielt die Reihenfolge keine Rolle, wenn eine der Ableitungen existiert und stetig ist.

◇◆ Der Beweis geht in zwei Schritten: Man wendet den Mittelwertsatz einer Veränderlichen auf eine Hilfsfunktion an. Damit kann man die Differenz der Ableitungen abschätzen und zeigen, dass sie beliebig klein wird und damit gleich sind.

*Beweis.* Setze  $\varphi(x, y) := f(a + xe_i + ye_j)$  mit  $(x, y) \in V \subset \mathbb{R}^2$ . Bei hinreichend kleiner Wahl von  $V$  existieren die Ableitungen  $\frac{\partial}{\partial y} \varphi$ ,  $\frac{\partial}{\partial x} \varphi$  und  $\frac{\partial}{\partial y} \frac{\partial}{\partial x} \varphi$

existiert und ist stetig nach Voraussetzung an  $f$ . Es reicht nun zu zeigen, dass  $\frac{\partial}{\partial x} \frac{\partial}{\partial y} \varphi(0, 0)$  existiert und

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

gilt. Sei dazu  $\epsilon > 0$  und  $V' \subset V$  so gewählt, dass  $|\frac{\partial}{\partial y} \frac{\partial}{\partial x} \varphi(x, y) - \frac{\partial}{\partial y} \frac{\partial}{\partial x} \varphi(0, 0)| < \epsilon$  gilt für  $(x, y) \in V'$ . Innerhalb eines achsenparallelen Quaders  $Q \subset V'$  mit Ecken  $(0, 0)$  und  $(h, k)$  setzen wir  $u(x) := \varphi(x, k) - \varphi(x, 0)$ . Zweimaliges Anwenden des Mittelwertsatzes einer Veränderlichen liefert

$$\begin{aligned} u(h) - u(0) &= hu'(0) \\ &= h(\frac{\partial}{\partial x} \varphi(\xi, k) - \frac{\partial}{\partial x} \varphi(\xi, 0)) = hk \frac{\partial}{\partial y} \frac{\partial}{\partial x} \varphi(\xi, \eta). \end{aligned}$$

Damit erhalten wir die Abschätzung

$$\left| \frac{u(h) - u(0)}{hk} - \frac{\partial}{\partial y} \frac{\partial}{\partial x} \varphi(0, 0) \right| < \epsilon.$$

Da

$$\begin{aligned} \lim_{k \rightarrow 0} \frac{u(h) - u(0)}{hk} &= \lim_{k \rightarrow 0} \frac{1}{h} \left( \frac{\varphi(h, k) - \varphi(h, 0)}{k} - \frac{\varphi(0, k) - \varphi(0, 0)}{k} \right) \\ &= \left( \frac{\frac{\partial}{\partial y} \varphi(h, 0) - \frac{\partial}{\partial y} \varphi(0, 0)}{k} \right) \end{aligned}$$

und

$$\lim_{h \rightarrow 0} \left( \frac{\frac{\partial}{\partial y} \varphi(h, 0) - \frac{\partial}{\partial y} \varphi(0, 0)}{k} \right) = \frac{\partial}{\partial x} \frac{\partial}{\partial y} \varphi(0, 0)$$

folgt die Behauptung. □

### Anwendung: Taylorreihe

**Definition 9** ( $C^k$ -Funktion). Eine Funktion  $f : U \rightarrow \mathbb{R}$  für die alle partiellen Ableitungen

$$\frac{\partial}{\partial x_{i_1}} \cdots \frac{\partial}{\partial x_{i_k}} f(a)$$

mit  $i_1 + \cdots + i_k \leq k$  existieren und stetig sind heißt  $C^k$ -Funktion oder  $k$ -mal stetig differenzierbar. Eine  $C^1$ -Funktion ist also eine differenzierbare Funktion.

**Definition 10** (Höhere Ableitungen). Für eine Funktion  $f : U \rightarrow \mathbb{R}$ ,  $a \in U$  und Vektoren  $v^1, \dots, v^p \in \mathbb{R}^n$  heißt

$$d^p f(a)(v^1, \dots, v^p) := \partial_{v^1} \cdots \partial_{v^p} f(a)$$

die  $p$ -te Richtungsableitung von  $f$ . Sie ist wegen dem Satz von Schwarz wohldefiniert. Mit Bemerkung 1 ist

$$d^p f(a)(v^1, \dots, v^p) = \sum_{i_1=1}^n \dots \sum_{i_p=1}^n \frac{\partial}{\partial x_{i_1}} \dots \frac{\partial}{\partial x_{i_p}} f(a) \cdot v_{i_1}^1 \dots v_{i_p}^p.$$

Für einen Vektor  $z \in \mathbb{R}^n$  definieren wir

$$d^p f(a)z^p := d^p f(a) \underbrace{(z, \dots, z)}_{p-\text{mal}}.$$

**Bemerkung 3.** Für  $p = 2$  und  $u, v \in \mathbb{R}^n$  ist

$$d^2 f(a)(u, v) = \sum_{i=1}^n \sum_{j=1}^n \frac{\partial}{\partial x_i} \frac{\partial}{\partial x_j} f(a) v_i u_j$$

und mit

$$f''(a) := \begin{pmatrix} \frac{\partial}{\partial x_1} \frac{\partial}{\partial x_1} f(a) & \dots & \frac{\partial}{\partial x_1} \frac{\partial}{\partial x_n} f(a) \\ \vdots & & \vdots \\ \frac{\partial}{\partial x_n} \frac{\partial}{\partial x_1} f(a) & \dots & \frac{\partial}{\partial x_n} \frac{\partial}{\partial x_n} f(a) \end{pmatrix}$$

ist  $d^2 f(a)(u, v) = u^T \cdot f''(a) \cdot v$ . Die Matrix  $f''(a)$  wird auch Hesse-Matrix genannt. Nach dem Satz von Schwarz ist sie symmetrisch.

**Satz 6** (Taylorapproximation). Sei  $f : U \rightarrow \mathbb{R}$  eine  $\mathcal{C}^{p+1}$ -Funktion und  $x, a \in U$ , so dass die Verbindung zwischen  $x$  und  $a$  in  $U$  liegt. Dann gilt

$$f(x) = f(a) + \sum_{k=1}^p d^k f(a)(x - a)^k + R_{p+1}(x, a)$$

mit dem Restglied  $R_{p+1}(x, a) := \frac{1}{(p+1)!} d^{p+1} f(\xi)(x - a)^{p+1}$  für ein  $\xi \in [a, x]$ .

*Beweis.* Sei  $h := (x - a)$  und  $F(t) := f(a + th)$  mit  $t \in [0, 1]$ . Wiederholte Anwendung der (Baby) Kettenregel mit  $\gamma(t) := a + th$  ergibt

$$\begin{aligned} F'(t) &= \sum_{i=1}^n \frac{\partial}{\partial x_i} f(a + th) h_i \\ F''(t) &= \sum_{j=1}^n \sum_{i=1}^n \frac{\partial}{\partial x_j} \frac{\partial}{\partial x_i} f(a + th) h_i h_j \\ &\vdots \\ F^p(t) &= \sum_{i_1=1}^n \dots \sum_{i_p=1}^n \frac{\partial}{\partial x_{i_1}} \dots \frac{\partial}{\partial x_{i_p}} f(a + th) h_{i_1} \dots h_{i_p}. \end{aligned}$$

Mit der Taylorapproximation für Funktionen einer Veränderlichen folgt

$$F(1) = F(0) + F'(0) + \frac{1}{2!}F''(0) + \cdots + \frac{1}{p!}F^p(0) + R_{p+1}$$

mit dem Restglied  $R_{p+1} := \frac{1}{(p+1)!}F^{p+1}(\tau)$  mit  $\tau \in [0, 1]$ . Da nach Konstruktion  $F(0) = f(a)$  und  $F(1) = f(x)$  folgt insgesamt die Behauptung.  $\square$

**Satz 7** (Qualitative Taylorformel). *Sei  $T_p(x, a) = f(a) + \sum_{k=1}^p d^k f(a)(x-a)^k$  die Taylorapproximation einer  $C^p$ -Funktion. Dann gilt:*

$$\lim_{x \rightarrow a} \frac{f(x) - T_p(x, a)}{\|x - a\|^p} = 0.$$

*Beweis.* Da die partiellen Ableitungen stetig sind, gibt es für jedes  $\epsilon > 0$  ein Radius  $r > 0$ , so dass für alle  $y \in K_r(a)$  gilt

$$\frac{1}{p!}(d^p f(y) - d^p f(a))h^p \leq \epsilon \|h\|_\infty^p.$$

Mit der Taylorapproximation ist

$$f(x) = T_{p-1}(x, a) + \frac{1}{p!}d^p f(\xi)(x-a)^p = T_p(x, a) + \frac{1}{p!}(d^p f(\xi) - d^p f(a))h^p(x-a)^p$$

Mit obiger Abschätzung folgt die Behauptung.  $\square$

## 2.2 Extrema

**Definition 11** (Extremum). *Sei  $f : X \subset \mathbb{R}^n \rightarrow \mathbb{R}$  eine reelle Funktion. Ein Punkt  $a \in X$  heißt lokales Maximum bzw. Minimum, falls eine Umgebung  $U$  von  $a$  existiert, so dass  $f(x) \leq f(a)$  bzw.  $f(x) \geq f(a)$  für alle  $x \in U$  gilt. Liegt einer der beiden Fälle vor, so spricht man von einem lokalen Extremum. Gilt strikt  $f(x) < f(a)$  bzw.  $f(x) > f(a)$ , so nennt man das Extremum isoliert. Ist  $U = X$  so nennt man es auch globales Maximum bzw. Minimum.*



Abbildung 10: Quelle: Wikipedia: <https://en.wikipedia.org/wiki/File:MaximumParaboloid.png>



Abbildung 11: Quelle: Wikipedia: <https://en.wikipedia.org/wiki/File:MaximumCounterexample.png>

**Satz 8** (Notwendiges Kriterium). Ist  $f : U \rightarrow \mathbb{R}$  differenzierbar und hat  $f$  in  $a \in U$  ein lokales Extremum, so gilt

$$\frac{\partial}{\partial x_1} f(a) = \dots = \frac{\partial}{\partial x_n} f(a) = 0.$$

Dies ist gleichbedeutend mit  $df(a) = 0$ .

*Beweis.* Setze  $F_k(t) := f(a + t\mathbf{e}_k)$ . Da  $f$  ein Extremum in  $a$  hat, hat  $F_k$  in einer hinreichend kleinen Umgebung um 0 ein Extremum. Da  $F_k$  eine Funktion einer Veränderlichen ist, gilt  $F'(0) = 0$ . Da  $\frac{\partial}{\partial x_k} f(a) = F'_k(0)$  folgt die Behauptung.  $\square$

**Definition 12** (Kritischer Punkt). Ein Punkt  $a$  mit  $df(a) = 0$  wird als kritischer Punkt Bezeichnet.

**Satz 9** (Hinreichendes Kriterium). Ist  $f : U \rightarrow \mathbb{R}$  eine  $C^2$ -Funktion und ist  $f'(a) = 0$  ein kritischer Punkt für ein  $a \in U$ . Dann gilt:

- $f''(a) > 0 \Rightarrow f$  hat in  $a$  ein isoliertes lokales Minimum.
- $f''(a) < 0 \Rightarrow f$  hat in  $a$  ein isoliertes lokales Maximum.

- $f''(a) \geq 0 \Rightarrow f$  hat in  $a$  einen Sattelpunkt.

*Beweis.* Sei  $f'(a) = 0$  und  $f''(a) > 0$ . Mit der Taylorformel gilt für hinreichend kurze Vektoren  $h \in \mathbb{R}^n$

$$f(a+h) = f(a) + \frac{1}{2} h^T f''(a) h + R(h)$$

mit  $\lim_{h \rightarrow 0} \frac{R(h)}{\|h\|^2} = 0$ . Für  $\|h\| \leq 1$  hat  $h^T f''(a) h$  sein Maximum  $m$  auf dem Einheitskreis  $\{h \in \mathbb{R}^n \mid \|h\| = 1\}$  da  $f''(a) > 0$ .

$$h^T f''(a) h = \|h\| \frac{1}{\|h\|} h^T f''(a) \|h\| \frac{1}{\|h\|} h \geq m \|h\|^2.$$

Wir wählen  $\epsilon$  so klein, dass  $R(h) \leq \frac{m}{2} \|h\|^2$  gilt für  $\|h\| < \epsilon$  (was geht wegen Taylorformel). Damit erhalten wir

$$f(a+h) \geq f(a) + m \|h\|^2.$$

und damit hat  $f$  ein lokales Minimum in  $a$ .

Der Fall  $f''(a) < 0$  wird mit Betrachtung von  $-f$  durch den vorigen Fall bewiesen.

Es sei nun  $f''(a) \geq 0$  und  $v$  mit  $v^T f''(a) v > 0$  und  $w$  mit  $w^T f''(a) w > 0$ . Betrachten wir die Funktionen

$$\begin{aligned} F_v(t) &:= f(a + tv) \\ F_w(t) &:= f(a + tw) \end{aligned}$$

dann ist

$$\begin{aligned} F'_v(t) &= 0; F''_v(0) = v^T f''(a) v > 0 \\ F'_w(t) &= 0; F''_w(0) = w^T f''(a) w < 0 \end{aligned}$$

und somit hat  $F_v$  ein isoliertes lokales Maximum und  $F_w$  ein isoliertes lokales Minimum und damit  $f$  kein lokales Extremum in  $a$ .  $\square$



Abbildung 12: Quelle: Wikipedia: [https://commons.wikimedia.org/wiki/File:Saddle\\_point.svg](https://commons.wikimedia.org/wiki/File:Saddle_point.svg)

### Anwendung: Gradientenverfahren

Wie kann man Minima einer differenzierbaren Abbildung  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  finden? Wir wissen, dass an jedem Punkt  $x_k \in \mathbb{R}^n$  der negative Gradient  $d_k := -\nabla f(x_k)$  in die steilste Abstiegsrichtung zeigt. Die Idee des Gradientenabstieg ist, ein bestimmtes Stück in diese Richtung abzusteigen, damit der Funktionswert kleiner wird, also  $x_{k+1} = x_k + \alpha_k d_k$  zu setzen. Für hinreichend kleines  $\alpha_k$  folgt mit Satz 2.3 über die lokale Linearisierung  $f(x_{k+1}) = f(x_k) + \alpha_k d_k = f(x_k) + \alpha_k df(x_k)d_k + R(\alpha_k d_k)$ . Somit gilt  $f(x_{k+1}) \leq f(x_k)$ , falls  $\nabla f(x_k) \neq 0$  und falls die Folge  $f(x_k)$  beschränkt ist, so gibt es einen Fixpunkt  $x^*$  und dieser ist ein Minimum, da  $\nabla f(x^*) = 0$  gelten muss.

---

#### Algorithm 1 Gradienstabstieg

---

- 1: Initialisiere  $k := 0$  und zufälligen Startwert  $x_0$ .
  - 2: Initialisiere Genauigkeit  $\epsilon > 0$ .
  - 3: **while**  $\|\nabla f(x_k)\| > \epsilon$  do ► So lange kein Extrema vorliegt
  - 4:     Bestimme  $\alpha_k$  mit  $f(x_k + \alpha d_k) = f(x_k) + \alpha_k df(x_k)d_k + R(\alpha_k d_k)$  ► Schrittweite bestimmen
  - 5:     Setze  $x_{k+1} := x_k + \alpha_k d_k$ . ► Absteigen
  - 6:      $k \leftarrow k + 1$  ► Wiederholen
- 

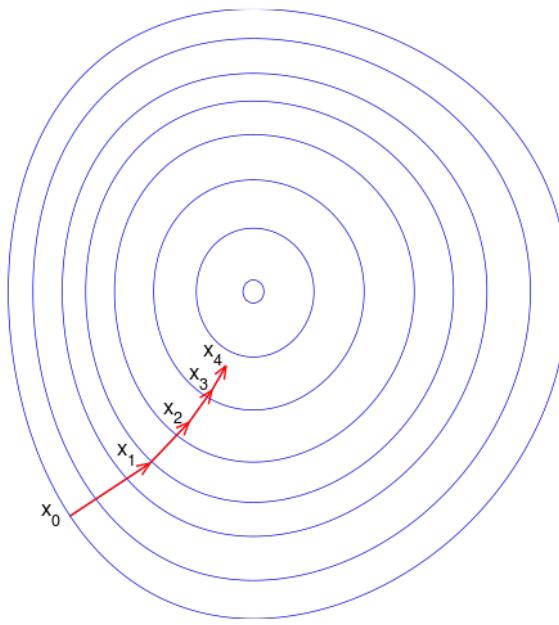


Abbildung 13: Quelle: Wikipedia

**Definition 13.** Sei  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  eine differenzierbare Funktion. Eine Kurve  $\gamma : I \rightarrow \mathbb{R}^n$ , auf der  $f$  konstant ist, also  $f(\gamma(t)) = c$  für eine festes  $c \in \mathbb{R}$  gilt, heißt Höhenlinie.



Abbildung 14: Quelle: <https://getoutside.ordnancesurvey.co.uk/guides/understanding-map-contour-lines-for-beginners/>

**Bemerkung 4.** Der Gradient steht senkrecht auf Höhenlinien. Dies bedeutet, dass

$$\langle \nabla f(\gamma(t)), \gamma'(t) \rangle = 0$$

gilt.

*Beweis.* Aus  $f(\gamma(t)) = c$  folgt  $\frac{d}{dt}f(\gamma(t)) = 0$ . Mit der Kettenregel folgt  $\frac{d}{dt}f(\gamma(t)) = f(\gamma(t)) \cdot \gamma'(t) = 0$  und damit  $\langle \nabla f(\gamma(t)), \gamma'(t) \rangle = 0$ .  $\square$

### Anwendung: Backpropagation

Ein Neuronales Netz ist eine Funktion  $f : \Omega \times \mathbb{R}^n \rightarrow \mathbb{R}^m$ , die einem Input (auch Feature genannt) in Abhängigkeit der Gewichte  $\Omega$  einen Ausgabewert zuordnet. Die Funktion ist dabei aus einfachen Bausteinen zusammengesetzt.



Abbildung 15: Quelle: Wikipedia

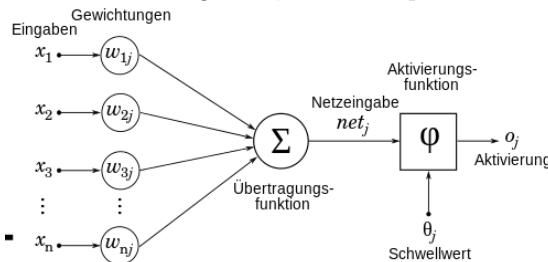


Abbildung 16: Quelle: Wikipedia

Gegeben ist ein Datensatz  $D := \{(x_i, y_i)\}$ . Finde Gewichte Omega, so dass Lossfunktion

$$L_D : \Omega \subset \mathbb{R}^n \rightarrow \mathbb{R}$$

minimal wird. Zum Beispiel  $L_D(\omega) := \sum_{(x_i, y_i) \in D} (f(\omega, x_i) - y_i)^2$ . Wende Gradientenverfahren an:

---

**Algorithm 2** Gradientabstieg

---

- 1: Initialisiere  $k := 0$  und zufällige Gewichte  $\omega_0$ .
  - 2: Initialisiere Genauigkeit  $\epsilon > 0$ .
  - 3: **while**  $\|\nabla L_D(\omega)\| > \epsilon$  **do** So lange kein Extrema vorliegt
  - 4:     Bestimme  $\alpha_k$  mit  $L_D(\omega_k + \alpha_k d_k) = L_D(\omega_k) + \alpha_k d L_D(\omega_k) d_k + R(\alpha_k d_k)$ 
    - Schrittweite bestimmen
    - 5:       Setze  $\omega_{k+1} := \omega_k + \alpha_k d_k$ . ► Absteigen
    - 6:        $k \leftarrow k + 1$  ► Wiederholen
- 

Wenn der Datensatz  $D$  sehr groß ist (Big Data), ist die Berechnung des Gradienten der Lossfunktion entsprechend aufwendig. Um diesen Aufwand zu reduzieren, kann man das Gradientenverfahren modifizieren, so dass man Gradienten nur auf Teilräumen berechnet. Man erhält somit das sogenannte Mini-Batch Gradientenverfahren und das stochastische Gradientenverfahren.

---

**Algorithm 3** Gradientabstieg

---

- 1: Initialisiere  $k := 0$  und zufällige Gewichte  $\mathbf{w}_0$ .
  - 2: Initialisiere Genauigkeit  $\epsilon > 0$ .
  - 3: Wähle Teilmenge  $D'_0 \subset D$
  - 4: **while**  $\|\nabla L_{D'_k}(\omega)\| > \epsilon$  **do** ► So lange kein Extrema vorliegt
  - 5:     Bestimme  $\alpha_k$  mit  $L_{D'_k}(\omega_k + \alpha_k d_k) = L_{D'_k}(\omega_k) + \alpha_k d L_{D'_k}(\omega_k) d_k + R(\alpha_k d_k)$  ► Schrittweite bestimmen
  - 6:     Setze  $\omega_{k+1} := \omega_k + \alpha_k d_k$ . ► Absteigen
  - 7:     Wähle neue Teilmenge  $D'_{k+1} \subset D$ .
  - 8:      $k \leftarrow k + 1$  ► Wiederholen
- 

Wenn man als Teilmenge immer nur eine einelementige Menge wählt, so spricht man vom stochastischen Gradientenverfahren.



Abbildung 17: Quelle: <https://towardsdatascience.com/batch-mini-batch-stochastic-gradient-descent-7a62ecba642a>

### 2.3 Gradient einer mehrdimensionalen Funktion

**Definition 14.** Eine Funktion  $F : U \rightarrow \mathbb{R}^m$  heißt differenzierbar, wenn es eine lineare Abbildung  $dF$  gibt, so dass

$$F(a + h) = F(a) + dF(a)h + R(h)$$

mit  $\lim_{h \rightarrow 0} \frac{R(h)}{\|h\|} = 0$  gilt für alle  $a \in U$  und  $h \in \mathbb{R}^n$ .

**Bemerkung 5.** Im Fall  $n = 1$  stimmt diese Definition mit der alten Definition 7 überein.

*Beweis.* Nach Satz gilt für eine differenzierbare Funktion  $f(a + th) = f(a) + df(a)h + R(th)$  mit  $\lim_{t \rightarrow 0} \frac{R(th)}{\|th\|} = 0$ . Umstellen ergibt

$$df(a)h = \lim_{t \rightarrow 0} \frac{f(a + th) - f(a)}{t}$$

und somit existieren alle linearen Abbildungen und wegen der Linearität von  $dF$  sind diese auch stetig.  $\square$

**Beispiel 1** (Affine Abbildung). Für  $A \in \mathbb{R}^{n \times n}$ ,  $b \in \mathbb{R}^n$  ist die Abbildung  $F(x) := Ax + b$  differenzierbar, da  $F(a+h) = A(a+h) + b = Aa + Ah + b = Aa + b + Ah = F(a) + Ah$  und damit für  $dF(a) := A$  und  $R(h) = 0$  die Definition 14 erfüllt ist.

**Satz 10** (Differenzierbarkeit von Produktfunktionen). Eine Funktion  $F := (F_1, F_2) : U \rightarrow \mathbb{R}^m \times \mathbb{R}^k$  ist genau dann differenzierbar, wenn  $F_1 : U \rightarrow \mathbb{R}^m$  und  $F_2 : U \rightarrow \mathbb{R}^k$  differenzierbar sind. In diesem Fall ist

$$dF(a) = (dF_1(a), dF_2(a)).$$

*Beweis.* Sind  $F_1$  und  $F_2$  differenzierbar, so gilt für  $i = 1, 2$

$$F_i(a+h) = F_i(a) + dF_i h + R_i(h)$$

Dann gilt mit  $dF(a) = (dF_1(a), dF_2(a))$  und  $R(h) := (R_1(h), R_2(h))$

$$F(a+h) = F(a) + dFh + R(h)$$

mit  $\lim_{h \rightarrow 0} \frac{R(h)}{\|h\|} = 0$  und damit ist  $F$  differenzierbar. Die Umkehrung folgt analog.  $\square$

**Bemerkung 6** (Differenzial von Produktfunktionen). Eine Abbildung  $F : U \rightarrow \mathbb{R}^m$  ist genau dann differenzierbar, wenn ihre Koordinaten-Funktionen  $F_1 : U \rightarrow \mathbb{R}, \dots, F_m : U \rightarrow \mathbb{R}$  mit  $F(a) = \begin{pmatrix} F_1(a) \\ \vdots \\ F_m(a) \end{pmatrix}$  differenzierbar sind. In diesem Fall gilt

$$dF(a) := \begin{pmatrix} \frac{\partial}{\partial x_1} F_1(a) & \cdots & \frac{\partial}{\partial x_n} F_1(a) \\ \vdots & & \vdots \\ \frac{\partial}{\partial x_1} F_m(a) & \cdots & \frac{\partial}{\partial x_n} F_m(a) \end{pmatrix}$$

**Bemerkung 7** (Differenzierbarkeit von Wegen). Ein Weg  $\gamma = \begin{pmatrix} \gamma_1 \\ \vdots \\ \gamma_m \end{pmatrix} : I \rightarrow U$  ist genau dann differenzierbar, wenn  $\gamma_i$  differenzierbar ist für  $i = 1, \dots, m$  und dann gilt  $\gamma'(t) = \begin{pmatrix} \gamma'_1(t) \\ \vdots \\ \gamma'_m(t) \end{pmatrix}$ .

**Satz 11** (Kettenregel). Seien  $G : U \subset \mathbb{R}^n \rightarrow V \subset \mathbb{R}^m$  und  $F : V \rightarrow Z \subset \mathbb{R}^k$  differenzierbar. Dann ist  $F \circ G$  differenzierbar und mit  $b := G(a)$  es gilt

$$d(F \circ G)(a) = dF(b) \cdot dG(a)$$

*Beweis.* Analog zu Baby Kettenregel.  $\square$

## Anwendung: Automatisches Ableiten



Abbildung 18: Quelle: Wikipedia

### Automatisches Ableiten in Pytorch

Mit Hilfe des Automatischen Ableitens kann man den Gradienten der Lossfunktion zu einem Neuronalen Netz einfach berechnen, da dieser sich aus einfachen Funktionen zusammensetzt.

## 3 Mehrdimensionale Integralrechnung

### 3.1 Lebesgue Maß

Für offene Intervalle  $(a_i, b_i) \subset \mathbb{R}$  mit  $a_i \leq b_i$  nennen wir  $I := (a_1, b_1) \times \cdots \times (a_n, b_n)$  einen  $n$ -dimensionalen Quader und  $\bar{I} := [a_1, b_1] \times \cdots \times [a_n, b_n]$  seinen Abschluss. Wir definieren das Volumen

$$\text{vol}(I) := \prod_{i=1}^n (b_i - a_i).$$

Mit

$$\mathbb{I}(n) := \{(a_1, b_1) \times \cdots \times (a_n, b_n) \mid (a_i, b_i) \subset \mathbb{R}\}$$

bezeichnen wir die Menge aller  $n$ -dimensionalen Quader und mit

$$\mathbb{I}^0(n) := \{(a_1, b_1) \times \cdots \times (a_n, b_n) \mid (a_i, b_i) \subset \mathbb{R} \text{ und } a_k = b_k \text{ für ein } k\}$$

die Menge der degenerierten Quader. Für eine Menge  $A \subset \mathbb{R}^n$  bezeichnen wir eine Menge von Quadern

$$\left\{ I_j \mid I_j \in \mathbb{I}(n) \text{ mit } A \subset \bigcup_j I_j \right\}$$

als Hüllquader für  $A$ .

**Definition 15** (Lebesguesche äußere Maß). *Für eine Menge  $A \subset \mathbb{R}^n$  definieren wir das Lebesguesche äußere Maß durch*

$$\mu(A) := \inf \left\{ \sum_{j=1}^{\infty} \text{vol}(I_j) ; I_j \in \mathbb{I}(n); A \subset \bigcup_{j=1}^{\infty} I_j \right\}$$

Falls für alle Hüllquader  $\text{vol}(I_j) = \infty$  gilt, so setzen wir  $\mu(A) = \infty$ .

**Definition 16** (Erinnerung Infimum).

**Definition 17** (Nullmenge). Eine Menge  $N \subset \mathbb{R}^n$  mit  $\mu(N) = 0$  heißt Nullmenge.

**Bemerkung 8.** Für  $A \subset B \subset \mathbb{R}^n$  ist  $\mu(A) \leq \mu(B)$

*Beweis.* Da  $A \subset B$  Teilmenge ist, sind Hüllquader von  $B$  auch Hüllquader von  $A$  und damit  $\mu(A) \leq \mu(B)$ .  $\square$

**Satz 12** ( $\sigma$ -subadditivität). Sei  $A_j \subset \mathbb{R}^n$  eine Folge von Mengen. Dann gilt

$$\mu\left(\bigcup_j A_j\right) \leq \sum_{j=1}^{\infty} \mu(A_j)$$

*Beweis.* Für jedes  $A_j$  und  $\epsilon > 0$  können wir eine geeignete Überdeckung  $A_j \subset \bigcup_k K_{j,k}$  mit Hüllquadern  $K_{j,k}$  finden, so dass  $\sum_k \text{vol}(K_{j,k}) \leq \mu(A_j) + \frac{\epsilon}{2^{j+1}}$ . Da  $\bigcup_j A_j \subset \bigcup_j \bigcup_k K_{j,k}$  eine Überdeckung mit Hüllquadern ist, folgt

$$\mu\left(\bigcup_j A_j\right) \leq \sum_j \sum_k \text{vol}(K_{j,k}) \leq \left(\sum_j \mu(A_j) + \frac{\epsilon}{2^{j+1}}\right) = \left(\sum_j \mu(A_j)\right) + \epsilon$$

(Die letzte Gleichung beruht auf dem Wert der geometrischen Reihe). Da die letzte Aussage für beliebiges  $\epsilon > 0$  gilt, folgt die Behauptung.  $\square$

**Bemerkung 9.** Für  $I \in \mathbb{I}(n)$  gilt  $\mu(I) = \text{vol}(I)$ .

*Beweis.* Seien  $I_j \in \mathbb{I}(n)$  mit  $I \subset \bigcup_j I_j$ . Da  $I$  beschränkt und abgeschlossen ist, ist  $I$  kompakt. Damit reichen endlich viele Intervalle, um  $I \subset \bigcup_{j=1}^n I_j$  zu überdecken. Für endlich viele Intervalle ist es einfach zu zeigen, dass

$$\text{vol}(I) \leq \sum_{j=1}^n \text{vol}(I_j)$$

gilt. Damit folgt die Behauptung.  $\square$

**Satz 13.** Für  $I \in \mathbb{I}(n)$  und  $A \subset \mathbb{R}^n$  mit  $I \subset A \subset \bar{I}$  gilt  $\mu(A) = \text{vol}(I)$ .

*Beweis.* Mit  $I_0 := I$  und  $I_j := \emptyset$  ist  $I \subset \bigcup_j I_j$  und damit gilt

$$\mu(I) \leq \sum_j \text{vol}(I_j) = \text{vol}(I)$$

Es sei  $\mathcal{A}_0 := \{A \in \mathbb{R}^n \mid I^0 \subset A \subset \bar{I}^0 \text{ mit } I^0 \in \mathbb{I}^0(n)\}$ . Für  $A_0 \in \mathcal{A}_0$  gibt es  $\epsilon > 0$  und  $I_\epsilon \in \mathbb{I}(n)$  mit  $A \subset I_\epsilon$  und  $\text{vol}(I_\epsilon) \leq \epsilon$  und damit  $\mu(A_0) = 0$ . Zu  $I \in \mathbb{I}(n)$  gibt es  $2n$ -Seiten  $J_j \in \mathcal{A}_0$  mit  $\bar{I} = I \cup \bigcup_{j=1}^{2n} J_j$ . Aus der  $\sigma$ -subadditivität folgt

$$\mu(\bar{I}) \leq \mu(I) + \sum_{j=1}^{2n} \mu(J_j) = \mu(I)$$



Abbildung 19: Äuferes Maß



Abbildung 20: Inneres Maß

Für  $I \subset A \subset \bar{I}$  folgt damit und mit der Monotonie

$$\mu(I) = \mu(A) = \mu(\bar{I}).$$

Mit Bemerkung 9 folgt die Behauptung.  $\square$

Es sei  $A, D \subset \mathbb{R}^n$  beschränkte Teilmenge mit  $A \subset D$ . Man kann einerseits das Volumen  $\mu(A)$ , als auch das Volumen  $\mu_D(A) := \mu(D) - \mu(D \setminus A)$  berechnen. Man approximiert hierbei das Volumen von  $A$  einmal mit Hüllquadern von außen und einmal von innen.

Es ist daher sinnvoll von einer messbaren Menge zu sprechen, wenn diese beiden Volumen übereinstimmen, also  $\mu_D(A) = \mu(A)$  gilt. Dies ist gleichbedeutend mit der Bedingung  $\mu(D) = \mu(A) + \mu(D \setminus A)$ . Für festes  $D$  wären somit alle Mengen  $A$  messbar, für die das äußere Lebesgue Maß additiv ist auf der Zerlegung  $D = A \cup D \setminus A$ . Lässt man die Bedingungen der Beschränktheit und  $A \subset D$  fallen, gelangt man zu folgender Definition.

**Definition 18** ( $\mu$ -messbare Menge). Eine Menge  $A \subset \mathbb{R}^n$  heißt messbar, falls für alle  $D \subset \mathbb{R}^n$  gilt

$$\mu(D) = \mu(A \cap D) + \mu(A^c \cap D)$$

Da wir die  $\sigma$ -subadditivität bereits nachgewiesen haben, können wir diese Bedingung auf

$$\mu(D) \geq \mu(A \cap D) + \mu(A^c \cap D)$$

reduzieren. Die Menge aller messbaren Mengen bezeichnen wir mit  $\mathcal{A}$ .

**Bemerkung 10.** Nullmengen sind messbar.

**Beweis.** Es sei  $N, D \subset \mathbb{R}^n$  mit  $\mu(N) = 0$ . Wegen der Monotonie von  $\mu$  ist  $0 \leq \mu(N \cap D) \leq \mu(N) = 0$  und somit ist  $N \cap D$  auch eine Nullmenge. Wir erhalten damit und nochmaliger Monotonie von  $\mu$

$$\mu(N \cap D) + \mu(N^c \cap D) = \mu(N^c \cap D) \leq \mu(D)$$

und damit ist  $N$  messbar.  $\square$

### $\sigma$ -Algebra der messbaren Mengen

Die Menge der messbaren Mengen  $\mathcal{A}$  hat einige besondere Eigenschaften. Wir geben hier nur diese Eigenschaften ohne Beweis an. Die Beweise verwenden im wesentlichen Techniken, die wir bereits kennengelernt haben.

- Für eine Folge  $A_i \in \mathcal{A}$  von messbaren Mengen mit  $A_i \cap A_j = \emptyset$  ist  $\mu(\bigcup_i A_i) = \sum_i \mu(A_i)$ .
- Ist  $A \in \mathcal{A}$ , so ist auch  $A^c \in \mathcal{A}$ .
- Für eine Folge  $A_i \in \mathcal{A}$  messbarer Mengen ist  $\bigcup_i A_i \in \mathcal{A}$  messbar.

**Satz 14** (Charakterisierung messbarer Mengen). *Eine Menge  $A$  ist genau dann messbar, wenn  $A = S \cup N$ , wobei  $N$  eine Nullmenge und  $S$  Vereinigung von kompakten Mengen ist.*

### 3.2 Lebesgue Integral

**Definition 19.** Für eine Teilmenge  $A \subset \mathbb{R}^n$  heißt

$$1_A(x) := \begin{cases} 1 & \text{falls } x \in A \\ 0 & \text{sonst} \end{cases}$$

Indikatorfunktion.

**Definition 20.** Eine Funktion

$$\varphi(x) := \sum_{k=1}^m c_k 1_{I_k}$$

mit  $c_k \in \mathbb{R}$  und  $I_k \in \mathbb{I}(n)$  mit  $I_l \cap I_h = \emptyset$  für  $i \neq j$  heißt Treppenfunktion.



Abbildung 21: Quelle: Wikipedia: <https://commons.wikimedia.org/wiki/File:Stepfunction1.png>

**Bemerkung 11.** Seien  $\varphi(x) = \sum_{k=1}^m c_k 1_{I_k}$  und  $\psi(x) = \sum_{j=1}^l u_j 1_{I_j}$ . Dann definiert  $(\varphi + \psi)(x) := \sum_{k=1}^m \sum_{j=1}^l (c_k + u_j) 1_{I_{k,j}}$  mit  $I_{k,j} := I_k \cap I_j$  eine Treppenfunktion (nach entsprechender Ummumerierung zu einem einzigen Summenzeichen).

**Definition 21.** Für eine Treppenfunktion  $\varphi(x) := \sum_{k=1}^m c_k 1_{I_k}$  definieren wir das Integral durch

$$\int_{\mathbb{R}^n} \varphi d\mu := \sum_{k=1}^m c_k \mu(I_k).$$

**Satz 15.** Seien  $\varphi(x) = \sum_{k=1}^m c_k 1_{I_k}$  und  $\psi(x) = \sum_{j=1}^l u_j 1_{I_j}$  zwei Treppenfunktionen. Für das Integral von Treppenfunktion gilt:

- Ist  $\varphi(x) = \psi(x)$  für alle  $x$ , dann ist  $\int_{\mathbb{R}^n} \varphi d\mu = \int_{\mathbb{R}^n} \psi d\mu$  (Das Integral hängt nicht von der Zerlegung der Treppenfunktion ab und ist wohldefiniert)
- $\int_{\mathbb{R}^n} \alpha \varphi + \beta \psi d\mu = \alpha \int_{\mathbb{R}^n} \varphi d\mu + \beta \int_{\mathbb{R}^n} \psi d\mu$
- $\left| \int_{\mathbb{R}^n} \varphi d\mu \right| \leq \int_{\mathbb{R}^n} |\varphi| d\mu$
- Ist  $\varphi(x) \leq \psi(x)$  für alle  $x$ , so ist  $\int_{\mathbb{R}^n} \varphi d\mu \leq \int_{\mathbb{R}^n} \psi d\mu$

*Beweis.* Der Beweis wird über eine vollständige Induktion geführt. Der Induktionsanfang ist einfach zu zeigen. Wir nehmen an, die Aussage gilt für alle Dimensionen  $k < n$ . Zerlege  $\mathbb{R}^n = \mathbb{R}^p \times \mathbb{R}^{n-p}$ . Jeder Quader  $I \in \mathbb{I}(n)$  zerlegt sich damit ebenfalls in ein Produkt  $I = I' \times I''$  mit  $I' \in \mathbb{I}(p)$  und  $I'' \in \mathbb{I}(n-p)$  und für  $z = (x, y) \in \mathbb{R}^p \times \mathbb{R}^{n-p}$  gilt  $1_I(z) = 1_{I'}(x) \cdot 1_{I''}(y)$ . Es sei nun  $\varphi(z) := \sum_{k=1}^m c_k 1_{I_k}(z)$  eine Treppenfunktion auf  $\mathbb{R}^p \times \mathbb{R}^{n-p}$ . Für jedes  $y \in \mathbb{R}^{n-p}$  definiert  $\varphi_y(x) = \sum_{k=1}^m c_k 1_{I'_k}(y) \cdot 1_{I'_k}(x)$  eine Treppenfunktion auf  $\mathbb{R}^{n-p}$ . Nach Induktionsvoraussetzung hängt das Integral

$$\int_{\mathbb{R}^p} \varphi_y(x) d\mu' = \sum_{k=1}^m c_k \mu'(I'_k) \cdot 1_{I'_k}(y) =: \phi(y)$$

nicht von der Zerlegung der Treppenfunktion ab.  $\phi(y)$  ist wiederum eine Treppenfunktion auf  $\mathbb{R}^{n-p}$  und nach Induktionsvoraussetzung hängt das Integral

$$\int_{\mathbb{R}^{n-p}} \phi(y) d\mu'' = \sum_{k=1}^m c_k \mu'(I'_k) \cdot \mu''(I''_k)(y)$$

nicht von der Zerlegung der Treppenfunktion ab. Somit gilt

$$\int_{\mathbb{R}^{n-p}} \int_{\mathbb{R}^p} \varphi_y(x) d\mu' d\mu'' = \sum_{k=1}^m c_k \mu'(I'_k) \cdot \mu''(I''_k)(y) = \sum_{k=1}^m c_k \mu(I_k) = \int_{\mathbb{R}^n} \varphi(z) d\mu.$$

Die linke Seite hängt damit nicht von der Zerlegung der Treppenfunktion ab und alle Behauptungen können so auf den Fall  $n = 1$  zurückgeführt werden.  $\square$

**Bemerkung 12** (Satz von Fubini für Treppenfunktionen). *Es gilt*

$$\int_{\mathbb{R}^n} \varphi(x, y) d\mu = \int_{\mathbb{R}^{n-p}} \left( \int_{\mathbb{R}^p} \varphi(x, y) d\mu' \right) d\mu''$$

*Beweis.* Siehe Beweis des letzten Satzes.  $\square$

### 3.2.1 Die $L^1$ -Halbnorm

Der Fall  $\infty$  ist bei den folgenden Betrachtungen immer als möglicher Fall zu berücksichtigen.

**Definition 22.** Eine Hüllereihe zu einer Funktion  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  ist eine Reihe  $\phi(x) := \sum_{k=1}^{\infty} c_k 1_{I_k}(x)$  mit den folgenden Eigenschaften:

- $c_k \in \mathbb{R}$  sind positive reelle Zahlen  $c_k > 0$ .
- $I_k \subset \mathbb{R}^n$  sind offene Quadrate.
- Für alle  $x \in \mathbb{R}^n$  gilt  $|f(x)| \leq \phi(x)$ .

**Definition 23.** Der Inhalt einer Hüllereihe  $\phi(x) := \sum_{k=1}^{\infty} c_k 1_{I_k}(x)$  ist definiert durch

$$I(\phi) := \sum_{k=1}^{\infty} c_k \mu(I_k).$$

**Definition 24.** Die  $L^1$ -Halbnorm einer Funktion  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  ist definiert durch das Infimum der Inhalte der Hüllereihen zu  $f$

$$\|f\|_1 := \inf \left\{ I(\phi) \mid \phi \text{ ist Hüllereihe zu } f \right\}.$$

**Lemma 1** (Verallgemeinerte Dreiecksungleichung). *Für nicht negative Funktionen  $f_k : \mathbb{R}^n \rightarrow \mathbb{R}_{\geq 0}$  gilt*

$$\left\| \sum_{k=1}^{\infty} f_k \right\|_1 \leq \sum_{k=1}^{\infty} \|f_k\|_1.$$

*Beweis.* Zu  $\epsilon > 0$  und  $f_k$  wählen wir eine Hüllereihe  $\phi_k = \sum_i c_{ik} 1_{I_{ik}}$  mit Inhalt

$$I(\phi_k) = \sum_i c_{ik} \mu(1_{I_{ik}}) = \|f_k\|_1 + \frac{\epsilon}{2^k}.$$

Dann ist  $\phi := \sum_{k,i} c_{ik} 1_{I_{ik}}$  eine Hüllereihe der Funktion  $\sum_k f_k$  mit

$$I(\phi) = \sum_{k,i} c_{ik} \mu(I_{ik}) = \sum_k \left( \sum_i c_{ik} \mu(I_{ik}) \right) \leq \sum_k \|f_k\|_1 + \epsilon$$

Mit  $\|\sum_{k=1}^{\infty} f_k\|_1 \leq I(\phi)$  und  $\epsilon \rightarrow 0$  folgt die Behauptung.  $\square$

**Bemerkung 13** (Rechenregeln). *Für  $f, g : \mathbb{R}^n \rightarrow \mathbb{R}$  und  $c \in \mathbb{R}$  gilt:*

- $\|cf\|_1 \leq |c| \|f\|_1$ .

- $\|f + g\|_1 \leq \|f\|_1 + \|g\|_1$
- Aus  $\|f(x)\|_1 \leq g(x)$  für alle  $x$  folgt  $\|f\|_1 \leq \|g\|_1$ .

*Beweis.* • Für eine Hüllreihe  $\varphi$  von  $f$  ist  $|c| \cdot \varphi$  eine Hüllreihe von  $c \cdot f$ .

- Da  $|f+g| \leq |f|+|g|$  folgt Behauptung aus (iii) und der verallgemeinerten Dreiecksungleichung.
- Hullreihen sind immer größer-gleich der Funktion und damit haben größere Funktionen größere Hullreihen.

□

**Lemma 2** (Fundamentallemma). *Für die charakteristische Funktion  $\mathbf{1}_I$  eines Quaders  $I$  gilt*

$$\|\mathbf{1}_I\|_1 = \mu(I) = \text{vol}(I) = \int \mathbf{1}_I d\mu$$

*Beweis.*  $\mathbf{1} \cdot \mathbf{1}_I$  ist eine Hülfunktion von  $\mathbf{1}_I$  und damit gilt  $\|\mathbf{1}_I\| \leq \mu(I)$ . Sei  $\varphi(x) = \sum_k c_k \mathbf{1}_{I_k}$  eine Hüllreihe von  $\mathbf{1}_I$  und  $\epsilon > 0$ . Da  $\varphi(x) \geq 1$  gibt es für jedes  $x$  einen Index  $N(x)$  mit  $\sum_{k=1}^{N(x)} c_k \mathbf{1}_{I_k} \geq 1 - \epsilon$ . Da die  $I_k$  offen sind, gibt es für jedes  $x$  eine Umgebung  $U(x)$ , so dass letztere Gleichung gilt. Da  $\bar{I}$  kompakt ist (beschränkt und abgeschlossen), überdecken endlich viele  $U(x_1), \dots, U(x_n)$  den Quader  $I$ . Mit  $N := \max\{N(x_1), \dots, N(x_n)\}$  folgt  $\sum_{k=1}^N c_k \mathbf{1}_{I_k} \geq (1 - \epsilon) \mathbf{1}_I$ . Aus Satz 15 (iii) folgt

$$I(\varphi) = \sum_k c_k \mu(I_k) \geq \sum_{k=1}^N c_k \mu(I_k) \geq (1 - \epsilon) \mu(I).$$

Mit  $\epsilon \rightarrow 0$  folgt  $I(\varphi) \geq \mu(I)$  und damit insgesamt die Behauptung. □

**Lemma 3.** *Für jede Treppenfunktion  $\varphi$  auf  $\mathbb{R}^n$  gilt*

$$\|\varphi\|_1 = \int |\varphi| d\mu.$$

### 3.2.2 Integration über den $\mathbb{R}^n$

**Definition 25.** Eine Funktion  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  heißt integrierbar, falls eine Folge von Treppenfunktionen  $\varphi_k$  existiert mit

$$\|f - \varphi_k\|_1 \rightarrow 0 \text{ für } k \rightarrow \infty.$$

In diesem Fall heißt

$$\int_{\mathbb{R}^n} f(x) d\mu := \lim_{k \rightarrow \infty} \varphi_k d\mu$$

das Integral von  $f$  über  $\mathbb{R}^n$ .

**Bemerkung 14.**

- Die reelle Zahlenfolge  $\int \varphi_k d\mu$  ist eine Cauchyfolge und damit konvergent.

- Der Grenzwert ist unabhängig von der Folge  $\varphi_k$ .

*Beweis.* Für Treppenfunktionen  $\psi$  und  $\xi$  gilt

$$\begin{aligned} \left| \int \psi d\mu - \int \xi d\mu \right| &\leq \int |\psi - \xi| d\mu = \|\psi - \xi\|_1 \\ &\leq \|\psi - f\|_1 + \|f - \xi\|_1 \end{aligned}$$

woraus die Behauptungen folgen.  $\square$

**Satz 16.** Ist  $f$  über  $\mathbb{R}^n$  integrierbar, so auch  $|f|$  und es gilt

$$\left| \int f d\mu \right| \leq \int |f| d\mu = \|f\|_1.$$

*Beweis.* Sei  $f$  integrierbar und  $\varphi_k$  eine Folge von Treppenfunktionen mit  $\|f - \varphi_k\|_1 \rightarrow 0$ . Aus  $\| |f| - |\varphi_k| \|_1 \leq \|f - \varphi_k\|_1$  ergibt sich wegen der Monotonie der  $L^1$ -Norm

$$\| |f| - |\varphi_k| \|_1 \leq \|f - \varphi_k\|_1.$$

Damit gilt  $\| |f| - |\varphi_k| \|_1 \rightarrow 0$  und somit ist  $|f|$  integrierbar und mit der Abschätzung von Beträgen für Treppenfunktionen aus Satz 15 gilt

$$\left| \int f d\mu \right| = \left| \lim_k \int \varphi_k d\mu \right| \leq \int |\varphi_k| d\mu = \int |f| d\mu$$

und damit der erste Teil der Behauptung. Mit der Dreiecksungleichung erhalten wir

$$\|f\| - \|f - \varphi_k\|_1 \leq \|\varphi_k\|_1 \leq \|f\|_1 + \|f - \varphi_k\|_1$$

und wegen  $\|\varphi_k\|_1 = \int |\varphi_k| d\mu \rightarrow \int |f| d\mu$  folgt die Behauptung.  $\square$

**Bemerkung 15** (Rechenregeln). Sind  $f$  und  $g$  integrierbar, so gilt

- $\alpha f + \beta g$  mit  $\alpha, \beta \in \mathbb{R}$  ist integrierbar mit

$$\int \alpha f + \beta g d\mu = \alpha \int f d\mu + \beta \int g d\mu.$$

- Aus  $f(x) \leq g(x)$  für alle  $x \in \mathbb{R}^n$  folgt  $\int f d\mu \leq \int g d\mu$ .

- Ist  $g$  zusätzlich beschränkt, so ist auch  $f \cdot g$  integrierbar.

*Beweis.* • Sind  $\phi_k$  und  $\psi_k$  approximierende Folge von Treppenfunktionen von  $f$  und  $g$ , so ist  $\alpha\phi_k + \beta\psi_k$  eine approximierende Folge von  $\alpha f + \beta g$ .

- Nach Satz 16 ist  $\int (g - f) d\mu = \|g - f\|_1 \geq 0$ .  $\square$

**Bemerkung 16.** Ist  $f$  integrierbar, so auch  $f^+ := \max(f, 0)$  und  $f^- := \min(f, 0)$ . Damit ist  $f = f^+ + f^-$  genau dann integrierbar, wenn  $f^+$  und  $f^-$  integrierbar sind. Da  $-f^- \geq 0$  ist, kann man sich in Beweisen häufig auf den Fall  $f \geq 0$  beschränken.

*Beweis.* Es ist  $\max(f, 0) = \frac{1}{2}(f + |f|)$  und  $\min(f, 0) = \frac{1}{2}(f - |f|)$  und die Behauptung folgt aus den Rechenregeln.  $\square$

**Definition 26** (Integration über Teilmengen). *Für eine Teilmenge  $A \subset \mathbb{R}^n$  und eine Funktion  $f : A \rightarrow \mathbb{R}$  heißt*

$$f_A(x) := \begin{cases} f(x) & \text{für } x \in A \\ 0 & \text{für } x \in \mathbb{R}^n \setminus A \end{cases}$$

*die triviale Fortsetzung von  $f$  auf  $\mathbb{R}^n$ .  $f$  heißt integrierbar über  $A$ , falls  $f_A$  über  $\mathbb{R}^n$  integrierbar ist und in diesem Fall bezeichnen wir mit*

$$\int_A f(x) d\mu := \int f_A(x) d\mu$$

*als das Integral von  $f$  über  $A$ .*

**Satz 17** (Kleiner Satz von B. Levi). *Zu  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  gebe es eine monoton wachsende Folge  $\varphi_k$  von Treppenfunktionen mit*

- Für alles  $x \in \mathbb{R}^n$  gilt  $\lim_{k \rightarrow \infty} \varphi_k(x) = f(x)$ .  $f$  ist also die punktweise gebildete Grenzfunktion der  $\varphi_k$ .
- Die reelle Folge der Integrale  $\int \varphi_k d\mu$  ist beschränkt.

*Dann ist  $f$  integrierbar und es gilt*

$$\int f d\mu = \lim_{k \rightarrow \infty} \int \varphi_k d\mu.$$

*Beweis.* Aus  $f - \varphi_k = \sum_{i=k}^{\infty} (\varphi_{i+1} - \varphi_i)$  folgt mit der verallgemeinerten Dreiecksungleichung Lemma 1 und Lemma 3

$$\|f - \varphi_k\|_1 \leq \sum_{i=k}^{\infty} \int |\varphi_{i+1} - \varphi_i| d\mu = \sum_{i=k}^{\infty} \left( \int \varphi_{i+1} d\mu - \int \varphi_i d\mu \right).$$

Die Folge  $\int \varphi_k$  ist monoton wachsend und beschränkt und damit konvergent. Bezeichnen wir mit  $I$  den Grenzwert, so folgt  $\|f - \varphi_k\|_1 \leq I - \int \varphi_k d\mu$ . Also gilt  $\|f - \varphi_k\|_1 \rightarrow 0$  für  $k \rightarrow \infty$  und damit ist  $f$  integrierbar und mit der Definition des Integrals folgt die Behauptung.  $\square$

**Definition 27.** Eine Menge  $U \subset \mathbb{R}^n$  heißt offen, falls für alle  $a \in U$  ein Radius  $r > 0$  existiert, so dass der Ball  $B_r(a) := \{x \in \mathbb{R}^n \mid \|x - a\| \leq r\} \subset U$  in  $U$  enthalten ist.

**Satz 18.** Sei  $U \subset \mathbb{R}^n$  offen und beschränkt und  $f : U \rightarrow \mathbb{R}$  stetig und beschränkt. Dann ist  $f$  über  $U$  integrierbar.

*Beweis.* Da  $U$  offen ist, kann man um jeden Punkt  $a \in U$  einen Würfel  $W_r(a)$  finden, dessen Mittelpunkt  $a$  und dessen Kantenlänge  $r$  eine rationale Zahl ist. Damit kann man zu Punkten  $a_1, \dots, a_n$  Würfel wählen, so dass  $W_r(a_i) \cap W_r(a_j) = \emptyset$  für  $i \neq j$  und mit  $m_i := \min\{f(x) \mid x \in W_r(a_i)\}$  Hüllreihen  $\psi_{a_1, \dots, a_n} := \sum_{i=1}^n m_i 1_{W_r(a_i)}$  konstruieren mit  $\psi \leq f$ . Bezeichnen wir mit  $\mathcal{T} :=$

$\{\psi_k\}$  die abzählbare Menge dieser Treppenfunktionen, so ist  $f = \sup\{\psi \mid \psi \in \mathcal{T}\}$  und  $\varphi_k := \max\{\psi_1, \dots, \psi_k\}$  eine monoton wachsende Treppenfunktion mit  $\varphi_k \rightarrow f$ . Da  $U$  beschränkt ist, gibt es eine Quader  $I$  mit  $U \subset I$  und mit  $M := \max(f)$  ist  $\int \varphi_k d\mu \leq M\mu(I)$  beschränkt. Mit dem Satz von B. Levi folgt die Behauptung.  $\square$

**Satz 19** (Schnittmenge). Sei  $A \subset \mathbb{R}^n$ . Für  $y \in \mathbb{R}^p$  heißt

$$A_y := \left\{ x \in \mathbb{R}^{n-p} \mid (x, y) \in A \right\}$$

Schnittmenge von  $A$  zu  $y$ .

**Satz 20** (Kleiner Satz von Fubini). Sei  $A \subset \mathbb{R}^n$  eine offene und beschränkte Teilmenge und  $f : U \rightarrow \mathbb{R}$  eine stetige und beschränkte Funktion. Dann ist für jedes  $y \in \mathbb{R}^p$  mit  $A_y \neq \emptyset$  die Funktion  $f_A(x) := f(x, y)$  über  $A_y$  und

$$F(y) := \begin{cases} \int_{A_y} f(x, y) d\mu_p, & \text{falls } A_y \neq \emptyset \\ 0 & \text{sonst} \end{cases}$$

über  $\mathbb{R}^p$  integrierbar und es gilt

$$\int_A f(x, y) d\mu_n = \int_{\mathbb{R}^p} \int_{\mathbb{R}^{n-p}} F(y) d\mu_p .$$

Hierfür schreiben wir auch kurz

$$\int_A f(x, y) d\mu_n = \int_{\mathbb{R}^p} \left( \int_{A_y} f(x, y) d\mu_{n-p} \right) d\mu_p$$

*Beweis.* Wie in Beweis zu Satz sei  $\varphi_k$  eine monoton wachsende Folge von Treppenfunktionen auf  $\mathbb{R}^n$  mit  $\|f_A - \varphi_k\|_1 \rightarrow 0$ . Für jedes  $y \in \mathbb{R}^p$  bilden dann die Funktionen  $\varphi_k(x)_y := \varphi_k(x, y)$  eine monoton wachsende Folge von Treppenfunktionen auf  $\mathbb{R}^{n-p}$ , die gegen  $f_A(x) := f_A(x, y)$  konvergiert. Die Folge der Integrale  $\int_{\mathbb{R}^{n-p}} \varphi_k(x)_y d\mu_{n-p}$  ist beschränkt, da  $A$  beschränkt ist, und daher gibt es eine Quader  $I$  mit  $U \subset I$  und mit  $M := \max(f)$  ist  $\int \varphi_k d\mu \leq M\mu(I)$  beschränkt. Mit dem kleinen Satz von B. Levi 17 gilt

$$F(y) := \lim_{k \rightarrow \infty} \int_{\mathbb{R}^{n-p}} \varphi_k(x)_y d\mu_{n-p} .$$

Die Funktionen

$$\phi_k(y) := \int_{\mathbb{R}^{n-p}} \varphi_k(x, y) d\mu_{n-p}$$

sind Treppenfunktionen auf  $\mathbb{R}^p$  und die Folge  $\phi_k$  konvergiert monoton wachsend gegen  $F$  und die Folge der Integrale  $\int \phi_k(y) d\mu_p$  ist beschränkt, da mit dem Satz von Fubini für Treppenfunktionen 12 und  $\varphi_k \leq f_A$

$$\int_{\mathbb{R}^p} \phi_k(y) d\mu_p = \int_{\mathbb{R}^n} \varphi_k(x, y) d\mu_n \leq \int_{\mathbb{R}^n} f_A(x, y) d\mu_n .$$

Mit dem kleinen Satz von B. Levi ist  $F$  integrierbar und es gilt

$$\begin{aligned}\int_{\mathbb{R}^p} F(y) d\mu_p &= \lim_{k \rightarrow \infty} \int_{\mathbb{R}^p} \phi_k(y) d\mu_p = \lim_{k \rightarrow \infty} \int_{\mathbb{R}^n} \phi_k(x, y) d\mu_n \\ &= \int_{\mathbb{R}^n} f_A(x, y) d\mu_n.\end{aligned}$$

□

**Satz 21** (Lebesgue vs. Riemann). *Eine Regelfunktion  $f$  auf  $[a, b] \subset \mathbb{R}$  ist über  $[a, b]$  Lebesgueintegrierbar und es gilt*

$$\int_{[a, b]} f(x) d\mu_1 = \int_a^b f(x) dx.$$

*Beweis.* Sei  $\varphi_k$  eine Folge von Treppenfunktionen mit  $\|f - \varphi_k\|_\infty \rightarrow 0$ . Da für jede Funktion  $h$  auf  $[a, b]$  die Abschätzung  $|h| \leq \|h\|_\infty 1_{[a, b]}$  gilt, folgt auch

$$\|f_A - \varphi_{k,A}\|_1 \rightarrow 0$$

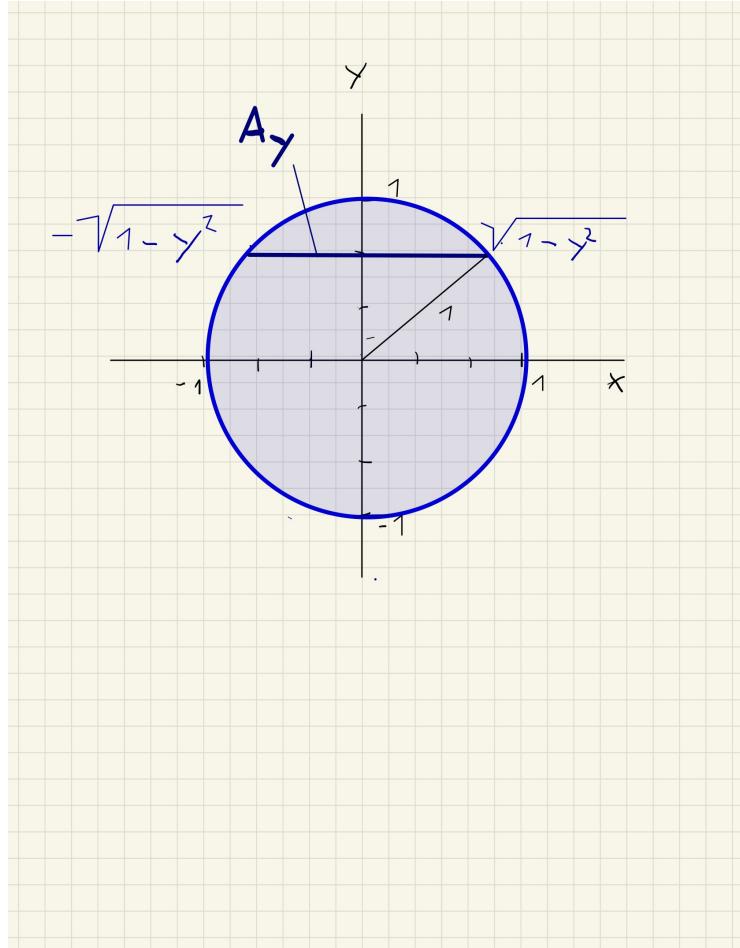
und damit ist  $f_{[a, b]}$  auch Lebesgueintegrierbar über  $[a, b]$  mit Lebesgueintegral

$$\int_{[a, b]} f(x) d\mu = \int_{\mathbb{R}} f_A(x) d\mu = \lim_k \varphi_{k,A} d\mu = \lim_k \int_a^b \varphi_k dx = \int_a^b f(x) dx$$

□

**Beispiel 2** (Integration über Kreis). *Sei  $K := B_1^2(0) := \{(x, y) \in \mathbb{R}^2 \mid \sqrt{x^2 + y^2} \leq 1\}$ . Mit dem kleinen Satz von Fubini erhalten wir*

$$\begin{aligned}\int_K 1 d\mu &= \int_{-1}^1 \left( \int_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}} 1 dx \right) dy = \\ &= 2 \int_{-1}^1 \sqrt{1-y^2} dy \\ (\text{substitution } y &= \sin(u)) = 2 \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cos(u)^2 du = 2 \cdot \frac{\pi}{2} = \pi\end{aligned}$$



**Satz 22** (Produktsatz). Für Funktionen  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  und  $g : \mathbb{R}^m \rightarrow \mathbb{R}$  definiert  $(f \otimes g)(x, y) := f(x) \cdot g(y)$  eine Abbildung  $f \otimes g : \mathbb{R}^{n+m} \rightarrow \mathbb{R}$ . Sind  $f$  und  $g$  integrierbar, so ist  $f \otimes g$  integrierbar und es gilt

$$\int_{\mathbb{R}^{n+m}} (f \otimes g)(x, y) d\mu_{n+m} = \left( \int_{\mathbb{R}^n} f(x) d\mu_n \right) \cdot \left( \int_{\mathbb{R}^m} g(y) d\mu_m \right)$$

### 3.2.3 Transformationssatz

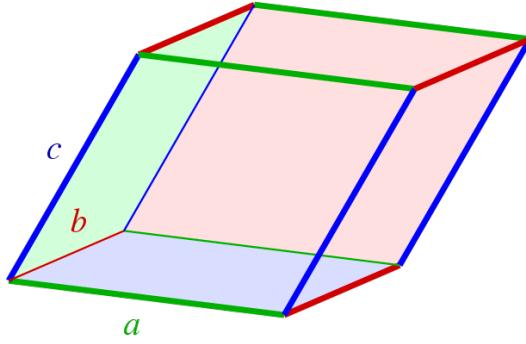
**Satz 23.** Seien  $U$  und  $V$  offene Teilmengen des  $\mathbb{R}^n$ ,  $T' : U \rightarrow V$  ein lineare Abbildung und  $Q \in \mathbb{I}(n)$  ein Quader. Dann gilt:

$$vol(T'(Q)) = \det(T') \cdot vol(Q).$$

*Beweis.* Für Vektoren  $a_1, \dots, a_n$  im  $\mathbb{R}^n$  heißt die Menge

$$P(a_1, \dots, a_n) := \left\{ x = \sum_{k=1}^n t_k a_k \mid t_1, \dots, t_n \in [0, 1] \right\}$$

Parallelotop.



Es gilt

$$\text{vol}(P(a_1, \dots, a_n)) = |\det(a_1, \dots, a_n)|.$$

Ausführlicher Beweis

□

**Definition 28** (Diffeomorphismus). Seien  $U$  und  $V$  offene Teilmengen des  $\mathbb{R}^n$ . Eine Abbildung  $T : U \rightarrow V$  heißt diffeomorphismus, wenn eine Umkehrfunktion  $T^{-1} : V \rightarrow U$  existiert, also  $T^{-1}(T(u)) = u$  gilt für alle  $u \in U$ , die ebenfalls differenzierbar ist.

**Beispiel 3.** Für eine invertiertere Matrix  $A$  ist  $T(x) := Ax$  ein Diffeomorphismus.

**Satz 24.** Seien  $U$  und  $V$  offene Teilmengen des  $\mathbb{R}^n$ ,  $T : U \rightarrow V$  ein Diffeomorphismus und  $f : V \rightarrow \mathbb{R}$  eine integrierbare Funktion. Dann gilt:

$$\int_V f(y) d\mu = \int_U f(T(x)) \cdot |\det(T'(x))| d\mu.$$

*Beweis.* Seien  $I_k \in \mathbb{I}(n)$  Quader,  $J_k := T(I_k)$  und  $b_k = T(c_k)$ . Dann ist

$$\sum_{k=1}^n b_k \text{vol}(J_k) \approx \sum_{k=1}^n T(c_k) \cdot |\det T'(c_k)| \text{vol}(I_k).$$

Die Behauptung folgt dann (nicht trivial) durch den Übergang zu Grenzwerten mit entsprechenden Konvergenzsätzen. □

### 3.2.4 Parameterabhängige Integrale

**Definition 29.** Eine Folge von Funktionen  $f_k$  konvergiert Punktweise fast überall gegen eine Funktion  $f$ , falls es eine Nullmenge  $N$  gibt, mit  $\lim_{k \rightarrow \infty} f_k(x) = f(x)$  für alle  $x \in \mathbb{R}^n \setminus N$ .

**Satz 25** (Satz von Lebesgue). Sei  $f_k$  eine Folge integrierbarer Funktionen auf  $\mathbb{R}^n$  die fast überall Punktweise gegen eine Funktion  $f$  konvergiert. Es gebe eine integrierbare Funktion  $F$  mit  $|f_k(x)| \leq F(x)$  für alle  $x \in \mathbb{R}^n$  und alles  $k$ . Dann ist  $f$  integrierbar und es gilt

$$\int f(x) d\mu = \lim_{k \rightarrow \infty} \int f_k(x) d\mu$$

Sei  $f : X \times T \subset \mathbb{R}^{n-p} \times \mathbb{R}^p$  eine Funktion, so dass für festes  $x \in X$  die Funktion  $f_x(t) := f(x, t)$  über  $T$  integrierbar ist. Durch Integration erhält man die Funktion

$$F(x) := \int_T f(x, t) d\mu_T$$

auf  $X$ .

**Satz 26** (Stetigkeitssatz).  $f$  habe zusätzlich die Eigenschaften:

- Für festes  $t$  ist  $f_t(x) := f(x, t)$  stetig.
- Es gibt auf  $T$  eine integrierbare Funktion  $\phi$  mit  $\phi(t) \geq 0$  und  $|f(x, t)| \leq \phi(t)$  für alle  $(x, t) \in X \times T$ .

Dann ist die oben definierte Funktion  $F$  stetig.

*Beweis.* Sei  $x_k \rightarrow x$  eine konvergente Folge in  $X$  und  $f_k(t) := f(x_k, t)$ . Nach Voraussetzung konvergiert diese Folge Punktweise gegen die Funktion  $f_t(x)$  und  $|f_k(x)| \leq \phi(x)$ . Mit dem Satz von Lebesgue folgt

$$\lim_{k \rightarrow \infty} \int_T f_k(t) d\mu_T = \int_T f(x, t) d\mu_T$$

und damit  $\lim_{k \rightarrow \infty} F(x_k) = F(x)$ .  $\square$

**Satz 27** (Differentiationssatz).  $f$  habe zusätzlich die Eigenschaften:

- Für festes  $t$  ist  $f_t(x) := f(x, t)$  stetig differenzierbar.
- Es gibt auf  $T$  eine integrierbare Funktion  $\phi$  mit  $\phi(t) \geq 0$  und  $|\frac{\partial}{\partial x_i} f(x, t)| \leq \phi(t)$  für alle  $(x, t) \in X \times T$  und  $i = 1, \dots, n-p$ .

Dann ist die oben definierte Funktion  $F$  stetig differenzierbar und es gilt

$$\frac{\partial}{\partial x_k} F(x) = \int_T \frac{\partial}{\partial x_k} f(x, t) d\mu_T .$$

*Beweis.* Sei  $x_k := x_0 + h_k e_i$  und

$$\varphi_k(t) := \frac{f(x, t) - f(x_0, t)}{h_k} .$$

Damit sind die Funktionen  $\varphi_k$  integrierbar und für jedes  $t \in T$  gilt

$$\lim_{k \rightarrow \infty} \varphi_k(t) = \frac{\partial}{\partial x_i} f(x_0, t) .$$

Mit dem Satz von Lebesgue gilt

$$\lim_{k \rightarrow \infty} \int_T \varphi_k(t) d\mu_T = \int_T \frac{\partial}{\partial x_i} f(x_0, t) d\mu_T$$

und da

$$\int_T \varphi_k(t) d\mu_T = \frac{F(x_k) - F(x_0)}{h_k}$$

ist, folgt die Behauptung.  $\square$

### 3.2.5 Faltung

**Definition 30.** Für integrierbare Funktionen  $f$  und  $g$  auf  $\mathbb{R}^n$  ist die Faltung definiert durch

$$(f * g)(x) := \int_{\mathbb{R}^n} f(x - y) \cdot g(y) d\mu_y . \quad (2)$$

Das Integral existiert wegen dem Satz von Fubini. Nach Anwendung der Transformation  $y \rightarrow x - y$  lässt sich die Faltung auch definieren durch

$$(f * g)(x) := \int_{\mathbb{R}^n} f(y) \cdot g(x - y) d\mu_y . \quad (3)$$

**Satz 28.** Für die Faltung gilt bei entsprechenden Voraussetzungen der Differenzierbarkeit der Funktionen

$$\partial^\alpha (f * g) = f * \partial^\alpha g .$$

*Beweis.* Nach Anwendung der Transformation  $y \rightarrow x - y$  folgt die Aussage direkt aus dem Differentiationssatz.  $\square$

## **Tabellenverzeichnis**

## Abbildungsverzeichnis

1	Quelle: Wikipedia: <a href="https://de.wikipedia.org/wiki/Datei:Diferencial_quotient_of_a_function.svg">https://de.wikipedia.org/wiki/Datei:Diferencial_quotient_of_a_function.svg</a>	
2	Quelle: Wikipedia: <a href="https://commons.wikimedia.org/wiki/File:Mittelwertsatz3.svg">https://commons.wikimedia.org/wiki/File:Mittelwertsatz3.svg</a>	6
3	Quelle: Wikipedia: <a href="https://commons.wikimedia.org/wiki/File:EpsilonSchlauch_klein.svg">https://commons.wikimedia.org/wiki/File:EpsilonSchlauch_klein.svg</a>	8
4	Quelle: Wikipedia: <a href="https://de.wikipedia.org/wiki/Datei:Limes_Definition_Vektorgrafik.svg">https://de.wikipedia.org/wiki/Datei:Limes_Definition_Vektorgrafik.svg</a>	9
5	Quelle: Wikipedia: <a href="https://commons.wikimedia.org/wiki/File:Upper_semi.svg">https://commons.wikimedia.org/wiki/File:Upper_semi.svg</a>	9
6	Quelle: Wikipedia: <a href="https://commons.wikimedia.org/wiki/File:Gradient2.svg">https://commons.wikimedia.org/wiki/File:Gradient2.svg</a>	10
7	Kantenzug mit achsenparallelen Kanten . . . . .	12
8	Quelle: Wikipedia: <a href="https://de.wikipedia.org/wiki/Datei:EbeneKurve.png">https://de.wikipedia.org/wiki/Datei:EbeneKurve.png</a>	13
9	Quelle: Wikipedia . . . . .	14
10	Quelle: Wikipedia: <a href="https://en.wikipedia.org/wiki/File:MaximumParaboloid.png">https://en.wikipedia.org/wiki/File:MaximumParaboloid.png</a>	18
11	Quelle: Wikipedia: <a href="https://en.wikipedia.org/wiki/File:MaximumCounterexample.png">https://en.wikipedia.org/wiki/File:MaximumCounterexample.png</a>	18
12	Quelle: Wikipedia: <a href="https://commons.wikimedia.org/wiki/File:Saddle_point.svg">https://commons.wikimedia.org/wiki/File:Saddle_point.svg</a>	19
13	Quelle: Wikipedia . . . . .	20
14	Quelle: <a href="https://getoutside.ordnancesurvey.co.uk/guides/understanding-map-contour-lines-for-beginners/">https://getoutside.ordnancesurvey.co.uk/guides/understanding-map-contour-lines-for-beginners/</a> . . . . .	21
15	Quelle: Wikipedia . . . . .	22
16	Quelle: Wikipedia . . . . .	22
17	Quelle: <a href="https://towardsdatascience.com/batch-mini-batch-stochastic-gradient-descent-7a62ecba642a">https://towardsdatascience.com/batch-mini-batch-stochastic-gradient-descent-7a62ecba642a</a> . . . . .	23
18	Quelle: Wikipedia . . . . .	25
19	Äußeres Maß . . . . .	27
20	Inneres Maß . . . . .	27
21	Quelle: Wikipedia: <a href="https://commons.wikimedia.org/wiki/File:Stepfunction1.png">https://commons.wikimedia.org/wiki/File:Stepfunction1.png</a>	28