Kurz-Skript zu „Differential- und Integralrechnung 2“

Thomas Schick*

Last compiled 7. August 2003; last edited 5.Mai 2003 or later

Inhaltsverzeichnis

*email: thomas.schick@uni-math.gwdg.de

Hinweis: dieses Skript ist wurde nicht korrekturgelesen. Es gibt mit Sicherheit eine Menge Fehler. Einige davon konnten unter anderem Dank der Hinweise von Denise Nakiboglu und Malte Schmidt behoben werden. Für weitere Hinweise auf Fehler schreiben Sie bitte eine email an schick@uni-math.gwdg.de. Nicht alle Beweise werden vorgeführt, nicht alle behandelten Sätze und Beispiele sind hier notiert.

1 Wiederholung: Topologie des $\mathbb{R}^n$

Der Begriff *Topologie* ist wahrscheinlich noch nicht gefallen. Anschaulich gesprochen, ist Topologie alles, was sich mit "Nachbarschaftsbeziehungen" — wann sind zwei Punkte "nahe beieinander" beschäftigt. Wichtige Fragestellungen sind insbesondere die Konvergenz von Folgen, die Stetigkeit von Funktionen, etc., die innerhalb der Topologie behandelt werden.

Zur Erinnerung:

1.1 Definition. Seien A, B Mengen. Eine Funktion $f\colon A \to B$ ordnet jedem Element $a \in A$ (genau) ein Element $f(a) \in B$ zu.

Insbesondere gibt es Funktionen $f\colon \mathbb{R}^n \to \mathbb{R}^m$.

1.2 Beispiel. Idealisiert kann man zu jedem Zeitpunkt jedem Punkt der Atmosphäre die jeweils dort herrschende Windgeschwindigkeit zuordnet (hier ignoriert man "atomare" Phänome). Solch eine Geschwindigkeit hat eine Richtung und einen Betrag, ist also ein (dreidimensionaler) Vektor. Die Punkte der Atmosphäre kann man idealisiert ebenfalls durch $\mathbb{R}^3$ repräsentieren, die Zeitpunkte durch Punkte auf $\mathbb{R}$. Man erhält also insgesamt eine Funktion

$$v\colon \mathbb{R}^4 = \mathbb{R} \times \mathbb{R}^3 \to \mathbb{R}^3; (x,t) \mapsto v(x,t),$$

wobei $v(x,t)$ die Windgeschwindigkeit zum Zeitpunkt t am Ort x ist.

Viele weitere solche Beispiele aus den verschiedensten Bereichen kommen vor.

Beachte, dass man solche Funktionen nur schwer zeichnen kann. Möglich ist dies noch für $f\colon \mathbb{R}^2 \to \mathbb{R}$, wo man über jedem Punkt x der Ebene die entsprechende Höhe $f(x)$ einzeichnet. Dies kann man auch recht gut durch entsprechende "Höhenlinien" visualisieren, wo man für gewisse Werte $c_1, c_2, \dots$ alle Punkte x mit $f(x) = c_1, \dots$ markiert. Ist f die Höhe über dem Meeresspiegel auf einer Landkarte, erhält man genau die Höhenlinien einer Landkarte (auch Geodäten benutzen Mathematik, zum Teil sogar recht intensiv — der Mathematiker Gauß hat erhebliche Arbeit in die Vermessung des Königreichs Hannover investiert; theoretisch und praktisch).

Funktionen von (Teilmengen von) $\mathbb{R}^n$ und $\mathbb{C}^n$ nach $\mathbb{R}^m$ und $\mathbb{C}^m$ haben wir bereits im letzten Semester kennen gelernt. Ein Großteil dieses Semesters besteht darin, sich intensiv mit ihren Eigenschaften auseinander zu setzen. Dabei werden wir auch einige Konzepte in abstrakterer Form kennen lernen.

1.3 Definition. Auf $\mathbb{C}^n$ definieren wir die *Normen* $|(x_1, \dots, x_n)|_1 := \sum_{k=1}^n |x_k|$, $|(x_1, \dots, x_n)|_2 := \sqrt{\sum_{k=1}^n |x_k|^2}$ und $|(x_1, \dots, x_n)|_\infty := \sup_{k=1}^n |x_k|$.

Es gilt (offensichtlich)

$$|v|_\infty \le |v|_1 \le n\,|v|_\infty \text{ für alle } v \in \mathbb{C}^n \tag{1.4}$$

und

$$|v|_\infty \le |v|_2 \le n\,|v|_\infty \text{ für alle } v \in \mathbb{C}^n. \tag{1.5}$$

Dies impliziert, dass für die Konvergenz- und Stetigkeitsfragen die Wahl der Norm unter diesen dreien keine Rolle spielt.

1.6 Definition. Sei $(v_k)_{k\in\mathbb{N}}$ eine Folge von Vektoren $v_k \in \mathbb{C}^n$. Sei $v \in \mathbb{C}^n$. Wir sagen $\lim_{k\to\infty} v_k = v$, falls $|v_k - v|_\infty \xrightarrow{k\to\infty} 0$. Äquivalent heißt dies, dass jede der Komponentenfolgen gegen die entsprechende Komponente von v konvergiert: falls $v_k = (x_k^1, \dots, x_k^n)$ und $v = (x^1, \dots, x^n)$, dann konvergiert v_k gegen v genau dann, wenn $x_k^1 \xrightarrow{k\to\infty} x^1, \dots, x_k^n \xrightarrow{k\to\infty} x^n$.

1.7 Lemma. *Äquivalent zur Konvergenz ist (wie gesagt), dass* $|v_k - v|_2 \xrightarrow{k\to\infty} 0$ *und dass* $|x_k - v|_1 \xrightarrow{k\to\infty} 0$.

Beweis. Da $n < \infty$ und $\left|v_k^i - v^i\right| \geq 0$ für alle i und k gilt $|v_k - v|_1 = \sum_{i=1}^n \left|v_k^i - v^i\right| \xrightarrow{k\to\infty} 0$ genau dann wenn $\left|v_k^i - v^i\right| \xrightarrow{k\to\infty} 0$ für $i = 1, \dots, n$. Die Ungleichungen (1.4) und (1.5) zeigen (da alle Normen immer ≥ 0 sind), dass

$$|v_k - v|_1 \xrightarrow{k\to\infty} 0 \quad \iff \quad |v_k - v|_\infty \xrightarrow{k\to\infty} 0$$

sowie

$$|v_k - v|_2 \xrightarrow{k\to\infty} 0 \quad \iff \quad |v_k - v|_\infty \xrightarrow{k\to\infty} 0.$$

□

1.8 Definition. Sei $D \subset \mathbb{C}^n$. Eine Funktion $f\colon D \to \mathbb{C}^m$ heißt *stetig an* $a \in D$, falls folgende äquivalenten Bedingungen erfüllt sind:

(1)
$$f(x_k) \xrightarrow{k\to\infty} f(a) \qquad \forall (x_k)_{k\in\mathbb{N}} \text{ mit } x_k \in D,\ x_k \xrightarrow{k\to\infty} a \in D.$$

Falls a ein *Häufungspunkt* von D ist, es also mindestens eine Folge $(x_k)_{k\in\mathbb{N}}$ wie oben gibt, schreiben wir:

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

(Beachte, dass wir Konvergenz von $f(x_k)$ gegen $f(a)$ für *jede* Folge (x_k) mit $x_k \xrightarrow{k\to\infty} a$ fordert!)

(2) Für jedes $\epsilon > 0$ existiert ein $\delta > 0$, so dass $|f(x) - f(a)|_1 < \epsilon$ für alle $x \in D$ mit $|x - a|_1 < \delta$.

(3) Die gleiche Bedingung wie in (2), aber $|\cdot|_1$ ersetzt durch entweder $|\cdot|_2$ oder $|\cdot|_\infty$.

(4) Schreibt man $f(x)$ als Vektor $f(x) = (f_1(x), \dots, f_m(x)) \in \mathbb{C}^m$, so erhält man m Funktionen $f_1, \dots, f_m\colon \mathbb{C}^n \to \mathbb{C}$. f ist stetig genau dann, wenn all die Funktionen $f_1, \dots, f_m$ stetig sind.

1.9 Aufgabe. Beweise, dass die Bedingungen in Definition 1.8 äquivalent zueinander sind.

1.10 Beispiel. Sei $A \in \mathrm{Hom}(\mathbb{R}^n, \mathbb{R}^m)$ eine lineare Abbildung. Dann ist die Abbildung $f\colon \mathbb{R}^n \to \mathbb{R}^m;\ f(x) := Ax$ stetig.

Beweis. Wir zeigen zunächst, dass f stetig an Null. Aus der linearen Algebra ist bekannt, dass f gegeben ist durch Multiplikation mit einer Matrix (a_{ij}), also $f(x_1, \ldots, x_n) = (\sum_{j=1}^n a_{1j}x_j, \ldots, \sum_{j=1}^n a_{mj}x_j)$.

Somit $|f(h) - f(0)|_1 = |f(h)|_1 \leq \sum_{i,j} |a_{ij}| \cdot |h_j| \leq (\sum_{ij} |a_{ij}|) \cdot \sum_{j=1}^n |h_j| = (\sum_{ij} |a_{ij}|) \cdot |h|_1 \xrightarrow{|h|_1 \to 0} 0$.

Sei nun $v \in \mathbb{R}^n$ beliebig. Dann gilt $|f(v+h) - f(v)|_1 = |A(v+h) - Av|_1 = |Av + Ah - Av|_1 = |Ah|_1 \xrightarrow{|h|_1 \to 0} 0$. f ist also überall stetig. □

1.11 Definition. Eine Teilmenge $U \subset \mathbb{R}^n$ heißt *offen*, falls für jedes $x \in U$ ein $\epsilon > 0$ existiert, so dass für jedes $y \in \mathbb{R}^n$ mit $|x - y|_1 < \epsilon$ gilt: $y \in U$. Anschaulich: um jedem Punkt $x \in U$ gibt es einen "Ball" vom Radius $\epsilon > 0$, der ganz in U enthalten ist.

Hier kann äquivalent auch $|\cdot|_2$ oder $|\cdot|_\infty$ benutzt werden.

1.12 Lemma. *Sei $U \subset \mathbb{R}^n$ offen. Sei $a = (a_1, \ldots, a_n) \in U$. Dann ist die Menge*

$$U_{a,k} := \{x \in \mathbb{R} \mid (a_1, \ldots, a_{k-1}, x, a_{k+1}, \ldots, a_n) \in U\} \subset \mathbb{R}$$

ebenfalls offen. Es handelt sich hier um die Menge, die man erhält, wenn man U mit der zur x_k-Achse parallelen Geraden durch a schneidet (genauer mit den k-ten Komponenten der Punkte auf dieser Geraden).

Beweis. Sei $x \in U_{a,k}$. Sei $\epsilon > 0$ so, dass $b \in U$, falls $|(a_1, \ldots, a_{k-1}, x, a_{k+1}, a_n) - b|_\infty < \epsilon$ (solches ϵ existiert, da U offen ist und nach Voraussetzung $(a_1, \ldots, a_{k-1}, x, a_{k+1}, a_n) \in U$). Falls $y \in \mathbb{R}$ und $|x - y| < \epsilon$, folgt dann

$$|(a_1, \ldots, a_{k-1}, x, a_{k+1}, a_n) - (a_1, \ldots, a_{k-1}, y, a_{k+1}, a_n)|_\infty = |x - y| < \epsilon,$$

also nach Definition $y \in U_{a,k}$. Also ist $U_{a,k}$ offen. □

2 Partielle Ableitungen

Sei $f\colon \mathbb{R}^n \to \mathbb{C}$ eine Funktion. Sei $a = (a_1, \ldots, a_n) \in \mathbb{R}^n$. Definiere die Funktion

$$f_{a,k}\colon \mathbb{R} \to \mathbb{C}; x \mapsto f_{a,k}(x) := f(a_1, \ldots, a_{k-1}, x, a_{k+1}, \ldots, a_m).$$

Wir setzen also x in die k-te Komponente von f ein, und "füllen" die anderen Komponenten mit den entsprechenden Komponenten von a.

2.1 Definition. Falls $f_{a,k}$ ableitbar an $x = a_k$, dann sagen wir, dass f an der Stelle a *partiell nach der k-ten Komponente ableitbar* ist. Wir schreiben

$$\frac{\partial f}{\partial x_k}(a) := f'_{a,k}(a_k).$$

Wenn dies für alle $a \in \mathbb{R}^n$ erfüllt ist, und die sich ergebende Funktion

$$\frac{\partial f}{\partial x_k}\colon \mathbb{R}^n \to \mathbb{C}; a \mapsto \frac{\partial f}{\partial x_k}(a)$$

stetig ist, so heißt f *stetig nach der k-ten Komponente ableitbar.*

Gilt dies für alle $k = 1, \ldots, n$, so heißt f *stetig partiell ableitbar.*

Die Definitionen verallgemeinern sich auf $f\colon D \to \mathbb{C}$, wenn D eine offene Teilmenge von $\mathbb{R}^n$ ist.

Manchmal schreibt man auch $\partial_k f$ oder $D_k f$ anstelle von $\partial f/\partial x_k$.

Falls $f\colon \mathbb{R}^n \to \mathbb{C}$ partiell ableitbar ist, definieren wir den Gradienten

$$\operatorname{grad}(f) := (\partial_1 f, \dots, \partial_n f)\colon \mathbb{R}^n \to \mathbb{C}^n.$$

Dies ist eine vektorwertige Funktion.

2.2 Bemerkung. Anschauliche Bedeutung des Gradienten von $f\colon \mathbb{R}^n \to \mathbb{R}$: die Richtung von $\operatorname{grad}(f)(x)$ ist die Richtung, in der f am schnellsten steigt, der Betrag mißt die Steilheit des Anstiegs.

Zusammengefasst: zur Bestimmung der partiellen Ableitungen $\partial_k f$ an der Stelle $(x_1, \dots, x_n)$ fasst man $x_1, \dots, x_{k-1}, x_{k+1}, x_n$ als Konstanten auf, ersetzt x_k durch x und leitet die sich ergebende Funktion nach x ab. In der Ableitung muss man dann x_k für x einsetzen.

2.3 Beispiel. (1) $f\colon \mathbb{R}^2 \to \mathbb{C};\ f(x_1, x_2) = \exp(x_1 x_2^3)$ hat partielle Ableitungen

$$\partial_1 f(x_1, x_2) = x_2^3 \exp(x_1 x_2^3); \quad \partial_2 f(x_1, x_2) = x_1 \cdot 3x_2^2 \exp(x_1 x_2^3).$$

(2) Betrachte $r\colon \mathbb{R}^n \to \mathbb{R}; r(x_1, \dots, x_n) := \sqrt{x_1^2 + \dots + x_n^2} = |(x_1, \dots, x_n)|_2$, also den euklidischen Abstand zum Ursprung diese Funktion ist auf $\mathbb{R}^n \setminus \{0\}$ stetig partiell ableitbar mit

$$\partial_k r(x_1, \dots, x_n) = \frac{x_k}{r(x_1, \dots, x_n)}.$$

Damit ergibt sich: $\operatorname{grad}(r)(x) = \frac{1}{r(x)} x$ für jedes $x \in \mathbb{R}^n \setminus \{0\}$.

(3) Die Funktion

$$g\colon \mathbb{R}^2 \to \mathbb{R}; f(x_1, x_2) = \begin{cases} \frac{x_1 x_2}{x_1^2 + x_2^2}; & (x_1, x_2) \neq (0, 0) \\ 0; & (x_1, x_2) = (0, 0) \end{cases}$$

ist auf ganz $\mathbb{R}^2$ partiell diffbar, aber an Null nicht stetig.

Beweis. Übungsaufgabe. □

2.4 Bemerkung. Das zweite Beispiel zeigt, dass aus partieller Differenzierbarkeit nicht die Stetigkeit folgt. Dies liegt daran, dass die partielle Diffbarkeit die Koordinatenrichtung gegenüber allen anderen Richtungen auszeichnet, und eigentlich nicht ganz das "richtige" Differenzierbarkeitskonzept ist.

2.5 Definition. Sei $f\colon \mathbb{R}^n \to \mathbb{C}^m; f(x) = (f_1(x), \dots, f_m(x))$ eine vektorwertige Funktion mit Komponentenfunktionen $f_1, \dots, f_m$. Wir nennen f partiell ableitbar, falls jede Komponentenfunktion f_k partiell ableitbar ist, entsprechend für stetig partiell ableitbar.

2.1 Höhere partielle Ableigungen

2.6 Definition. Sei $f\colon \mathbb{R}^n \to \mathbb{C}$ partiell differenzierbar. Wir erhalten die Funktionen $\partial_1 f, \dots, \partial_n f\colon \mathbb{R}^n \to \mathbb{C}$. Falls all diese Funktionen selbst (in allen Richtungen) partiell differenzierbar sind, so nennen wir f *zweimal partiell ableitbar*. Wir erhalten dann Funktionen $\partial_i(\partial_j f)$ $(i, j = 1, \dots, n)$. Man schreibt auch

$$\frac{\partial^2 f}{\partial x_i \partial x_j} := \partial_i(\partial_j f).$$

Beachte, dass es auf die Reihenfolge von i und j ankommt.

Induktiv definiert man, dass f n-mal partiell diffbar ist, wenn f $n-1$-mal partiell diffbar ist, und alle $n-1$-ten partiellen Ableitungen

$$\partial_{i_{n-1}}(\partial_{i_{n-2}}(\dots \partial_{i_1} f)\dots)$$

selbst partiell diffbar sind. Falls all diese partiellen Ableitung stetig sind, heißt f n-mal stetig partiell diffbar.

Entsprechende Definitionen für $f\colon \mathbb{R}^n \to \mathbb{C}^m$, indem man die einzelnen Komponentenfunktionen anschaut.

2.7 Bemerkung. Beachte, dass $f\colon \mathbb{R}^n \to \mathbb{C}$ n^k verschiedene k-te partielle Ableitungen besitzt.

2.8 Satz (von Schwarz). *Sei $f\colon \mathbb{R}^n \to \mathbb{C}$ zweimal stetig partiell differenzierbar. Dann gilt*

$$\partial_i(\partial_j f) = \partial_j(\partial_i f).$$

Aus Stetigkeit *der partiellen Ableitungen folgt also, dass die Reihenfolge keine Rolle spielt.*

Beweis. Um Schreibarbeit zu sparen betrachte $n = 2$. Wir beweisen Gleichheit an $(x_0, y_0) \in \mathbb{R}^2$. Dazu zeigen wir

$$\begin{aligned}
&\partial_x(\partial_y f)(x_0, y_0) \\
&\quad = \lim_{h\to 0} \frac{f(x_0+h, y_0+h) - f(x_0+h, y_0) - f(x_0, y_0+h) + f(x_0, y_0)}{h^2} \\
&\qquad\qquad = \partial_y(\partial_x f)(x_0, y_0).
\end{aligned}$$

Für $h \neq 0$ wende den Mittelwertsatz auf die stetig differenzierbare Funktion $y \mapsto f(x_0+h, y) - f(x_0, y)$ an. Damit erhält man

$$\frac{f(x_0+h, y_0+h) - f(x_0, y_0+h) - (f(x_0+h, y_0) - f(x_0, y_0))}{h} = \partial_y f(x_0+h, \xi_h) - \partial_y f(x_0, \xi_h),$$

mit geeignetem ξ_h zwischen y_0 und $y_0 + h$. Der Mittelwertsatz für die stetig differenzierbare Funktion $x \mapsto \partial_y f(x, \xi_h)$ liefert

$$\begin{aligned}
\frac{f(x_0+h, y_0+h) - f(x_0+h, y_0) - f(x_0, y_0+h) + f(x_0, y_0)}{h^2} &= \frac{\partial_y f(x_0+h, \xi_h) - \partial_y f(x_0, \xi_h)}{h} \\
&= \partial_x(\partial_y f)(\zeta_h, \xi_h) \xrightarrow{h\to 0} \partial_x(\partial_y f)(x_0, y_0),
\end{aligned}$$

da ζ_h zwischen x_0 und $x_0 + h$ liegt.

Der identische Ausdruck ergibt sich für $\partial_y(\partial_x f)(x_0, y_0)$. Damit folgt der Satz. □

2.9 Definition. Sei $f\colon \mathbb{R}^n \to \mathbb{C}^n$, $f = (f_1, \dots, f_n)$ mit Komponentenfunktionen f_k. Sei f partiell diffbar. Wir definieren die *Divergenz*

$$\operatorname{div}(f) := \sum_{k=1}^n \partial_k f_k.$$

Falls f zweimal diffbar ist, definiert man

$$\Delta f := \sum_{i=1}^n \partial_i(\partial_i f) = \operatorname{div}(\operatorname{grad} f).$$

$\Delta := \sum_{i=1}^n \frac{\partial^2}{\partial_i^2}$ heißt der *Laplace*-Operator.

Falls $n = 3$ definiert man außerdem die *Rotation*

$$\operatorname{rot}(f) = (\partial_2 f_3 - \partial_3 f_2, \partial_3 f_1 - \partial_1 f_3, \partial_1 f_2 - \partial_2 f_1)$$

2.10 Lemma. *Falls $f\colon \mathbb{R}^3 \to \mathbb{C}$ 2-mal stetig partiell diffbar ist, gilt*

$$\operatorname{rot}(\operatorname{grad}(f)) = (0, 0, 0).$$

Beweis. Dies folgt durch Einsetzen aus dem Satz von Schwarz 2.8. □

2.11 Definition. Seien $c, k > 0$. Eine zweimal partiell diffbare Funktion $f\colon \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$, für welche gilt

$$\Delta f - \frac{1}{c^2}\frac{\partial^2}{\partial t^2} f(x,t) = 0 \qquad \forall (x,t) \in \mathbb{R}^n \times \mathbb{R}$$

heißt Lösung der *Wellengleichung* mit *Ausbreitungsgeschwindigkeit* $c > 0$..

Falls für f gilt

$$\Delta f - \frac{1}{k}\frac{\partial}{\partial t} f = 0 \qquad \forall (x,t) \in \mathbb{R}^n \times \mathbb{R},$$

so heißt f Lösung der *Wärmeleitungsgleichung* mit Temperaturleitfähigkeit k. Hierbei setzt man in beiden Fällen (nicht ganz konform mit Definition 2.9)

$$\Delta f := \sum_{i=1}^n \partial_i(\partial_i f),$$

d.h. Δf enthält keine Ableitung in Richtung der $n+1$-ten Koordinate (der t-Koordinate).

2.12 Bemerkung. Bei den beschriebenen Gleichungen handelt es sich um *partielle Differentialgleichungen*. Entscheidender Unterschied zu den im letzten Semester betrachteten *gewöhnlichen Differentialgleichungen* ist, dass Ableitungen nach verschiedenen Richtungen vorkommen. Dies macht enorme Unterschiede, z.B. was die "Anzahl" der Lösungen betrifft.

2.13 Beispiel. Sei $n = 1$ und $\phi_1, \phi_2\colon \mathbb{R} \to \mathbb{R}$ beliebige zweimal differenzierbare Funktionen. Dann gilt

$$f(x,t) := \phi_1(x-t) + \phi_2(x+t)$$

ist eine Lösung der Wellengleichung. Sie beschreibt die Überlagerung eines nach rechts laufenden Wellenpakets (gegeben durch ϕ_1) und eines nach links laufenden Wellenpakets (gegeben durch ϕ_2).

Beweis. Ableiten und Einsetzen. □

3 Totale Ableitung

Wir wollen nun den "richtigen" Ableitungsbegriff für auf $\mathbb{R}^n$ definierte Funktionen erarbeiten. Dazu folgende Erinnerung:

Falls $f\colon \mathbb{R} \to \mathbb{R}$ gegeben ist, und $a \in \mathbb{R}$, so beschreibt $f'(a)$ die bestmögliche "lineare" Funktion, die f in der Nähe von a approximiert, indem nämlich (per Definition!)

$$f(a+h) = f(a) + f'(a)h + \rho(h)$$

mit einem "kleinen" Fehler $\rho(h)$ (nämlich so, dass gilt $\rho(h)/h \xrightarrow{h\to 0} 0$). Dies kann man jetzt sofort auf Funktionen mehrerer Veränderlicher verallgemeinern.

3.1 Definition. Sei $f\colon \mathbb{R}^n \to \mathbb{R}^m$ gegeben, und $a \in \mathbb{R}^n$. Eine lineare Abbildung $A\colon \mathbb{R}^n \to \mathbb{R}^m$ heißt *Ableitung* (oder *totale Ableitung*) von f an der Stelle a, falls gilt

$$f(a+h) = f(a) + A \cdot h + \rho(h) \qquad \forall h \in \mathbb{R}^n, \tag{3.2}$$

wobei $\rho(h) \in \mathbb{R}^m$ erfüllt

$$\lim_{h\to 0} \frac{|\rho(h)|_1}{|h|_1} = 0. \tag{3.3}$$

Wir schreiben dann $Df(a) := A$ oder $f'(a) := A$. Beachte: als lineare Abbildung $Df(a) \in \mathrm{Hom}(\mathbb{R}^n, \mathbb{R}^m)$ kann $Df(a)$ auch als $n \times m$-Matrix aufgefasst werden:

$$Df(a) = (a_{ij}) \tag{3.4}$$

3.5 Beispiel. Sei $f\colon \mathbb{R}^n \to \mathbb{R}^m$ linear, also gegeben durch Multiplikation mit einer Matrix A: $f(v) = A \cdot v$. Dann ist diese Abbildung an jeder Stelle $a \in \mathbb{R}^n$ total diffbar, und es gilt

$$Df(a) = A.$$

Beweis. Einsetzen und Linearität ausnutzen. □

Wir wollen nun die Komponenten der Matrix $Df(a)$ berechnen.

3.6 Satz. *Sei $f\colon \mathbb{R}^n \to \mathbb{R}^m$ an $a \in \mathbb{R}^n$ total differenzierbar. Dann ist f (also jede Komponentenfunktion von f) an a partiell diffbar, und es gilt*

$$Df(a) = (\partial_j f_i(a))_{i=1,\ldots,m;\ j=1,\ldots,n}, \tag{3.7}$$

also

$$Df(a) \cdot h = \Big(\sum_{j=1}^n \partial_j f_1(a) h_j, \ldots, \sum_{j=1}^n \partial_j f_m(a) h_i\Big) \tag{3.8}$$

für jeden Vektor $h = (h_1, \ldots, h_n) \in \mathbb{R}^n$.

Außerdem ist f an a stetig.

Beweis. Nach Beispiel 1.10 gilt

$$|f(a+h) - f(a)|_1 = |Df(a)h + \rho(h)|_1 \le |Df(a)h|_1 + |\rho(h)|_1 \xrightarrow{h\to 0} 0,$$

also folgt die Stetigkeit.

Für die erste Behauptung betrachten wir die i-te Komponente der Gleichung (3.2), und speziell den Vektor $h = s \cdot e_j$, wobei e_j der j-te Standardbasisvektor ist. Es gilt dann (mit $Df(a) = (a_{ij})$)

$$f_i(a + se_j) = f_i(a) + a_{ij}s + \rho_i(se_j).$$

Dabei gilt $\frac{|\rho_i(se_j)|}{s} \le \frac{|\rho(se_j)|_1}{|se_j|_1} \xrightarrow{s\to 0} 0$. Also ist nach Definition die Funktion $x \mapsto f_i(a + xe_j)$ an der Stelle 0 ableitbar, mit Ableitung a_{ij}. Dies ist aber gerade die j-te partielle Ableitung von f_i an der Stelle a, also

$$a_{ij} = \partial_j f_i(a),$$

wie behauptet. □

3.9 Satz. *Sei $f\colon \mathbb{R}^n \to \mathbb{R}^m$ stetig partiell diffbar. Dann ist f auch total diffbar. Die Beziehung zwischen der totalen und den partiellen Ableitungen ist dann durch Satz 3.6 gegeben.*

Beweis. Sei $a \in \mathbb{R}^n$. Definiere

$$\rho(h) := f(a + h) - f(a) - (\partial_j f_i(a))_{ij} \cdot h \qquad \forall h \in \mathbb{R}^n.$$

Wir müssen zeigen:

$$\frac{|\rho(h)|_1}{|h|_1} \xrightarrow{h\to 0} 0.$$

Dazu genügt es, für jede Komponente von $\rho(h)$ zu zeigen, dass sie gegen Null konvergiert. Sei $h = (h_1, \dots, h_n)$. Setze $a_0 := a$, $a_1 := a + h_1e_1$, $a_2 := a + h_1e_1 + h_2e_2$, …, $a_n := a + h_1e_1 + \dots h_ne_n = a + h$. Nun gilt

$$\begin{aligned} f_i(a+h) - f_i(a) &= (f_i(a_n) - f_i(a_{n-1})) + (f_i(a_{n-1}) - f_i(a_{n-2})) + \dots + (f_i(a_1) - f_i(a_0)) \\ &= \partial_n f_i(\xi_n)h_n + \dots + \partial_1 f_i(\xi_1)h_1 \end{aligned}$$

unter Anwendung des Zwischenwertsatzes auf die Funktionen $x \mapsto f_i(a_{i-1}+xe_i)$ für x zwischen 0 und h_i. Beachte, dass all diese Funktionen nach Voraussetzung (stetige partielle Diffbarkeit) stetig diffbar sind, mit Ableitungen $\partial_1 f_i, \dots, \partial_n f_i$. Die Zwischenpunkte ξ_i liegen auf der Strecke zwischen a_{i-1} und a_i. Wenn $h \to 0$, konvergiert also ξ_i gegen a für jedes i

Also

$$\begin{aligned} \rho_i(h) &= \left| f_i(a+h) - f_i(a) - \sum_{j=1}^n \partial_j f_i(a)h_j \right| \\ &= \left| \sum_{j=1}^n (\partial_j f_i(\xi_i) - \partial_j f_i(a))h_j \right| \\ &\le \sum_{j=1}^n |\partial_j f_i(\xi_i) - \partial_j f_i(a)|\, |h|_1 \end{aligned}$$

Wegen der Stetigkeit von $\partial_j f_i$ folgt also

$$\frac{|\rho_i(h)|}{|h|_1} \xrightarrow{h\to 0} 0.$$

Damit folgt die Behauptung. □

3.10 Bemerkung. Der Beweis von Satz 3.9 zeigt, dass in der Behauptung die Stetigkeit der partiellen Ableitungen nur am jeweils betrachteten Punkt notwendig ist.

3.11 Lemma. ***Mehr zur stetigkeit linearer Abbildungen:***
Sei $A \in Hom(\mathbb{R}^n, \mathbb{R}^m)$ eine linerae Abbildung, also $A\colon \mathbb{R}^n \to \mathbb{R}^m; x \mapsto A \cdot x$. Dann gibt es eine (von A abhängige) Konstante $C > 0$, so dass

$$|A \cdot x|_1 \leq C \cdot |x|_1 \qquad \forall x \in \mathbb{R}^n.$$

Beweis. Dies haben wir im Beweis von Beispiel 1.10 bereits gesehen. □

3.12 Satz. ***Kettenregel:***
Sei $f\colon \mathbb{R}^n \to \mathbb{R}^m$ total ableitbar an $a \in \mathbb{R}^n$ und $g\colon \mathbb{R}^m \to \mathbb{R}^k$ total ableitbar an $f(a)$. Dann ist auch die Verknüpfung $g \circ f\colon \mathbb{R}^n \to \mathbb{R}^k$ an a total ableitbar, und es gilt

$$D(g \circ f)(a) = Dg(f(a)) \cdot Df(a).$$

Beweis. Nach Definition gilt

$$f(a+h) = f(a)+Df(a)\cdot h+\rho_f(h) \qquad g(f(a)+u) = g(f(a))+Dg(f(a))\cdot u+\rho_g(u),$$

wobei

$$\lim_{h\to 0} \frac{|\rho_f(h)|_1}{|h|_1} = 0, \qquad \lim_{u\to 0} \frac{|\rho_g(u)|_1}{|u|_1} = 0.$$

Folglich

$$\begin{aligned} g \circ f(a+h) =& g(f(a) + Df(a) \cdot h + \rho_f(h)) \\ =& g(f(a)) + Dg(f(a)) \cdot (Df(a) \cdot h + \rho_f(h)) + \rho_g(Df(a) \cdot h + \rho_f(h)) \\ =& g(f(a)) + Dg(f(a)) \cdot Df(a) \cdot h + Dg(f(a)) \cdot \rho_f(h) + \rho_g(Df(a) \cdot h + \rho_f(h)). \end{aligned}$$

Wir setzen also

$$\rho_{g\circ f}(h) := Dg(f(a)) \cdot \rho_f(h) + \rho_g(Df(a) \cdot h + \rho_f(h)),$$

und müssen nur noch zeigen, dass

$$\lim_{h\to 0} \frac{|\rho_{g\circ f}(h)|_1}{|h|_1} = 0.$$

Da $Dg(f(a))$ eine lineare Abbildung, und $Df(a)$ eine lineare Abbildung, existiert nach Lemma 3.11 $C > 0$, so dass $|Dg(f(a))x|_1 \leq C \cdot |x|_1$ und $|Df(a) \cdot y|_1 \leq C \cdot |y|_1$ für alle $x \in \mathbb{R}^m$ und $y \in \mathbb{R}^n$.

Somit gilt

$$\frac{|Dg(f(a)) \cdot \rho_f(h)|_1}{|h|_1} \leq C \frac{|\rho_f(h)_1|}{|h|_1} \xrightarrow{h\to 0} 0,$$

sowie

$$\begin{aligned} \frac{|\rho_g(Df(a) \cdot h + \rho_f(h))|_1}{|h|_1} =& \frac{|\rho_g(Df(a) \cdot h + \rho_f(h))|_1}{|Df(a) \cdot h + \rho_f(h)|_1} \cdot \frac{|Df(a) \cdot h + \rho_f(h)|_1}{|h|_1} \\ \leq& \frac{|\rho_g(Df(a) \cdot h + \rho_f(h))|_1}{|Df(a) \cdot h + \rho_f(h)|_1} \cdot \left(C\frac{|h|_1}{|h|_1} + \frac{|\rho_f(h)|_1}{|h|_1} \right) \xrightarrow{h\to 0} 0 \end{aligned}$$

□

3.13 Korollar. ***Kettenregel für partielle Ableitungen***
Seien $f\colon \mathbb{R}^n \to \mathbb{R}^m$ und $g\colon \mathbb{R}^m \to \mathbb{R}^k$ einmal stetig partiell ableitbar. Dann ist auch die Verknüpfung $g \circ f$ stetig partiell ableitbar, und es gilt für die r-te Komponentenfunktion

$$\partial_i (g \circ f)^r(a) = \sum_{j=1}^m ((\partial_j g^r) \circ f(a)) \cdot \left(\partial_i f^j(a)\right).$$

Beweis. Wegen Satz 3.9 sind f und g total differenzierbar, und wegen Satz 3.12 auch die Verknüpfung $g \circ f$. Wegen Satz 3.6 ist $g \circ f$ dann auch partiell ableitbar. Ausserdem sind die partiellen Ableitungen durch die Matrix $D(g \circ f)$ gegeben. Wir erhalten für $a \in \mathbb{R}^n$

$$\begin{aligned}\left(\partial_i (g \circ f)^j(a)\right)_{i,j} =& D(g \circ f) = Dg(f(a)) \cdot Df(a) \\ &= \left(\partial_\alpha g^\beta(f(a))\right)_{\alpha,\beta} \cdot \left(\partial_\gamma f^\delta(a)\right)_{\gamma,\delta}.\end{aligned}$$

Ausmultiplizieren liefert die Behauptung. Die erhaltenen partiellen Ableitungen von $g \circ f$ sind lineare Funktionen von Verknüpfungen der partiellen Ableitungen von f und g, und daher wieder stetig. □

3.14 Beispiel. Wenn man höhere partielle Ableitungen von Verknüpfungen von Funktionen berechnen will, muss man die Kettenregel mehrfach anwenden, und dazu noch die Produktregel. Die sich ergebenenden Terme werden recht lang. Falls f und g wie in Korollar 3.13, erhalten wir z.B. für $k = 1$

$$\begin{aligned}\partial_i \partial_i (g \circ f)(a) =& \partial_i \left(\sum_{j=1}^m ((\partial_j g) \circ f(a)) \cdot \left(\partial_i f^j(a)\right) \right) \\ =& \sum_{k=1}^m \sum_{j=1}^m \partial_k \partial_j g(f(a)) \cdot \partial_i f^j(a) \partial_i f^k(a) + \sum_{j=1}^m (\partial_j g)(f(a)) \partial_i \partial_i f^j(a).\end{aligned}$$

3.15 Aufgabe. Zeige durch ein Beispiel, dass die Kettenregel für partielle Ableitungen nicht gilt, wenn die totalen Ableitungen nicht existieren.

3.16 Satz. ***Summen-, Produkt- und Quotientenregel*** *Seien $f, g\colon \mathbb{R}^n \to \mathbb{R}^m$ (total) diffbar an $a \in \mathbb{R}^n$. Dann gilt*

(1) $D(\lambda f + \mu g)(a) = \lambda Df(a) + \mu Dg(a)\ forall \lambda, mu \in \mathbb{R}$.

(2) $D(fg)(a) = Df(a) \cdot g(a) + f(a) \cdot Dg(a)$, falls $m = 1$

(3) $D(g^{-1})(a) = -g(a)^{-1} Dg(a) g(a)^{-1}$, falls $m = 1$ und $g(a) \neq 0$.

Beweis. Zunächst beobachten wir, dass auch $f \oplus g\colon \mathbb{R}^n \to \mathbb{R}^m \oplus \mathbb{R}^m; x \mapsto (f(x), g(x))$ an a diffbar ist, mit $D(f \oplus g)(a) = Df(a) \oplus Dg(a)\colon \mathbb{R}^n \to \mathbb{R}^m \oplus \mathbb{R}^m$, da

$$\begin{aligned}f \oplus g(a+h) =& f(a+h), g(a+h)) = (f(a) + Df(a) \cdot h + \rho_f(h), g(a) + Dg(a) \cdot h + \rho_g(h)) \\ =& f \oplus g(a) + (Df(a) \oplus Dg(a)) \cdot h + (\rho_f(h), \rho_g(h)).\end{aligned}$$

Da $|(\rho_f(h), \rho_g(h))|_1 = |\rho_f(h)|_1 + |\rho_g(h)|_1$, konvergiert der Fehlerterm auch nach Division durch $|h|_1$ gegen Null, also folgt die Hilfsbehauptung.

Als nächstes betrachten wir die Abbildung

$$V\colon \mathbb{R}^m \oplus \mathbb{R}^m\colon (v,w) \mapsto \lambda v + \mu w.$$

Diese Abbildung ist linear, also nach Beispiel 3.5 differenzierbar, und die Ableitung stimmt an jedem Punkt mit der linearen Funktion überein.

Weiter gilt $\lambda f + \mu g = V \circ (f \oplus g)$. Nach Kettenregel ist also auch $\lambda f + \mu g$ an a diffbar, mit

$$D(\lambda f + \mu g)(a) = DV \circ D(f \oplus g) = \lambda Df(a) + \mu Dg(a).$$

Alternativ können wir direkt die Definition verwenden:

$$\begin{aligned}(\lambda f + \mu g)(a+h) =& \lambda f(a+h) + \mu g(a+h)\\ =& \lambda(f(a) + Df(a)\cdot h + \rho_f(h)) + \mu(g(a) + Df(a)\cdot h + \rho_g(h))\\ =& (\lambda f + \mu g)(a) + (\lambda Df(a) + \mu Dg(a))\cdot h + \lambda\rho_f(h) + \mu\rho_g(h).\end{aligned}$$

Da

$$\frac{|\lambda\rho_f(h) + \mu\rho_g(h)|_1}{|h|_1} \leq |\lambda|\,\frac{|\rho_f(h)|_1}{|h|_1} + |\mu|\,\frac{|\rho_g(h)|_1}{|h|_1} \xrightarrow{h\to 0} 0,$$

folgt die Behauptung.

Für das Produkt betrachte die Abbildung $m\colon \mathbb{R}^2 \to \mathbb{R}; (x,y) \mapsto xy$. Diese Abbildung ist an jeder Stelle diffbar, mit $Dm(a,b) = (b,a)$, was aus Satz 3.9 und Satz 3.6 folgt, da die partiellen Ableitung $\partial_x m(a,b) = b$ und $\partial_y m(a,b) = a$ stetig sind.

Weiter gilt $fg = m \circ (f \oplus g)$, und wegen der Kettenregel ist auch fg an a total diffbar, mit

$$\begin{aligned}D(fg)(a) =& Dm(f(a),g(a))\cdot Df(a) \oplus Dg(a) = (g(a),f(a))\cdot Df(a) \oplus Dg(a)\\ =& g(a)Df(a) + f(a)Dg(a).\end{aligned}$$

Alternativ kann man auch hier direkt aus der Definition sehen:

$$\begin{aligned}(fg)(a+h) =& f(a+h)g(a+h) = (f(a) + Df(a)\cdot h + \rho_f(a))(g(a) + Dg(a)\cdot h\\ &+ \rho_g(h))\\ =& fg(a) + (f(a)Dg(a) + Df(a)g(a))\cdot h + (f(a) + Df(a)\cdot h)\rho_g(h)\\ &+ \rho_f(a)(g(a) + Dg(a)\cdot h + \rho_g(h)).\end{aligned}$$

Nun gilt

$$\begin{aligned}&\frac{|(f(a) + Df(a)\cdot h)\rho_g(h) + \rho_f(a)(g(a) + Dg(a)\cdot h + \rho_g(h))|_1}{|h|_1}\\ &\leq \frac{|(f(a) + Df(a)\cdot h)\rho_g(h)|}{|h|_1} + \frac{|\rho_f(a)(g(a) + Dg(a)\cdot h + \rho_g(h))|}{|h|_1}\\ &\leq |f(a) + Df(a)\cdot h|\,\frac{|\rho_g(h)|}{|h|_1} + |g(a) + Dg(a)\cdot h + \rho_g(h)|\,\frac{|\rho_f(a)|}{|h|_1} \xrightarrow{h\to 0} 0,\end{aligned}$$

da die vorderen Faktoren beschränkt sind, und die zweiten Faktoren wegen der Differenzierbarkeit gegen Null konvergieren. Es folgt, dass fg total diffbar ist mit $D(fg)(a) = Df(a)g(a) + f(a)Dg(a)$.

Quotientenregel: Übungsaufgabe. □

3.17 Definition. Richtungsableitung:
Sei $f\colon \mathbb{R}^n \to \mathbb{R}$ eine Funktion, $a \in \mathbb{R}^n$. Sei $\xi \in \mathbb{R}^n$ ein "Richtungsvektor". Die *Richtungsableitung* von f im Punkt a in Richtung ξ ist, falls folgender Grenzwert existiert, definiert als

$$\partial_\xi f(a) := \lim_{h\to 0} \frac{f(a+h\xi) - f(a)}{h}.$$

Falls man für ξ den i-ten Einheitsvektor einsetzt, erhält man also die i-te partielle Ableitung.

Üblicherweise wählt man für die Richtung Einheitsvektoren, d.h. $|\xi|_2 = 1$.

3.18 Satz. *Sei $f\colon \mathbb{R}^n \to \mathbb{R}$ im Punkt a total diffbar mit Ableitung $Df(a)$. Sei $\xi = (\xi_1, \dots, \xi_n) \in \mathbb{R}^n$. Dann existiert die Richtungsableitung $\partial_\xi f(a)$, und es gilt*

$$\partial_\xi f(a) = Df(a) \cdot \xi = \sum_{i=1}^n \partial_i f(a) \cdot \xi_i.$$

Beweis. Übungsaufgabe, entsprechend dem Beweis von Satz 3.6. □

3.19 Bemerkung. Sämtliche Definitionen und Resultate dieses Abschnitts verallgemeinern sich, wenn f nur auf einer offenen Teilmenge $U \subset \mathbb{R}^n$ definiert ist. Die Beweise bleiben im Prinzip unverändert.

4 Lokale Extrema, Taylorformel

4.1 Notation. Wenn man Funktionen $f\colon \mathbb{R}^n \to \mathbb{R}$ untersucht, und mehrfach partiell ableitet, gerät man schnell in die Situation, dass die Notation nicht mehr überschaubar wird. Um dies zu verbessern, führen wir nun eine Abkürzende Schreibweise ein. Wegen des Satzes von Schwarz 2.8 kommt es auf die Reihenfolge nicht an, wir müssen uns also nur merken, wie oft in die verschiedenen Richtungen abgeleitet wird. Sei dazu $\alpha = (\alpha_1, \dots, \alpha_n) \in (\mathbb{N} \cup \{0\})^n$. Wir setzten

$$\partial^\alpha := \partial_1^{\alpha_1} \partial_2^{\alpha_2} \cdots \partial_n^{\alpha^n}.$$

Wichtig ist, wieviele Abeitungen insgesamt vorkommen. Hierzu setze

$$|\alpha| := \alpha_1 + \cdots + \alpha_n.$$

Was bei Ableitungen ebenfalls häufig vorkommt, sind Fakultäten. Wir setzten

$$\alpha! := \alpha_1! \cdot \alpha_2! \cdots \alpha_n!.$$

Für einen Vektor $\xi = (\xi_1, \dots, \xi_n) \in \mathbb{R}^n$ setze

$$\xi^\alpha := \xi_1^{\alpha_1} \cdot \xi_2^{\alpha_2} \cdots \xi_n^{\alpha_n}.$$

4.2 Satz. *Sei $f\colon U \to \mathbb{R}$ $(k+1)$-mal stetig differenzierbar, $U \subset \mathbb{R}^n$ offen und $\phi\colon \mathbb{R} \to \mathbb{R}^n$ gegeben durch $f(t) = a + t\xi$ für einen Vektor $\xi \in \mathbb{R}^n$. Die Verbindungsstrecke zwischen a und $a + \xi$ liege ganz in U. Dann ist $f \circ \phi\colon \mathbb{R} \to \mathbb{R}$ $(k+1)$-mal stetig differenzierbar und es gibt $\theta \in [0,1]$, so dass folgende Taylorformel gilt:*

$$\begin{aligned}
f(a+\xi) =& \sum_{i=0}^{k} \sum_{j_1=1}^{n} \cdots \sum_{j_i=1}^{n} \frac{1}{i!} \partial_{j_1} \cdots \partial_{j_i} f(a) \xi_{j_1} \cdots \xi_{j_i} \\
&+ \sum_{j_1=1}^{k+1} \cdots \sum_{j_i=1}^{k+1} \frac{1}{(k+1)!} \partial_{j_1} \cdots \partial_{j_i} f(a+\theta\xi) \xi_{j_1} \cdots \xi_{j_i} \\
=& \sum_{i=0}^{k} \sum_{|\alpha|=i} \frac{\partial^\alpha f(a)}{\alpha!} \xi^\alpha + \sum_{|\alpha|=k+1} \frac{D^\alpha f(a+\theta\xi)}{\alpha!} \xi^\alpha.
\end{aligned}$$

In der zweiten Summe werden die Terme der ersten Summe, die wegen des Satzes von Schwarz 2.8 übereinstimmen, da nur die Differentiationsreihenfolge verschieden ist, zusammengefasst. Man kann dadurch die Schreibweise etwas abkürzen.

Beweis. Nach Kettenregel ist $g = f \circ \phi$ $k+1$-mal differenzierbar. Insbesondere gilt die gewöhnliche Taylorformel (mit geeignetem θ)

$$g(1) = \sum_{i=0}^{k} g^{(i)}(0)/i! + g^{(k+1)}(\theta)/(k+1)!.$$

Wir zeigen nun, dass

$$\begin{aligned}
g^{(i)}(t) &= \sum_{j_1=1}^{n} \cdots \sum_{j_i=1}^{n} \partial_{j_1} \cdots \partial_{j_i} f(a+t\xi) \xi_{j_1} \cdots \xi_{j_i} \\
&= \sum_{|\alpha|=i} \frac{i!}{\alpha!} \partial^\alpha f(a+t\xi) \cdot \xi^\alpha.
\end{aligned}$$

Für die obere Formelmuss man nur induktiv immer wieder die Kettenregel und den Satz von Schwarz 2.8 anwenden. Mit der Induktionsvoraussetzung erhalten wir nämlich

$$\begin{aligned}
g^{(i+1)}(t) =& \sum_{j=1}^{n} \left(\partial_j \sum_{j_1=1}^{n} \cdots \sum_{j_i=1}^{n} \partial_{j_1} \cdots \partial_{j_i} f(a+t\xi) \xi_{j_1} \cdots \xi_{j_i} \right) \cdot \phi_j'(t) \\
=& \sum_{j_1=1}^{n} \cdots \sum_{j_{i+1}=1}^{n} \partial_{j_1} \cdots \partial_{j_{i+1}} f(a+t\xi) \xi_{j_1} \cdots \xi_{j_{i+1}},
\end{aligned}$$

da $\phi_j'(t) = \xi_j$. Die zweite Formel für $g^{(i)}$ ergibt sich aus der ersten, da zu vorgegebenem $\alpha = (\alpha_1, \ldots, \alpha_i)$ vom $|\alpha| = i$ genau $\frac{i!}{\alpha!}$ Summanden in der ersten Summe vorkommen, die denselben Beitrag

$$\partial_1^{\alpha_1} \cdots \partial_n^{\alpha^n} f(a+t\xi) \xi^\alpha$$

liefern. Zum Schluss muss man nur noch beachten, dass $f(a+\xi) = g(1)$, und die Taylorformel folgt. $\square$

4.3 Definition. Wir haben bereits die ersten partiellen Ableitungen zum Gradienten zusammengefasst. Die zweiten partiellen Ableitungen bilden die sogenannte *Hesse-Matrix*

$$Hess f(a) = (\partial_i \partial_j f(a))_{i,j=1,\dots,n}.$$

Wenn f zweimal stetig partiell ableitbar ist, ist dies wegen des Satzes von Schwarz 2.8 eine symmetrische Matrix.

4.4 Korollar. *Sei $f\colon \mathbb{R}^n \to \mathbb{R}$ zweimal stetig diffbar. Dann gilt*

$$f(a+\xi) = f(a) + \langle \operatorname{grad} f(a), \xi\rangle + \frac{1}{2}\langle \xi, Hess f(a)\xi\rangle + \rho(\xi),$$

wobei $\langle (v_1,\dots,v_n),(w_1,\dots,w_n)\rangle = \sum_{i=1}^n v_i\overline{w_i}$, *und* $\frac{\rho(\xi)}{|\xi|_1^2} \xrightarrow{\xi\to 0} 0$.

Beweis. Man wende die Taylorformel für $k=2$ an. Der Korrekturterm $\partial_{j_1}\partial_{j_2} f(a+\theta\xi)$ wird durch $\partial_{j_1}\partial_{j_2} f(a)$ ersetzt. Wegen der Stetigkeit der 2. partiellen Ableitungen geht die Differenz gegen Null. Da man den Fehlerterm $\phi(\xi)$ erhält, indem man diese Differenz mit $\xi_{j_1}\xi_{j_2}$ multipliziert (und dann aufsummiert), und da $|\xi_{j_1}\xi_{j_2}/\,|\xi|_1| \le 1$, konvergiert $\rho(\xi)/\,|\xi|_1^2$ gegen Null für $\xi \to 0$. □

4.5 Definition. Für eine Funktion $f\colon \mathbb{R}^n \to \mathbb{R}$ definiert man *Maximum*, *Minimum*, *lokales Maximum*, *lokales Minimum* wie für Funktionen von $\mathbb{R}$ nach $\mathbb{R}$. Ein Punkt $a \in \mathbb{R}$ heißt insbesondere *lokale Maximumstelle*, wenn es $\epsilon > 0$ gibt, so dass $f(x) \le f(a)$ für alle x mit $|x-a|_1 < \epsilon$.

Beachte, dass es keinen Sinn macht, bei vektorwertigen Funktionen von Extrema zu sprechen. Es macht aber Sinn, statt $\mathbb{R}^n$ Teilmengen als Definitionsbereich für f zuzulassen.

4.6 Satz. *Sei $U \subset \mathbb{R}^n$ offen, $f\colon U \to \mathbb{R}$ partiell diffbar. Sei $a \in U$ eine lokale Extremstelle von f. Dann gilt* $\operatorname{grad} f(a) = 0$.

Beweis. Für jede Richtung gilt dann, dass $\phi_i(t) := f(a+te_i)$ 0 als lokale Extremstelle hat. Also verschwindet die Ableitung von ϕ_i an der Stelle Null. Dies ist nach Definition $\partial_i f(a)$. Weiter gilt $\operatorname{grad} f(a) = (\partial_1 f(a),\dots,\partial_n f(a))$. □

4.7 Definition. Eine (symmetrische) $n\times n$-Matrix A heißt

(1) *positive definit*, falls $c>0$ existiert mit $\langle A\xi,\xi\rangle \ge c\langle \xi,\xi\rangle$ für alle $\xi \in \mathbb{R}^n$.

(2) *negativ definit*, falls $c>0$ existiert mit $\langle A\xi,\xi\rangle < -c\langle \xi,\xi\rangle$ für alle $\xi \in \mathbb{R}^n$.

(3) *indefinit*, falls es ξ_1,ξ_2 gibt, mit $\langle A\xi_1,\xi_1\rangle > 0$ und $\langle A\xi_2,\xi_2\rangle < 0$.

(4) *positiv semidefinit*, falls $\langle A\xi,\xi\rangle \ge 0$ für alle $\xi \ne 0$, entsprechend *negativ semidefinit*.

Es gibt Matrizen, die keine dieser Eigenschaften haben.

4.8 Bemerkung. Falls in Definition 4.7 $\langle A\xi,\xi\rangle > 0$ für alle $\xi \ne 0$, so folgt bereits, dass A positiv definit ist, entsprechend für negativ definit. Eine symmetrische Matrix ist genau dann positiv definit, wenn all ihre Eigenwerte positiv sind.

Beweis. Es ist bekannt, dass zu einer symmetrische Matrix eine Orthonormalbasis des $\mathbb{R}^n$ von Eigenvektoren existiert. Sei $\{v_1, \dots, v_n\}$ solch eine Orthonormalbasis mit Eigenwerten $\lambda_1 \le \dots \le \lambda_n$, also $Av_i = \lambda_i v_i$, und $\langle v_i, v_j\rangle = 0$ falls $i \ne j$, $\langle v_i, v_i\rangle = 1$.

Sei $v \in \mathbb{R}^n$. Dann gilt $v = \sum \alpha_i v_i$ mit geeigneten $\alpha_i \in \mathbb{R}$. Also

$$\langle Av, v\rangle = \sum_{i,j=1}^{n} \langle A\alpha_i v_i, \alpha_j v_j\rangle = \sum_{i=1}^{n} \lambda_i \alpha_i^2.$$

Wir erkennen: $\langle Av, v\rangle > 0$ für alle $v \ne 0$ genau dann, wenn $\lambda_i > 0$ für alle i. In diesem Fall gilt aber, falls $c > 0$ so gewählt ist dass $\lambda_i \ge c > 0$ für $i = 1, \dots, n$, dass

$$\langle Av, v\rangle = \sum_{i=1}^{n} \lambda_i \alpha_i^2 \ge c \sum \alpha_i^2 = c\langle v, v\rangle.$$

Entsprechend für negativ definit. □

4.9 Satz. *Sei $U \subset \mathbb{R}^n$ offen und $f \colon U \to \mathbb{R}$ zweimal stetig differenzierbar. Sei $a \in U$ mit* $\operatorname{grad} f(a) = 0$. *Falls gilt*

(1) $Hessf(a)$ ist positiv definit, so hat f an der Stelle a ein lokales Minimum.

(2) $Hessf(a)$ ist negativ definit, so hat f an der Stelle a ein lokales Maximum.

(3) $Hessf(a)$ ist indefinit, so hat f an a weder ein lokales Maximum, noch ein lokales Minimum (statt dessen liegt ein sogenannter Sattelpunkt *vor).*

Beweis. Benutze Korollar. Danach gilt

$$f(a+\xi) = f(a) + \langle \operatorname{grad} f(a), \xi\rangle + \langle Hessf(a)\xi, \xi\rangle + \rho(\xi) = f(a) + \langle Hessf(a)\xi, \xi\rangle + \rho(\xi).$$

Ist $Hessf(a)$ positiv definit, gibt es also nach Voraussetzung $c > 0$ so dass $\langle Hessf(a)\xi, \xi\rangle \ge c\langle \xi, \xi\rangle = |\xi|_2^2$. Andererseits gilt $\rho(\xi)/\,|\xi|_1^2 \xrightarrow{\xi\to 0} 0$, also existiert $\epsilon > 0$, so dass $|\rho(\xi)| \le c/2\,|\xi|_2^2$ falls $|x|_2^2 < \epsilon$. Insgesamt gilt für solche ξ

$$f(a+\xi) \le f(a) + c\,|\xi|_2^2 + \rho(\xi) \le f(a) + c/2\,|\xi|_2^2 > f(a).$$

f hat also an a ein lokales Minimum. Entsprechend behandelt man den Fall, dass $Hessf(a)$ negativ definit ist.

Sei zuletzt $Hessf(a)$ indefinit und $\xi_1, \xi_2 \in \mathbb{R}^n$ mit $\langle Hessf(a)\xi_1, \xi_1\rangle > 0$, $\langle Hessf(a)\xi_2, \xi_2\rangle < 0$. Wie oben sieht man dann, dass $f(a + t\xi_1) < f(a)$ für $t > 0$ genügend klein, aber $f(a + t\xi_2) > f(a)$ für $t > 0$ genügend klein. Damit folgt die Aussage. □

4.10 Beispiel. $f(x, y) = x^2 + y^2$ hat lokales Minimum bei 0, $g(x, y) = x^2 - y^2$ hat Sattelpunkt bei 0, und $h(x, y) = -x^2 - y^2$ hat lokales Maximum bei 0.

5 Skalarprodukte, Normen, Metriken und Topologien

5.1 Skalarprodukt, Norm und Metrik

Wir kehren zurück zur Frage der Nachbarschaftsbeziehungen. Zunächst betrachten wir diese Frage auf ganz allgemeinen Mengen. Dazu benutzt man das Konzept der Metrik.

5.1 Definition. Sei M eine Menge. Eine Metrik auf M ist eine Funktion

$$d\colon M \times M \to \mathbb{R};\ (x,y) \mapsto d(x,y)$$

mit folgenden Eigenschaften:

- Positivität: $d(x,y) \geq 0\ \forall x,y \in M$. $d(x,y) = 0$ genau dann wenn $x = y$.
- Symmetrie: $d(x,y) = d(y,x)\ \forall x,y \in M$.
- Dreiecksungleichung: $d(x,z) \leq d(x,y) + d(y,z)\ \forall x,y,z \in M$.

5.2 Beispiel. Auf $\mathbb{R}$ und $\mathbb{C}$ ist $d(x,y) := |x-y|$ eine Metrik.

5.3 Definition. Im folgenden werden einige Eigenschaften eingeführt, die sowohl für $\mathbb{R}$- als auch für $\mathbb{C}$-Vektorräume relevant sind. Wir werden deshalb im folgenden hin und wieder von $\mathbb{K}$-Vektorräumen reden, wobei $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$.

5.4 Definition. Sei V ein $\mathbb{R}$-Vektorraum. Eine Norm auf V ist eine Funktion

$$|\colon| V \to \mathbb{R};\ v \mapsto |v|$$

mit folgenden Eigenschaften:

- Positivität: $|v| \geq 0\ \forall v \in V$, $|v| = 0$ genau dann wenn $v = 0$.
- Linearität: $|\lambda v| = |\lambda| \cdot |v|\ \forall \lambda \in \mathbb{R},\ v \in V$.
- Dreiecksungleichung: $|v+w| \leq |v| + |w|\ \forall v,w \in V$.

Entsprechend definiert man eine Norm auf einem $\mathbb{C}$-Vektorraum W, wobei die $|\lambda w| = |\lambda| \cdot |w|$ für alle $\lambda \in \mathbb{C}$ und alle $w \in W$ gelten muss.

5.5 Beispiel. Sei $D \subset \mathbb{C}$ eine Menge. Setze $V := \{f\colon D \to \mathbb{C} \mid f \text{ beschränkt}\}$. Auf V definierten wir die Norm $|f|_\infty := \sup\{|f(x)| \mid x \in D\}$.

Es ist leicht nachzuprüfen, und wir haben auch schon benutzt, dass dies wirklich eine Norm liefert.

Beachte, dass f beschränkt sein muss, damit $|f|_\infty \in \mathbb{R}$. Sei $D \subset \mathbb{C}$ abgeschlossen und beschränkt. Betrachten wir dann nur die stetigen Funktionen, also $C^0(D;\mathbb{C})$ so sind diese automatisch beschränkt.

5.6 Definition. Sei V ein $\mathbb{R}$- oder $\mathbb{C}$-Vektorraum. Ein *Skalarprodukt* auf V ist eine Abbildung $V \times V \to \mathbb{K}; (v,x) \mapsto \langle v,w \rangle$, welche folgende Eigenschaften erfüllt:

(1) $\langle \lambda v_1 + v_2, w \rangle = \lambda \langle v_1, w \rangle + \langle v_2, w \rangle$ für alle $\lambda \in \mathbb{K}$, $v_1, v_2, w \in V$.

(2) $\langle v, w\rangle = \overline{\langle w, v\rangle}$ für alle $v, w \in V$ (falls $\langle w, v\rangle \in \mathbb{R}$, so gilt für das komplex konjugierte natürlich $\overline{\langle w, v\rangle} = \langle w, v\rangle$).

(3) $\langle v, v\rangle \in \mathbb{R}$ und $\langle v, v\rangle \geq 0$ für alle $v \in V$, $\langle v, v\rangle = 0$ genau wenn $v = 0$.

5.7 Satz. ***(Cauchy-Schwarz Ungleichung)***
Sei V ein $\mathbb{R}$- oder $\mathbb{C}$-Vektorraum mit einem Skalarprodukt $\langle\cdot,\cdot\rangle$. Dann gilt für alle $v, w \in V$

$$|\langle v, w\rangle| \leq \sqrt{\langle v, v\rangle} \cdot \sqrt{\langle w, w\rangle}.$$

Falls $v \neq 0$, gilt Gleichheit genau dann, wenn es ein $t \in \mathbb{R}$ gibt mit $tv = w$.

Beweis. Es genügt, dies für $\langle v, w\rangle \neq 0$ zu beweisen. Fixiere solche v und $w \in V$. Setze $\lambda := \overline{\langle v, w\rangle} / |\langle v, w\rangle|$, dies ist eine komplexe Zahl vom Betrag 1. Betrachte dann das quadratische Polynom

$$\langle t\lambda v - w, t\lambda v - w\rangle = \lambda\overline{\lambda}\langle v, v\rangle t^2 - 2\Re\langle \lambda v, w\rangle t + \langle w, w\rangle = \langle v, v\rangle t^2 - 2\,|\langle v, w\rangle|\,t + \langle w, w\rangle =: p(t),$$

da für eine komplexe Zahl z gilt $z\overline{z} = |z|^2$. Dieser Ausdruck wird Null genau dann, wenn $t\lambda v = w$, hat also höchstens eine reelle Nullstelle. Wir wissen aber, dass es zwei reelle Nullstellen gibt, wenn

$$4(|\langle v, w\rangle|^2 - 4\langle v, v\rangle\langle w, w\rangle > 0.$$

Dieser Fall kann also nicht auftreten. Dass Polynom hat genau eine Nullstelle genau dann, wenn für diese Diskriminante gilt $|\langle v, w\rangle|^2 = \langle v, v\rangle\langle w, w\rangle$. Hier steht $\Re$ für den Realteil.

Damit ist der Satz bewiesen. □

5.8 Satz. *Sei V ein $\mathbb{R}$-Vektorraum mit einem Skalarprodukt $\langle\cdot,\cdot\rangle$. Dann definiert*

$$\|v\| := \sqrt{\langle v, v\rangle} \qquad v \in V$$

eine Norm auf V.

Beweis. Es muss nur die Dreiecksungleichung nachgeprüft werden. Sei also $v, w \in V$. Dann gilt $|v + w| \leq |v| + |w|$ genau dann, wenn

$$\langle v + w, v + w\rangle = |v + w|^2 \leq |v|^2 + 2\,|v|\,|w| + |w|^2\,,$$

also

$$\langle v, v\rangle + 2\langle v, w\rangle + \langle w, w\rangle \leq \langle v, v\rangle + 2\sqrt{\langle v, v\rangle}\sqrt{\langle w, w\rangle} + \langle w, w\rangle.$$

Diese Ungleichung folgt aber aus der Cauchy-Schwarz Ungleichung. □

Das wichtigste Beispiel für uns ist durch den euklidischen Raum $\mathbb{R}^n$ gegeben:

5.9 Beispiel. Auf $\mathbb{R}^n$ definieren wir das *euklidische Skalarprodukt*

$$\langle (x_1, \ldots, x_n), (y_n, \ldots, y_n)\rangle := \sum_{i=1}^{n} x_i y_i.$$

Man prüft sofort nach, dass dies ein Skalarprodukt ist. Insbesondere erhält $\mathbb{R}^n$ damit eine Norm $|(x_1, \ldots, x_n)| := \sqrt{x_1^2 + \cdots + x_n^2}$. Diese Norm wird *euklidische Norm* genannt.

Entsprechend definiert man auf $\mathbb{C}^n$

$$\langle (x_1,\dots,x_n),(y_n,\dots,y_n)\rangle := \sum_{i=1}^n x_i\overline{y_i}.$$

Eine weitere Norm auf $\mathbb{R}^n$ oder $\mathbb{C}^n$ erhält man durch

$$|(x_1,\dots,x_n)|_\infty := \max\{|x_1|,\dots,|x_n|\},$$

eine dritte durch

$$|(x_1,\dots,x_n)|_1 := \sum_{k=1}^n |x_k|.$$

5.10 Beispiel. Für $f,g\in C([-1,1],\mathbb{R})$ setze

$$\langle f,g\rangle_{L^2} := \int_0^1 f(x)g(x)\,dx.$$

Dies definiert ein Skalarprodukt auf $C([-1,1];\mathbb{R})$.

Beweis. Übungsaufgabe. □

5.11 Beispiel. In Vorlesung weggelassen Sei $p\ge 1$. Definiere den Raum *(klein)* l^p *der* p*-summierbaren Folgen*

$$l^p := \{z=(z_0,z_1,z_2,\dots)\mid z_i\in\mathbb{C};\ |z|_{l^p} := \left(\sum_{j=0}^\infty |z_j|^p\right)^{1/p} < \infty\}.$$

Für $p=2$ definiert man auf l^2 das Skalarprodukt

$$\langle z,w\rangle_{l^2} := \sum_{k=0}^\infty z_k\overline{w_k}.$$

Beweis. Der Beweis dass es sich um normierte bzw. Innenprodukträume handelt, für $p=1$ und $p=2$ ist Übungsaufgabe, für die übrigen $p\in(1,\infty)$ wird er später geführt (für die Dreiecksungleichung braucht man noch eine weitere fundamental Ungleichung, die Hölder-Ungleichung). □

5.12 Korollar. *Cauchy-Schwarz Ungleichung für* $\mathbb{C}^n$*, Folgenräume und Integrale* *Nach Satz 5.7 gilt in jedem Skalarproduktraum die Cauchy-Schwarz Ungleichung, also auch in den Skalarprodukträumen aus Beispiel 5.9, Beispiel 5.10 und Beispiel 5.11. Somit gilt*

$$\left|\sum_{i=1}^n x_iy_i\right| \le \sqrt{\sum_{i=1}^n |x_i|^2}\cdot\sqrt{\sum_{i=1}^n |y_i|^2}$$

$$\left|\int_a^b f(x)g(x)\,dx\right| \le \sqrt{\int_a^b f(x)^2\,dx}\sqrt{\int_a^b g^2}$$

$$\left|\sum_{i=1}^\infty x_iy_i\right| \le \sqrt{\sum_{i=1}^\infty |x_i|^2}\cdot\sqrt{\sum_{i=1}^\infty |y_i|^2}$$

5.13 Lemma. *Ist V ein Vektorraum mit Norm $|\cdot|$, so definiert $d(v,w) := |v-w|$ eine Metrik auf V.*

Beweis. Übungsaufgabe. $\square$

Wie wir gesehen haben, verallgemeinert der Begriff der Metrik (und der Norm) den Abstandsbegriff, den wir für $\mathbb{R}$ und $\mathbb{C}$ bisher benutzt haben. Insbesondere liefert die euklidische Norm den üblichen Abstandsbegriff, der in $\mathbb{R}^3$ benutzt wird.

Beachte, dass für die Begriffe Konvergenz und Stetigkeit von Folgen und Funktionen in $\mathbb{C}$ *nur* die oben erwähnten elementaren Eigenschaften des Abstandes bzw. der Norm eine Rolle spielen. All diese Konzepte, und die zugehörigen grundlegenden Sätze, verallgemeinern sich also auf Folgen in beliebigen Metrischen Räumen, und auf Abbildungen zwischen metrischen Räumen. Wir werden uns im folgenden insbesondere um Folgen von Funktionen zu kümmern haben, und erklären hier deshalb exemplarisch einiges zu Folgen in metrischen Räumen.

5.14 Beispiel. Sei (M,d) ein metrischer Raum und $U \subset M$. Dann ist die Einschränkung von d auf $U \times U$ eine Metrik auf U, die sogenannte *induzierte* Metrik.

Sei konkret $M = \mathbb{R}^2$ mit der euklidischen Metrik und $U = S^1 := \{v \in \mathbb{R}^2 \mid |v|_2 = 1\}$. Die induzierte Metrik auf S^1 ist *nicht* durch die Länge des Pfades auf S^1 gegeben, sondern durch die Länge der „Abkürzung“ in $\mathbb{R}^2$.

5.15 Definition. Sei M eine Menge mit Metrik d. Eine Folge $(x_n)_{n\in\mathbb{N}}$ mit $x_n \in M$ nennen wir *konvergent mit Grenzwert $x \in M$*, falls für jedes $\epsilon > 0$ ein $N_\epsilon \in \mathbb{N}$ existiert, so dass $d(x_n, x) < \epsilon$ für alle $n \geq N_\epsilon$.

5.16 Satz. *Sei M eine Menge mit Metrik d, $x, x_n \in M$. Dann gilt*

(1) Jede Folge hat höchstens einen Grenzwert: wenn ein Grenzwert existiert, ist er eindeutig.

(2) $x_n \xrightarrow{n\to\infty} x \iff d(x_n - x) \xrightarrow{n\to\infty} 0.$

(3) Teilfolgen und Umordnungen einer konvergenten Folge sind wieder konvergent, und zwar gegen denselben Grenzwert wie vorher. Teilfolgen und Umordnungen werden genauso definiert wie in Diff I für Folgen komplexer Zahlen.

(4) Falls (x_n) eine konvergente Folge, so ist die Folge beschränkt*, d.h. für jedes $x \in M$ gibt es $R > 0$, so dass $d(x_n, x) \leq R$ für alle $n \in \mathbb{N}$.*

Beweis. Der Beweis geht wie die entsprechenden Aussagen für Folgen komplexer Zahlen, die wir in Diff I kennen gelernt haben. Damals wurden nämlich nur die in den Axiomen festgelegten Eigenschaften einer Metrik benutzt. Exemplarisch soll gezeigt werden, dass der Grenzwert eindeutig ist. Sei also $y \in M$, so dass $x_n \xrightarrow{n\to\infty} y$, und $x_n \xrightarrow{n\to\infty} x$. Dann gibt es für jedes $\epsilon > 0$ ein $n \in \mathbb{N}$, so dass $d(x_n, x) < \epsilon$ und $d(x_n, y) < \epsilon$. Wegen der Dreiecksungleichung also $d(x,y) < 2\epsilon$. Da dies für jedes $\epsilon > 0$ gilt, folgt $d(x,y) = 0$. Es folgt dass $x = y$. $\square$

5.17 Definition. Seien X und Y metrische Räume. Eine Funktion $f\colon X \to Y$ heißt *stetig* am Punkt $x \in X$, falls für jedes $\epsilon > 0$ ein $\delta > 0$ existiert, so dass

$$d(f(x), f(y)) < \epsilon \quad \text{für alle } y \in X \text{ mit } d(x,y) < \delta.$$

5.18 Satz. *Sei V ein $\mathbb{C}$-Vektorraum mit Norm $|\cdot|$ und zugehöriger Metrik d. Seien $(v_n)_{n\in\mathbb{N}}$ und $(w_n)_{n\in\mathbb{N}}$ Folgen mit $v_n, w_n \in V$ und mit $\lim_{n\to\infty} v_n = v$, $\lim_{n\to\infty} w_n = w$. Sei $(\lambda_n)_{n\in\mathbb{N}}$ eine Folge komplexer Zahlen mit Grenzwert λ. Dann gilt:*

- $\lim_{n\to\infty}(v_n + w_n) = v + w$.
- $\lim_{n\to\infty}(\lambda_n v_n) = \lambda v$.

Beweis. Wörtlich wie für Folgen komplexer Zahlen. □

5.19 Definition. Sei X ein metrischer Raum. Eine Folge $(x_n)_{n\in\mathbb{N}}$ in X heißt *Cauchyfolge*, falls für jedes $\epsilon > 0$ ein $N \in \mathbb{N}$ existiert, so dass für alle $n > m > N$ gilt $d(x_n, x_m) < \epsilon$.

Falls jede Cauchyfolge in X konvergiert, so nennen wir X *vollständigen metrischen Raum*.

Ist X ein normierter Vektorraum und vollständig, so wird X *Banachraum* genannt.

Hat X ein Skalarprodukt und ist vollständig, so wird X *Hilbertraum* genannt.

5.20 Beispiel. $\mathbb{R}$ und $\mathbb{R}^n$ (mit dem euklidischen Skalarprodukt) sind vollständig, also Hilberträume.

$\mathbb{Q}$ und $\mathbb{Q}^n$ sind (mit derselben Metrik) metrische Räume welche nicht vollständig sind.

5.21 Beispiel. Sei X ein metrischer Raum. Wir definieren $C_b(X;\mathbb{C}) := \{f\colon X \to \mathbb{C} \mid f \text{ stetig und beschränkt}\}$.

$C_b(X)$ mit der Supremumsnorm ist ein Banachraum. Für $f_n, f \in C_b(X)$ gilt $|f_n - f|_\infty \xrightarrow{n\to\infty} 0$ genau dann, wenn die Funktionenfolge f_n gleichmäßig gegen f konvergiert.

Außerdem gilt: Falls $f_n \in C_b(X)$ gleichmäßig gegen eine Funktion $g\colon X \to \mathbb{C}$ konvergiert, so ist g automatisch stetig.

Beweis. Übungsaufgabe. □

5.22 Bemerkung. Punktweise Konvergenz kann auch für Funktionen definiert werden, die nicht beschränkt sind. Außerdem gibt es (beschränkte) Funktionen, die punktweise, aber nicht gleichmäßig konvergieren.

Leider gibt es im allgemeinen keine Metrik auf der Menge aller Funktionen, mit der punktweise Konvergenz beschrieben werden kann. Dies ist einer der Gründe, den Begriff des metrischen Raums noch zu Verallgemeinern, zum *topologschen Raum*. Hierauf werden wir in einem späteren Abschnitt eingehen.

5.23 Beispiel. Die Folgenräume l^p aus Beispiel 5.11 erfüllen: l^2 ist ein Hilbertraum, und l^p für jedes $p \in [1,\infty)$ ein Banachraum.

Beweis. Wir wollen uns hier um die Vollständigkeit kümmern. Sei also $(v_j)_{j\in\mathbb{N}}$ eine Folge von Elementen $v_j = (v_j^0, v_j^1, \dots) \in l^p$, die die Cauchy-Eigenschaft erfüllen, also

$$|v_j - v_k|_{l^p} \xrightarrow{j,k\to\infty} 0.$$

Insbesondere gibt es $C > 0$, so dass $|v_j|_{l^p} \le C$ für alle j.

Beachte zunächst, dass für jede Komponente $v_j^n - v_k^n$ gilt

$$\left|v_j^n - v_k^n\right| \leq |v_j - v_k|_{l^p} \xrightarrow{j,k\to\infty} 0.$$

Da $\mathbb{C}$ vollständig ist, kann man also für jedes $n \in \mathbb{N}$ $v^n := \lim_{j\to\infty} v_j^n \in \mathbb{C}$ definieren. Definiere $v := (v^0, v^1, \dots)$. Wir müssen zeigen, dass $v \in l^p$, und dass $|v_j - v|_{l^p} \xrightarrow{j\to\infty} 0$. Für die erste Aussage findet man für jedes $n \in \mathbb{N}$ ein j, so dass

$$\left|v^1\right|^p + \cdots + |v^n|^p \leq \left|v_j^1\right|^p + \cdots + \left|v_j^n\right|^p + 1 \leq |v_j|_{l^p}^p + 1 \leq C + 1.$$

Da dies für jedes $n \in \mathbb{N}$ gilt, ist $v \in l^p$.

Um Konvergenz zu zeigen, beachte, dass für jedes $\epsilon > 0$ ein $m \in \mathbb{N}$ existiert, so dass $|v_j - v_k|_{l^p} < \epsilon$ für alle $j, k \geq m$. Wähle nun $n \in \mathbb{N}$ so dass $\sum_{k=n}^{\infty} \left|v^k\right|^p < \epsilon$ und $\sum_{k=n}^{\infty} \left|v_m^k\right|^p < \epsilon$. Wähle danach $j > m$ so groß, dass außerdem $\sum_{k=0}^{n-1} \left|v^k - v_j^k\right|^p < \epsilon$. Dann erhält man insgesamt

$$|v - v_j|_{l^p}^p \leq \epsilon + \epsilon + \sum_{k=n}^{\infty} \left|v_j^k\right|^p \leq 2\epsilon + \sum_{k=n}^{\infty} \left|f_m^k\right|^p + |v_m - v_j|_{l^p}^p \leq 3\epsilon + \epsilon^p.$$

Damit folgt die Konvergenz in l^p. □

5.24 Beispiel. Auf dem Raum $C([-1,1];\mathbb{C})$ haben wir in Beispiel 5.10 ein Skalarprodukt definiert. Aber $C([-1,1];\mathbb{R})$ ist bezüglich der zugehörigen Norm nicht vollständig. Betrachte dazu die Funktionenfolge

$$f_n(x) := \begin{cases} -1; & -1 \leq x \leq -1/n \\ nx & -1/n \leq x \leq 1/n \\ 1 & x \geq 1/n. \end{cases}$$

All diese Funktionen sind stetig, und für $m \geq n$ gilt

$$|f_m - f_n|_{L^2} = \sqrt{\int_{-1}^{1} |f_n - f_m|^2} \leq \sqrt{\int_{-1/n}^{1/n} 8} \leq 4/\sqrt{n}.$$

Falls $\epsilon > 0$ und $m > n > 16/\epsilon^2$, gilt also $|f_m - f_n|_{L^2} < \epsilon$. Somit ist die Folge eine Cauchyfolge.

Wäre aber $f \in C([-1,1];\mathbb{R})$ Grenzwert der Folge f_n, so folgt $f(x) = -1$ für $x < 0$ und $f(x) = 1$ für $x > 0$. Sonst gäbe es etwa $x > 0$ mit $f(x) \neq 1$, somit wegen Stetigkeit $|f(t) - 1| > c > 0$ für alle $t \in [x, x+\epsilon)$ für ein $\epsilon > 0$. Falls $n > 1/\epsilon$, somit $\int_{-1}^{1} |f_n - f|^2 \geq \int_x^{x+\epsilon} |1 - f|^2 \geq \epsilon c^2$. Somit kann der Limes nicht 0 sein.

5.25 Satz. ***Banachscher Fixpunktsatz***
Sei $\emptyset \neq M$ ein vollständiger metrischer Raum, z.B. ein Banachraum. Sei $f\colon M \to M$ eine kontrahierende Abbildung*, d.h. es gebe $0 \leq c < 1$, so dass*

$$d(f(x), f(y)) \leq c d(x,y) \qquad \forall x, y \in M.$$

Dann ist f stetig, und es genau einen Fixpunkt *von f, d.h. genau einen Punkt $z \in M$ mit $f(z) = z$.*

Beweis. Wenn $d(x,y) < \epsilon$, so gilt $d(f(x), f(y)) \le c\epsilon < \epsilon$, also ist f stetig.

Seien $z_1, z_2 \in M$ mit $f(z_1) = z_1$ und $f(z_2) = z_2$. Wegen der Kontraktionseigenschaft folgt

$$d(z_1, z_2) = d(f(z_1), f(z_2)) \le c d(z_1, z_2). \tag{5.26}$$

Da $0 \le c < 1$, folgt $d(z_1, z_2) = 0$, also $z_1 = z_2$, es gibt also höchstens einen Fixpunkt.

Um zu beweisen, dass ein Fixpunkt existiert, konstruieren wir eine Cauchyfolge, so dass jeder Grenzwert dieser Folge ein Fixpunkt sein muss. Da M vollständig ist, hat diese Cauchyfolge tatsächlich einen Fixpunkt, die Existenz folgt.

Wähle einen beliebigen Startpunkt $x_0 \in M$. Definiere induktiv für $n > 0$ $x_n := f(x_{n-1})$. Dann gilt, wie sofort mit Induktion aus der Kontraktionsungleichung (5.26) folgt

$$d(x_n, x_{n+1}) \le c^n d(x_0, x_1).$$

Mit der Dreiecksungleichung erhält man, indem man diese Ungleichungen aufsummert, falls $m > n \ge 0$,

$$d(x_m, x_n) \le \sum_{k=1}^{m-n} d(x_{n+k}, x_{n+k-1}) \le c^n \sum_{k=1}^{m-n} c^k d(x_0, x_1) \le c^n/(1-c) d(x_0, x_1).$$

Wir sehen, dass die Folge $(x_n)_{n\in\mathbb{N}}$ eine Cauchyfolge ist. Sie hat also einen Grenzwert z. Für diesen Grenzwert gilt

$$f(z) = f(\lim_{n\to\infty} x_n) = \lim_{n\to\infty} f(x_n) = \lim_{n\to\infty} x_{n+1} = z.$$

□

5.27 Bemerkung. Beim Beweis von Picard-Lindelöff über die Existenz von Lösungen von Differentialgleichungen haben wir (insgeheim) schon den Banachschen Fixpunktsatz bewiesen. Wir haben nämlich auch dort eine Folge von Funktionen konstruiert, indem wir eine kontrahierende Abbildung (mittels Integration) immer wieder auf die vorher erhaltene Funktion anwendeten.

Wie in dieser Anwendung, kann man den Fixpunktsatz oft auch anwenden, wenn man Lösungen von Gleichungen $f(x) = 0$ sucht, nämlich immer dann, wenn $f\colon V \to V$ eine Selbstabbildung eines Banachraums ist (oder eines Vektorraums mit einer Metrik). Man bestimmt dann Fixpunkte der Funktion $g(x) := f(x) - x$.

5.1.1 Operatornorm

5.28 Definition. Seien V, W normierte Vektorräume (über $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$). Wir definieren die Menge der beschränkten linearen Abbildungen

$$\mathcal{B}(V, W) := \{A\colon V \to W \mid A \text{ linear und stetig}\}.$$

Für $A \in \mathcal{B}(V, W)$ definieren wir die *Operatornorm*

$$\|A\| := \sup_{v\in V\setminus\{0\}} \frac{|Av|_W}{|v|_V}.$$

Beachte, dass dies insbesondere auf $A \in \mathrm{Hom}_{\mathbb{R}}(\mathbb{R}^n, \mathbb{R}^m)$ angewendet werden kann (in der Regel werden wir hierbei $\mathbb{R}^n$ mit der euklidischen Norm versehen).

5.29 Lemma. *Seien V, W normierte Vektorräume, $A \in \mathrm{Hom}(V, W)$. A ist genau dann stetig, wenn A beschränkt ist, d.h. wenn $C > 0$ existiert, so dass $|Av| \leq C\,|v|\ \forall v \in V$.*

Beweis. Sei A beschränkt. Sei $v_0 \in V$ und $\epsilon > 0$. Setze $\delta := \epsilon/C$. Falls $v \in V$ mit $|v - v_0| < \delta$, so gilt $|Av - Av_0| = |A(v - v_0)| \leq C\,|v - v_0| < C\delta = \epsilon$, A ist also stetig an v_0.

Ist umgekehrt A stetig an 0, so gibt es zu $\epsilon = 1$ ein $\delta > 0$, so dass $|Av - 0| < 1$ für alle $v \in V$ mit $|v - 0| < \delta$. Für beliebiges $v \neq 0$ heißt dies, da $w := v \cdot \delta/2\,|v|$ Norm $|w| = \delta/2$ hat,

$$|Av| = \left|A((2\,|v|\,/\delta)(v\dot{\delta}/2\,|v|))\right| = \frac{2\,|v|}{\delta}\,|A(v \cdot \delta/2\,|v|)| \leq \frac{2}{\delta}\,|v|\,,$$

also ist A beschränkt. $\square$

5.30 Lemma. *Seien V, W, U normierte Vektorräume. Die Operatornorm ist eine Norm auf $\mathcal{B}(V, W)$. Außerdem gilt $\|AB\| \leq \|A\| \cdot \|B\|\ \forall A \in \mathcal{B}(V, W), B \in \mathcal{B}(U, V)$.*

Beweis. Übungsaufgabe, die sofort aus der Definition folgt. $\square$

5.31 Satz. *Sei V ein normierter Vektorraum und W ein Banachraum. Dann ist $\mathcal{B}(V, W)$ mit der Operatornorm ein Banachraum.*

Beweis. Es muss gezeigt werden, dass $\mathcal{B}(V, W)$ vollständig ist. Sei also $T_n \in \mathcal{B}(V, W)$ eine Cauchyfolge. Sei $x \in V$. Dann ist $(T_n x)_{n \in \mathbb{N}}$ eine Cauchyfolge in W, denn $|T_n x - T_m x| \leq \|T_n - T_m\| \cdot |x| \xrightarrow{n,m\to\infty} 0$. Da W vollständig ist, gibt es $Tx \in W$ mit $T_n x \xrightarrow{n\to\infty} Tx$. Die zuordnung $x \mapsto Tx$ definiert eine lineare Abbildung von V nach W, da für $x, y \in V$ und $\lambda \in \mathbb{K}$

$$T(\lambda x + y) = \lim_{n\to\infty} T_n(\lambda x + y) = \lambda \lim_{n\to\infty} T_n x + \lim_{n\to\infty} T_n y = \lambda Tx + Ty.$$

T ist stetig, und es gilt $\|T - T_n\| \xrightarrow{n\to\infty} 0$, da es zu jedem $\epsilon > 0$ ein $N(\epsilon) \in \mathbb{N}$ gibt, so dass für alle $x \in V$ mit $|x| = 1$ und alle $n \geq N(\epsilon)$ gilt

$$|T_n x - Tx| = \lim_{m\to\infty} |T_n x - T_m x| \leq \epsilon.$$

Es folgt $\|T_n - T\| \xrightarrow{n\to\infty} 0$, insbesondere ist für $n \geq N(\epsilon)$ $T_n - T \in \mathcal{B}(V, W)$, und damit auch $T = (T - T_n) + T_n \in \mathcal{B}(V, W)$. $\square$

5.32 Bemerkung. Identifiziert man $\mathrm{Hom}(\mathbb{R}^n, \mathbb{R}^m)$ mit $\mathbb{R}^{nm}$, über die Koordinaten der den Homomorphismus beschreibenden Matrix, so erhält man auf $\mathbb{R}^{mn}$ eine weitere Norm, die im allgemeinen mit keiner der bisher definierten Normen übereinstimmt.

5.1.2 Offene und abgeschlossene Mengen

5.33 Definition. Sei (M, d) eine Menge mit Metrik. Sei $U \subset M$. Wir nennen U *offen*, falls für jedes $x \in U$ ein $\epsilon > 0$ existiert, so dass der *ϵ-Ball* um x $B_x(\epsilon) := \{y \in M \mid d(x, y) < \epsilon\}$ ganz in U enthalten ist: $B_x(\epsilon) \subset U$.

Ist $x \in M$ und $U \subset M$ offen, so dass $x \in U$, so heißt U *offene Umgebung* von x.

Wir nennen $A \subset M$ abgeschlossen, wenn das Komplement $M \setminus A$ offen ist.

Sei $X \subset M$. Der *Abschluss* von X $\overline{X} \subset M$ ist definiert als die Schnittmenge aller abgeschlossenen Mengen, welche X enthalten. Äquivalent gilt: $x \notin \overline{M}$ genau dann wenn es keine offene Umgebung U von x gibt, so dass $U \cap X = \emptyset$.

Sei $X \subset M$. Das *Innere* X° von X ist die Vereinigung aller offenen Mengen, welche in X enthalten sind. Äquivalent gilt: $x \in X^\circ$ genau dann, wenn es eine offene Umgebung U von x gibt, so dass $U \subset X$.

5.34 Lemma. *Sei (M, d) eine Menge mit Metrik. Sei I eine beliebige Indexmenge und $U_i \subset M$ für $i \in I$ seien offen. Dann ist auch $U := \bigcup_{i\in I} U_i$ offen.*

Seien U_1 und $U_2 \subset M$ offen, dann ist auch $U_1 \cap U_2$ offen.

Es gilt M ist offen, und $\emptyset$ ist offen.

Beweis. Sei $x \in U$. Dann gibt es $i \in I$, so dass $x \in U_i$. Da U_i offen ist, gibt es $\epsilon > 0$, so dass $B_x(\epsilon) \subset U_i \subset U$. Also ist U offen.

Sei $x \in U_1 \cap U_2$. Dann gibt es, da sowohl U_1 als auch U_2 offen sind, $\epsilon_1, \epsilon_2 > 0$ so dass $B_x(\epsilon_1) \subset U_1$ und $B_x(\epsilon_2) \subset U_2$. Sei $\epsilon = \min\{\epsilon_1, \epsilon_2\}$. Dann ist $B_x(\epsilon) \subset U_1 \cap U_2$. □

5.2 Topologische Räume

5.35 Definition. Sei M eine Menge. Eine *Topologie* auf M ist eine Menge τ von Teilmengen von M (genannt die *offenen Mengen* der Topologie), so dass folgende Eigenschaften erfüllt sind:

(1) $\emptyset \in \tau$, $M \in \tau$.

(2) Falls $U, V \in \tau$, dann auch $U \cap V \in \tau$

(3) Falls I eine beliebige Indexmenge ist und $U_i \in \tau$ für jedes $i \in I$, so gilt auch $\bigcup_{i\in I} U_i \in \tau$.

Also: die leere Menge und die Menge selbst sind immer offen, der Schnitt zweier offener Mengen ist wieder offen, die Vereinigung beliebiger offener Mengen ist wieder offen.

5.36 Beispiel. Sei M eine Menge mit Metrik d, und τ die bezüglich der Metrik offenen Mengen nach Definition 5.33. Dann ist nach Lemma 5.34 τ eine Topologie auf M.

5.37 Beispiel. Extrembeispiele: für jede Menge M ist $\{\emptyset, M\}$ eine Topologie auf M, genauso die Menge aller Teilmengen auf M. die zweite Topologie heißt die *diskrete Topologie* auf M.

5.38 Definition. Seien (M, τ_M) und (N, τ_N) topologische Räume (also Mengen mit einer Topologie). Eine Abbildung $f \colon M \to N$ heißt *stetig*, falls das Urbild $f^{-1}(U)$ offen ist für jede offene Teilmenge U von N.

5.39 Lemma. *Dies ist für metrische Räume äquivalent zum bisherigen Stetigkeitsbegriff.*

Beweis. Übungsaufgabe. □

5.40 Beispiel. Für die Stetigkeit kommt es natürlich auf die verwendete Topologie an. Ist M eine Menge und τ_d die diskrete Topologie auf M (also alle Teilmengen sind offen), so ist *jede* Abbildung $f\colon M \to X$ stetig (wenn alle Mengen offen sind, ist auch das Urbild jeder offenen Menge offen). Wählt man umgekehrt die total indiskrete Topologie $\tau = \{\emptyset, \tau\}$, so ist jede Abbildung $f\colon X \to M$ stetig.

Wir wissen, dass es für die gewöhnliche Topologie von $\mathbb{R}$ auch Funktionen $f\colon \mathbb{R} \to \mathbb{R}$ gibt, die nicht stetig sind.

5.41 Definition. Sei M ein topologischer Raum mit Topologie τ und $N \subset M$. Wir definieren die *Teilraumtopologie* $\tau_N := \{U \cap N \mid U \in \tau\}$. Dies ist eine Topologie auf N. Die Inklusion $N \to M$ ist bezüglich τ_N und τ stetig.

5.3 Kompaktheit

5.42 Definition. Sei M eine Menge mit Topologie τ und sei $K \subset M$. K heißt *kompakt* (in M), falls für jede Kollektion von offenen Teilmengen $(U_i \in \tau)_{i\in I}$ (I eine beliebige Indexmenge), so dass $K \subset \bigcup_{i\in I} U_i$ endlich viele dieser Teilmengen $U_{i_1}, \dots, U_{i_n}$ gefunden werden können, so dass schon gilt

$$K \subset \bigcup_{k=1}^{n} U_{i_k}.$$

Anders ausgedrückt: wenn immer man K mit (einer beliebigen Zahl) offener Mengen überdecken kann, reichen schon endlich viele dieser Mengen aus, um K zu überdecken.

5.43 Beispiel. Jede Teilmenge, die nur einen Punkt enthält, ist kompakt. Das gleiche gilt für Teilmengen mit endlich vielen Punkten.

5.44 Satz. *Sei (M, d) ein metrischer Raum, $K \subset M$ eine Teilmenge. Es gilt: K ist kompakt genau dann, wenn jede Folge $(x_n) \subset K$ eine konvergente Teilfolge mit Grenzwert in K besitzt.*

Beweis. Sei $(x_n)_{n\in\mathbb{N}}$ eine Folge in M, welche keine konvergente Teilfolge besitzt. Für jedes $x \in M$ wähle eine offene Umgebung U_x von x, so dass nur endlich viele Folgenglieder in U_x liegen (dies geht, da keine Teilfolge gegen x konvergiert, es könnte etwa $U_x = B_x(\epsilon_x)$ für geeignetes $\epsilon_x > 0$ sein). Dann ist $\{U_x \mid x \in M\}$ eine offene Überdeckung von M. Da in jedem U_x nur endlich viele der Folgenglieder liegen, liegen auch in jeder endlichen Vereinigung nur endlich viele Folgenglieder, also insbesondere nicht jeder Punkt aus M, so dass keine endliche Teilmenge von $\{U_x\}$ schon ganz M überdeckt. Somit ist M nicht kompakt.

Die Umkehrung wird hier nicht bewiesen. Sie ist recht kompliziert. □

5.45 Korollar. *Sei M ein metrischer Raum und $K \subset M$ kompakt. Dann ist K abgeschlossen.*

Beweis. Übungsaufgabe. Lösung: Sei $(x_n)_{n\in\mathbb{N}}$ eine Folge in K, welche in M konvergiert. Wir müssen zeigen, dass der Grenzwert zu K gehört. Da K kompakt ist, hat (x_n) eine in K konvergente Teilfolge. Der Grenzwert dieser Teilfolge stimmt mit dem Grenzwert von (x_n) überein, also liegt auch der letztere in K. □

5.46 Korollar. *Sei K kompakter topologischer (z.B. metrischer) Raum und $L \subset K$ abgeschlossen. Dann ist auch L kompakt.*

Beweis. Übungsaufgabe. Lösung: Sei $(U_i)_{i\in I}$ eine offene Überdeckung von L. Da L abgeschlossen, ist $K \setminus L$ eine offene Teilmenge von L. Somit bildet $(U_i)_{i\in I}$ zusammen mit $K \setminus L$ eine offene Überdeckung von K. Dies hat eine endliche Teilüberdeckung, etwas $U_{i_1}, \dots, U_{i_n}$, eventuell zusammen mit $K \setminus L$. Auf jeden Fall ist dann $U_{i_1}, \dots, U_{i_n}$ eine endliche Teilüberdeckung von L. □

5.47 Bemerkung. Der Beweis von Satz 5.44 zeigt, dass in jeder kompakten Teilmenge K eines beliebigen topologischen Raumes jede Folge eine konvergente Teilfolge besitzt.

Es kann aber Teilmengen geben, in denen jede Folge eine konvergente Teilfolge besitzt, und die trotzdem nicht kompakt sind (also die Überdeckungs-Eigenschaft nicht erfüllen). Aber natürlich nicht in metrischen Räumen.

5.48 Satz. *Sei $K \subset \mathbb{R}^n$, wobei wir auf $\mathbb{R}^n$ die Topologie von der euklidischen Norm benutzen. Dann gilt:*

K ist kompakt genau dann, wenn K abgeschlossen und beschränkt ist.

Beweis. Sei zunächst $K \subset \mathbb{C}^n$ abgeschlossen und beschränkt. Sei $(U_i)_{i\in I}$ eine offene Überdeckung von K. Wir wollen das Halbierungsverfahren anwenden (welches wir ja schon für die Existenz einer konvergenten Teilfolge einer beschränkten Folge in $\mathbb{R}$ benutzt haben), um eine endliche Teilüberdeckung zu finden. Genauer gesagt: wir nehmen an, dass es keine endliche Teilüberdeckung gibt, und führen dies zum Widerspruch. Da K beschränkt ist, gibt es einen Würfel $W = [-R, R]^n$ ($R > 0$ geeignet), so dass $K \subset W$. W teilen wir nun in 2^n abgeschlossene Teilwürfel mit jeweils halber Kantenlänge (nicht disjunkt, an Kanten schneiden sich diese Teilwürfel). Mindestens einer dieser Teilwürfel (nenne ihn W^1) erfüllt, dass $K \cap W^1$ nicht von endlich vielen Teilmengen aus $\{U_i\}$ überdeckt wird (wenn für jeden der endlich vielen Teilwürfel endlich viele Mengen reichen, dann auch insgesamt).

Durch schrittweises halbieren erhält man so eine geschachtelte Folge $W \supset W^1 \supset W^2 \supset \cdot$ von (abgeschlossenen) Würfeln, deren Kantenlänge jeweils halbiert wird, und so dass $K \cap W^k$ für kein $k \in \mathbb{N}$ von nur endlich vielen der $\{U_i\}$ überdeckt wird (insbesondere ist $K \cap W^k$ nie leer).

Man wähle aus jedem $K \cap W^k$ einen Punkt x_k aus. Die Folge (x_k) ist eine Cauchyfolge, hat also einen Grenzwert x. Da K abgeschlossen, gilt $x \in K$. Also gibt es $i \in I$, so dass $x \in U_i$. Für genügend grosses k gilt dann aber auch $W^k \in U_i$ (da U_i offen, und die Kantenlänge von W^k gegen Null konvergiert, aber $x \in W^k$).

Dies ist ein Widerspruch dazu, dass $K \cap W^k$ nicht von endlich vielen der U_i überdeckt wird, es reicht ja sogar eine einzige dieser Mengen. □

5.49 Satz. *Seien M, N topologische Räume (z.B. metrische Räume), $f \colon M \to N$ stetig und $K \subset M$ kompakt. Dann ist die Bildmenge $f(K)$ ebenfalls kompakt.*

Beweis. Sei I eine Indexmenge und $(V_i)_{i\in I}$ eine offene Überdeckung von $f(K)$. Setze $U_i := f^{-1}(V_i)$. Da f stetig ist, ist U_i offen für jedes $i \in I$. Also ist U_i eine offene Überdeckung von K. Es gibt also eine endliche Teilüberdeckung $U_{i_1}, \dots, U_{i_n}$. Dann ist $V_{i_1}, \dots, V_{i_n}$ eine endliche Teilüberdeckung von $f(K)$. □

5.50 Satz. *Seien K, M kompakte metrische Räume, $f\colon K \to M$ eine stetige bijektive Abbildung. Dann ist die Umkehrabbildung $g\colon L \to M$ ebenfalls stetig.*

Beweis. Es genügt, für jede abgeschlossene Teilmenge $L \subset K$ zu zeigen, dass $g^{-1}(L) = f(L)$ abgeschlossen ist. Nach Korollar 5.46 ist L kompakt, daher wegen Satz 5.49 auch $f(L)$, und nach Korollar 5.45 ist $f(L) = g^{-1}(L)$ somit abgeschlossen. □

5.51 Satz. *Sei M topologischr Raum (z.B. eub metrischer Raum), $f\colon M \to \mathbb{R}$ stetig und $K \subset M$ kompakt. Dann gilt:*

es gibt $x \in K$, so dass $f(x) \le f(y)$ für alle Punkte $y \in K$. Die Einschränkung von f auf K nimmt also auf K ihr Minimum an. Insbesondere ist f auf K nach unten beschränkt.

Entsprechend für Maximum und nach oben beschränkt.

Beweis. Da f stetig, ist nach Satz 5.49 $f(K)$ eine kompakte Teilmenge von $\mathbb{R}$, also nach Satz 5.48 eine abgeschlossene beschränkte Teilmenge von $\mathbb{R}$. Insbesondere ist $f|_K$ beschränkt. Da $f(K)$ abgeschlossen ist, liegen $\sup\{f(x) \mid x \in K\}$ und $\operatorname{ind}\{f(x) \mid x \in K\}$ in der Bildmenge $f(K)$, werden also angenommen.

Man kann auch folgendermaßen direkt argumentieren:

Die Betrachtung des Maximum reduziert sich auf die betrachtung des Minimum von $-f$.

Die Mengen $V_n := (-n, n) \subset \mathbb{R}$ sind offen, also auch $U_n := f^{-1}(V_n)$. Weiter gilt für jedes $x \in M$, dass $n \in \mathbb{N}$ existiert mit $-n < f(x) < n$. Also gilt $\bigcup_{n\in\mathbb{N}} U_n = M$. Da K kompakt ist, gibt es endlich viele Teilmengen $U_1, \dots, U_N$, so dass $K \subset U_1 \cup \dots \cup U_N$. Demzufolge gilt $-N < f(x) < N$ für alle $xinK$, also ist $f|_K$ beschränkt. Sei $R := \inf\{f(x) \mid x \in K\}$. Wir wollen zeigen, dass $x \in K$ existiert mit $f(x) = R$. Wäre dies nicht der Fall, so würden die Mengen $W_n := f^{-1}(R+1/n, \infty)$ eine offene Überdeckung von K bilden (da dann $f(x) > R$ für alle $x \in K$). Da bereits eine endliche Teilüberdeckung ausreichen würde, gäbe es $N \in \mathbb{N}$ so dass $f(x) > R + 1/N$ für alle $x \in K$. Dann folgt aber auch $\inf\{f(x) \mid x \in K\} > R + 1/N$, ein Widerspruch. □

5.52 Korollar. *Sei $A \subset \mathbb{R}^n$ abgeschlossen und beschränkt und $f\colon A \to \mathbb{R}$ stetig. Dann nimmt f auf A sein Maximum und Minimum an, ist insbesondere beschränkt.*

Beweis. Nach Satz 5.48 ist A kompakt, die Aussage folgt also aus Satz 5.51.

Alternativ kann man folgenden direkten Beweis geben, der den Beweis für $\mathbb{C}$ imitiert.

Angenommen, f wäre unbeschränkt. Dann gibt es eine Folge (x_k) in K, so dass $|f(x_k)| \xrightarrow{k\to\infty} +\infty$. Indem wir in n Schritten zu Teilfolgen übergehen, können wir eine Teilfolge (x_{k_j}) erhalten, so dass jede Komponente $(x^i_{k_j})$ eine konvergente Folge reeller Zahlen bildet (da jede beschränkte Folge reeller Zahlen eine konvergente Teilfolge hat). Dann ist aber auch die Folge der Vektoren (x_{k_j}) konvergent gegen einen Grenzwert $x \in \mathbb{R}^n$. Da K abgeschlossen, gilt $x \in K$. Da f stetig ist, gilt $f(x) = \lim_{j\to\infty} f(x_{k_j})$. Andererseits sollte $f(x_k)$ gegen $+\infty$ divergieren. Dies ist ein Widerspruch, also ist die Funktion beschränkt. □

5.53 Lemma. *Sei A eine symmetrische $n \times n$-Matrix über $\mathbb{R}$ mit $\langle Ax, x\rangle > 0$ $\forall x \in \mathbb{R}^n \setminus \{0\}$. Dann gibt es $c > 0$, so dass $\langle Ax, x\rangle \ge c\langle x, x\rangle$ $\forall x \in \mathbb{R}^n$.*

Beweis. Übungsaufgabe. Betrachte die Einschränkung der Funktion $x \mapsto \langle Ax, x\rangle$ auf $S^{n-1} := \{x \in \mathbb{R}^n \mid |x|_2 = 1\}$. □

5.54 Definition. Sei V ein $\mathbb{K}$-Vektorraum und $|\cdot|, |\cdot|'$ zwei Normen auf V. Diese Normen heißen *äquivalent*, falls es $C > 0$ gibt, so dass $C^{-1}|v| \leq |v|' \leq C|v|$ $\forall v \in V$.

5.55 Lemma. *Sei V ein Vektorraum mit zwei Normen $|\cdot|$ und $|\cdot|'$. Diese sind äquivalent genau dann, wenn die identische Abbildung* $\mathrm{id}\colon V \to V$ *stetig sowohl als Abbildung von $(V, |\cdot|)$ nach $(V, |\cdot|')$ als auch als Abbildung von $(V, |\cdot|')$ nach $(V, |\cdot|)$ ist (beachte, dass es für Stetigkeit auf die Topologie, also auf die Norm ankommt).*

Insbesondere sind die induzierten Topologien gleich, für topologische Fragen (wie Stetigkeit, Grenzwert,...) kann man also eine Norm durch eine äquivalente Norm ersetzen.

Beweis. Übungsaufgabe. □

5.56 Satz. *Sei $(V, ||)$ ein $\mathbb{K}$-Vektorraum der Dimension n. Sei $T\colon \mathbb{R}^n \to V$ ein Vektorraumisomorphismus. Dann sind T und T^{-1} stetig.*

Folgerung: zwei beliebige Normen auf $\mathbb{R}^n$ sind äquivalent. Zwei beliebige Normen auf $\mathbb{C}^n$ sind äquivalent.

Beweis. Sei $v = \sum_{i=1}^n \lambda_i e_i$. Es gilt

$$|Tv| = |T(\lambda_1 e_1 + \cdots + \lambda_n e_n)| \leq |\lambda_1|\,|Te_1| + \cdots + |\lambda_n|\,|Te_n| \leq |v|_1 \max\{|Te_1|, \ldots, |Te_n|\}.$$

Damit sieht man, dass $T\colon \mathbb{R}^n \to V$ stetig ist. Da auch die Abbildung $x \mapsto |x|; (V, ||) \to \mathbb{R}_{\geq 0}$ stetig ist, ist die Verknüpfung $v \mapsto |Tv|$ stetig. Auf der kompakten Menge $K := \{v \in \mathbb{R}^n \mid |v|_1 \leq 1\}$ gibt es also $x_m \in K$ mit $|Tx| \geq |Tx_m| =: c > 0$ $\forall x \in K$.

Somit $|Tx| = \left|T(|x|_1 \frac{x}{|x|_1})\right| = |x|_1 \left|T(\frac{x}{|x|_1})\right| \geq |x|_1 \cdot c$, oder $\left|T^{-1}v\right| \leq \frac{1}{c}|v|_1$ für alle $v \in V$ (benutze die obige Ungleichung mit $x = T^{-1}v$). □

6 Mächtigkeit von Mengen

Wir haben im letzten Abschnitt mit beliebigen Indexmengen gearbeitet. Hier wollen wir nun ein klein wenig über die "Größe" solcher Mengen lernen.

6.1 Definition. Seien M_1 und M_2 zwei Mengen. Wir sagen, dass M_1 und M_2 *gleiche Mächtigkeit* besitzen, wenn es eine Bijektion $f\colon M_1 \to M_2$ gibt (beachte: es wird nur die Existenz einer beliebigen Abbildung gefordert, die injektiv und surjektiv ist. Linearität, Stetigkeit etc. werden hier *nicht* verlangt). Dann schreiben wir $|M_1| = |M_2|$.

Wir sagen, M_1 hat *Mächtigkeit nicht größer* als M_2, falls es eine injektive Abbildung $g\colon M_1 \to M_2$ gibt. In diesem Fall schreiben wir $|M_1| \leq |M_2|$.

6.2 Beispiel. Für jede endliche Menge M gibt es genau ein k, so dass M die gleiche Mächtigkeit wie wie $\{1, \ldots, k\}$. Wir schreiben dann $|M| = k$.

Beweis. Wir erinnern uns, dass eine Menge genau dann endlich ist, wenn sie zu einer der Mengen $\{1, \ldots, k\}$ bijektiv ist (dies ist die Definition). □

6.3 Satz. *Seien M und N Mengen, und $f\colon M \to N$ injektiv, sowie $g\colon N \to M$ injektiv. Dann gibt es eine bijektive Abbildung $h\colon M \to N$.*

Das heißt: wenn sowohl $|M| \le |N|$ also auch $|N| \le |M|$, dann gilt $|M| = |N|$.

Beweis. Nicht so einfach, vergleiche Klaua: Mengenlehre, V.14, Satz 6. □

6.4 Satz. *Sei M eine Menge. Mit $P(M)$ bezeichnen wir die* Potenzmenge $\{U \subset M\}$*, die Menge aller Teilmenge von M. Dann gibt es keine surjektive Abbildung $M \to P(M)$. Insbesondere hat $P(M)$ größere Mächtigkeit als M.*

Beweis. Sei $f\colon M \to P(M)$ eine Abbildung. Wir zeigen, dass es mindestens eine Teilmenge von M gibt, welche nicht im Bild von f liegt. Definiere dazu $U := \{x \in M \mid x \notin f(x)\}$. Für beliebiges $x \in M$ gilt dann $f(x) \neq U$, da entweder $x \in f(x)$, dann $x \notin U$, oder $x \notin f(x)$, dann $x \in U$. Also ist f nicht surjektiv. □

6.5 Definition. Eine Menge heißt abzählbar, wenn sie dieselbe Mächtigkeit wie $\mathbb{N}$ hat.

6.6 Satz. *Eine abzählbare Vereinigung endlicher oder abzählbarer Mengen ist endlich oder abzählbar.*

Beweis. Seien $M_0, M_1, M_2, \dots$ abzählbar viele endliche oder abzählbare Mengen. Definiere eine diagonale Bijektion f von $\mathbb{N}$ nach $\bigcup_{n\in\mathbb{N}} M_n$ (falls diese Vereinigung nicht endlich ist): $f(0)$ sei das erste Element von M_0 (falls M_0 nicht leer, ansonsten wird M_0 einfach weggelassen). $f(1)$ sei das erste Element von M_1, dass noch nicht im Bild von f liegt (falls solch ein Element existiert, ansonsten wird (ab jetzt) auch M_1 ausgelassen). $f(2)$ sei das zweite noch nicht gezählte Element von M_0, $f(3)$ das zweite Element von M_1, $f(4)$ das erste Element von M_3. Fährt man so fort, erhält man die gewünschte Bijektion. Vergleiche Bild in Vorlesung. □

6.7 Beispiel. $\mathbb{Z}$ ist abzählbar, eine mögliche Bijektion ist gegeben durch $f\colon \mathbb{Z} \to \mathbb{N}$ mit $f(x) = 2x$ falls $x \ge 0$ und $f(x) = -2x - 1$ falls $x < 0$.

$\mathbb{N}\times\mathbb{N}$ ist abzählbar mit dem Diagonaltrick, da Vereinigung abzählbarer vieler Mengen $M_n := \{(x, n) \mid x \in \mathbb{N}, n \in \mathbb{N}$.

$\mathbb{Q}$ ist abzählbar, da es abzählbare Vereinigung der abzählbaren Mengen $M_n := \{z/n \mid z \in \mathbb{Z}\}$, $n \in \mathbb{N} \setminus \{0\}$ ist.

6.8 Satz. $\mathbb{R}$ *ist nicht abzählbar.*

Beweis. Angenommen, dass Intervall $(0, 1]$ wäre abzählbar. Jedes Element $x \in (0, 1]$ kann in eindeutiger Weise als Dezimalbruch $0, a_1a_2a_3 \dots$ geschrieben werden, wenn man voraussetzt, dass nicht alle $a_i = 0$ für $i > N$ ($0,9999999 \dots = 1$, $0,4999999 \dots = 1/2$).

Wir erhalten also abzählbar viele solche Dezimalbrüche $z_1, z_2, \dots$. Konstruiere einen neuen Dezimalbruch $z = 0, b_0b_1b_2 \dots$, mit $b_i = 0$, falls die i-te Stelle von z_i ungleich 9, und $b_i = 1$, falls die i-te Stelle von z_i gleich 9. Dann gilt $z \in (0, 1]$, aber $z \neq z_i$ für jedes $i \in \mathbb{N} \setminus \{0\}$, da sich die beiden Zahlen ja zumindest in der i-ten Nachkommastelle unterscheiden.

Dies ist ein Widerspruch zu der Annahme, dass $(z_i)_{i=1,2,\dots}$ eine Abzählung von $(0, 1]$ ist. □

6.9 Bemerkung. Es gibt eine surjektive Abbildung $\mathbb{R} \to \mathbb{R} \times \mathbb{R}$.

Beweis. Übungsaufgabe. □

7 Implizite Funktionen, lokale Umkehrfunktion

Wir kehren nun wieder zurück zur Untersuchung von Funktionen $f\colon \mathbb{R}^n \to \mathbb{R}^m$. Wichtig für viele Anwendung ist das Auflösen von Gleichungen: gegeben solch eine Funktion f, finde alle $x \in \mathbb{R}^n$ mit $f(x) = 0$. In dieser Allgemeinheit gibt es keine Chance, Lösungsverfahren anzugeben. Wir werden in diesem Kapitel Bedingungen finden, wie mit einer Lösung gleich eine ganze "Familie" weiterer Lösungen gefunden wird (genauer gesagt: wir werden ihre Existenz nachweisen, ohne die Lösungen konkret berechnen zu können).

Die Größe dieser Familie wird durch die Situation vorgegeben, unter geeigneten Voraussetzungen gibt es genau eine Lösung, daraus folgt dann die (lokale) Umkehrbarkeit der Funktion.

7.1 Satz. ***Satz von der lokalen Umkehrfunktion***
Sei $U \subset \mathbb{R}^n$ offen und $f\colon U \to \mathbb{R}^n$ stetig partiell diffbar. Sei $a \in U$ und die lineare Abbildung

$$Df(a) \in \mathrm{Hom}_{\mathbb{R}}(\mathbb{R}^n, \mathbb{R}^n)$$

sein invertierbar (also $\det Df(a) \neq 0$). Dann gibt es eine offene Teilmenge $V \subset U$ mit $a \in V$, so dass gilt:

(1) Die Einschränkung von f auf V ist injektiv.

(2) Die Bildmenge $f(V) \subset \mathbb{R}^n$ ist offen.

(3) Die Umkehrabbildung $g\colon f(V) \to V$ von $f|_V$ ist stetig diffbar. Es gilt

$$Dg(f(a)) = Df(a)^{-1}$$

Beweis. Gegeben unten, siehe Seite 34. □

Für das Auflösen von Gleichungen erhalten wir folgenden Satz:

7.2 Satz. ***Satz über implizite Funktionen***
Sei $f\colon \mathbb{R}^{n+k} \to \mathbb{R}^k$ stetig diffbar. Schreibe $(v, w) \in \mathbb{R}^{n+k}$, mit $v \in \mathbb{R}^n$ und $w \in \mathbb{R}^k$. Es sei $(v_0, w_0) \in \mathbb{R}^{n+k}$ mit $f(v_0, w_0) = 0$.

Die Ableitung $Df(v_0, w_0) \in \mathrm{Hom}(\mathbb{R}^{n+k}, \mathbb{R}^k)$ ist eine $(n+k) \times k$-Matrix, die man Zerlegen kann in eine $n \times k$ Matrix A und ein $k \times k$-Matrix B. Die Matrix B sei invertierbar (also $\det(B) \neq 0$). Dann gibt es eine offene Umgebung $U \subset \mathbb{R}^n$ von v_0, $V \subset \mathbb{R}^k$ von w_0 und eine stetig diffbare Funktion $g\colon U \to V$, so dass

$$f(v, g(v)) = 0 \qquad \forall v \in U,$$

und $f(v, w) = 0$ für $v \in V$, $w \in U$ genau dann wenn $w = g(v)$. Für die totale Ableitung von g an v_0 gilt

$$Dg(w_0) = -B^{-1}A.$$

Beweis. Diese Aussage kann auf folgende Weise auf den Satz über die lokale Existenz einer Umkehrfunktion zurückgeführt werden:

Wir betrachten die Funktion $F\colon \mathbb{R}^{n+k} \to \mathbb{R}^{n+k}; (v, w) \mapsto (v, f(v, w))$. Da F stetig partiell diffbar ist, gilt dies auch für F, und es gilt $DF(v_0, w_0) = \begin{pmatrix} 1_k & A \\ 0 & B \end{pmatrix}$.

Insbesondere gilt $\det(DF(v_0, w_0)) = \det B \neq 0$. Also gibt es eine offene Umgebungen $W_1, W_2 \subset \mathbb{R}^{n+k}$ von (v_0, w_0) und $(v_0, 0)$, so dass $F\colon W_1 \to W_2$ bijektiv mit stetig diffbarer Umkehrfunktion $G\colon W_2 \to W_1$ ist. Da $G(v, f(v,w)) = G(F(v,w)) = (v,w)$ und $(x,y) = F(G(x,y)) = (G_1(x,y), f(G(x,y))$, hat G die Gestalt $G(x,y) = (x, G_2(x,y))$ für alle $(x,y) \in W_2$, und es gilt $y = f(x, G_2(x,y))$ für alle $(x,y) \in W_2$. Setze nun $U := \{v \mid (v,0) \in W_2\} \subset \mathbb{R}^n$. Dies ist eine offene Menge mit $v_0 \in U$. Definiere $g(v) := G_2(v,0)$ für $v \in U$. Damit gilt $f(v, g(v)) = 0$ für alle $v \in U$.

Für $(v,w) \in W_1$ gilt $f(v,w) = 0$ genau wenn $F(v,w) = (v,0)$, also genau wenn $w = G_2(v,0) = g(v)$. Indem man U gegebenenfalls verkleinert, findet man eine offene Umgebung V von w_0 so dass $U \times V \subset W_1$ (da W_1 offen ist) und so dass $g(U) \subset V$ (da g stetig und $g(v_0) = w_0$). Es gilt natürlich weiterhin $f(v, g(v)) = 0$, für $(v,w) \in U \times V$ folgt aus $f(v,w) = 0$ nun auch $w = g(v)$.

Nach der Kettenregel ist g stetig diffbar, und es ergibt sich die beschriebene Ableitung, da g die Komposition von $\mathbb{R}^n \to \mathbb{R}^{n+k}; v \mapsto (v,0)$ mit G ist. □

Für den Beweis des Satzes über die lokale Umkehrfunktion benötigen wir Information über die Ableitung (und Ableitbarkeit) der Umkehrfunktion einer differenzierbaren Funktion. Im 1-dimensionalen Fall ist die Umkehrfunktion diffbar genau dann, wenn die Ableitung nicht verschwindet. Im mehrdimensionalen verallgemeinert man $\neq 0$ zu "invertierbar", braucht aber noch zusätzliche Voraussetzunen, nämlich, dass die Umkehrfunktion stetig ist.

7.3 Satz. *Sei $U \subset \mathbb{R}^n$ offen, $f\colon U \to V$ diffbar und bijektiv, $V \subset \mathbb{R}^n$ offen. Die Umkehrfunktion $g\colon V \to U$ sei stetig an $f(a)$.*

Es sei $a \in U$ und $Df(a) \in Hom_{\mathbb{R}}(\mathbb{R}^n, \mathbb{R}^n)$ sei invertierbar. Dann ist g an $f(a) \in V$ diffbar, mit Ableitung

$$Dg(f(a)) = Df(a)^{-1}.$$

Beweis. Die Formel für die Ableitung folgt direkt aus der Kettenregel, da $g \circ f = \mathrm{id}$. Entscheidend ist, die Differenzierbarkeit zu beweisen. Durch Translationen und Multiplikation mit der Matrix $Df(a)^{-1}$ können wir annehmen, dass $a = 0 = f(a)$ und $Df(a) = 1$. Es gilt wegen der Diffbarkeit von f an 0

$$f(h) = h + \rho(h), \qquad \rho(h)/\left|h\right|_2 \xrightarrow{h \to 0} 0.$$

Also, da $f(g(h)) = h$, (indem man oben statt h $g(\xi)$ einsetzt)

$$g(\xi) = \xi - \rho(g(\xi)).$$

Um die Diffbarkeit von g an 0 nachzuprüfen, müssen wir also nur noch beachten

$$\frac{\rho(g(\xi))}{|\xi|_2} = \frac{\rho(g(\xi))}{|g(\xi)|_2} \frac{|g(\xi)|_2}{|\xi|_2}.$$

Da g bijektiv ist, gilt $g(\xi) = 0$ nur für $\xi = 0$). Da g stetig an Null, gilt $\frac{\rho(g(\xi))}{|g(\xi)|_2} \xrightarrow{\xi \to 0} 0$, da ρ der Fehlerterm der Ableitung von f.

Es bleibt noch zu zeigen, dass es $K > 0$ gibt, so dass für ξ genügend klein gilt $|g(\xi)|_2 \leq K\,|\xi|_2$. Nun gilt $g(\xi) = \xi - \rho(g(\xi))$, somit $|g(\xi)|_2 \leq |\xi|_2 + |\rho(g(\xi))|_2$. Ist ξ genügend klein, ist wegen der Stetigkeit von g an 0 auch $g(\xi)$ klein, und somit wegen der Eigenschaft von ρ $|\rho(g(\xi))| \leq \frac{1}{2}\,|g(\xi)|_2$. Für diese ξ gilt somit $|g(\xi)|_2 \leq 2\,|\xi|_2$. Insgesamt folgt $\frac{\rho(g(\xi))}{|\xi|_2} \xrightarrow{\xi \to 0} 0$. □

Beweis von Satz 7.1. Nach Translation und Multiplikation mit $Df(a)^{-1}$ kann man annehmen, dass $a = 0 = f(a)$ und $Df(0) = 1$.

Um für $y \neq 0$ Lösungen der Gleichung $f(x) = y$ zu finden, und zu zeigen, dass diese eindeutig sind (natürlich nur, solange x und y nahe 0), wollen wir den Banachschen Fixpunktsatz anwenden. Betrachte daher für $y \neq 0$ die Funktion T_y mit

$$T_y(x) := x - f(x) + y.$$

Ihre Fixpunkte x sind genau die Punkte mit $f(x) = y$. Wir müssen jetzt natürlich noch einen geeigneten (kleinen) Definitionsbereich finden, der auf sich selbst abgebildet wird, so dass T_y auf diesem Definitionsbereich kontrahierend ist.

Da f stetig diffbar und $Df(0) = 1$, gibt es $r > 0$, so dass $\|1 - Df(x)\| \leq 1/2$ für alle x mit $|x| < r$.

Da $DT_0(x) = 1 - Df(x)$ folgt für $|x| < r$ somit aus dem Satz von Tayler (nur bis zur ersten Ordnung)

$$|x - f(x)| = |T_0(x) - T_0(0)| = |DT_0(\theta x)x| \leq \|DT_0(\theta x)\| \cdot |x| \leq |x|/2,$$

wobei $\theta \in [0, 1]$. Somit

$$|T_x(y)| = |x - f(x) + y| \leq |x - f(x)| + |y| < 2r,$$

solange $|y| < r$ und $|x| \leq 2r$, also

$$T_x \colon B_0(2r) \to B_0(2r).$$

Der Banachsche Fixpunktsatz kann ja nur auf eine Selbstabbildung angewendet werden. Beachte, dass das Bild im Inneren von $D_{2r}(0)$ liegt. Als nächstes zeigen wir, dass $T_y \colon B_0(2r) \to B_0(2r)$ eine Kontraktion ist. Die Taylorformel (Approximation 1. Grades) liefert nun für $x_1, x_2 \in B_0(2r)$ für ein $\theta \in [0, 1]$

$$\begin{aligned} |T_y(x_2) - T_y(x_1)| &= |DT_y(x_1 + \theta(x_2 - x_1))(x_2 - x_1)| \\ &\leq \|DT_0(x_1 + \theta(x_2 - x_1))\| \cdot |x_2 - x_1| \leq \frac{1}{2} |x_2 - x_1|. \end{aligned}$$

Der Banachsche Fixpunktsatz 5.25 liefert also für jedes y mit $|y| < r$ ein eindeutiges $x \in B_0(2r)$, so dass $T_y(x) = x$, also $f(x) = y$. Genauer genommen gilt $|x| < 2r$, da das Bild von T_y im Inneren von $B_0(2r)$ liegt. Setze $U \subset B_0(2r) := f^{-1}(B_0(r)^\circ)$. Da das Urbild unter der stetigen Abbildung f einer offenen Menge offen ist, ist U offen, und wie gerade dargelebt, ist $f \colon U \to B_0(r)^\circ$ eine Bijektion. Es ist noch zu zeigen, dass die Umkehrabbildung stetig ist, nach Satz 7.3 folgt dann ihre Differenzierbarkeit, und aus der Kettenregel die Formel für die Ableitung.

Nun ist für $r' < r$ $K_{r'} := f^{-1}(B_0(r')) \subset U$ abgeschlossen (da f stetig) und beschränkt, also nach Satz 5.48 kompakt. Also ist $f|_{K_{r'}} \colon K \to B_0(r')$ stetige Bijektion zwischen kompakten Mengen. Nach Satz 5.50 ist die Umkehrung $g|_{B_0(r')} \colon B_0(r') \to K_{r'}$ stetig. Da Stetigkeit eine lokale Eigenschaft ist, folgt die Stetigkeit von g auf $B_0(r')^\circ$. Da $B_0(r)^\circ = \bigcup_{r'<r} B_0(r')^\circ$, folgt die Stetigkeit von g. □

8 Extremwerte unter Nebenbedingungen

In einem Optimierungsfragen in der Praxis sei eine Funktion $f\colon \mathbb{R}^n \to \mathbb{R}$ gegeben.

Sehr oft werden aber nur ganz bestimmte Parameter erlaubt sein, also gar nicht die Extrema von f auf ganz $\mathbb{R}$, sondern nur die Extrema der Einschränkung von f auf eine Teilmenge gesucht sind. Diese Teilmenge wird oft durch eine Gleichung beschrieben sein.

8.1 Beispiel. Was sind die Extrema der Funktion $f(x,y) = x^2 + y^4$ auf dem Kreis $\{(x,y) \mid x^2 + y^2 = 1\}$.

Will man hier mit dem Gradienten arbeiten, so ist folgende Beobachtung anschaulich plausibel: an einer lokalen Extremstelle der eingeschränkten Funktion wird die Richtungsableitung in alle Richtungen, die tangential zur Untermenge liegen, verschwinden, aber nicht notwendigerweise in die anderen Richtungen.

Die Richtungsableitung ist das Skalarprodukt des Gradienten mit der Richtung, der Gradient muss also senkrecht auf allen Richtungen tangential zur Untermenge stehen.

8.2 Satz. *Sei $U \subset \mathbb{R}^n$ offen und $f\colon U \to \mathbb{R}$ stetig diffbar. Sei $g\colon U \to \mathbb{R}^r$ stetig diffbar. Definiere $M := \{x \in U \mid f(x) = 0\}$.*

Sei $a \in M$, und $f|_M$ habe in a ein lokales Extremum (beachte: das heißt nicht, dass $f\colon U \to \mathbb{R}$ in a ein lokales Extremum hat).

Um sinnvoll von Tangenten an M reden zu können, habe $Dg(x) \in \mathrm{Hom}(\mathbb{R}^n, \mathbb{R}^r)$ für jedes $x \in M$ den Rang r. (Tangential zu M am Punkt $a \in M$ sind dann (per Definition) alle Vektoren $v \in \mathbb{R}^n$, die senkrecht auf allen Gradienten $\nabla g_i(a)$, also den Zeilen von $Dg(a)$, stehen.)

Dann gilt, dass $\nabla f(a)$ eine Linearkombination von $\nabla g_1(a), \dots, \nabla g_r(a)$, ist, d.h. es gibt $\lambda_1, \dots, \lambda_r \in \mathbb{R}$, so dass $\nabla f(a) = \sum_{i=1}^r \lambda_i \nabla g_i(a)$.

Beweis. Da $Dg(a)$ Rang r hat, gibt es r linear unabhängige Spalten der Matrix $Dg(a)$. Durch Umnummerieren der Koordinaten können wir annehmen, dass es sich um die letzten r Koordinaten handelt. Dann schreiben wir $Dg(a) = \left(\begin{smallmatrix} A & B \end{smallmatrix}\right)$, mit invertierbarer $r \times r$-Matrix B. Nach dem Satz über implizite Funktionen 7.2 gibt es also eine offene Umgebung U von $a = (a_1, a_2)$ und $V \subset \mathbb{R}^{n-r}$ offen mit einer stetig diffbaren Funktion $\phi\colon V \to U$, der Form $\phi(v) = (v, \psi(v))$, so dass für $x = (v,w) \in U$ gilt: $g(x) = g(v,w) = 0 \iff x = (v,w) = (v, \psi(v)) = \phi(v)$ für $v \in V$.

Es folgt, dass a_1 lokales Extremum von $f \circ \phi\colon U \to \mathbb{R}$ ist. Da nach Kettenregel $f \circ \phi$ stetig diffbar ist, hat $f \circ \phi$ an a_1 verschwindenden Gradienten. Nach Kettenregel und Satz gilt

$$0 = \operatorname{grad} f \circ \phi(a_1) = Df(\phi(a_1)) \cdot D\phi(a_1) = \operatorname{grad} f(a) \begin{pmatrix} 1_{n-r} \\ D\psi(a_1) \end{pmatrix} = \operatorname{grad} f(a) \begin{pmatrix} 1_{n-r} \\ -B^{-1}A \end{pmatrix}$$

$\operatorname{grad} f(a)$ steht also auf allen $(n-r)$ Spalten der Matrix $\begin{pmatrix} 1 \\ -B^{-1}A \end{pmatrix}$ senkrecht.

Außerdem gilt auch

$$\begin{pmatrix} A & B \end{pmatrix} \begin{pmatrix} 1 \\ -B^{-1}A \end{pmatrix} = (A - A) = 0,$$

also steht jede Zeile $\operatorname{grad} g_i(a)$ von $Dg(a) = \begin{pmatrix} A & B \end{pmatrix}$ auf jeder Spalte von $D\phi(a_1)$ senkrecht. Der von $\{\operatorname{grad} g_i(a)\}$ erzeugte Vektorraum ist nach Annahme r-dimensional (da der Rang von $Dg(a)$ gerade r ist), die Spalten von $D\phi(a_1)$ sind linear unabhängig, erzeugen somit einen $n-r$-dimensionalen Unterraum. Damit sind diese beiden Vektorräume gegenseitig orthogonale Komplemente, und insbesondere ist $\operatorname{grad} f(a)$, da senkrecht zu den Spalten von $D\phi(a_1)$, enthalten im Vektorraum, der von den $\operatorname{grad} g_i(a)$ aufgespannt wird, ist somit Linearkombination dieser Vektoren. □

8.3 Beispiel. Für unser Beispiel 8.1 ergibt sich: $\nabla f(x,y) = (2x, +4y^3)$, und $\nabla g(x,y) = (2x, 2y)$. Auf der Menge $M = S^1 = \{(x,y) \mid x^2+y^2=1\}$ gilt also $\operatorname{grad} g(x,y) \neq 0$, so dass wir Satz 8.2 anwenden können.

Korrektur!! Ist also $(x,y) \in S^1$ ein lokales Extremum von $f|_{S^1}$, so ist $\nabla f(x,y)$ linear abhängig von $\nabla g(x,y)$. Man erkennt sofort, dass $\nabla f(x,y)$ nur dann linear abhängig (also ein Vielfaches) von $\nabla g(x,y)$ sein kann, wenn $\nabla f(x,y) = \nabla g(x,y)$, oder wenn $x=0$. Dies bedeutet im ersten Fall, dass $4y^3 = 2y$, also $y^2 = 1/2$ oder $y=0$. Da gleichzeitig $(x,y) \in S^1$, folgt, dass

$$(x,y) \in \{(\pm 1, 0), (\pm\sqrt{2}/2, \pm\sqrt{2}/2)\},$$

wobei alle möglichen Vorzeichenkombinationen zugelassen sind. Für den zweiten Fall $x=0$ folgt $y=\pm 1$ (da $(x,y) \in S^1$), und auch in diesen beiden Fällen ist der Gradient $\nabla g(0,\pm 1)$ linear abhängig von $\nabla f(0,\pm 1)$.

Es folgt natürlich noch nicht, dass alle diese 8 Punkte wirklich lokale Extrema sind, dazu muss die Funktion weiter untersucht werden. Da s^1 abgeschlossen und beschränkt, also kompakt, nimmt f auf S^1 aber Maximum und Minimum an, und wir können folgern, dass dies an den 8 berechneten Punkten geschieht.

Durch Einsetzen erhält man die Extrempunkte und Extremwerte (Übungsaufgabe).

9 Kurven und Kurvenintegral

9.1 Definition. Sei $I \subset \mathbb{R}$ ein Interval (mit mehr als einem Punkt), X ein metrischer Raum. Eine stetige Funktion $f\colon I \to X$ heißt Kurve (in Parameterdarstellung). Oft interessiert einen nur das Bild $f(I)$ (eine Punktmenge in X), in der Regel wird dies aber nicht alles über die Kurve aussagen.

Wir interessieren uns insbesondere für den Spezialfall $X = \mathbb{R}^n$. In diesem Fall sprechen wir von C^k-Kurven, wenn die Abbildung f k-mal stetig differenzierbar ist. Insbesondere sind C^1-Kurven wichtig. Falls f eine C^1-Kurve ist, und $Df(t) \neq 0$ für alle $t \in I$, so heißt die Kurve *regulär.* $Df(t) \in \mathbb{R}^n$ wird der *Tangentialvektor* der Kurve an t genannt (er hängt nicht nur von der Bildmenge, sondern von der ganzen Funktion f ab).

Zur Kurve $f\colon [a,b] \to X$ definiert man die *Umgekehrte Kurve* $f^-\colon [a,b] \to X$ mit $f^-(t) := f(a+b-t)$.

9.2 Beispiel. (1) Strecken $f(t) = v_0 + (w_0 - v_0)t$

(2) Einheitskreis $f(t) = (\cos(t), \sin(t))$.

(3) Schraubenlinie $f(t) = (\cos(t), \sin(t), at)$.

(4) $h(t) := \begin{cases} -t^2; & t < 0 \\ t^2; & t \geq 0 \end{cases}$, $f(t) = (h(t), t^2)$: die Funktion ist C^1, der Graph hat einen "Knick", $Df(0) = 0$, also die Kurve nicht regulär.

9.3 Definition. Eine Kurve Γ, $f\colon [a,b] \to \mathbb{R}^n$ heißt *rektifizierbar*, falls $L(\Gamma) < \infty$, wobei

$$L(\Gamma) = \sup\{\sum_{i=1}^{n} |f(t_i) - f(t_{i-1})|_2 \mid a = t_0 \leq t_1 \leq \cdots \leq t_n = b.\}$$

9.4 Proposition. *Ist $f\colon [a,b] \to \mathbb{R}^n$ eine C^1-Kurve, so ist f rektifizierbar und es gilt*

$$L(f) = \int_a^b |Df(t)|_2 \; dt$$

9.5 Definition. Seien $f\colon [a,b] \to \mathbb{R}^n$ und $g\colon [c,d] \to \mathbb{R}^n$ C^1-Kurven. Sie heißen *äquivalent*, falls es eine *Parametertransformation*, also eine bijektive C^1-Abbildung

$$u\colon [c,d] \to [a,b] \qquad \text{mit } u'(t) \neq 0 \; \forall t \in [c,d]$$

gibt, so dass $g = f \circ u$.

Falls $u' > 0$, so sagen wir: die beiden äquivalenten Kurven haben *dieselbe Orientierung*. Falls $u' < 0$: *umgekehrte Orientierung.*

9.6 Lemma. *Äquivalenz von Kurven ist eine Äquivalenzrelation, d.h. $f \sim f$, $f \sim g \implies g \sim f$, $f \sim g$ und $g \sim h$ impliziert $f \sim h$.*

9.7 Satz. *Sind f, g zwei äquivalente Kurven, so gilt $L(f) = L(g)$*

Beweis. Kettenregel. □

9.8 Satz. *$f\colon [a,b] \to \mathbb{R}$ reguläre C^1-Kurve (also $f'(t) \neq 0$ $\forall t$). Definiere die* Bogenlänge

$$s(t) := \int_a^t |Df(\tau)|_2 \; d\tau.$$

Es gilt $s'(t) = |f'(t)|_2 > 0$, also hat $s\colon [a,b] \to [0, L(f)]$ eine C^1-Umkehrfunktion $t\colon [0, L(f)] \to [a,b]$, $t'(s) = 1/|f'(t(s))|$.

Definiere $g(s) := f(t(s))$. Dann ist g äquivalent zu f mit derselben Orientierung, und $|g'(s)| = |f'(t(s))t'(s)| = 1$.

g nennt man Darstellung von f bezüglich der Bogenlänge als Parameter.

9.9 Definition. Sei Γ C^1-Kurve im $\mathbb{R}^n$, $f\colon \operatorname{im}(\Gamma) \to \mathbb{R}$ stetig. Definiere

$$\Gamma - \int f(x) \; ds = \int_a^b f(\Gamma(t)) \, |\Gamma'(t)|_2 \; dt,$$

das Kurvenintegral von f über Γ bezüglich Bogenlänge.

(Formal $ds = |\Gamma'(t)| \; dt$.)

9.10 Satz. *(1) Kurvenintegral ist linear in f.*

(2) Kurvenintegral ist additiv unter Verkettung von Wegen.

(3) *Kurvenintegral ist unabhängig von der Parametrisierung des Weges.*

(4) *Abschätzung* $\left|\Gamma - \int f\right| \leq \sup f \cdot L(\Gamma)$.

9.11 Definition. Sei Γ C^1-Kurve in $\mathbb{R}^n$ und $f\colon \operatorname{im}(\Gamma) \to \mathbb{R}^n$ stetig. Definiere das *Kurvenintegral der Vektorfunktion f über Γ* durch

$$\Gamma - \int f(x)\, dx := \int_a^b \langle f(\Gamma(t)), D\Gamma(t)\rangle\, dt.$$

Formal $f(x)dx$ als Skalarprodukt $f_1dx_1 + \dots f_ndx_n$, und mit $x_i = \Gamma_i(t)$, $dx_i = \Gamma_i'(t)\, dt$.

9.12 Satz. (1) *Vektor-Kurvenintegral ist linear in f.*

(2) *Vektor-Kurvenintegral ist additiv unter Verkettung von Wegen.*

(3) *Vektor-Kurvenintegral ändert Vorzeichen bei Orientierungsumkehr*

(4) *Vektor-Kurvenintegral ist unabhängig von Parametrisierung derselben Orientierung.*

(5) *Abschätzung* $\left|\Gamma - \int f\right| \leq \sup |f|_2 \cdot L(\Gamma)$.

9.13 Definition. $f\colon [a,b] \to \mathbb{R}^n$ heißt *stückweise C^1-Kurve* falls f stetig, und $a = t_0 < \cdot < t_n = b$ existiert, so dass $f|_{[t_{i-1},t_i]}$ C^1-Kurve für jedes dieser Teilintervalle (die Ableitungen müssen an den t_i nicht zusammenpassen).

Alles setzt sich offensichtlich auf stückweise diffbare Kurven fort, indem man ggf. Summen über die C^1-Teilstücke bildet.

9.14 Definition. $U \subset \mathbb{R}^n$ offen, $f\colon U \to \mathbb{R}^n$ stetig.

Kurvenintegrale von f heißen *vom Weg unabhängig*, wenn ihr Wert nur von den Endpunkten des Weges abhängt, also $\Gamma - int f = \Gamma' - \int f$ für zwei beliebige Wege $\Gamma\colon [a,b] \to U$ und $\Gamma'\colon [c,d] \to U$ mit $\Gamma(a) = \Gamma'(c)$, $\Gamma(b) = \Gamma'(d)$.

Eine Funktion $F\colon U \to \mathbb{R}$ heißt *Stammfunktion* von f in U, falls F einmal stetig diffbar mit $\operatorname{grad} F = f$. $U := -F$ heißt *Potential* von f.

9.15 Satz. *Sei $U \subset \mathbb{R}^n$ offen und zusammenhängend, $f\colon U \to \mathbb{R}^n$ stetig. Es gilt*

(1) *f hat Stammfunktion genau dann, wenn Kurvenintegrale über f vom Weg unabhängig sind.*

(2) *Stammfunktionen sind eindeutig bis auf Addition einer Konstanten*

(3) *Eine Stammfunktion ist (für festes $a \in Y$) gegeben durch*

$$F(y) := \int_a^y f(x)\, dx,$$

umgekehrt gilt

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

für jede Stammfunktion.

Hier schreiben wir $\int_a^b f\, dx = \Gamma - \int f$, für jeden Weg Γ in U von a nach b, falls das Kurvenintegral vom Weg unabhängig ist.

10 Zusammenfassung zum (Lebesgue)-Integral

10.1 Motivation für Integrale

Differentialrechnung: Wichtig immer wenn veränderliche Prozesse beschrieben werden sollen (und in weiteren Zusammenhängen).
Integralrechnung: Wichtig immer, wenn Mittelwerte gebildet werden sollen (und in weiteren Zusammenhängen). Dies ist auch wichtig für Funktionen, die auf (Teilmengen des) $\mathbb{R}^n$ definiert sind.

Wenn man an die Stochastik und ihre Anwendungen denkt, wird klar, dass man Mittelwerte sogar in noch viel allgemeinerem Zusammenhang definieren will, für Funktionen, die auf ganz anderen Definitionsbereichen definiert sind.

Weitere Anwendungen:

- Bildung von Mittelwerten (s.o.)
- Lösung von Differentialgleichungen
- Berechnung von Volumina

Wir wollen hier zunächst einen "axiomatischen" Ansatz gehen, und eine "Wunschliste" von Eigenschaften des Intergrals aufstellen, die uns erlauben, es in vielen Fällen zu berechnen und damit umzugehen. Die grundlegende Konstruktion wird dann später nachgereicht.

10.2 Grundlegende Eigenschaften des Lebesgue-Integrals

10.1 Definition. Sei M ein metrischer Raum und $f\colon M \to \mathbb{C}$ eine Funktion. Der *Träger* von f ist der Abschluß der Menge $\{x \in M \mid f(x) \neq 0\}$. Beachte: per Definition ist der Träger abgeschlossen.

Diese Definition wird auf beliebige (nicht nur stetige) Funktionen angewandt.

10.2 Theorem. *Es gibt eine Vektorraum $\mathcal{L}^1(\mathbb{R}^n)$ von Funktionen $f\colon \mathbb{R}^n \to \mathbb{C}$, der* Raum der Lebesgue-integrierbaren Funktionen *und eine lineare Abbildung*

$$\int : \mathcal{L}^1(\mathbb{R}^n) \to \mathbb{C},$$

so dass unter anderem folgende Eigenschaften erfüllt sind:

(1) Jede stetige Funktion mit kompaktem Träger gehört zu $\mathcal{L}^1(\mathbb{R}^n)$.

(2) Falls $n = 1$ und f eine Regelfunktionen mit kompaktem Träger (enthalten im Interval $[a, b]$), so gilt $f \in \mathcal{L}^1(\mathbb{R})$, und

$$\int f = \int_a^b f.$$

(3) Monotonie*: Falls $f, g \in \mathcal{L}^1(\mathbb{R}^n)$ mit $f(x) \leq g(x)\ \forall x$, so gilt*

$$\int f \leq \int g.$$

(4) $f \in \mathcal{L}^1(\mathbb{R}^n)$ *genau dann wenn* $|f| \in \mathcal{L}^1(\mathbb{R}^n)$*, und es gilt in diesem Fall)*

$$\left|\int f\right| \le \int |f|$$

(5) Monotone Konvergenz: Seien $f_0, f_1, f_2, \dots \in \mathcal{L}^1(\mathbb{R}^n)$*,* $f_k\colon \mathbb{R}^n \to [0,\infty)$ *und*

$$f_1(x) \le f_2(x) \le \cdots \qquad \forall x \in \mathbb{R}^n.$$

Definiere

$$f(x) := \begin{cases} \lim_{k\to\infty} f_k(x); & \textit{Limes existiert} \\ 0; & \textit{sonst.} \end{cases}$$

Es gebe $C > 0$ *so dass* $\int_{\mathbb{R}^n} f_k < C \ \forall k \in \mathbb{N}$*. Dann gilt:*

$$f \in \mathcal{L}^1(\mathbb{R}^n) \qquad \textit{und} \qquad \int_{\mathbb{R}^n} f \le C.$$

10.3 Satz. ***(Lemma von Fatou)***
Seien $f_0, f_1, \dots \in \mathcal{L}^1(\mathbb{R}^n)$*,* $f_k\colon \mathbb{R}^n \to [0,\infty)$*. Es gelte*

$$\liminf_{k\to\infty} \int_{\mathbb{R}^n} f_k < \infty.$$

Dann gilt:

$$\int_{\mathbb{R}^n} \liminf_{k\to\infty} f_k \le \liminf_{k\to\infty} \int_{\mathbb{R}^n} f_k.$$

Dabei gehört die Funktion auf der linken Seite zu $\mathcal{L}^1(\mathbb{R}^n)$*, wobei wir die Konvention benutzten, dass der Wert* 0 *gesetzt wird, wenn der* $\liminf_{k\to\infty} f_k(x)$ *nicht existiert (also die Folge reeller Zahlen gegen* $+\infty$ *divergiert).*

10.3 Einschränken auf Untermengen

10.4 Definition. Sei $X \subset \mathbb{R}^n$ eine Menge. Definiere die *charakteristische Funktion*

$$\chi_X\colon \mathbb{R}^n \to \mathbb{R};\ \chi_X(x) = \begin{cases} 1; & x \in X \\ 0; & x \notin X. \end{cases}$$

10.5 Definition. Sei $U \subset \mathbb{R}^n$. Wir nennen U *messbar mit endlichem Maß*, falls die charakteristische Funktion $\chi_U \in \mathcal{L}^1(\mathbb{R}^n)$. In diesem Fall definieren wir

$$\mathrm{vol}(U) := \int_U \chi_U.$$

Wir nennen U *messbar*, wenn es abzählbar viele messbare Teilmengen mit endlichem Maß $U_0, U_1, U_2, \dots$ gibt, so dass U gleich der Vereinigung ist:

$$U = \bigcup_{i\in\mathbb{N}} U_i.$$

Eine messbare Menge V mit $\mathrm{vol}(V) = 0$ wird *Nullmenge* genannt.

10.6 Bemerkung. Integration und Volumenberechung hängen sehr eng zusammen. Tatsächlich wird bei der Definition des Integrals zunächst das Volumen (geeigneter) Mengen definiert, und dann das Integral über eine Treppenfunktion als das Volumen der entsprechenden Menge. Daraus werden dann die anderen Integrale hergeleitet.

Diesen Ansatz werden wir bei der Konstruktion des Integrals im nächsten Semester gehen.

10.7 Definition. Sei für jedes $x \in \mathbb{R}^n$ A_x eine Aussage (z.B. die Aussage: $(f_n(x))$ konvergiert, wenn $f_n \colon \mathbb{R}^n \to \mathbb{C}$ eine Folge von Funktionen bildet).

Wir sagen, dass die Aussage A_x *für fast alle* $x \in \mathbb{R}^n$ *gilt*, wenn es eine Nullmenge $U \subset \mathbb{R}^n$ gibt, so dass A_x für alle $x \in \mathbb{R}^n \setminus U$ gilt.

10.8 Beispiel. Falls $U \subset \mathbb{R}$ eindimensional, spricht man statt von Volumen von "*Länge*". Insbesondere, wenn $U = [a, b]$ ein Intervall (offen oder abgeschlossen). Es gilt dann $\mathrm{vol}([a, b]) = b - a$.

Im zweidimensionalen Fall ist das von uns definierte Volumen eine *Fläche.*

Allgemein: $\mathrm{vol}([a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_n, b_n]) = (b_1 - a_1) \cdot (b_n - a_n)$.

Insbesondere ist jede einpunktige Menge eine Nullmenge, wegen der sigma-Additivität also jede abzählbare Menge (beachte: auch $\mathbb{Q} \subset \mathbb{R}$ ist ein Nullmenge, obwohl der Abschluss von $\mathbb{Q}$ ganz $\mathbb{R}$ ist).

10.9 Satz. *Sei $f \in \mathcal{L}^1(\mathbb{R}^n)$ und $U \subset \mathbb{R}^n$ messbar. Dann ist auch $f\chi_U \in \mathcal{L}^1(\mathbb{R}^n)$. Wir schreiben*

$$\int_U f := \int_{\mathbb{R}^n} f\chi_U.$$

Dies erleichtert natürlich ein wenig die Schreibarbeit, da man das "Einschränken" auf U jetzt sehr schön formulieren kann.

Beweis. Dies ist eine weitere der Eigenschaften des Lebesgue-Integrals, die wir dieses Semester nicht beweisen wollen. □

10.10 Definition. Sei $U \subset \mathbb{R}^n$ eine messbare Teilmenge. Wir setzen

$$\mathcal{L}^1(U) := \{f \colon U \to \mathbb{C} \mid \overline{f} \in \mathcal{L}^1(\mathbb{R}^n)\}$$

Hierbei sei $\overline{f} \colon \mathbb{R}^n \to \mathbb{C}$ die durch Null fortgesetzte Funktion, also

$$\overline{f}(x) = \begin{cases} f(x); & x \in U \\ 0; & x \notin U. \end{cases}$$

10.11 Lemma. *Wegen der Additivität des Integrals folgt für $f \in \mathcal{L}^1(\mathbb{R}^n)$ und zwei disjunkte messbare Teilmengen $U, V \subset \mathbb{R}^n$*

$$\int_U f + \int_V f = \int_{U \cup V} f.$$

10.12 Satz. ***(von Fubini)*** *Sei $f \colon \mathbb{R}^n \to \mathbb{C}$ eine Funktion. Schreibe $\mathbb{R}^n = \mathbb{R}^k \times \mathbb{R}^{n-k}$ mit Elementen $(x, y) \in \mathbb{R}^k \times \mathbb{R}^{n-k}$. Definiere $F_x(y) := f(x, y)$ und $G_y(x) := f(x, y)$ für $x \in \mathbb{R}^k$, $y \in \mathbb{R}^{n-k}$, also $F_x \colon \mathbb{R}^{n-k} \to \mathbb{C}$ und $G_y \colon \mathbb{R}^k \to \mathbb{C}$.*

Betrachte folgende drei Aussagen:

(1) $F_x \in \mathcal{L}^1(\mathbb{R}^{n-k})$ *für fast alle* x*, und*

$$u_a(x) := \begin{cases} \int_{\mathbb{R}^{n-k}} |F_x(y)| \; dy; & F_x \in \mathcal{L}^1(\mathbb{R}^{n-k}) \\ 0; & sonst \end{cases}$$

erfüllt $u \in \mathcal{L}^1(\mathbb{R}^k)$*.*

(2) $G_y \in \mathcal{L}^1(\mathbb{R}^k)$ *für fast alle* x*, und*

$$v_a(y) := \begin{cases} \int_{\mathbb{R}^{n-k}} |G_y(x)| \; dx; & G_y \in \mathcal{L}^1(\mathbb{R}^k) \\ 0; & sonst \end{cases}$$

erfüllt $v \in \mathcal{L}^1(\mathbb{R}^{n-k})$*.*

(3) $f \in \mathcal{L}^1(\mathbb{R}^n)$*.*

Es gilt: diese drei Aussagen sind äquivalent. Falls sie wahr sind, gilt

$$\int_{\mathbb{R}^n} f = \int_{\mathbb{R}}^k u = \int_{\mathbb{R}^{n-k}} v,$$

wobei

$$u_a(x) := \begin{cases} \int_{\mathbb{R}^{n-k}} F_x(y) \; dy; & F_x \in \mathcal{L}^1(\mathbb{R}^{n-k}) \\ 0; & sonst \end{cases}$$

und

$$v(y) := \begin{cases} \int_{\mathbb{R}^{n-k}} G_y(x) \; dx; & G_y \in \mathcal{L}^1(\mathbb{R}^k) \\ 0; & sonst \end{cases}$$

Für die iterierten Integrale schreibt man in der Regel

$$\int_{\mathbb{R}^k} u = \int_{\mathbb{R}^k} \left(\int_{\mathbb{R}^{n-k}} f(x,y) \; dy \right) \; dx; \quad \int_{\mathbb{R}^{n-k}} v = \int_{\mathbb{R}^{n-k}} \left(\int_{\mathbb{R}^k} f(x,y) \; dx \right) \; dy; \quad . \tag{10.13}$$

10.14 Bemerkung. In in der Formel (10.13), und auch in anderen Zusammenhängen, hängt man manchmal (insbesondere wenn es mehrere Variablen gibt) das Symbol "dx" hinter das Integral, um deutlich zu machen, über welche Variablen integriert wird.

10.15 Definition. Seien $U, V \subset \mathbb{R}^n$ offen und $\phi\colon U \to V$ eine stetig partiell diffbare Bijektion mit stetig partiell diffbarer Umkehrung. Dann heißt ϕ C^1-Diffeomorphismus zwischen U und V.

10.16 Satz. ***Transformationsformel*** *Seien* $U, V \subset \mathbb{R}^n$ *offen und* $\phi\colon U \to V$ *ein* C^1*-Diffeomorphismus. Dann gilt*

$$f \in \mathcal{L}^1(V) \iff f \circ \phi \cdot |\det(D\phi)| \in \mathcal{L}^1(U),$$

und wenn dies der Fall ist, so ist

$$\int_V f = \int_U f(\phi(x)) \, |\det(D\phi(x))| \; dx. \tag{10.17}$$

10.18 Beispiel. Affine Koordinatentransformationen: Sei $A \in Gl(n, \mathbb{R})$ und $v_0 \in \mathbb{R}$.

Dann ist $\phi \colon \mathbb{R}^n \to \mathbb{R}^n; \phi(v) = v_0 + Av$ eine affin lineare Transformation, insbesondere ein C^1-Diffeomorphismus. Es gilt $D\phi(v) = A$ für jedes $v \in \mathbb{R}^n$, also $\det(D\phi(v)) = \det A$. Somit erhalten wir

$$\int_{\mathbb{R}}^{n} (f(x))\, dx = \int_{\mathbb{R}^n} f(v_0 + Ax) \left|\det(A)\right|\, dx. \tag{10.19}$$

Sei als Beispiel $f = \chi_U$ die charakteristische Funktion einer messbaren Menge U mit endlichem Maß. Dann ist die Funktion $x \mapsto f(v_0 + Ax) = \chi(v_0 + Ax)$ die charakteristische Funktion von $A^{-1}(U - v_0)$, also der um v_0 verschobenen und dann mit A^{-1} "verdrehstreckten" Menge U.

Bei der "Verdrehstreckung" ändert sich das Volumen (insbesondere ist das für höherdimensionale Parallelogramme klar), und zwar wird es mit $\left|\det(A^{-1})\right|$ multipliziert.

Damit die Formel (10.19) dieser Effekt wieder ausgeglichen wird, muss man den Faktor mit $\det(A)$ wieder wettmachen.

Hier liegt auch der Grund für die Transformationsformel: jede C^1-Abbildung ϕ kann lokal durch eine affine Abbildung approximiert werden (konstanter und linearer Term in der Taylorentwicklung), das Volumen wird also in der Nähe eines Punktes ungefähr so verzerrt wie durch die entsprechende affine Transformation, also mit $\left|\det(D\phi(x))^{-1}\right|$. Dies wird durch den Faktor in der Transformationsformel (10.17) wieder wettgemacht.

Anschaulich: das Volumenelement in neuen Koordinaten muss mit einem geeigneten Faktor verändert werden, damit sein Volumen dem alten Volumen (in den alten Koordinaten) entspricht.

10.20 Beispiel. Polarkoordinaten entsprechen der Transformation

$$\phi \colon (0, \infty) \times (0, 2\pi) \to \operatorname{im}(\phi) \subset \mathbb{R}; \phi(r, \theta) = (r\cos(\theta), r\sin(\theta)).$$

Hierfür gilt

$$D\phi(r, \theta) = \begin{pmatrix} \cos(\theta) & -r\sin(\theta) \\ \sin(\theta) & r\cos(\theta) \end{pmatrix}; \qquad \det(D\phi(r, \theta)) = r.$$

Also

$$\int_{\mathbb{R}^2} f(x, y)\, dxdy = \int_{(0,\infty)\times(0,2\pi)} f(\phi(r, \theta)) r\, drd\theta.$$

Schaut man sich insbesondere $f(x, y) = \chi_S$ an, wobei S ein Sektor mit zwischen den Winkeln 0 und α in der Kreisscheibe vom Radius R ist, so erhält man

$$\operatorname{vol}(S) = \int_S 1\, dxdy = \int_{\mathbb{R}^2} \chi_S(x, y)\, dxdy = \int_{\mathbb{R}^2} \chi_S(\phi(r, \theta)) \cdot r\, drd\theta = \int_0^R \int_0^\alpha r\, d\theta dr = \alpha R^2/2.$$

Hierbei wird der Winkel natürlich im Bogenmaß berechnet, und wir erhalten genau die klassische Fläche dieses Sektors.

10.4 Messbare Menge und Nullmengen

10.21 Bemerkung. Nullmengen sind deshalb wichtig, weil man sie beim integrieren völlig vernachlässigen kann (ihr Volumen ist ja Null), so dass man bei den Voraussetzungen für Sätze oft auf einer Nullmenge keine Kenntnisse haben muss (z.B. über Konvergenz), und die Sätze gelten trotzdem.

Das besondere am Lebesgue-Integral sind die zwei folgenden Eigenschaften:

(1) Ist U die Vereinigung von abzählbar vielen Nullmengen, so ist U selbst eine Nullmenge.

(2) Ist U eine Nullmenge, und $V \subset U$ eine beliebige Teilmenge, so ist auch V eine Nullmenge.

Die erste Eigenschaft wird *σ-Additivität* genannt, die zweite *Vollständigkeit.*

10.22 Lemma. *Sei $U \subset \mathbb{R}^n$ eine Nullmenge und $f\colon U \to \mathbb{C}$ eine beliebige Funktion. Dann gilt: $f \in \mathcal{L}^1(U)$ und*

$$\int_U f = 0.$$

Beweis. Dies ist eine der charakteristischen Eigenschaften des Lebesgue-Integrals, die mit der Vollständigkeit zusammen hängt, und hier nicht bewiesen wird. □

10.23 Lemma. *Sei $f \in \mathcal{L}^1(\mathbb{R}^n)$ und $U \subset \mathbb{R}^n$ eine Nullmenge. Sei $g\colon \mathbb{R}^n \to \mathbb{C}$ eine Funktion, so dass*

$$f(x) = g(x) \qquad \forall x \notin U.$$

Dann gilt: $g \in \mathcal{L}^1(\mathbb{R}^n)$ und

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

Beweis. Die Menge $U^c = \mathbb{R}^n \setminus U$ ist messbar, also gilt $f\chi_{U^c} \in \mathcal{L}^1(\mathbb{R}^n)$. Nach Lemma 10.22 gilt $g\chi_U \in \mathcal{L}^1(\mathbb{R}^n)$, also auch

$$g = f\chi_{U^c} + g\chi_U.$$

Ausserdem

$$\begin{aligned}\int_{\mathbb{R}^n} g = \int_{\mathbb{R}^n} f\chi_{U^c} + \int_{\mathbb{R}^n} g\chi_U &= \int_{\mathbb{R}^n\setminus U} f + \int_U g \\ = \int_{\mathbb{R}^n\setminus U} f + 0 = \int_{\mathbb{R}^n\setminus U} f + \int_U f &= \int_{\mathbb{R}^n} f\end{aligned}$$

□

10.24 Bemerkung. Wegen Lemma 10.23 kann man, ohne das Integral und die Integrierbarkeit zu ändern, eine gegebene Funktion auf einer Nullmenge beliebig abändern. Wir haben dies bisher schon genutzt, wenn wir Funktionen an manchen Stellen willkürlich auf Null gesetzt haben (z.B. wo kein Limes existierte). Statt dies zu tun, geht man manchmal den pragmatischen Weg, Funktionen zuzulassen, die auf einer Nullmenge gar nicht definiert sind. Diese kann man trotzdem integrieren: man kann sie auf der Nullmenge mit beliebige Werten fortsetzen (z.B. durch Null), und danach das Integral bilden (falls es existiert), es ist dann unabhängig von der gewählten Fortsetzung.

10.25 Lemma. *Die Kollektion der messbaren Teilmenge des $\mathbb{R}^n$ hat einige interessante Eigenschaften, die an die Eigenschaften einer Topologie erinnern (ohne damit überein zu stimmen), nämlich*

(1) $\emptyset$ ist messbar.

(2) Ist $U \subset \mathbb{R}^n$ messbar, so auch $\mathbb{R}^n \setminus U$.

(3) Jede abzählbare Vereinigung und jeder abzählbare Schnitt messbarer Mengen ist messbar.

Beweis. $\chi_\emptyset = 0$, gehört also zu $\mathcal{L}^1(\mathbb{R}^n)$. Jede messbare Menge ist nach Voraussetzung abzählbare Vereinigung von messbaren Mengen von endlichem Maß, eine abzählbare Vereinigung messbarer Mengen ist also eine abzählbare Vereinigung von abzählbaren Vereinigungen. Zusammengenommen ist dies immer noch abzählbar, also ist die resultierende Menge messbar.

Komplement: Übungsaufgabe. Für jedes $R > 0$ ist $[-R, R]^n \subset \mathbb{R}^n$ messbar mit endlichem Ma"s. Sei χ_R die charakteristische Funktion dieses n-dimensionalen Würfels. Immer wenn U messbar ist, so gilt $\chi_U \cdot \chi_R \in \mathcal{L}^1(\mathbb{R}^n)$. Dies ist die charakteristische Funktion von $U \cap [-R, R]^n$, welches also messbar mit endlichem Maß ist. Damit ist auch $\chi_R - \chi_R\chi_U \in \mathcal{L}^1(\mathbb{R}^n)$, es handelt sich nun um die charakteristische Funktion von $[-R, R]^n \setminus U$, welches somit messbar von endlichem Maß. Schließlich gilt

$$\mathbb{R}^n \setminus U = \bigcup_{k\in\mathbb{N}} [-, k]^n \setminus U,$$

also ist das Komplement messbar.

Zum Schluss gilt $\bigcap_{k\in\mathbb{N}} U_k = \mathbb{R}^n \setminus \bigcup_{k\in\natural}(\mathbb{R}^n \setminus U_k)$, so dass die Aussage über Schnittmengen aus derjenigen über Vereinigungen und über Komplemente folgt. □

10.26 Definition. Sei Ω eine (beliebige) Menge, $S \subset \mathcal{P}(\Omega)$ eine Kollektion von Teilmengen von Ω. S heißt *sigma-Algebra* oder *σ-Algebra*, falls

(1) $\emptyset \in S$

(2) Falls $U \in S$, dann auch $\Omega \setminus U \in S$.

(3) S ist abgeschlossen unter abzählbaren Vereinigungen: falls $U_0, U_1, \dots \in S$, dann auch $U = \bigcup_{k\in\mathbb{N}} U_k$.

Eine Funktion $\mu\colon S \to [0, \infty]$ heißt *Maß* auf der σ-Algebra S, falls für eine disjunkte Vereinigung

$$U = U_0 \cup U_1 \cup U_2 \cup \dots \qquad U_i \in S$$

gilt

$$\mu(U) = \sum_{k=0}^{\infty} \mu(U_i).$$

10.27 Bemerkung. Die oben definiere Menge der messbaren Mengen auf $\mathbb{R}^n$ ist eine σ-Algebra. Definiert man $\mathrm{vol}(U) = +\infty$ für eine messbare Menge, welche nicht von endlichem Maß ist, so definiert vol ein Maß auf dieser σ-Algebra.

10.28 Lemma. *Sei $U \subset \mathbb{R}^n$ offen. Dann ist U messbar. Im Beweis wird benutzt, dass jeder Würfel mit endlicher Kantenlänge messbar ist.*

Beweis. Wir müssen U als abzählbare Vereinigung von messbaren Mengen darstellen. Wir wissen, dass für jedes $r > 0$ und $v \in \mathbb{R}^n$ der Würfel $W_{v,r} := \{v + x \mid x \in [-r, r]^n\}$ messbar ist. Wir werden zeigen, dass U Vereinigung abzählbar vieler solcher Würfel ist. Dazu wollen wir zunächst abzählbar viele Zentren auswählen: die Punkte $v \in U$, so dass alle Koordinaten rational sind (also $U \cap \mathbb{Q}^n$). Für jedes $v \in U \cap \mathbb{Q}^n$ betrachte nun alle Würfel $W_{u,r}$, so dass $r \in \mathbb{Q}$ und $W_{u,r} \subset U$. Es handelt sich um insgesamt abzählbar viele Würfel, welche alle in U enthalten sind.

Wir müssen nur noch zeigen, dass jedes $x \in U$ in mindestens einem dieser Würfel enthalten ist, und haben dann U als Vereinigung abzählbar vieler messbarer Mengen mit endlichem Maß identifiziert.

Da U offen, gibt es zu $x \in U$ ein $\epsilon > 0$, so dass der Würfel $W_{x,\epsilon} \subset U$. Wähle $0 < q < \epsilon/2$ mit $q \in \mathbb{Q}$. Wähle in $W_{x,q}$ einen Punkt v mit rationalen Koordinaten. Dann gilt $x \in W_{v,q}$, wegen der Dreiecksungleichung aber auch $W_{v,q} \subset W_{x,2q} \subset U$. □

Es gibt auch eine verbesserte Version des Satzes von Lebesgue über dominierte Konvergenz:

10.29 Satz. ***Satz von Lebesgue über dominierte Konvergenz***
Seien $g, f_k \in \mathcal{L}^1(\mathbb{R}^n)$. Für jedes $k \in \mathbb{N}$ gelte

$$|f_k(x)| \leq g(x) \qquad \textit{für fast alle } x \in \mathbb{R}^n.$$

Sei $f\colon \mathbb{R}^n \to \mathbb{C}$ eine Funktion, und es gelte

$$f(x) = \lim_{k\to\infty} f_k(x) \qquad \textit{für fast alle } x \in \mathbb{R}^n.$$

Dann gilt:

$$f \in \mathcal{L}^1(\mathbb{R}^n) \qquad \textit{mit} \qquad \int_{\mathbb{R}^n} f = \lim_{k\to\infty} \int_{\mathbb{R}^n} f_k$$

Beweis. Wir wollen beweisen, dass die schwächeren Voraussetzungen ausreichen.

Sei N_k die Nullmenge, auf der nicht gilt: $|f_k(x)| \leq g(x)$. Sei M die Nullmenge, auf der nicht gilt: $f(x) = \lim_{k\to\infty f_k(x)}$. Setze $N := M \cup \bigcup_{k\in\mathbb{N}} N_k$. Dies ist abzählbare Vereinigung von Nullmengen, also selbst eine Nullmenge. Ersetze f_k und f durch die Funktionen h_k und h, die auf M den Wert Null haben, und außerhalb die Werte $f_k(x)$ oder $f(x)$. Dann sind auch alle $h_k \in \mathcal{L}^1(\mathbb{R}^n)$, und das Integral hat sich nicht geändert. Für die Funktionen h_k und h gilt aber: $h(x) = \lim_{k\to\infty} h_k(x)$ und $|h_k(x)| \leq g(x)$ für *alle* $x \in \mathbb{R}^n$. Nach der schon bewiesenen Version des Satzes über dominierte Konvergenz gilt also $h \in \mathcal{L}^1(\mathbb{R}^n)$ und $\int_{\mathbb{R}^n} h = \lim_{k\to\infty} \int_{\mathbb{R}^n} h_k$. Da h und f sich nur auf einer Nullmenge unterscheiden, ist dann auch $f \in \mathcal{L}^1(\mathbb{R}^n)$ und

$$\int_{\mathbb{R}^n} f = \lim_{k\to\infty} \int_{\mathbb{R}^n} f_k$$

(da $\int f_k = \int h_k$). □

11 Parameterabhängige Integrale

11.1 Satz. *Sei $D \subset \mathbb{R}^p$ offen und $f\colon \mathbb{R}^n \times D \to \mathbb{C}$ eine Funktion, so dass für jedes $t \in D$ die Funktion $f_t(x) := f(x,t)$ zu $\mathcal{L}^1(\mathbb{R}^n)$ gehört, außerdem die Funktion $G_x(t) := f(x,t)$ für fast alle $x \in \mathbb{R}^n$ eine stetige Funktion $G_x\colon D \to \mathbb{C}$.*

Es gebe eine Majorante $g \in \mathcal{L}^1(\mathbb{R}^n)$, d.h. es gelte

$$|f(x,t)| \leq g(x) \qquad \textit{für alle } (x,y) \in \mathbb{R}^n \times D.$$

Dann ist die Funktion $F\colon D \to \mathbb{C}$ mit

$$F(t) = \int_{\mathbb{R}^n} f(x,t)\, dx$$

eine stetige Funktion auf D.

Beweis. Sei $t_k \in D$ mit $t_k \xrightarrow{k\to\infty} t \in D$. Wir müssen zeigen, dass $\lim_{k\to\infty} F(t_k) = F(t)$. Nach der Stetigkeitsvoraussetzung konvergiert $f_{t_k}(x)$ punktweise fast überall gegen $f_t(x)$. Außerdem ist g eine Majorante für die Funktionenfolge $(f_{t_k})_{k\in\mathbb{N}}$. Nach dem Satz von Lebesgue über dominierte Konvergenz gilt daher

$$\lim_{k\to\infty} F(t_k) = \lim_{k\to\infty} \int_{\mathbb{R}^n} f_{t_k}(x) = \int_{\mathbb{R}^n} f_t(x) = F(t).$$

Da dies für jede gegen t konvergierende Folge t_k gilt, ist F stetig an $t \in D$. □

11.2 Beispiel. Die Funktion $f_0(x) := f(x) = 1/\sqrt{x}\colon (0,1] \to \mathbb{R}$ ist monotoner Limes der Funktionen

$$f_k(x) := \begin{cases} 1/\sqrt{x}; & 1 \geq x > 1/k \\ 0; & 0 \leq x \leq 1/k. \end{cases}$$

Die Funktionen f_k sind R-integrierbar, also in $\mathcal{L}^1((0,1])$, und es gilt $\int_0^1 f_k \leq [2\sqrt{(}x)]_0^1 = 2$. Nach dem Satz über monotone Konvergenz gilt also $f \in \mathcal{L}^1((0,1])$. Für $t \neq 0$ ist die Funktion $f_t(x) = 1/(t+\sqrt{x})$, $0 < x \leq 1$ R-integrierbar, also $f_t \in \mathcal{L}^1(0,1])$ für $t \neq 0$. Außerdem gilt $f_t(x) \leq f(x)$, somit ist f Majorante für die f_t. Somit ist die Funktion

$$g\colon \mathbb{R} \to \mathbb{R}; g(t) = \int_{(0,1]} f_t(x)\, dx$$

an jedem Punkt (insbesondere auch an Null) stetig.

11.3 Satz. *Sei $D \subset \mathbb{R}$ offen und $f\colon \mathbb{R}^n \times D \to \mathbb{C}$ eine Funktion, so dass für jedes $t \in D$ die Funktion $f_t(x) := f(x,t)$ zu $\mathcal{L}^1(\mathbb{R}^n)$ gehört, außerdem die Funktion $G_x(t) := f(x,t)$ für fast alle $x \in \mathbb{R}^n$ eine differenzierbare Funktion $G_x\colon D \to \mathbb{C}$.*

Es gebe eine Majorante $g \in \mathcal{L}^1(\mathbb{R}^n)$ der Ableitung, d.h. es gelte

$$|\partial_t f(x,t)| \leq g(x)$$

für alle $t \in D$ und für fast alle $x \in \mathbb{R}^n$.

Dann ist die Funktion $F\colon D \to \mathbb{C}$ mit

$$F(t) = \int_{\mathbb{R}^n} f(x,t)\, dx$$

eine differenzierbare Funktion auf D.

Außerdem ist für jedes $t \in D$ die Funktion $x \mapsto \partial_t f(x,t)$ aus $\mathcal{L}^1(\mathbb{R}^n)$, und es gilt

$$\partial_t F(t) = \int_{\mathbb{R}^n} \partial_t f(x,t)\, dx$$

Beweis. Vergleiche die ähnliche Übungsaufgabe (mit stärkeren Voraussetzungen)! Wir geben hier einen vollständigen Beweis mit Hilfe des Satzes von Lebesgue über dominierte Konvergenz.

Integrierbarkeit der Ableitung $\partial_t f(x,t)$:

Für vorgegebenes t und Nullfolge $(t_k)_{k\in\mathbb{N}}$ reeller Zahlen (ungleich Null) betrachte die Funktionen

$$h_{t,k}(x) := \frac{f(x,t+t_k) - f(x,t)}{t_k} \left(= \partial_t f(x,\tau_k)\right). \tag{11.4}$$

Dann gilt $h_{t,n} \in \mathcal{L}^1(\mathbb{R}^n)$. Für fast alle x konvergiert $h_{t,n}(x)$ gegen $\partial_t f(x,t)$. Wegen des Mittelwertsatzes ist $g(x)$ auch Majorante von $h_{t,n}(x)$ (für fast alle x, benutze dazu den in Klammern stehenden Ausdruck in Gleichung (11.4)). Nach dem Satz über dominierte Konvergenz ist also $x \mapsto \partial_t f(x,t)$ Lebesgue integrierbar.

Weiter gilt nach diesem Satz:

$$\begin{aligned}\int_{\mathbb{R}^n} \partial_t f(x,t)\, dx &= \int_{\mathbb{R}^n} \lim_{k\to\infty} h_{t,k} = \lim_{k\to\infty} \int_{\mathbb{R}^n} \frac{f(x,t+t_k) - f(x,t)}{t_k} dx \\ &= \lim_{k\to\infty} \frac{F(t+t_k) - F(t)}{t_k} = F'(t).\end{aligned}$$

Beachte: wegen der Existenz des Grenzwerts (die Nullfolge $(t_k)_{k\in\mathbb{N}}$ war beliebig) folgt insbesondere die Ableitbarkeit von F.

Wir haben Nullmengen, auf denen Ableitungen etc. nicht existiert, ignoriert, wie wir es ja auch tun können. □

11.5 Bemerkung. Satz 11.3 verallgemeinert sich sofort auf partielle Ableitungen, wenn $D \subset \mathbb{R}^p$.

11.6 Beispiel. Regularisierung mit Hilfe von Mollifyern.

Definiere die Funktion $\rho\colon \mathbb{R}^n \to \mathbb{R}$ mit

$$\rho(x) = \begin{cases} \exp(\frac{1}{|x|_2^2 - 1}); & |x|_2 < 1 \\ 0; & |x|_2 \geq 1. \end{cases}$$

Diese Funktion ist beliebig oft differenzierbar, und hat kompakten Träger D^n, dort ist sie positiv.

Definiere für $a > 0$ $\rho_a(x) := c_a \cdot \rho(x/a)$, so dass

(1) $\rho_a(x) = 0$ falls $|x|_2 \geq a$, $\rho_a(x) > 0$ falls $|x|_2 < a$.

(2) $\int_{\mathbb{R}^n} \rho_a = 1$ (wähle c_a geeignet).

Ist nun $f \in \mathcal{L}^1(\mathbb{R}^n)$, definiere

$$F_a(x) := \int_{\mathbb{R}^n} f(y)\rho_a(x-y)\, dy.$$

Der Satz impliziert, dass $F_a(x)$ für jedes $a > 0$ eine beliebig oft differenzierbare Funktion ist.

Ausserdem: für (kleines) a wird der Wert $F_a(x)$ nur durch die Werte $f(y)$ für $|x-y|_2 < a$ bestimmt: man erhält auf diese Weise eine glatte Approximation von f.

Präziser:

11.7 Satz. *Wenn $f\colon \mathbb{R}^n \to \mathbb{C}$ stetig mit kompaktem Träger, dann konvergieren die eben definierten Funktionen F_a (welche dann ebenfalls kompakten Träger haben) gleichmäßig gegen f.*

Beweis. Die Funktion f ist gleichmäßig stetig, da ihr Träger kompakt ist. Zu $\epsilon > 0$ existiert also $\delta > 0$, so dass $|f(x) - f(y)| < \epsilon$ wenn immer $|x-y| < \delta$. Es folgt sofort $|F_a(x) - f(x)| < \epsilon$. □

11.8 Korollar. *Die Menge der glatten Funktionen mit kompaktem Träger liegt dicht im normierten Vektorraum der stetigen Funktionen mit kompaktem Träger.*

11.1 Die Γ-Funktion

11.9 Definition. Für $t > 0$ definiere

$$\Gamma(t) := \int_{(0,\infty)} x^{t-1} \exp(-x)\, dx$$

11.10 Satz. *Es gilt:*

(1) $\Gamma(t)$ *ist definiert für jedes* $t > 0$, *d.h.* $f_t \in \mathcal{L}^1(0,\infty)$ *für* $t > 0$, *wobei* $f_t(x) = x^{t-1}\exp(-x)$.

(2) $\Gamma(1) = 1$, $\Gamma(1/2) = \sqrt{\pi}$.

(3) Für jedes $t > 0$ gilt

$$\Gamma(t+1) = t\Gamma(t),$$

insbesondere

$$\Gamma(n+1) = n!$$

(4) Sei $D^n := \{x \in \mathbb{R}^n \mid |x|_2 \le 1\}$. Dann gilt

$$\mathrm{vol}(D^n) = \frac{\pi^{n/2}}{\Gamma(\frac{n}{2}+1)}.$$

(5) $\Gamma \in C^\infty((0,\infty))$ *mit*

$$\Gamma^{(n)}(t) = \int_{(0,\infty)} x^{t-1} \exp(-x)(\log(x))^n\, dx.$$

(6) Für $t > 0$ gilt

$$\Gamma(t) = \lim_{k\to\infty} \frac{k!k^t}{t(t+1)\cdot(t+k)}.$$

Beweis. Übungsaufgabe. Je nach Zeit werden Teile des Beweises noch vorgeführt. Beachte insbesondere: Differenzieren unter dem Integral ist erlaubt wegen Satz 11.3 (Voraussetzungen nachprüfen!). Die Funktionen sind aus $\mathcal{L}^1((0,\infty))$, weil man auf $[1/k, k]$ einschränken, dort normale Integrationsabschätzungen anwenden und dann monotone Konvergenz anwenden kann. Für (3) muss man partiell integrieren.

Die Berechnung des Volumens der Einheitsbälle wird induktiv mittels (Transformationsformel und) Satz von Fubini durchgeführt. Man erhält explizite Formeln, die nachträglich durch die Γ-Funktion wie angegeben ausgedrückt werden können.

(6) wird Produktdarstellung der Γ-Funktion genannt. Diese erhält man, indem man $\int_0^\infty x^{t-1}\exp(-x) = \lim_{k\to\infty}\int_0^k x^{t-1}(1-x/k)^k$ beweist, und die Integrale auf der rechten Seite durch Produktintegration explizit berechnet. $\square$