VI Analysis in metrischen Räumen

Welche Motive veranlassen Mathematiker, Räume von unendlicher Dimension einzuführen, eine Folge reeller Zahlen als einen Punkt in einem Folgenraum und eine Funktion als einen Punkt in einem Funktionenraum anzusehen? Zwei Problemkreise bildeten die treibende Kraft für die Entwicklung dieser Konzepte: zum Einen handelt es sich um Integralgleichungen der Form

$$u(t) + \int_0^1 k(t,s)u(s)ds = f(s), \quad s \in [0,1]$$

mit gegebenem Kern k und rechter Seite f und dem Ziel eine Lösung u zu finden. Zum Anderen sind es Variationsprobleme der Form

$$F(u) = \int_0^1 f(s, u(s), u'(s))ds = \min, \quad u \in X,$$

für eine gegebene Funktion f und eine gegebene Menge X von Funktionen mit dem Ziel eine Funktion $\overline{u} \in X$ zu finden, für welche F ihr Minimum annimmt. DAVID HILBERT (1862-1943) erkannte, dass sich obige Integralgleichungen als lineare Gleichungssysteme im – heute sogenananten – Hilbertschen Folgenraum ℓ^2 behandeln lassen.
Ein anderer Ansatz zur Lösung obiger Integralgleichungen besteht darin, diese Gleichungen iterativ zu lösen, also als Limes einer als $u_{j+1}(t) = f(s) - \int_0^1 k(t,s)u_j(s)ds$ definierten Folge. Bei den anstehenden Konvergenzbetrachtungen ist es dann ganz natürlich und auch sehr hilfreich, eine Funktion als Element eines Raumes zu verstehen. Diese Vorstellung ist auch bei Variationsproblemen ganz natürlich: hier setzt man bei der Suche nach einem Minimum in F Argumente u ein, mit dem Unterschied, dass, im Gegensatz zu bisher, die Argumente u nun selbst Funktionen sind.
MAURICE FRECHET (1878-1937) führte in seiner Doktorarbeit 1906 den abstrakten Begriff des metrischen Raumes ein, ein Begriff, welcher heute noch von großer Bedeutung ist, da er es erlaubt Konvergenz- und Stetigkeitsbetrachtungen auf einheitliche und anschauliche Art und Weise zu behandeln. Die Konvergenztheorie führt dann auf den Begriff des vollständigen, metrischen Raumes, der auf Fréchet und FELIX HAUSDORFF (1868-1942) zurückgeht.
Eine besonders wichtige Klasse vollständiger, metrischer Räume sind Banachräume. Dieser auf STEFAN BANACH (1892-1945) zurückgehende Begriff ist in der heutigen,

modernen Analysis von enormer Wichtigkeit und fußt auf dem Begriff des normierten Vektorraums. Unter den Banachräumen spielen diejenigen Räume, deren Norm durch ein Skalarprodunkt definiert werden kann, eine wichtige Sonderrolle. Diese werden heute Hilberträume genannt und wurden axiomatisch erstmals von JOHN VON NEUMANN (1903-1957) im Jahre 1929 eingeführt. Sie spielen insbesondere in der Quantenmechanik eine zentrale Rolle.
Im folgenden Kapitel beschäftigen wir uns zunächst mit den grundlegenden topologischen Grundbegriffen, wie Umgebungen und offenen Mengen in metrischen Räumen. Abschnitt 2 dehnt diese Betrachtungen in natürlicher Weise auf Konvergenz von Folgen und Stetigkeit von Funktionen in metrischen Räumen aus. Hier wird auch der Begriff eines vollständingen metrischen oder normierten Raumes eingeführt. In Abschnitt 3 werden kompakte Mengen via der Überdeckungseigenschaft eingeführt; ferner wird gezeigt dass in metrischen Räumen die Begriffe „überdeckungskompakt“ und „folgenkompakt“ übereinstimmen. Somit lassen sich die grundlegenden Eigenschaften stetiger Funktionen definiert auf kompakten Teilmengen des $\mathbb{R}^n$ problemlos auf metrische Räume übertragen. Den Abschluss dieses Kapitels bildet ein Abschnitt über stetige Funktionen definiert auf zusammenhängenden Mengen. Der Zusammenhang ist nichtzuletzt von Wichtigkeit, da er eine topologische Invariante darstellt.

1 Topologie metrischer Räume

In der Analysis treten neben den bisher untersuchten $\mathbb{R}^n$ und $\mathbb{C}^n$ noch viele weitere Räume mit einer Umgebungs- bzw. Abstandsstruktur auf. Wichtige Klassen in diesem Zusammenhang bilden die normierten und metrischen Räume. In letzteren wird der „Abstand“ zweier Punkte x und y durch die Zahl $d(x, y)$ beschrieben, wobei d eine Metrik ist, ein Abstandsbegriff, welchen wir im Folgenden axiomatisch definieren. Unser Ziel ist es, die schon bekannten topologischen Grundbegriffe des $\mathbb{R}^n$ auf einen metrischen Raum auszudehnen. Hierbei wollen wir bei der Definition der Begriffe „Umgebung“, „offene Menge“, „Konvergenz“ und „Häufungspunkt“ die bekannte Begriffsbildung für den $\mathbb{R}^n$ nachahmen.

Wir beginnen mit der Definition der Metrik und des metrischen Raums.

1.1 Definition. (Metrischer Raum). Es sei $M \neq \emptyset$ eine Menge. Eine Funktion

$$d : M \times M \to \mathbb{R}$$

heißt *Metrik* auf M, falls für alle $x, y, z \in M$ die folgenden Bedingungen gelten:

M1) $d(x, y) = 0 \Leftrightarrow x = y$ (Definitheit)

M2) $d(x, y) = d(y, x)$ (Symmetrie)

M3) $d(x, z) \leq d(x, y) + d(y, z)$ (Dreiecksungleichung)

Das Paar (M, d) heißt *metrischer Raum*. Die Zahl $d(x, y)$ heißt *Abstand* der Punkte x und y.

Bemerkung. Es gilt $d(x, y) \geq 0$ für alle $x, y \in M$, denn es ist

$$0 = d(x, x) \leq d(x, y) + d(y, x) = 2d(x, y).$$

Im Folgenden erläutern wir die Definition des metrischen Raumes anhand von Beispielen.

1.2 Beispiele. a) Für $x, y \in \mathbb{K}$ definieren wir durch

$$d(x, y) := |x - y|$$

die *Euklidische Metrik*.
b) Ist M eine beliebige Menge, so definiert die Vorschrift

$$d(x, y) := \begin{cases} 0 & x = y \\ 1 & x \neq y \end{cases}$$

eine Metrik auf M, die sogenannte *diskrete Metrik*.
c) Ist $X \subset M$ und (M, d) ein metrischer Raum, so definiert

$$d_X : X \times X \to \mathbb{R}, \quad d_X(x, y) := d(x, y),\ x, y \in X,$$

eine Metrik auf X, die sogenannte *induzierte Metrik*.

Eine sehr wichtige Unterklasse der metrischen Räume sind *normierte Räume*. Um diese genauer zu studieren, beginnen wir zunächst mit dem Begriff der Norm auf einem Vektorraum.

1.3 Definition. Es sei V ein Vektoraum über $\mathbb{K}$. Die Abbildung

$$\|\cdot\| : V \to \mathbb{R}$$

heißt *Norm* auf V, falls für alle $x, y \in V$ und alle $\lambda \in \mathbb{K}$ die folgenden Bedingungen gelten:

N1) $\|x\| = 0 \Leftrightarrow x = 0$ (Definitheit)

N2) $\|\lambda x\| = |\lambda| \|x\|$ (Homogenität)

N3) $\|x + y\| \leq \|x\| + \|y\|$ (Dreiecksungleichung).

Das Paar $(V, \|\cdot\|)$ heißt *normierter Vektorraum*.

1.4 Bemerkungen. a) Ist $(V, \|\cdot\|)$ ein normierter Vektoraum, so definiert

$$d(x,y) := \|x-y\|, \quad x,y \in V,$$

eine Metrik auf V. Versehen mit dieser kanonischen Metrik wird also jeder normierte Vektorraum zum metrischen Raum.
b) Die folgenden Eigenschaften der Norm ergeben sich unmitelbar aus der Definition:

i) $\|x\| \geq 0$ für alle $x \in V$.

ii) $\|x-y\| \geq |\|x\| - \|y\||, \quad x,y \in V$.

1.5 Beispiele. a) Es sei $1 \leq p \leq \infty$ und $n \geq 1$. Definiert man die sogenannte *p-Norm* durch

$$\|x\|_p := \|(x_1, \ldots, x_n)\|_p := \begin{cases} (\sum_{i=1}^n |x_i|^p)^{\frac{1}{p}} & ,1 \leq p < \infty \\ \max_{1\leq i\leq n} |x_i| & ,p = \infty \end{cases},$$

so ist $(\mathbb{K}^n, \|\cdot\|_p)$ ein normierter Vektorraum.
Die Norm $\|\cdot\|_2$ heißt auch *euklidische Norm*; die Norm $\|\cdot\|_\infty$ wird oft auch *Maximumsnorm* genannt.
Die geometrische Gestalt einer Kugel in einem normierten Raum hängt natürlich von der gewählten Norm ab. Die unten stehenden Abbildungen zeigen die Einheitskugel in $\mathbb{R}^2$, d.h. $B_1(0) := \{x \in \mathbb{R}^2; \|x\|_p < 1\}$ bezüglich der p-Normen für $p = 1, 2$ und $p = \infty$.

b) Auf dem Vektorraum V der stetigen Funktionen auf einem kompakten Intervall $[a,b]$, d.h. $V = C[a,b] = \{f : [a,b] \to \mathbb{K} : f \text{ stetig}\}$ definiert die Vorschrift

$$\|f\|_p := \begin{cases} \left(\int\limits_a^b |f(x)|^p dx\right)^{\frac{1}{p}}, & 1 \leq p < \infty \\ \sup\limits_{x\in[a,b]} |f(x)|, & p = \infty \end{cases}$$

eine Norm auf V. Der erste Ausdruck wird als *L^p-Norm* bezeichnet, der zweite als *Supremumsnorm.* Die L^p-Norm spielt eine zentrale Rolle in der harmonischen Analysis.
c) Es sei $c := \{(x_n)_{n\in\mathbb{N}} : (x_n)_{n\in\mathbb{N}} \text{ konvergent}\}$ der Raum der konvergenten Folgen und $x := (x_n)$. Dann definiert

$$\|x\|_\infty := \|(x_n)_{n\in\mathbb{N}}\|_\infty := \sup_{n\in\mathbb{N}} |x_n|$$

eine Norm auf c.

Im Folgenden übertragen wir die uns aus dem $\mathbb{R}^n$ schon bekannten topologischen Grundbegriffe wie etwa Umgebungen, offene bzw. abgeschlossene Mengen auf allgemeine metrische Räume (M, d).

1.6 Definition. Es sei (M, d) ein metrischer Raum. Dann heißt

a) für $x \in M$ und $\varepsilon > 0$ die Menge $U_\varepsilon(x) := \{y \in M : d(x, y) < \varepsilon\}$ *ε-Umgebung* von x.

b) $U \subset M$ *Umgebung* von x, falls ein $\varepsilon > 0$ existiert mit $U_\varepsilon(x) \subset U$.

c) $O \subset M$ *offen*, falls für jedes $x \in O$ ein $\varepsilon_x > 0$ existiert mit $U_{\varepsilon_x}(x) \subset O$.

d) $A \subset M$ *abgeschlossen*, falls $M \backslash A$ offen ist.

e) $X \subset M$ *beschränkt*, falls ein $x \in M$ und ein $r > 0$ existiert mit $X \subset U_r(x)$.

f) $x \in M$ *Randpunkt* der Menge $X \subset M$, falls jede Umgebung von x sowohl einen Punkt aus X als auch einen Punkt aus $M \backslash X$ enthält. Setzt man

   i) $\partial X := \{x \in M : x \text{ ist Randpunkt von } X\}$, so heißt ∂X der *Rand* von X,

   ii) $\mathring{X} := X \backslash \partial X$, so heißt $\mathring{X}$ das *Innere* von X. Ein Element $x \in \mathring{X}$ heißt *innerer Punkt* von X.

g) $x \in M$ *Häufungspunkt* der Menge $X \subset M$, falls jede Umgebung von x unendlich viele Punkte aus X enthält. Setzt man

$$\overline{X} := \{x \in M : x \in X \text{ oder } x \text{ ist Häufungspunkt von } X\},$$

so heißt $\overline{X}$ der *Abschluß* von X.

Die obigen Begriffsbildungen sind konsistent mit den entsprechenden für $\mathbb{K}^n$, die in Kapitel III.2 in Analysis I getroffen wurden. Dies bedeutet insbesondere, dass sich viele der dortigen Beweise über topologische Gegebenheiten direkt auf metrische Räume übertragen lassen, indem man nur $|x - y|$ durch $d(x, y)$ ersetzt.

1.7 Satz. (Hausdorffsches Trennungsaxiom). *Ist (M, d) ein metrischer Raum und sind $x, y \in M$ mit $x \neq y$, so existieren Umgebungen U_x und U_y von x bzw. y mit $U_x \cap U_y = \emptyset$.*

Beweis. Für $\varepsilon := \frac{1}{2}d(x,y)$ setzen wir $U_x := U_\varepsilon(x)$ und $U_y := U_\varepsilon(y)$. Falls ein $z \in U_x \cap U_y$ existieren würde, so wäre

$$2\varepsilon = d(x,y) \leq d(x,z) + d(z,y) < \varepsilon + \varepsilon = 2\varepsilon.$$

Widerspruch.

$\square$

Ganz analog wie in $\mathbb{K}^n$ beweisen wir das folgende Resultat.

1.8 Lemma. *In einem metrischen Raum (M,d) gilt:*
a) Beliebige Vereinigungen und endliche Schnitte offener Mengen sind offen.
b) Beliebige Schnitte und endliche Vereinigungen abgeschlossener Mengen sind abgeschlossen.

Der Beweis verläuft ganz analog zum Beweis von Satz III.2.6. Wir ersetzen in diesem Fall nur $|x-y|$ durch $d(x,y)$.

1.9 Definition. Zwei Normen $\|\cdot\|_1$ und $\|\cdot\|_2$ auf einem normierten Vektorraum V heißen *äquivalent*, falls zwei Konstanten $c, C > 0$ existieren, derart dass

$$c\|x\|_1 \leq \|x\|_2 \leq C\|x\|_1, \quad \text{für alle } x \in V \tag{1.1}$$

gilt.

Offensichtlich sind die oben betrachteten Normen $\|f\|_1$ und $\|f\|_\infty$ auf $C[0,1]$ nicht äquivalent, denn betrachtet man die Funktionenfolgen $(f_n)_{n\in\mathbb{N}}$ gegeben durch

so gilt $\|f_n\|_\infty = 1$ für alle $n \in \mathbb{N}$, aber $\|f_n\|_1 = \frac{1}{2n}$ für alle $n \in \mathbb{N}$.
Andererseits sind die euklidische Norm und die Maximumsnorm auf $\mathbb{R}^n$ äquivalent, was unmittelbar aus der Abschätzung

$$\|x\|_\infty \leq \|x\|_2 \leq \sqrt{n}\|x\|_\infty$$

folgt.
Wir beweisen im Folgenden, dass auf einem endlichdimensionalen Vektorraum alle Normen äquivalent sind.

1.10 Theorem. *Auf einem endlichdimensionalen $\mathbb{K}$-Vektorraum sind alle Normen äquivalent.*

Beweis. a) Wir beweisen die Aussage zunächst für $\mathbb{R}^n$ und notieren, dass es genügt zu zeigen, dass jede Norm $\|\cdot\|$ auf $\mathbb{R}^n$ äquivalent zur euklidischen Norm $\|\cdot\|_2$ ist.
Sei also $\|\cdot\|$ eine Norm auf $\mathbb{R}^n$ und e_j der j-te Einheitsvektor. Dann ist $x = (x_1, \ldots, x_n)^T = \sum_{j=1}^n x_j e_j$ und die Cauchy-Schwarzsche Ungleichung impliziert

$$\|x\| = \|\sum_{j=1}^n x_j e_j\| \leq \sum_{j=1}^n |x_j| \|e_j\| \leq C\|x\|_2, \tag{1.2}$$

wobei $C := (\sum_{j=1}^n \|e_j\|^2)^{\frac{1}{2}}$ gilt.
Um die Abschätzung nach links in der Ungleichung (1.1) zu zeigen, setze

$$c := \inf\{\|x\| : x \in S\},$$

wobei $S := \{x \in \mathbb{R}^n : \|x\|_2 = 1\}$ die euklidische Einheitssphäre bezeichnet. Im Folgenden zeigen wir, dass $c > 0$ gilt. Nehmen wir an, dass $c = 0$ ist, so gibt es in S eine Folge $(x_k)_{k\in\mathbb{N}}$ mit $\lim_{k\to\infty} \|x_k\| = 0$. Nach dem Satz von Bolzano-Weierstraß aus Analysis I (vgl. Kapitel II.2) besitzt $(x_k)_{k\in\mathbb{N}}$ eine bezüglich der euklidischen Norm konvergente Teilfolge, ebenfalls bezeichnet mit $(x_k)_{k\in\mathbb{N}}$, mit Grenzwert $a \in S$, da $1 = \lim_{k\to\infty}(x_{k1}^2 + \ldots + x_{kn}^2) = a_1^2 + \ldots + a_n^2$.

Andererseits folgt aus (1.2), dass für jedes $k \in \mathbb{N}$ gilt

$$\|a\| \leq \|a - x_k\| + \|x_k\| \leq C\|a - x_k\|_2 + \|x_k\|.$$

Für $k \to \infty$ folgt $\|a\| = 0$, also $a = 0$, im Widerspruch dazu, dass $a \in S$ gilt. Somit ist also $c > 0$.

Für $x \neq 0$ ist $x/\|x\|_2 \in S$; somit gilt $c \leq \|x/\|x\|_2\|$ und daher

$$c\|x\|_2 \leq \|x\|.$$

Da diese Ungleichung offensichtlich auch für $x = 0$ gilt, folgt die Behauptung für den Fall $\mathbb{R}^n$.

b) Es seien nun V ein beliebiger $\mathbb{K}$-Vektorraum und $\|\cdot\|$ bzw. $\|\cdot\|^\sim$ zwei Normen auf V. Ist $\phi : \mathbb{R}^n \to V$ ein Isomorphismus von $\mathbb{R}^n$ auf V, und setzt man

$$\|x\|_\phi := \|\phi(x)\| \text{ und } \|x\|_\phi^\sim := \|\phi(x)\|^\sim,$$

so folgt die Behauptung aus Teil a).

□

.

1.11 Bemerkung. Das obige Theorem hat wichtige Konsequenzen. Es besagt insbesondere, dass die oben eingeführten topologischen Grundbegriffe wie Umgebung, offene Menge und Häufungspunkt in einem endlichdimensionalen normierten Vektorraum nicht von der gewählten Norm abhängen.

2 Konvergenz und Stetigkeit

Im Folgenden übertragen wir die Begriffe der Konvergenz von Folgen und der Stetigkeit von Funktionen von $\mathbb{K}^n$ auf allgemeine metrische Räume (M, d).

2.1 Definition. Es sei (M, d) ein metrischer Raum.
a) Eine Folge $(x_j)_{j \in \mathbb{N}} \subset M$ heißt *konvergent* gegen $x \in M$, falls für jede Umgebung U von x ein $N_0 \in \mathbb{N}$ existiert mit

$$x_j \in U, \quad \text{für alle } j \geq N_0.$$

In diesem Fall schreibt man $x = \lim_{j \to \infty} x_j$ und nennt x den *Limes* der Folge $(x_j)_{j \in \mathbb{N}}$.
b) Eine Folge $(x_j)_{j \in \mathbb{N}} \subset M$ heißt *beschränkt*, falls die Menge $\{x_j : j \in \mathbb{N}\} \subset M$ beschränkt ist.
c) Ein Element $x \in M$ heißt *Häufungspunkt* der Folge $(x_j)_{j \in \mathbb{N}}$ in M, falls jede Umgebung von x unendlich viele Folgenglieder enthält.
d) Eine Folge $(x_j)_{j \in \mathbb{N}} \subset M$ heißt *Cauchyfolge*, falls für jedes $\varepsilon > 0$ ein $N_0 \in \mathbb{N}$ existiert mit der Eigenschaft

$$d(x_n, x_m) < \varepsilon, \quad \text{für alle } n, m \geq N_0.$$

Wir bemerken an dieser Stelle, dass ein Häufungspunkt der Menge $\{x_k : k \in \mathbb{N}\}$ auch ein Häufungspunkt der Folge $(x_k)_{k \in \mathbb{N}}$ ist. Die Umkehrung gilt jedoch im Allgemeinen wie schon in der Situation von $\mathbb{R}$ nicht.

Die folgenden Aussagen können analog zu den jeweiligen Aussagen aus Analysis I leicht aus den obigen Definitionen hergeleitet werden.

2.2 Satz. *Es gelten die folgenden Aussagen:*

a) *Der Limes einer konvergenten Folge ist eindeutig.*

b) *Jede konvergente Folge ist eine Cauchyfolge.*

c) *Jede Cauchyfolge ist beschränkt.*

d) *x ist genau dann Häufungspunkt einer Folge, wenn diese eine gegen x konvergente Teilfolge besitzt.*

e) *Eine Menge $A \subset M$ ist genau dann abgeschlossen, wenn für jede Folge $(x_j)_{j \in \mathbb{N}} \subset A$, welche in M konvergiert, gilt* $\lim_{j \to \infty} x_j \in A$.

Wir übertragen nun den Begriff der Stetigkeit einer Funktion definiert auf einer Teilmenge des $\mathbb{R}^n$ auf Abbildungen eines metrischen Raumes in einen anderen.

2.3 Definition. (Stetigkeit). Es seien (M, d_M) und (N, d_N) metrische Räume. Eine Abbildung $f : M \to N$ heißt *stetig in* $x_0 \in M$, falls für jede Umgebung V von $f(x_0) \in N$ eine Umgebung U von x_0 existiert mit $f(U) \subset V$.

Das folgende Theorem charakterisiert stetige Funktionen in Termen der $\varepsilon-\delta$-Stetigkeit und der Folgenstetigkeit.

2.4 Theorem. (Charakterisierung stetiger Funktionen). *Es sei $f : M \to N$ eine Funktion zwischen zwei metrischen Räumen (M, d_M) und (N, d_N). Die folgenden Aussagen sind äquivalent:*

a) f ist stetig in $x_0 \in M$

b) ($\varepsilon - \delta$-Stetigkeit). Für alle $\varepsilon > 0$ existiert ein $\delta > 0$ mit der Eigenschaft:

$$d_M(x, x_0) < \delta \Rightarrow d_N(f(x), f(x_0)) < \varepsilon.$$

c) (Folgenstetigkeit). Für jede Folge $(x_n)_{n\geq 1} \subset M$ mit $\lim_{n\to\infty} x_n = x_0$ gilt

$$\lim_{n\to\infty} f(x_n) = f(x_0).$$

Beweis. Die Äquivalenz der Aussagen a) und b) folgt unmittelbar aus der Definition des Begriffs der Umgebung.
Ferner folgt die Äquivalenz der Aussagen b) und c) wie im Beweis von Satz III.1.2 in Analysis I, indem wir $|x - y|$ durch $d_M(x, y)$ und $|f(x) - f(y)|$ durch $d_N(f(x), f(y))$ ersetzen.

□

2.5 Theorem. *Es seien (M, d_M) und (N, d_N) metrische Räume und $f : M \to N$ eine Funktion. Die folgenden Aussagen sind äquivalent:*

i) f ist stetig.

ii) $f^{-1}(A)$ ist abgeschlossen in M für alle in N abgeschlossenen Mengen $A \subset N$.

iii) $f^{-1}(O)$ ist offen in M für alle in N offenen Mengen $O \subset N$.

Der Beweis verläuft wortwörtlich zu demjenigen von Theorem III.2.17, indem wir wiederum $|x-y|$ durch $d_M(x,y)$ und $|f(x)-f(y)|$ durch $d_N(x,y)$ ersetzen.

Wir untersuchen nun speziell die Stetigkeit von linearen Abbildungen zwischen normierten Vektorräumen. Es seien also V und W zwei $\mathbb{K}$-Vektorräume. Eine Abbildung $T: V \to W$ heißt *lineare Abbildung* oder *linearer Operator*, falls

$$T(\alpha x + \beta y) = \alpha T(x) + \beta T(y), \quad \alpha, \beta \in \mathbb{K}, x, y \in V$$

gilt. Existiert für eine lineare Abbildung $T: V \to W$ zwischen zwei normierten Vektorräumen $(V, \|\cdot\|_V)$ und $(W, \|\cdot\|_W)$ eine Konstante $M \geq 0$ mit

$$\|T(x)\|_W \leq C, \quad \text{für alle } x \in V \text{ mit } \|x\|_V \leq 1,$$

so heißt T *beschränkt*. Üblicherweise schreiben wir Tx anstelle von $T(x)$.

2.6 Satz. *Für eine lineare Abbildung $T: V \to W$ zwischen zwei normierten Vektorräumen $(V, \|\cdot\|_V)$ und $(W, \|\cdot\|_W)$ sind folgende Aussagen äquivalent:*
a) T ist stetig
b) T ist stetig in einem $x_0 \in V$.
c) Es existiert eine Konstante $L > 0$ mit $\|Tx - Ty\|_W \leq L\|x-y\|_V$ für alle $x, y \in V$.
d) T ist beschränkt.

Beweis. Wir zeigen die folgenden Implikationen: $a) \Rightarrow b) \Rightarrow d) \Rightarrow c) \Rightarrow b) \Rightarrow a)$.
Die Aussage $a) \Rightarrow b)$ ist klar.
$b) \Rightarrow d)$: Nach Voraussetzung existiert zu $\varepsilon = 1$ ein $\delta > 0$ mit

$$\|T(x-x_0)\|_W = \|Tx - Tx_0\|_W \leq 1 \text{ für alle } x \in \overline{U_\delta}(x_0) := \{y \in V : \|x_0 - y\| \leq \delta\}$$

Setzt man $h := (x - x_0) \in \overline{U_\delta}(0)$, so ist diese Aussage äquivalent zu $\|Th\|_W \leq 1$ für alle $h \in \overline{U_\delta}(0)$, was wiederum äquivalent zu $\|T(\delta h)\|_W \leq 1$ für alle $h \in \overline{U_1}(0)$ und $\|T(h)\|_W \leq 1/\delta$ für alle $h \in \overline{U_1}(0)$ umformuliert werden kann. Dies bedeutet, dass T beschränkt ist.
$d) \Rightarrow c)$: Es gelte $\|Tx\|_W \leq M$ für alle $x \in \overline{U_1}(0)$. Dann folgt

$$\|T(\frac{x}{\|x\|_V})\|_W \leq M \qquad \text{für alle } x \in V \text{ mit } x \neq 0,$$

also $\|Tx\|_W \leq M\|x\|_V$ für alle $x \in V$ und somit

$$\|Tx - Ty\|_W \leq M\|x-y\|_V \text{ für alle } x, y \in V.$$

Die Aussage $c) \Rightarrow b)$ ist klar; ferner beweisen wir die Aussage $b) \Rightarrow a)$ in den Übungen. $\square$

2.7 Bemerkung. Im obigen Beweis der Implikation $d) \Rightarrow c)$ haben wir auch bewiesen, dass

$$\|T_x\|_W \leq M \text{ für alle } x \in \overline{U_1}(0) \Leftrightarrow \|Tx\|_W \leq M\|x\|_V \text{ für alle } x \in V$$

gilt. Das Infimum über alle obigen Konstanten M, d.h.

$$\|T\| = \inf\{M \geq 0 : \|Tx\|_W \leq M\|x\|_V \text{ für alle } x \in V\}$$

heißt *Operatornorm von* T. Man kann ferner zeigen, dass

$$\begin{aligned} \|T\| &= \sup\{\|Tx\|_W : \|x\|_V \leq 1\} \quad \text{und} \\ \|Tx\|_W &\leq \|T\|\|x\|_V, \quad x \in V \end{aligned}$$

gilt. Setzt man ferner

$$\mathcal{L}(V, W) := \{T : V \to W : T \text{ ist linear und beschränkt}\},$$

so wird $(\mathcal{L}(V, W), \|\cdot\|)$ ein normierter Vektorraum.

2.8 Beispiel. (Die Zeilensummennorm).Versieht man $\mathbb{K}^n$ mit der Maximumsnorm und betrachtet stetige lineare Abbildungen T auf $\mathbb{K}^n$ in sich, so ist die zugehörige Operatornorm gegeben durch

$$\|T\| = \max_i \sum_{j=1}^n |a_{ij}|,$$

wobei wir T durch die Matrix $(a_{ij})_{i,j=1}^n$ darstellen.

Die Vollständigkeit der reellen Zahlen $\mathbb{R}$ war für viele Resultate der Analysis I von zentraler Bedeutung. Wir definieren jetzt den Begriff der Vollständigkeit eines metrischen Raumes analog zur Formulierung der Vollständigkeit von $\mathbb{R}$ via Cauchyfolgen.

2.9 Definition. Ein metrischer Raum (M, d) heißt *vollständig*, falls jede Cauchyfolge in M konvergiert. Ferner wird ein vollständiger, normierter Vektorraum *Banachraum* genannt.

Die obige Definition eines vollständigen, normierten Vektoraums geht auf STEFAN BANACH (1892-1945) zurück, einem polnischen Mathematiker, von welchem grundlegende Beiträge zur Funktionalanalysis stammen. Er sah, dass die meisten in diesem Zusammenhang wichtige Resultate sich in Räumen abspielten, die neben der metrischen Eigenschaft eine Vektorraumstruktur besitzen und dass der Abstand zweier Elemente x und y sich aus der Differenz $x - y$ ableitet.
Der weiter unten folgende Banachsche Fixpunktsatz ist auch heute noch einer der bedeutendstenden Fixpunktsätze.

2.10 Beispiele. a) Der Raum $(\mathbb{K}^n, \|\cdot\|_p)$ versehen mit der $\|\cdot\|_p$-Norm ist ein Banachraum.
b) Der Funktionenraum $C[a,b]$ versehen mit der Supremumsnorm $\|\cdot\|_\infty$ ist ein Banachraum für beliebige $-\infty < a < b < \infty$, da jede auf dem kompakten Intervall $[a,b]$ gleichmäßig konvergente Folge stetiger Funktionen nach Theorem IV.4.6 eine stetige Grenzfunktion besitzt.
c) Versieht man den Raum $C^1[a,b]$ der stetig differenzierbaren Funktionen ebenfalls mit der Supremumsnorm, so entsteht ein normierter Raum, der jedoch *kein* Banachraum ist; vgl. die Übungen.
d) Versieht man hingegen $C^1[a,b]$ mit der Norm

$$\|f\| := \|f\|_\infty + \|f'\|_\infty,$$

so ist $(C^1[a,b], \|\cdot\|)$ vollständig, also ein Banachraum. Der Beweis hiervon beruht auf Theorem IV.4.7; vgl. die Übungen.
e) Der *Hilbertsche Folgenraum* ℓ^2 bestehend aus allen quadratsummierbaren Folgen komplexer Zahlen, d.h. aus allen Folgen $a = (a)_{j\in\mathbb{N}} = (a_1, a_2, \ldots)$ mit

$$\|a\|_2 := (\sum_{j=1}^{\infty} |a_j|^2)^{1/2}, \quad a \in \ell^2$$

ist ein vollständiger normierter Vektorraum, also ein Banachraum. Für den Beweis verweisen wir auf die Übungen. Die Vorschrift

$$< a, b >:= \sum_{j=1}^{\infty} a_j \overline{b_j}, \quad a, b \in \ell^2$$

definiert auf ℓ^2 sogar ein Skalarprodukt. Vektorräume versehen mit einem Skalarprodukt heißen *Hilberträume*, falls sie mit der vom Skalarprodukt induzierten Norm vollständig sind. Hilberträume spielen in der Quantenmechanik eine massgebliche Rolle; der Begriff des Hilbertraums geht auf DAVID HILBERT (1862-1943) zurück, der erkannte, dass gewisse Integralgleichungen sich in lineare Gleichungssysteme in ℓ^2 übersetzen lassen.

Fixpunktsätze haben vielfältige Anwendungen in der Mathematik. Der folgende Banachsche Fixpunktsatz besagt, dass eine strikte Kontraktion auf einem vollständigen metrischen Raum genau einen Fixpunkt besitzt.

2.11 Theorem. (Banachscher Fixpunktsatz). *Es sei (M,d) ein vollständiger, metrischer Raum und $F : M \to M$ eine strikte Kontraktion, d.h. es existiere eine Konstante $q < 1$ mit*

$$d(F(x), F(y)) \leq q\, d(x,y), \quad x, y \in M.$$

Dann besitzt F genau einen Fixpunkt $r \in M$, d.h. es existiert genau ein $r \in M$ mit $f(r) = r$.
Ferner konvergiert die rekursiv definierte Folge

$$x_{n+1} := F(x_n), \quad n \in \mathbb{N}_0$$

für beliebig gewähltes $x_0 \in M$ gegen r.

Der Beweis verläuft wortwörtlich analog zu Theorem II.2.13 aus Analysis I, indem man $|x - y|$ durch $d(x, y)$ ersetzt.

3 Kompaktheit

Die von uns im Abschnitt III.3 getroffene Definition der Kompaktheit einer Menge, nämlich die Überdeckungskompaktheit, läßt sich auch auf metrische Räume übertragen. Hierbei verstehen wir analog zur Situation in Abschnitt III.3 unter einer *offenen Überdeckung* einer Teilmenge $K \subset M$ eines metrischen Raumes M eine Familie $\{O_i\}_{i\in I}$ offener Mengen in M derart, dass jedes $x \in M$ in mindestens einem der O_i liegt, d.h. dass gilt $K \subset \bigcup_{i\in I} O_i$.

3.1 Definition. Eine Teilmenge $K \subset M$ eines metrischen Raums M heißt *kompakt*, falls jede offene Überdeckung $\{O_i\}_{i\in I}$ von K eine endliche Teilüberdeckung besitzt, d.h. falls $i_1, \ldots, i_r$ existieren mit

$$K \subset O_{i_1} \cup \ldots \cup O_{i_r}.$$

3.2 Beispiele. a) Jede endliche Teilmenge eines metrischen Raums ist kompakt
b) Es sei $(x_n)_{n\in\mathbb{N}}$ eine konvergente Folge in M mit $x = \lim_{n\to\infty} x_n$. Dann ist

$$K := \{x_n : n \in \mathbb{N}\} \cup \{x\}$$

kompakt in M.
c) Eine abgeschlossene Teilmenge einer kompakten Menge ist wiederum kompakt, d.h. ist K kompakt in M und $A \subset K$ abgeschlossen, so ist A kompakt. Der Beweis verläuft wortwörtlich analog zu demjenigen von Lemma III.3.4.

Ausgehend von dieser abstrakten Definition wollen wir nun kompakte Mengen in metrischen Räumen charakterisieren. Wir erinnern uns zunächst an die Situation des $\mathbb{R}^n$, in welcher wir kompakte Teilmengen des $\mathbb{R}^n$ nach dem Satz von Heine-Borel als genau diejenigen charakterisiert hatten, welche abgeschlossen und beschränkt sind.
Wir werden in diesem Abschnitt beweisen, dass kompakte Teilmengen eines metrischen Raums weiterhin abgeschlossen und beschränkt sind. Die Umkehrung ist jedoch in allgemeinen metrischen Räumen nicht mehr gültig. Ein Beispiel hierfür ist der Raum $C[0, \pi]$ versehen mit der Supremumsnorm, vgl. das Beispiel 3.5 c) unten.
Wir zeigen im folgenden Theorem jedoch, dass in metrischen Räumen eine Menge genau dann kompakt ist, wenn sie folgenkompakt ist.

3.3 Theorem. *Für eine nichtleere Teilmenge K eines metrischen Raumes M sind die folgenden Ausagen äquivalent:*

i) K ist kompakt (Überdeckungskompaktheit).

ii) Jede Folge $(x_n)_{n\in\mathbb{N}} \subset K$ besitzt eine konvergente Teilfolge $(x_{n_j})_{j\in\mathbb{N}}$ mit $\lim_{j\to\infty} x_{n_j} \in K$ (Folgenkompaktheit).

iii) Jede Folge $(x_n)_{n\in\mathbb{N}} \subset K$ besitzt einen Häufungspunkt in K.

Beweis. *i)* $\Rightarrow$ *ii)*: wird exakt wie in Theorem III.3.11 bewiesen.
ii) $\Leftrightarrow$ *iii)*: ist bereits in Satz 2.2 d) bewiesen.
ii) $\Rightarrow$ *i)*: Es sei $(O_i)_{i\in I}$ eine offene Überdeckung von K.
Schritt 1: es existiert ein $\delta > 0$, so daß für alle $x \in K$ ein $i \in I$ existiert mit $U_\delta(x) \subset O_i$. Wie nehmen an, die Behauptung sei falsch. Dann existiert für alle $n \in \mathbb{N}$ ein $x_n \in K$ mit $U_{1/n}(x_n) \not\subset O_i$ für alle $i \in I$. Nach Voraussetzung besitzt die Folge $(x_n)_{n\in\mathbb{N}}$ eine konvergente Teilfolge mit $x := \lim_{j\to\infty} x_{n_j} \in K$, also $x \in O_j$ für ein $j \in I$. Da O_j offen ist gilt $U_\varepsilon(x) \subset O_j$ für geeignetes $\varepsilon > 0$. Wir wählen nun $l \in \mathbb{N}$ mit $l > 2/\varepsilon$ und

$$d(x_l, x) < \varepsilon/2.$$

Dann ist $U_{1/l}(x_l) \subset U_\varepsilon(x) \subset O_j$ im Widerspruch zur obigen Eigenschaft, dass $U_{1/n}(x_n) \not\subset O_i$ für alle $i \in I$.
Schritt 2: Für alle $\delta > 0$ existieren $x_0, \ldots, x_n \in K$ mit $K \subset \bigcup_{l=0}^n U_\delta(x_l)$.
Wir nehmen wiederum an, dass die Behauptung falsch sei. Dann existiert $\delta > 0$ mit $K \not\subset \bigcup_{l=0}^n U_\delta(x_l)$ für beliebiges $n \in \mathbb{N}$ und beliebige Punkte $x_0, x_1, \ldots, x_n \in K$. Wähle nun $x_0 \in K$; dann ist $K \not\subset U_\delta(x_0)$. Es existiert daher ein $x_1 \in K \backslash U_\delta(x_0)$ und es gilt

$$K \not\subset U_\delta(x_0) \cup U_\delta(x_1)$$

Wir erhalten auf diese Weise rekursiv eine Folge $(x_n)_{n\in\mathbb{N}} \subset K$ mit

$$x_{n+1} \in K\backslash(\bigcup_{l=0}^n U_\delta(x_l)).$$

Diese Folge erfüllt nach Konstruktion

$$d(x_n, x_m) \geq \delta, \quad \text{für alle } n, m \in \mathbb{N} \text{ mit } n \neq m;$$

daher kann keine Teilfolge von $(x_n)_{n\in\mathbb{N}}$ eine Cauchyfolge sein und daher kann auch keine Teilfolge der Folge $(x_n)_{n\in\mathbb{N}}$ konvergent sein. Widerspruch!
Schritt 3: Wir wählen nun $\delta > 0$ wie in Schritt 1 und Punkte $x_0, \ldots, x_n \in K$ wie in Schritt 2. Dann gilt

$$K \subset \bigcup_{k=0}^n U_\delta(x_k) \subset \bigcup_{l=0}^n O_{i_l}$$

für geeignete $i_0, \ldots, i_l \in I$.

$\square$

Die obige Charakterisierung kompakter Mengen in metrischen Räumen durch die Folgenkompaktheit erlaubt es uns wortwörtlich analog zur Situation von $\mathbb{R}^n$ zu zeigen, dass kompakte Mengen immer abgeschlossen und beschränkt sind. Wir halten diese wichtige Beobachtung explizit im folgenden Korollar fest.

3.4 Korollar. *Eine kompakte Teilmenge eines metrischen Raums ist abgeschlossen und beschränkt.*

3.5 Beispiele. a) Es sei M eine unendliche Menge versehen mit der diskreten Metrik. Dann ist M abgeschlossen und beschränkt, aber nicht kompakt.
b) Wie in Beispiel 1.5 c) betrachten wir den Raum c der konvergenten Folgen versehen mit der Supremumsnorm. Dann ist die abgeschlossene Einheitskugel in c, d.h.

$$\overline{B_1(0)} := \{(x_n)_{n\in\mathbb{N}} \in c : |x_n| \leq 1 \text{ für alle } n \in \mathbb{N}\}$$

abgeschlossen und beschränkt, aber *nicht* kompakt. Zum Beweis sei $e_n := (0, 0, \ldots, 1, 0, \ldots)$ der n-te Einheitsvektor in c mit 1 an der n-ten Stelle. Dann ist $\|e_n - e_m\|_\infty = 1$ für alle $n \neq m$, welches bedeutet dass die Folge $(e_n)_{n\in\mathbb{N}} \subset \overline{B_1(0)}$ keine konvergente Teilfolge besitzt und somit $\overline{B_1(0)}$ nicht kompakt sein kann.
c) Die abgeschlossene Einheitskugel

$$\overline{B_1(0)} := \{f \in C[0,\pi] : \|f\|_\infty \leq 1\}$$

des Banachraums $(C[0,\pi], \|\cdot\|_\infty)$ ist nicht folgenkompakt und daher auch nicht kompakt. Ansonsten hätte die Folge der Funktionen $(f_j)_{j\in\mathbb{N}} \subset \overline{B_1(0)}$ definiert durch $f_j : [0,\pi] \to \mathbb{C}, f_j(x) := e^{2jix}$ eine konvergente Teilfolge, was aber wegen $\|f_k - f_l\|_\infty = 2$ für alle $l \neq k$ unmöglich ist.
d) Allgemeiner kann man zeigen, dass die abgeschlossene Einheitskugel $\overline{B_1(0)} := \{x \in V : \|x\| \leq 1\}$ eines normierten Vektorraums genau dann kompakt ist, wenn $\dim V < \infty$ gilt.

3.6 Korollar. *Ein kompakter, metrischer Raum* (M,d) *ist vollständig.*

Beweis. Es sei $(x_n)_{n\in\mathbb{N}} \subset M$ eine Cauchyfolge. Nach obigem Theorem 3.3 besitzt diese eine Teilfolge $(x_{n_j})_{j\in\mathbb{N}}$, welche in M konvergiert. Setzt man $x := \lim_{j\to\infty} x_{n_j}$, so existiert zu gegebenem $\varepsilon > 0$ ein $N_0 \in \mathbb{N}$ mit

$$\begin{aligned} d(x_{n_j}, x_m) &\leq \varepsilon/2 \quad \text{für alle } m, j \geq N_0 \text{ und} \\ d(x, x_{n_j}) &\leq \varepsilon/2 \quad \text{für alle } j \geq N_0; \end{aligned}$$

also gilt

$$d(x, x_m) \leq d(x, x_{n_{N_0}}) + d(x_{n_{N_0}}, x_m) \leq \varepsilon/2 + \varepsilon/2 = \varepsilon \quad \text{für alle } m \geq N_0,$$

welches $x = \lim_{m\to\infty} x_m \in M$ bedeutet.

□

In Abschnitt III.3 studierten wir im Detail Eigenschaften stetiger Funktionen welche auf einer kompakten Menge des $\mathbb{R}^n$ definiert waren, wie etwa den Satz über die Annahme eines Maximums (vgl. Satz III.3.9) oder den Satz über die gleichmäßige Stetigkeit (vgl. III.3.13). Im Folgenden übertragen wir diese Eigenschaften nun auf stetige Funktionen, welche auf einer kompakten Teilmenge eines beliebigen metrischen Raums definiert sind. Wichtige Aussagen der heutigen Analysis beruhen auf diesen Eigenschaften.

3.7 Satz. (Stetige Bilder kompakter Mengen sind kompakt). *Es seien M und N metrische Räume und $f : M \to N$ eine stetige Funktion. Ist M kompakt, so ist auch $f(M) \subset N$ kompakt.*

Der Beweis verläuft ganz analog zum Beweis von Theorem III.3.7.

3.8 Korollar. *Ist $f : M \to \mathbb{R}$ stetig und M kompakt, so nimmt f ein Minimum bzw. Maximum an.*

Der Beweis verläuft ganz analog zum Beweis von Theorem III.3.9.

3.9 Korollar. *Ist $f : M \to N$ stetig und bijektiv und M kompakt, so ist $f^{-1} : N \to M$ ebenfalls stetig.*

Beweis. Nach Theorem 2.5 genügt es zu zeigen, dass $f(A)$ abgeschlossen ist für alle in M abgeschlossenen Teilmengen $A \subset M$. Nach Beispiel 3.2 c) ist A als abgeschlossene Teilmenge einer kompakten Menge wiederum kompakt. Der obige Satz 3.7 impliziert, dass $f(A)$ ebenfalls kompakt ist; wegen des obigen Korollars 3.4 also inbesondere auch abgeschlossen.

$\square$

Zum Abschluss dieses Abschnitts betrachten wir noch gleichmäßig stetige Funktionen auf metrischen Räumen. In Analogie zur Situation von Analysis I, nennen wir eine Funktion $f : M \to N$ zwischen zwei metrischen Räumen (M, d_M) und (N, d_N) *gleichmäßig stetig*, falls zu jedem $\varepsilon > 0$ ein $\delta > 0$ existiert mit

$$d_N(f(x), f(y)) < \varepsilon \quad \text{für alle } x, y \in M \text{ mit } d_M(x, y) < \delta.$$

Es gilt dann das folgende Resultat.

3.10 Satz. *Es sei $f : M \to N$ eine stetige Funktion zwischen zwei metrischen Räumen M und N. Ist M kompakt, so ist f gleichmäßig stetig.*

Der Beweis verläuft wiederum wortwörtlich wie in Satz III.3.13.

4 Zusammenhang

Der Zwischenwertsatz aus Analysis I war ein wichtiger Bestandteil im Aufbau der Analysis. Wir verallgemeinern diesen Satz, welchen wir in Abschnitt III.1 für Intervalle formuliert hatten, nun auf stetige Funktionen deren Definitionsbereich eine zusammenhängende Teilmenge eines metrischen Raumes ist.

4.1 Definition. Ein metrischer Raum M heißt *zusammenhängend*, falls keine Zerlegung $M = X \cup Y$ existiert, in der X und Y disjunkt, offen und nicht leer sind.
Eine Menge $X \subset M$ heißt *zusammenhängend*, falls (X, d_X) als metrischer Raum zusammenhängend ist.

4.2 Beispiele. a) $\mathbb{R}^n$ ist zusammenhängend.
b) Eine Menge $M \subset \mathbb{R}$ mit mindestens zwei Elementen ist genau dann zusammenhängend, wenn sie ein Intervall ist. Für den Beweis verweisen wir auf die Übungen.
c) $\mathbb{Q} \subset \mathbb{R}$ ist nicht zusammenhängend, denn es gilt

$$\mathbb{Q} = (-\infty, \sqrt{2}) \cup (\sqrt{2}, \infty).$$

d) Die Menge $M = \{x \in \mathbb{R}^2 : x_1^2 - x_2^2 = 1\}$ ist nicht zusammenhängend.

Wir verallgemeinern nun den Satz III.1.11 über stetige Bilder von Intervallen auf zusammenhängende Mengen.

4.3 Satz. *Es sei f eine stetige Abbbildung zwischen zwei metrischen Räumen M und N. Ist M zusammenhängend, so ist auch $f(M)$ zusammenhängend.*

Beweis. Nehmen wir an die Behauptung sei falsch. Dann gäbe es disjunkte, nicht leere, offene Mengen X und Y mit $f(M) = X \cup Y$ und man erhielte $M = f^{-1}(X) \cup f^{-1}(Y)$, also einen Widerspruch.

□

Der Zwischenwertsatz in metrischen Räumen lautet wie folgt.

4.4 Satz. *Es seien M ein zusammenhängender metrischer Raum, $f : M \to \mathbb{R}$ eine stetige Funktion und $a, b \in M$. Dann nimmt f jeden Wert zwischen $f(a)$ und $f(b)$ an.*

Der Beweis ist einfach: falls $f(a) \neq f(b)$ gilt, so ist $f(M)$ nach Beispiel 4.2 b) ein Intervall.

4.5 Beispiel. Wir betrachten die Gruppe

$$O(n,\mathbb{R}) \text{ der orthogonalen } n \times n\text{-Matrizen},$$

wobei $A \in O(n,\mathbb{R})$ genau dann, wenn $A^{-1} = A^T$ gilt. Dann ist $O(n,\mathbb{R}) \subset \mathbb{R}^{n^2}$ nicht zusammenhängend.
In der Tat ist $\det : O(n,\mathbb{R}) \to \mathbb{R}$ ein Polynom in n^2 Variablen, also insbesondere eine stetige Funktion. Ferner gilt $\det diag(1,1,\ldots,1) = 1$ und $\det diag(-1,1,\ldots,1) = -1$. Wäre $O(n,\mathbb{R})$ zusammenhängend, so gäbe es nach Satz 4.4 eine Matrix $A \in O(n,\mathbb{R})$ mit $\det A = 0$ im Widerspruch dazu, dass A invertierbar ist.

Ein weiterer Zusammenhangsbegriff ist ebenfalls von Bedeutung.

4.6 Definition. Ein metrischer Raum M heißt *wegzusammenhängend*, falls es zu je zwei Punkten $a, b \in M$ eine stetige Abbildung $\gamma : [\alpha, \beta] \to M$ mit $\gamma(\alpha) = a$ und $\gamma(\beta) = b$ gibt.

Natürliche Beispiele von wegzusammenhängenden Mengen in $\mathbb{R}^n$ für $n \geq 2$ sind
a) konvexe Mengen,
b) die Einheitssphäre $S^{n-1} = \{x \in \mathbb{R}^n : \|x\|_2 = 1\}$, sowie
c) $\mathbb{R}^n \backslash \{0\}$.

Im Folgenden untersuchen wir die Verbindung zwischen zusammenhängenden und wegzusammenhängeden Mengen.

4.7 Satz. *a) Ein wegzusammenhängender metrischer Raum ist zusammenhängend.*
b) Eine zusammenhängende offene Menge X eines normierten Vektorraums ist wegzusammenhängend.

Den nicht schwierigen Beweis überlassen wir dem Leser als Übungsaufgabe.

Wir beenden diesen kurzen Abschnitt mit einem Beispiel, welches für den Orientierungsbegriff von Vektorräumen über $\mathbb{R}$ von großer Bedeutung ist.

4.8 Beispiel. a) Es sei

$$G = GL(n,\mathbb{R}) \text{ die Gruppe der reellen } n \times n\text{-Matrizen } A \text{ mit } \det A \neq 0.$$

Fassen wir G als Teilraum von $\mathbb{R}^{n^2}$ auf, so ist die Gruppe G nicht zusammenhängend. Nehmen wir an, die Behautung wäre falsch, so wäre das Bild der stetigen Abbbildung $\det : G \to \mathbb{R}$ zusammenhängend; tatsächlich gilt aber $\det(G) = \mathbb{R}\backslash\{0\}$. Widerspruch!
b) Bedeutend schwieriger zu beweisen ist die Tatsache, dass die Untergruppe

$$GL^+(n,\mathbb{R}) \text{ der reellen } n \times n\text{-Matrizen } A \text{ mit } \det A > 0$$

hingegen zusammenhängend ist.

Der Zusammenhang von metrischen Räumen stellt eine wichtige topologische Invariante dar. Sind M und N homöomorphe Räume, so ist M genau dann zusammenhängend, wenn dies auch für N zutrifft. Hierbei heißen M und N *homöomorph*, falls eine bijektive, stetige Abbildung von M auf N existiert, deren Umkehrabbildung ebenfalls stetig ist. Cantor bewies schon im Jahre 1878, dass $\mathbb{R}$ bijektiv auf $\mathbb{R}^2$ abgebildet werden kann; ferner zeigte Peano im Jahre 1890, dass es eine stetige Surjektion des Intervalls $[0,1]$ auf das Quadrat $[0,1] \times [0,1]$ gibt. Da diese Abbildung nicht bijektiv und Cantor's Konstruktion nicht stetig ist, stellt sich die Frage nach einer homöomorphen Abbildung von $\mathbb{R}^n$ auf $\mathbb{R}^m$ für $n \neq m$. L. Brouwer bewies dann im Jahe 1911, dass es eine solche Abbildung nicht geben kann. Wir beweisen hier nur den Spezialfall $m = 1$.

4.9 Satz. *Es sei $n \geq 2$. Dann ist ist $\mathbb{R}^n$ nicht homöomorph zu $\mathbb{R}$.*

Beweis. Für $n \geq 2$ ist $\mathbb{R}^n \backslash \{0\}$ zusammenhängend. Andererseits ist die Menge $\mathbb{R} \backslash \{x\}$ für kein $x \in \mathbb{R}$ nach Beispiel 4.2 b) zusammenhängend.
Gäbe es einen Homöomorphismus $f : \mathbb{R}^n \to \mathbb{R}$, so gäbe es auch einen solchen zwischen $\mathbb{R}^n \backslash \{0\}$ und $\mathbb{R} \backslash \{f(0)\}$ im Widerspruch dazu, dass nach Satz 4.3 stetige Bilder zusammenhänger Mengen wiederum zusammenhängend sind.

$\square$

VII Differentialrechnung mehrerer Veränderlicher

In diesem Kapitel erweitern wir die Differentialrechnung von Funktionen einer Variablen auf solche mit mehreren Veränderlichen. Wiederum lassen wir uns von der zentralen Idee der linearen Approximierbarkeit leiten. Im Vergleich zu unseren bisherigen Untersuchungen ist die Situation im Falle von Funktionen mehrerer Variablen nun hingegen deutlich komplizierter, da die linearen Abbildungen in der mehrdimensionalen Situation eine deutlich reichhalterigere Struktur besitzen als dies bei nur einer Variablen der Fall ist.

Wir beginnen in Abschnitt 1 mit dem Begriff der Differenzierbarkeit.

1 Differenzierbare Abbildungen

In diesem Abschnitt betrachten wir $\mathbb{R}^n$ und $\mathbb{R}^m$ als normierte Vektoräume und $f : U \to \mathbb{R}^m$ eine Abbildung definiert auf einer offenen Menge $U \subset \mathbb{R}^n$. Von besonderer Wichtigkeit wird dann der Vektorraum $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$, der Raum aller linearen Abbildungen von $\mathbb{R}^n$ in $\mathbb{R}^m$, sein, welchen wir mit der in Abschnitt VI.2 definierten Operatornorm versehen. Die Betrachtungen in Abschnitt VI.2 implizieren, aufgrund der Endlichkeit der Dimensionen, dass jede lineare Abbildung $\mathbb{R}^n \to \mathbb{R}^m$ stetig und dass $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ ein Banachraum ist.

Im Folgenden definieren wir die Differenzierbarkeit einer Funktion in einem Punkt als Approximierbarkeit durch eine lineare Abbildung.

1.1 Definition. Es seien $U \subset \mathbb{R}^n$, $x_0 \in U$ ein innerer Punkt von U und $f : U \to \mathbb{R}^m$ eine Funktion. Dann heißt f *differenzierbar* in $x_0 \in U$, falls Abbildungen $A \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ und $r : U \to \mathbb{R}^m$ existieren mit

$$f(x) = f(x_0) + A(x - x_0) + r(x), \quad x \in U,$$

und

$$\lim_{x \to x_0} \frac{r(x)}{\|x - x_0\|} = 0$$

gilt.

1.2 Bemerkungen. a) Als Norm auf $\mathbb{R}^n$ wählen wir die Euklidische Norm. Wir notieren jedoch, dass die obige Definition nach Theorem VI.1.10 unabhängig von der gewählten Norm ist.
b) Gilt $n = m = 1$, so ist die obige Definition konsistent mit derjenigen aus Kapitel IV.1.
c) Ist $f : U \subset \mathbb{R}^n \to \mathbb{R}^m$ in $x_0 \in U$ differenzierbar, so ist die lineare Abbildung A in Definition 1.1 eindeutig bestimmt, vgl. die Übungen.
d) Die eindeutig bestimmte Abbildung A aus Defintion 1.1 heißt die *Ableitung* oder das *Differential* von f in x_0. Wir schreiben

$$A = f'(x_0) \quad \text{oder } A = Df(x_0).$$

e) Wählt man in $\mathbb{R}^n$ und $\mathbb{R}^m$ kanonische Basen, so wird $A \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ durch eine Matrix $(a_{ij})_{m\times n}$ beschrieben. Für $x \in \mathbb{R}^n$ gilt dann

$$Ax = \begin{pmatrix} a_{11} & \cdot & \cdot & a_{1n} \\ \cdot & & & \\ \cdot & & & \\ \cdot & & & \\ a_{m1} & \cdot & \cdot & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{pmatrix}.$$

Im Folgenden identifizieren wir stillschweigend eine lineare Abbildung A mit der sie darstellenden Matrix (a_{ij}).

1.3 Beispiel. Es sei $B = (b_{ij}) \in M_n(\mathbb{R})$ eine symmetrische Matrix und $f : \mathbb{R}^n \to \mathbb{R}$ definiert durch

$$f(x) := (x|Bx) := \sum_{i,j=1}^{n} b_{ij} x_i x_j,$$

wobei $(\cdot|\cdot)$ das Skalarprodukt auf $\mathbb{R}^n$ bezeichnet. Dann gilt für $x_0, h \in \mathbb{R}^n$ und $x = x_0 + h \in \mathbb{R}^n$

$$\begin{aligned} f(x) = f(x_0+h) &= (x_0+h|Bx_0+Bh) = (x_0|Bx_0) + \underbrace{(x_0|Bh) + (h|Bx_0)}_{=2(Bx_0|h)} + (h|Bh) \\ &= f(x_0) + 2(Bx_0|x-x_0) + \underbrace{(x-x_0|B(x-x_0))}_{=r(x)} \\ &= f(x_0) + (2Bx_0)^T \cdot (x-x_0) + r(x). \end{aligned}$$

Mit der Cauchy-Schwarzschen Ungleichung folgt

$$|r(x)| \le \|x-x_0\| \cdot \|B(x-x_0)\|$$

und es gilt

$$\frac{|r(x)|}{\|x-x_0\|} \leq \|B(x-x_0)\| \overset{x\to x_0}{\longrightarrow} 0,$$

da $B: \mathbb{R}^n \to \mathbb{R}^n$ stetig ist; vgl. die Übungen. Daher ist f in $x_0 \in \mathbb{R}^n$ differenzierbar und es gilt

$$f'(x_0) = (2Bx_0)^T \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}) = (\mathbb{R}^n)'.$$

1.4 Satz. *Ist $U \subset \mathbb{R}^n$ offen und $f: U \to \mathbb{R}^m$ eine in $x_0 \in U$ differenzierbare Funktion, so ist f in x_0 stetig.*

Beweis. Für $x \in U$ gilt

$$f(x) = f(x_0) + f'(x_0)\cdot(x-x_0) + r(x)$$

Da die lienare Abbildung $f'(x_0)$ nach Satz VI.2.6 stetig ist und da $\lim_{x\to x_0} r(x) = 0$ gilt, folgt $\lim_{x\to x_0} f(x) = f(x_0)$, welches nach Satz VI.2.4 bedeutet, dass f in x_0 stetig ist.

□

Betrachtet man eine in $x_0 \in U$ differenzierbare Funktion $f: U \to \mathbb{R}^m$, so interessiert man sich natürlich für die Frage, wie man die Ableitung $f'(x_0) \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ konkret berechnen kann. Wir verfolgen bei der Beantwortung dieser Frage die folgende Idee: Da $f'(x_0)$ linear ist, genügt es $f'(x_0)$ auf einer Basis $\{v_1, \dots, v_n\}$ des $\mathbb{R}^n$ zu kennen. Wir berechnen daher zunächst $f'(x_0)v$ für ein $v \in \mathbb{R}^n$ mit $v \neq 0$.
Hierzu setzen wir $x = x_0 + tv$ mit $t \in \mathbb{R}$. Da U offen ist, existiert ein $\delta > 0$, so dass $x \in U$ ist für alle t mit $|t| < \delta$. Daher gilt mit den Bezeichnungen aus Definition 1.1

$$f'(x_0)v = \frac{f(x_0+tv) - f(x_0)}{t} - \frac{r(x)}{t}.$$

Da $\lim_{t\to 0} \frac{r(x)}{t} = 0$ gilt, erhalten wir

$$f'(x_0)v = \lim_{t\to 0} \frac{f(x_0+tv) - f(x_0)}{t}.$$

Dies motiviert die folgende Definition.

1.5 Definition. Es sei $U \subset \mathbb{R}^n$ offen, $f: U \to \mathbb{R}^m$ eine Funktion, $x_0 \in U$ und $v \in \mathbb{R}^n \setminus \{0\}$. Existiert

$$D_v f(x_0) := \lim_{t\to 0} \frac{f(x_0+tv) - f(x_0)}{t} \in \mathbb{R}^m,$$

so heißt $D_v f(x_0)$ die *Richtungsableitung* von f in x_0 in Richtung v.

1.6 Satz. *Es sei $f : U \subset \mathbb{R}^n \to \mathbb{R}^m$ eine differenzierbare Funktion in $x_0 \in U$. Dann existiert $D_v f(x_0)$ für alle $v \in \mathbb{R}^n \setminus \{0\}$ und es gilt*

$$D_v f(x_0) = Df(x_0) \cdot v.$$

Beweis. Für $v \neq 0$ gilt

$$f(x_0 + tv) = f(x_0) + Df(x_0)(tv) + r(x_0 + tv)$$

mit $\lim_{t \to 0} \frac{r(x_0+tv)}{||tv||} = 0$. Daher gilt

$$\frac{f(x_0 + tv) - f(x_0)}{t} = Df(x_0)v + \frac{r(x_0 + tv)}{t},$$

welches für $t \to 0$ die Behaupting impliziert.

□

Es sei an dieser Stelle angemerkt, dass die Umkehrung des obigen Satzes im Allgemeinen falsch ist. Betrachte hierzu zum Beispiel die Funktion $f : \mathbb{R}^2 \to \mathbb{R}$ gegeben durch

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

Wir haben in Definition 1.5 die Richtungsableitung einer Funktion bezüglich einer beliebigen Richtung definiert. Die Ableitung in Richtung der Koordinatenachsen ist von besonderer Wichtigkeit.

1.7 Definition. a) Für die Ableitung in Richtung der Koordinatenachsen e_j für $j = 1, \ldots, n$ schreibt man

$$\partial_j f(x_0) := \frac{\partial f}{\partial x_j}(x_0) := D_{e_j} f(x_0) = \lim_{t \to 0} \frac{f(x_0 + te_j) - f(x_0)}{t}, \quad 1 \leq j \leq n,$$

und nennt $\partial_j f(x_0)$ die *partielle Ableitung* von f in x_0 bezüglich e_j.
b) Eine Funktion $f : U \subset \mathbb{R}^n \to \mathbb{R}^m$ heißt in x_0 *partiell differenzierbar*, wenn alle partiellen Ableitungen $\partial_1 f(x_0), \ldots, \partial_n f(x_0)$ in x_0 existieren; analog heißt f in x_0 *stetig partiell differenzierbar*, falls alle partiellen Ableitungen $\partial_j f$ in x_0 existieren und stetig sind.

Existiert $\partial_j f(x_0)$ für ein $j \in 1, \ldots, n$, so gilt

$$\partial_j f(x_0) = \lim_{h \to 0} \frac{1}{h}[f(x_0^1, \ldots, x_0^{j-1}, x_0^j + h, x_0^{j+1}, \ldots, x_0^n) - f(x_0)]$$

d.h. f ist genau dann in x_0 partiell bezüglich x_j differenzierbar, wenn die Abbildung

$$t \mapsto f(x_0^1, \ldots, x_0^{j-1}, t, x_0^{j+1}, \ldots, x_0^n)$$

als Funktion einer Variablen in x_0^j differenzierbar ist.

Wählt man in $\mathbb{R}^n$ und $\mathbb{R}^m$ die kanonischen Basen und identifiziert $A \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ mit der Matrix $(a_{ij})_{m\times n}$, so erhält man die folgende Darstellung der Ableitung.

1.8 Satz. *Ist $f = (f_1, \ldots, f_m) : U \subset \mathbb{R}^n \to \mathbb{R}^m$ eine in $x_0 \in U$ differenzierbare Funktion, so gilt*

$$f'(x_0) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(x_0) & \cdots & \frac{\partial f_1}{\partial x_n}(x_0) \\ \vdots & & \\ \vdots & & \\ \frac{\partial f_m}{\partial x_1}(x_0) & \cdots & \frac{\partial f_m}{\partial x_n}(x_0) \end{pmatrix}_{m\times n}.$$

Beweis. a) Für $f_1 : U \subset \mathbb{R}^n \to \mathbb{R}$ und $h = (h_1, \ldots, h_n) \in \mathbb{R}^n$ mit $h = \sum_{j=1}^n h_j e_j$ gilt die Darstellung

$$Df_1(x_0)h = \sum_{j=1}^n h_j Df_1(x_0)e_j \overset{1.6}{=} \sum_{j=1}^n \partial_j f_1(x_0)h_j,$$

und ebenso gilt eine analoge Darstellung für $f_2, \ldots, f_m$.
b) Die obige Darstellung folgt nun, da die Funktion $f = (f_1, \ldots, f_m)$ in x_0 genau dann differenzierbar ist, wenn jede Koordinatenfunktion f_j in x_0 differenzierbar ist.

□

1.9 Definition. a) Die in Satz 1.8 definierte Matrix heißt *Jacobimatrix* oder *Funktionalmatrix* von f in x_0.
b) Gilt $m = 1$, so heißt

$$\mathrm{grad} f(x_0) := \left(\frac{\partial f}{\partial x_1}(x_0), \ldots, \frac{\partial f}{\partial x_n}(x_0)\right),$$

der *Gradient* von f in x_0. Letzterer wird auch mit $\nabla f(x_0) := \mathrm{grad}\, f(x_0)$ bezeichnet.

1.10 Bemerkung. Ist $f : U \subset \mathbb{R}^n \to \mathbb{R}$ in $x_0 \in U$ differenzierbar, so zeigt $\nabla f(x_0)$ in Richtung des steilsten Anstiegs und $-\nabla f(x_0)$ in Richtung des steilsten Abfalls von f. Dies folgt aus der Cauchy-Schwarzschen Ungleichung, denn für $v \in \mathbb{R}^n \backslash \{0\}$ gilt

$$D_v f(x_0) = Df(x_0) \cdot v = (\mathrm{grad} f(x_0)|v) \leq \|\mathrm{grad} f(x_0)\| \|v\|$$

und das Gleichheitszeichen gilt genau dann, wenn $\mathrm{grad} f(x_0) = \lambda v$ für ein $\lambda \geq 0$ ist.

Betrachtet man zum Beispiel die Funktion $f : \mathbb{R}^3 \to \mathbb{R}$ gegeben durch

$$f(x, y, z) := x^2 \sin \frac{y}{2} + e^{3z},$$

so ist der Gradient von f gegeben durch

$$\operatorname{grad} f(x, y, z) = (2x \sin \frac{y}{2}, \frac{x^2}{2} \cos \frac{y}{2}, 3e^{3z}).$$

Im Folgenden wollen wir Kriterien für die Differenzierbarkeit von Funktionen $f = (f_1, \dots, f_m) : U \subset \mathbb{R}^n \to \mathbb{R}^m$ im Punkte $x_0 \in U$ entwickeln, die einfacher handzuhaben sind als Definition 1.1. Notwendigerweise müssen zunächst alle partiellen Ableitungen

$$\frac{\partial f_k}{\partial x_j}(x_0), \quad j = 1, \dots, n; k = 1, \dots, m,$$

existieren, ansonsten wäre f in x_0 nicht differenzierbar. Ferner ist es für die Differenzierbarkeit einer Funktion f ebenfalls notwendig, dass

$$\lim_{x \to x_0} \frac{r(x)}{\|x - x_0\|} = 0 \quad \text{für } r(x) := f(x) - f(x_0) - A(x - x_0)$$

gilt, wobei $A = f'(x_0)$ gesetzt ist.
Es ist interessant zu bemerken, dass die Existenz aller Richtungsableitungen $D_v f(x_0)$ für $v \in \mathbb{R}^n \backslash \{0\}$ noch nicht einmal die Stetigkeit von f in x_0 impliziert. Ein Gegenbeispiel hierfür ist durch die Funktion

$$f : \mathbb{R}^2 \to \mathbb{R}, f(x, y) = \begin{cases} 1, & \text{für } 0 < y < x^2 \\ 0, & \text{sonst} \end{cases}$$

gegeben. Dann existiert $D_v f(0, 0)$ für alle $v \in \mathbb{R}^2$ mit $v \neq 0$, aber f ist nicht stetig in $(0, 0)$; vgl. auch die Übungen.
Es gilt jedoch das folgende Kriterium. Hierbei nennen wir eine differenzierbare Funktion $f : U \to \mathbb{R}^m$ auf einer offenen Menge $U \subset \mathbb{R}^n$ *stetig differenzierbar* in $x_0 \in U$, falls ihre Ableitung $Df : U \to \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m), x \mapsto Df(x)$ in x_0 stetig ist.

1.11 Satz. *Es seien $U \subset \mathbb{R}^n$ offen, $x_0 \in U$ und $f : U \to \mathbb{R}^m$ eine Funktion. Die folgenden Aussagen sind äquivalent:*

i) f ist stetig differenzierbar in x_0.

ii) Alle partiellen Ableitungen $\frac{\partial f_j}{\partial x_i}(x_0)$ $j = 1, \dots, m,\ i = 1, \dots, n$ existieren und sind stetig in x_0.

Beweis.
i) $\Rightarrow$ ii): folgt direkt aus Satz 1.6.
ii) $\Rightarrow$ i): Wir bemerken zunächst, dass f genauu dann in x_0 differenzierbar ist, falls alle Funktionen $f_1, \ldots, f_m$ in x_0 differenzierbar sind. Wir betrachten daher oBdA Funktionen $f : U \subset \mathbb{R}^n \to \mathbb{R}$.
Für $h = (h_1, \ldots, h_n) \in \mathbb{R}^n$ setzen wir

$$z_0 := x_0, z_1 := z_0 + h_1 e_1, z_2 := z_1 + h_2 e_2, \ldots, z_n := z_{n-1} + h_n e_n = x_0 + h.$$

Dann ist $\|x_0 - z_j\| \leq \|h\|$ für $j = 0, \ldots, n$. Es gilt also $z_j \in U$ für alle $j = 0, \ldots, n$, falls h nur genügend klein ist und der Mittelwertsatz aus Kapitel IV.2 impliziert

$$\begin{aligned} f(x_0 + h) - f(x_0) &= f(z_n) - f(z_{n-1}) + f(z_{n-1}) - f(z_{n-2}) + \ldots + f(z_1) - f(z_0) \\ (1.1) \qquad &= \frac{\partial f}{\partial x_n}(\xi_n) \cdot h_n + \frac{\partial f}{\partial x_{n-1}}(\xi_{n-1}) h_{n-1} + \ldots + \frac{\partial f}{\partial x_1}(\xi_1) h_1 \end{aligned}$$

für geeignete $\xi_j \in (z_{j-1}, z_j)$.
Deswegen gilt

$$|f(x_0 + h) - f(x_0) - \text{grad } f(x_0) \cdot h| \leq \sum_{j=1}^{n} |\frac{\partial f}{\partial x_j}(\xi_j) - \frac{\partial f}{\partial x_j}(x_0)| \cdot |h_j|$$

und somit

$$\frac{1}{\|h\|} |f(x_0 + h) - f(x_0) - \text{grad } f(x_0) \cdot h| \leq \sum_{j=1}^{n} \left| \frac{\partial f(\xi_j)}{\partial x_j} - \frac{\partial f}{\partial x_j}(x_0) \right| \overset{h \to 0}{\longrightarrow} 0,$$

da $\frac{\partial f}{\partial x_j}$ für alle $j = 1, \ldots, n$ in x_0 stetig ist.

$\square$

1.12 Bemerkung. Die Gleichung (1.1) impliziert unmittelbar das folgende Resultat: Sind alle partiellen Ableitungen $\frac{\partial f_k}{\partial x_j} : U \to \mathbb{R}$ für alle $j = 1, \ldots, n$, $k = 1, \ldots, m$ beschränkt in einer Umgebung von x_0, so ist f stetig in x_0.

Zum Abschluss dieses Abschnitts fassen wir die bisher diskutierten Ergebnisse graphisch zusammen. Für eine Funktion $f : U \subset \mathbb{R}^n \to \mathbb{R}^m$ gilt:

$$\begin{array}{ccc}
f \text{ stetig differenzierbar} & \Leftrightarrow & f \text{ stetig partiell differenzierbar} \\
\Downarrow & & \Downarrow \\
f \text{ differenzierbar} & & f \text{ part. diff'bar mit lokal beschr. part. Abl.} \\
\Downarrow & & \Downarrow \\
\text{alle Richtungsableitungen } D_v f \text{ existieren} & & f \text{ ist stetig} \\
\Downarrow & & \Downarrow \\
f \text{ partiell differenzierbar} & \Rightarrow & f \text{ stetig bzgl. jeder Koordinate}
\end{array}$$

2 Ableitungsregeln

Wir beginnen diesen Abschnitt mit der Kettenregel für differenzierbare Funktionen.

2.1 Satz. (Kettenregel). *Es seien $U \subset \mathbb{R}^n, V \subset \mathbb{R}^m$ offene Mengen, $f : U \to \mathbb{R}^m$ und $g : V \to \mathbb{R}^l$ Abbildungen mit $f(U) \subset V$. Ferner sei f in $x_0 \in U$ und g in $y_0 := f(x_0)$ differenzierbar. Dann ist die Abbildung $g \circ f : U \to \mathbb{R}^l$ in x_0 differenzierbar und es gilt*

$$D(g \circ f)(x_0) = Dg(f(x_0)) \cdot Df(x_0),$$

bzw. in Matrixschreibweise

$$J_{g \circ f}(x_0) = J_g(f(x_0)) \cdot J_g(x_0).$$

Im obigen Satz bedeutet „$\cdot$" zum einen die Hintereinanderausführung von linearen Abbildungen und zum anderen die Multiplikation von Matrizen.

Beweis. Wir setzen $A := Df(x_0) \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ und $B := Dg(f(x_0)) \in \mathcal{L}(\mathbb{R}^m, \mathbb{R}^l)$. Nach Voraussetzung gilt für $x = x_0 + h \in U$ und $y = y_0 + k \in V$

$$\begin{aligned} f(x_0 + h) &= f(x_0) + Ah + r_f(x) \\ g(y_0 + k) &= g(y_0) + Bk + r_g(y), \end{aligned}$$

mit

$$\lim_{h \to 0} \frac{r_f(x)}{\|h\|} = 0 = \lim_{k \to 0} \frac{r_g(y)}{\|k\|}.$$

Setzt man $k := f(x_0 + h) - f(x_0) = Ah + r_f(x)$, so erhalten wir

$$\begin{aligned} (g \circ f)(x_0 + h) &= g(\underbrace{f(x_0)}_{=y_0} + \underbrace{Ah + r_f(x)}_{=:k}) \\ &= g(f(x_0)) + \underbrace{BAh + Br_f(x)}_{=Bk} + r_g(y), \end{aligned}$$

und es bleibt zu zeigen, dass

$$\lim_{h \to 0} \frac{Br_f(x)}{\|h\|} = 0 \quad \text{und} \quad \lim_{h \to 0} \frac{r_g(y)}{\|h\|} = 0$$

gilt. Da $B \in \mathcal{L}(\mathbb{R}^m, \mathbb{R}^l)$ nach Satz VI.2.6 stetig ist, folgt

$$\lim_{h \to 0} \frac{Br_f(x)}{\|h\|} = B \lim_{h \to 0} \frac{r_f(x)}{\|h\|} = B0 = 0,$$

und somit die erste Aussage.
Für den Beweis der zweiten Aussage notieren wir, dass $A \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ nach Satz

VI.2.6 wiederum stetig ist. Nach demselben Satz existiert eine Konstante $M > 0$ mit $\|Ah\|_{\mathbb{R}^m} \leq L\|h\|_{\mathbb{R}^n}$ für alle $h \in \mathbb{R}^n$ und daher gilt

$$\|k\| = \|Ah + r_f(x)\| \leq (M + \frac{\|r_f(x)\|}{\|h\|})\|h\|.$$

Für $h \to 0$ gilt auch $k = f(x_0 + h) - f(x_0) \to 0$ und somit ist

$$\frac{\|r_g(y)\|}{\|h\|} = \frac{\|r_g(y)\|}{\|k\|} \cdot \frac{\|k\|}{\|h\|} \leq \frac{\|r_g(y)\|}{\|k\|} \cdot (M + \frac{\|r_f(x)\|}{\|h\|}).$$

Also gilt

$$\lim_{h \to 0} \frac{r_g(y)}{\|h\|} = 0.$$

$\square$

2.2 Beispiel. Wir betrachten die Funktionen $f : \mathbb{R}^2 \to \mathbb{R}^3$ and $g : \mathbb{R}^3 \to \mathbb{R}^2$ gegeben durch

$$\begin{aligned} f(x, y) &:= (x^2, xy, xy^2) \\ g(u, v, w) &:= (\sin v, \cos(uvw)) \end{aligned}$$

Die Funktion $h := g \circ f : \mathbb{R}^2 \to \mathbb{R}^2$ gegeben durch $h(x, y) = (\sin x^2, \cos(x^4y^3))$ ist differenzierbar und es gilt

$$Dh(x, y) = \begin{pmatrix} 2x \cos x^2 & 0 \\ -4x^3y^3 \sin(x^4y^3) & -3x^4y^2 \sin(x^4y^3) \end{pmatrix}. \tag{2.1}$$

Für die Ableitungen von f bzw. g gilt

$$Dg(u, v, w) = \begin{pmatrix} \cos u & 0 & 0 \\ -vw \sin(uvw) & -uw \sin(uvw) & -uv \sin(uvw) \end{pmatrix}, \; Df(x, y) = \begin{pmatrix} 2x & 0 \\ y & x \\ y^2 & 2xy \end{pmatrix}$$

und wir verifizieren, dass das Produkt dieser Matrizen an der Stelle $(u, v, w) = f(x, y)$ mit (2.1) übereinstimmt.

Aus der obigen Kettenregel können wir nun relativ einfach Ableitungsregeln für Summen und Produkte differenzierbarer Funktionen ableiten.

2.3 Korollar. *Es seien $U \subset \mathbb{R}^n$ offen und $f, g : U \to \mathbb{R}^m$ in $x_0 \in U$ differenzierbare Funktionen. Dann ist $\alpha f + \beta g$ für beliebige $\alpha, \beta \in \mathbb{R}$ in x_0 differenzierbar und es gilt*

$$D(\alpha f + \beta g)(x_0) = \alpha Df(x_0) + \beta Dg(x_0).$$

Beweis. Setzt man $F := \binom{f}{g} : U \to \mathbb{R}^m \times \mathbb{R}^m$ und $G : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}^m, G(u,v) := \alpha u + \beta v$, so ist G linear und somit differenzierbar mit Ableitung $DG(u,v) = G$ für alle $u, v \in \mathbb{R}^m$ und F ist ebenfalls in x_0 differenzierbar mit Ableitung $DF(x_0) := \binom{Df(x_0)}{Dg(x_0)}$. Die Kettenregel impliziert daher, dass $G \circ F$ gegeben durch $G \circ F(x) = \alpha f(x) + \beta g(x)$ in x_0 differenzierbar ist mit der Ableitung

$$D(G \circ F)(x_0) = DG(F(x_0)) \cdot DF(x_0) = G\begin{pmatrix} Df(x_0) \\ Dg(x_0) \end{pmatrix} = \alpha \cdot Df(x_0) + \beta \cdot Dg(x_0).$$

□

.

2.4 Korollar. (Produktregel). *Es seien $U \subset \mathbb{R}^n$ offen und $f, g : U \to \mathbb{R}$ in $x_0 \in U$ differenzierbare Funktionen. Dann ist $f \cdot g$ in x_0 differenzierbar und es gilt*

$$D(f \cdot g)(x_0) = f(x_0)Dg(x_0) + g(x_0)Df(x_0).$$

Beweis. Setzt man $F := \binom{f}{g} : U \to \mathbb{R} \times \mathbb{R}$ und $G : \mathbb{R} \times \mathbb{R} \to \mathbb{R},\ G(\alpha, \beta) := \alpha \cdot \beta$, so ist $(f \cdot g)(x) = (G \circ F)(x)$ und die Behauptung folgt aus der obigen Kettenregel.

□

2.5 Korollar. *Es seien $J \subset \mathbb{R}$ ein Intervall, $V \subset \mathbb{R}^m$ eine offene Menge und $\gamma = (\gamma_1, \ldots, \gamma_m) : J \to V$ in $t_0 \in J$ und $f : V \to \mathbb{R}$ in $s_0 = \gamma(t_0)$ differenzierbare Funktionen. Dann ist $f \circ \gamma : J \to \mathbb{R}$ in t_0 differenzierbar und es gilt*

$$D(f \circ \gamma)(t_0) = (\mathrm{grad} f(\gamma(t_0)) | \gamma'(t_0)) = \sum_{j=1}^{m} \frac{\partial f}{\partial x_j}(\gamma(t_0))\gamma_j'(t_0).$$

Für den Beweis verweisen wir auf die Übungen.

2.6 Bemerkung. Stetige Abbildungen $\gamma : J \to \mathbb{R}^m$ werden auch *Kurven* genannt. Letztere werden wir später noch genauer untersuchen. Es sei an dieser Stelle jedoch erwähnt, dass, fasst man $t \in J$ als Zeit und $\gamma(t) \in \mathbb{R}^m$ als Ort auf, γ die zeitliche Bewegung eines Punktes in $\mathbb{R}^m$ beschreibt. Jede Kurve $\gamma : J \to \mathbb{R}^m$ kann durch ein m-Tupel $\gamma = (\gamma_1, \ldots, \gamma_m)$ beschrieben werden und für eine differenzierbare Kurve γ gilt

$$\gamma'(t_0) = (\gamma_1'(t_0), \ldots, \gamma_m'(t_0))^T \in \mathcal{L}(\mathbb{R}, \mathbb{R}^m) = \mathbb{R}^m.$$

Der Vektor $\gamma'(t_0)$ heißt der *Tangentialvektor* der Kurve γ in t_0.

Mithilfe der obigen Formulierung der Kettenregel lassen sich die Begriffe Gradient und Niveaumenge geometrisch wie folgt interpretieren: es sei $f : U \to \mathbb{R}$ eine differenzierbare Funktion, $U \subset \mathbb{R}^n$ offen und $\gamma : J \to U$ eine differenzierbar Kurve definiert auf einem Intervall $J \subset \mathbb{R}$. Verläuft γ auf einer Niveaumenge von f, d.h. gilt $f(\gamma(t)) = c$ für alle $t \in J$ und eine Konstante $c \in \mathbb{R}$, so steht der Gradient von f im Punkte $\gamma(t)$ senkrecht auf dem Tangentialvektor $\gamma'(t)$, d.h. es gilt

$$\text{grad}g(\gamma(t)) \perp \gamma'(t), \quad t \ \in J.$$

Gilt $U \subset \mathbb{R}^2$, so lässt sich der Graph von f als „Landschaft" über U mit $f(x)$ als „Höhe" über x interpretieren. Die Niveaumengen von f entsprechen dann den Höhenlinien des Graphen. Die obige Aussage besagt in diesem Bild also, dass der Gradient von f in x senkrecht auf der Höhenlinie durch x steht. Ferner zeigt $\text{grad}f(x)$ in die Richtung des stärkesten Anstiegs von f und $-\text{grad}f(x)$ in die Richtung des steilsten Abfalls.

2.7 Beispiel. Eine differenzierbare Funktion $f : \mathbb{R}^n \setminus \{0\} \to \mathbb{R}$ heißt *positiv homogen vom Grade* $\alpha \in \mathbb{R}$, falls

$$f(tx) = t^\alpha f(x), \qquad \text{für alle } x \in \mathbb{R}^n \setminus \{0\},\ t > 0$$

gilt. Das obige Korollar 2.5 impliziert, dass

$$(\text{grad } g(x)|x) = \alpha g(x), \quad x \in \mathbb{R}^n \setminus \{0\}$$

gilt. Diese Beziehung wird auch *Eulersche Relation* genannt. Für den Beweis verweisen wir auf die Übungen.

Eine weitere Folgerung aus der Kettenregel ist der folgende Mittelwertsatz. Wie im Fall einer Variablen kann man hiermit die Differenz von Funktionswerten durch die Ableitung ausdrücken.

2.8 Satz. (Mittelwertsatz). *Es seien* $U \subset \mathbb{R}^n$ *offen und* $f : U \to \mathbb{R}$ *eine differenzierbare Funktion. Ferner seien* $a, b \in U$ *so, dass die Verbindungssgtrecke* $\overline{ab} := \{a + t(b-a), t \in [0,1]\}$ *ganz in* U *liegt. Dann existiert ein* $\xi \in \overline{ab}$ *mit*

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

Beweis. Definiert man $g : [0,1] \to U$ durch $g(t) = a + t(b-a)$, so ist g differenzierbar mit $g'(t) = b - a$ für alle $t \in (0,1)$. Nach Korollar 2.5 ist $F = f \circ g : [0,1] \to \mathbb{R}$ ebenfalls differenzierbar mit $F'(t) = f'(g(t))(b-a)$ für alle $t \in (0,1)$. Nach dem klassischen Mittelsatz, Theorem IV.2.4 aus Analysis I, existiert ein $\tau \in (0,1)$ mit

$$f(b) - f(a) = F(1) - F(0) = F'(\tau) = f'(\xi)(b-a)$$

für $\xi := g(\tau) \in \overline{ab}$.

$\square$

In diesem Zusammenhang tritt in natürlicher Weise der Begriff der konvexen Menge auf. Eine Menge $U \subset \mathbb{R}^n$ heißt *konvex*, falls $\overline{ab} = \{a + t(b-a), t \in [0,1]\} \subset U$ gilt für alle $a, b \in U$.

Für konvexe Definitionsbereiche gilt die folgende Variante des Mittelwertsatzes.

2.9 Satz. (Schrankensatz). *Es sei $U \subset \mathbb{R}^n$ eine offene und konvexe Menge. Ist $f : U \to \mathbb{R}$ differenzierbar und existiert ein $L \geq 0$ mit $\|\text{grad} f(x)\| \leq L$ für alle $x \in U$, so gilt*

$$|f(x) - f(y)| \leq L\|x - y\|, \quad x, y \in U,$$

d.h. f ist Lipschitz-stetig mit Lipschitzkonstante L.

Beweis. Nach dem Mittelwertsatz und der Cauchy-Schwarzschen Ungleichung gilt

$$|f(x) - f(y)| = |(\text{grad} f(\xi)|(x-y))| \leq \|\text{grad} f(\xi)\| \, \|x - y\| \leq L\|x - y\|$$

für ein geeignetes $\xi \in \overline{ab}$.

□

2.10 Korollar. *Es sei $U \subset \mathbb{R}^n$ eine offene Menge, so dass für je zwei Punkte $x, y \in U$ ein Streckenzug $x = z_0, z_1, \ldots, z_l = y$ existiert mit $\overline{z_{k-1}z_k} \in U$ für alle $k = 1, \ldots, l$. Dann ist eine differenzierbare Funktion $f : U \to \mathbb{R}$ genau dann konstant auf U, wenn* $\text{grad} f(x) = 0$ *für alle $x \in U$ gilt.*

Für den Beweis verweisen wir auf die Übungen.

Wir beenden diesen Abschnitt mit einer nützlichen Variante des Mittelwertsatzes, die jedoch auf der stärkeren Voraussetzung der stetigen Differenzierbarkeit beruht.

2.11 Satz. (Mittelwertsatz in Integralform). *Es sei $U \subset \mathbb{R}^n$ offen und $f : U \to \mathbb{R}$ eine stetig differenzierbare Funktion. Dann gilt*

$$f(y) - f(x) = \int_0^1 Df(x + t(y-x))(y-x)dt$$

für alle $x, y \in U$ *mit* $\overline{xy} \subset U$.

Beweis. Für $t \in [0,1]$ definieren wir $\varphi(t) := f(x + t(y-x))$. Dann ist φ stetig differenzierbar und der Hauptsatz der Differential- und Integralrechnung impliziert

$$f(y) - f(x) = \varphi(1) - \varphi(0) = \int_0^1 \varphi'(t)dt = \int_0^1 Df(x + t(y-x))(y-x)dt.$$

$\square$

3 Höhere Ableitungen

Betrachtet man eine Funktion $f : U \subset \mathbb{R}^n \to \mathbb{R}$ mit partiellen Ableitungen $\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n}$, so können diese wiederum partiell differenzierbar sein. Die Funktionen $\frac{\partial}{\partial x_i}(\frac{\partial f}{\partial x_j})$ heißen *partielle Ableitungen 2. Ordnung* von f und werden oft auch als

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

geschrieben.

Allgemeiner heißt eine Funktion $f : U \to \mathbb{R}$ $(\ell+1)$*-mal (stetig) partiell differenzierbar*, falls f ℓ-mal partiell differenzierbar ist und alle Ableitungen $\ell-$ter Ordnung (stetig) partiell differenzierbar sind. In den folgenden Untersuchungen wird der Vektorraum

$$C^k(U) := \{f : U \to \mathbb{R}, f \text{ ist } k-\text{mal stetig partiell differenzierbar}\}$$

eine wichtige Rolle spielen. Ist f k-mal stetig partiell differenzierbar, so schreiben wir auch

$$\frac{\partial}{\partial x_{i_k}} \cdots \frac{\partial}{\partial x_{i_1}} f =: \frac{\partial^k f}{\partial x_{i_k} \ldots \partial x_{i_1}} =: f_{x_{i_n} \ldots x_{i_1}} \text{ und } \partial_i \ldots \partial_i f =: \frac{\partial^k}{\partial x_i^k}.$$

Im folgenden Beispiel berechnen wir die zweiten partiellen Ableitungen der Funktion $f : \mathbb{R}^2 \to \mathbb{R}$ definiert durch $f(x,y) := x^2 \sin y$. Es gilt dann

$$\begin{array}{rclrcl} f_x(x,y) &=& 2x \sin y & f_y(x,y) &=& x^2 \cos y \\ f_{xx}(x,y) &=& 2 \sin y & f_{xy}(x,y) &=& 2x \cos y \\ f_{yx}(x,y) &=& 2x \cos y & f_{yy}(x,y) &=& -x^2 \sin y \end{array}$$

und insbesondere ist $f_{xy} = f_{yx}$ in diesem Beispiel. Im Allgemeinen ist dies jedoch nicht der Fall wie das folgende Beispiel aufzeigt.

3.1 Beispiel. Es sei $f : \mathbb{R}^2 \to \mathbb{R}$ definiert durch

$$f(x,y) := \begin{cases} \frac{x^3 y - x y^3}{x^2+y^2} & (x,y) \neq (0,0) \\ 0 & (x,y) = (0,0) \end{cases}$$

Dann ist $f \in C^1(\mathbb{R}^2)$, die partiellen Ableitungen f_{xy} und f_{yx} existieren und sind stetig auf $\mathbb{R}^2 \backslash \{(0,0)\}$, aber es gilt

$$f_{xy}(0,0) = 1 \text{ und } f_{yx}(0,0) = -1.$$

Diese Tatsache verifizieren wir in den Übungen.

Es kann sogar die Situation eintreten, dass nur eine der beiden partiellen Ableitungen $\partial_{ij} f$ oder $\partial_{ji} f$ existieren. Der folgende Satz von H.A. SCHWARZ (1843-1921) besagt, dass derartiges nicht eintritt, wenn eine der beiden partiellen Ableitungen $\partial_{ij} f$ oder $\partial_{ji} f$ stetig ist.

3.2 Satz. (Satz von Schwarz). *Es sei $U \subset \mathbb{R}^n$ offen und $f : U \to \mathbb{R}$ besitze für eine Wahl von $i, j \in \{1, \dots, n\}$ in einer Umgebung von $x_0 \in U$ partielle Ableitungen $\partial_i f$, $\partial_j f$, $\partial_{ij} f$. Ist $\partial_{ij} f$ stetig in x_0, so existiert $\partial_{ji} f$ und es gilt*

$$\partial_{ij} f(x_0) = \partial_{ji} f(x_0).$$

Beweis. Wir wählen $\delta_i, \delta_j > 0$ so klein, dass $x_0 + se_i + te_j \in U$ für alle $(s,t) \in (-\delta_i, \delta_i) \times (-\delta_j, \delta_j) =: Q \subset \mathbb{R}^2$ gilt. Dann ist die Funktion

$$\varphi : Q \to \mathbb{R}, \ \varphi(s,t) := f(x_0 + se_i + te_j)$$

wohl definiert und partiell differenzierbar. Ferner existieren die partiellen Ableitungen $\partial_1 \partial_2 \varphi$ und diese sind auch stetig in $(0,0)$.

Es ist zu zeigen, dass $\partial_2 \partial_1 \varphi(0,0)$ existiert und mit $\partial_1 \partial_2 \varphi(0,0)$ übereinstimmt. Nach Definition gilt

$$\begin{aligned} \partial_2 \partial_1 \varphi(0,0) &= [\frac{d}{dt}(\lim_{s \to 0} \frac{\varphi(s,t) - \varphi(0,t)}{s})](0) \\ &= \lim_{t \to 0} \lim_{s \to 0} \frac{1}{s} \frac{[\varphi(s,t) - \varphi(0,t)] - [\varphi(s,0) - \varphi(0,0)]}{t}. \end{aligned}$$

Wenden wir den Mittelwertsatz auf den Differenzenquotienten bzgl. der zweiten Variablen an, so erhalten wir

$$\begin{aligned} \frac{1}{s} \frac{[\varphi(s,t) - \varphi(0,t)] - [\varphi(s,0) - \varphi(0,0)]}{t} &= \frac{1}{s} \partial_2 [\varphi(s, \xi t) - \varphi(0, \xi t)] \\ &= \frac{1}{s} [\partial_2 \varphi(s, \xi t) - \partial_2 \varphi(0, \xi t)] \end{aligned}$$

für ein $\xi \in (0,1)$. Dies ist jedoch ein Differenzenquotient von $\partial_2 \varphi$ bezüglich der ersten Variable s. Nach Voraussetzung ist $\partial_2 \varphi$ bzgl. der ersten Variablen differenzierbar und nach dem Mittelwertsatz existiert ein $\eta \in (0,1)$ mit $\frac{1}{s}[\partial_2 \varphi(s, \xi t) - \partial_2 \varphi(0, \xi t)] = \partial_1 \partial_2 \varphi(\eta s, \xi t)$. Ferner ist $\partial_1 \partial_2 \varphi$ nach Voraussetzung stetig in $(0,0)$ und daher gilt

$$\partial_2 \partial_1 \varphi(0,0) = \lim_{t \to 0} \lim_{s \to 0} \partial_1 \partial_2 \varphi(\eta s, \xi t) = \partial_1 \partial_2 \varphi(0,0).$$

□

3.3 Korollar. *Ist $U \subset \mathbb{R}^n$ eine offene Menge und $f \in C^k(U)$ für ein $k \in \mathbb{N}$, so gilt*

$$\frac{\partial}{\partial x_{i_k}} \cdots \frac{\partial f}{\partial x_{i_1}} = \frac{\partial}{\partial x_{\pi(k)}} \cdots \frac{\partial f}{\partial x_{\pi(1)}}$$

für jede Permutation $\pi : \{1, \dots, k\} \to \{1, \dots, k\}$.

Den Beweis via Induktion nach k überlassen wir dem Leser.

Im Folgenden wollen wir die Ableitungen k-mal stetig differenzierbarer Funktionen f in Verallgemeinerung des Differentials als symmetrische, k-fach lineare Abbildung

$$D^k f(x_0) : \mathbb{R}^n \times \cdots \times \mathbb{R}^n \to \mathbb{R}$$

auffassen.

Wir beginnen mit dem Fall $k = 2$ und betrachten 2-mal stetig differenzierbare Funktionen, d.h. differenzierbare Funktionen f für welche Df stetig differenzierbar ist und definieren eine Bilinearform $a(u, v)$ für $(u, v) \in \mathbb{R}^n \times \mathbb{R}^n$ durch

$$D^2 f(x_0)(u, v) := D_u(D_v f)(x_0), \quad x_0 \in \mathbb{R}^n.$$

Aufgrund von Satz 1.6 ist die Richtungsableitung $D_v f(x_0)$ von f in Richtung v gegeben durch $D_v f(x_0) = Df(x_0)v$. Ferner ist die Funktion $D_v f$ wiederum in x_0 differenzierbar, da dies für Df gilt. Nach Satz 1.6 besitzt daher $D_v f$ in x_0 Richtungsableitungen in Richtung u und es gilt

$$D_u(D_v f)(x_0) = D(D_v f(x_0)) \cdot u = \sum_{i,j=1}^{n} \partial_{ij} f(x_0) v_i u_j.$$

Die Abbildung

$$(u, v) \mapsto D_u D_v f(x_0)$$

ist linear in u und v und ferner nach dem obigen Satz von Schwarz auch symmetrisch. Sie wird als *Differential zweiter Ordnung* von f in x_0 bezeichnet. Bezüglich der kanonischen Basis des $\mathbb{R}^n$ ist ihre Matrixdarstellung gegeben durch

$$H_f(x_0) = \begin{pmatrix} \partial_{11} f(x_0) & \dots & \partial_{1n} f(x_0) \\ \vdots & & \vdots \\ \partial_{n1} f(x_0) & \dots & \partial_{nn} f(x_0) \end{pmatrix}.$$

Diese Matrix heißt *Hesse-Matrix*. Nach dem Satz von Schwarz ist sie eine symmetrische Matrix und es gilt

$$D^2 f(x_0)(u, v) = u^T \cdot H_f(x_0) \cdot v. \tag{3.1}$$

Wir werden im Anschluss an die Taylorapproximation noch genauer auf die geometrische Bedeutung der zweiten Ableitung eingehen.

Für beliebiges $k \in \mathbb{N}$ definieren wir $D^k f(x_0)$ analog zum Fall $k = 2$ als

$$D^k f(x_0)(v^1, \dots, v^k) := D_{v^1} \dots D_{v^k} f(x_0), \quad v^1, \dots, v^k \in \mathbb{R}^n.$$

Diese Abbildung ist wiederum linear in den Variablen $v^1, \dots, v^k$.

In diesem Zusammenhang erweist es sich als nützlich, den Begriff des *Multiindex* einzuführen. Darunter versteht man ein n-Tupel $\alpha = (\alpha_1, \dots, \alpha_n) \in \mathbb{N}^n$. Die natürliche Zahl

$$|\alpha| := \alpha_1 + \dots + \alpha_n$$

heißt *Ordnung* von α. Ferner definiert man

$$\alpha! := \alpha_1!\alpha_2! \dots \alpha_n!$$

und für $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ setzt man

$$\begin{aligned} x^\alpha &:= x_1^{\alpha_1} x_2^{\alpha_2} \dots x_n^{\alpha_n} \text{ und} \\ D^\alpha f &:= D_1^{\alpha_1} D_2^{\alpha_2} \dots D_n^{\alpha_n} f := \frac{\partial^{|\alpha|}}{\partial x_1^{\alpha_1} \cdots \partial x_n^{\alpha_n}} f, \\ D^0 f &:= f. \end{aligned}$$

Ersetzt man in einem Polynom p von n Variablen $\xi_1, \dots, \xi_n$ vom Grade $m \in \mathbb{N}$, d.h.

$$p(\xi) = \sum_{|\alpha| \le m} a_\alpha \xi^\alpha,$$

die Variablen ξ_i durch Ableitungsoperatoren ∂_i, so entsteht ein sogenannter *linearer Differentialoperatoren $P(D)$ der Form*

$$P(D) : C^m(\mathbb{R}^n) \to C(\mathbb{R}^n), \quad P(D) := \sum_{|\alpha| \le m} a_\alpha D^\alpha$$

mit Koeffizienten a_α.

3.4 Beispiele. a) Ein sehr wichtiges Beispiel eines Differentialoperators ist der *Laplace-Operator* definiert durch

$$\Delta := \partial_1^2 + \dots + \partial_n^2.$$

Das zugehörige Polynom p ist in diesem Fall gegeben durch $p(\xi) = \xi_1^2 + \dots + \xi_n^2$.
b) Für $h = (h_1, \dots, h_n) \in \mathbb{R}^n$ betrachten wir das Polynom $p(\xi) = h_1\xi_1 + \dots + h_n\xi_n$ und setzen

$$\nabla h := p(D) = h_1\partial_1 + \dots + h_n\partial_n.$$

c) Für $a = (a_1, \dots, a_n) \in \mathbb{R}^n$ und $\ell \in \mathbb{N}$ gilt

$$(a_1 + \dots + a_n)^\ell = \sum_{|\alpha| = \ell} \frac{\ell! a^\alpha}{\alpha!}.$$

Den Induktionsbeweis überlassen wir dem Leser.
d) Für $h = (h_1, \dots, h_n)$ und $\ell \in \mathbb{N}$ gilt

$$(\nabla h)^\ell = (h_1\partial_1 + \dots + h_n\partial_n)^\ell = \ell! \sum_{|\alpha| = \ell} \frac{h^\alpha \partial^\alpha}{\alpha!}.$$

Den Laplace-Operator kann man auch als die Spur der Hesse-Matrix $H_f(x)$ einer zweimal stetig differenzierbaren Funktion identifizieren, d.h. es gilt

$$\text{spur } H_f(x) = \sum_{i=1}^{n} \partial_i^2 f(x) = \Delta f(x). \tag{3.2}$$

Aufgrund dieser Beziehung, können wir nun die Drehinvarianz der Spur einer Matrix auf die Drehinvarianz des Laplace-Operators übertragen. Genauer gesagt, gilt der folgende Satz.

3.5 Satz. (Drehinvarianz des Laplace-Operators). *Für jede Orthonormalbasis* $v_1, \ldots, v_n$ *des* $\mathbb{R}^n$ *gilt*

$$\Delta = \partial_{v_1}^2 + \ldots + \partial_{v_n}^2.$$

Beweis. Nach Gleichung (3.1) gilt

$$\partial_{v_i}\partial_{v_i} f(x) = v_i^T H_f(x) v_i = e_i^T \tilde{H} e_i,$$

mit $\tilde{H} = V^T H_f V$, $V = (v_1, \ldots, v_n)$ und den kanonischen Basisvektoren $e_1, \ldots, e_n$. Somit gilt

$$\sum_{i=1}^{n} \partial_{v_i}^2 f = \text{spur } \tilde{H}.$$

Da V nach Voraussetung orthogonal ist, besitzen die Matrizen H_f und $\tilde{H}$ diesselbe Spur und die Behauptung folgt aus Gleichung (3.2).

□

Der Laplace Operator tritt in vielen Differentialgleichungen der Analysis und der Physik auf. Beispielhaft erwähnen wir an dieser Stelle die folgenden Gleichungen:

1. Die Potentialgleichung

$$\Delta u = 0.$$

Sie beschreibt Diffusionsprozesse und tritt auch in der Wahrscheinlichkeitstheorie auf. Ihre Lösungen heißen *harmonische* Funktionen. In der Dimension 2 bilden diese den Ausgangspunkt der Funktionentheorie.

2. Die Wellengleichung

$$u_{tt} = c\Delta u$$

beschreibt die Auslenkung eines elastischen Körpers.

3. Die Wärmeleitungsgleichung

$$u_t = c\Delta u$$

beschreibt die Wärmeleitung in homogenen Medien.

4. Die Schrödingergleichung

$$u_t = i\Delta u$$

ist die zentrale Gleichung der Quantenmechanik.

Zum Abschluss dieses Abschnitts berechnen wir noch Δf für eine rotationssymmetrische Funktion. Es sei $F \in C^2(J)$, wobei $J \subset (0, \infty)$ ein Intervall ist. Setzt man $f(x) := F(\|x\|_2)$ und $r := \|x\|_2$, so gilt

$$\partial_i f(x) = F'(r)\frac{x_i}{r}$$

und

$$\partial_i^2 f(x) = F''(r)\frac{x_i^2}{r^2} + F'(r)(\frac{1}{r} - \frac{x_i^2}{r^3}).$$

Somit ergibt sich

$$\Delta f(x) = F''(r) + \frac{n-1}{r}F'(r)$$

und es gilt $\Delta f = 0$ genau dann wenn die Gleichung $F''(r) + \frac{n-1}{r}F'(r) = 0$ erfüllt ist. Wir können leicht nachprüfen, dass für $n > 2$ die Funktion F gegeben durch $F(r) = r^{2-n}$ eine Lösung dieser Gleichung ist. Somit ist die durch

$$N(x) := \frac{1}{\|x\|_2^{n-2}}$$

auf $\mathbb{R}^n \backslash \{0\}$ definierte Funktion eine Lösung der Potentialgleichung $\Delta f = 0$. Die Funktion N stimmt bis auf einen Skalierungsfaktor mit dem sogenannten *Newton-Potential* auf $\mathbb{R}^n \backslash \{0\}$ überein.

4 Der Satz von Taylor

Wir erinnern zunächst an den Satz von Taylor in der eindimensionalen Situation. Ist $f \in C^{m+1}(J)$, $J \subset \mathbb{R}$ ein Intervall und $0, x \in J$, so gilt

$$f(x) = f(0) + \frac{f'(0)}{1!}x + \frac{f''(0)}{2!}x^2 + \ldots + \frac{f^{(m)}(0)}{m!}x^m + R_{m+1}f(x,0),$$

wobei $R_{m+1}f(x,0)$ das Restglied der Taylorapproximation bezeichnete. In der Lagrangeschen Darstellung hatte es die Form

$$R_{m+1}f(x,0) = \frac{f^{(m+1)}(\xi)}{(m+1)!}x^{m+1}$$

für ein $\xi \in (0,x)$. Die obige Approximation hatte das Ziel eine gegebene, glatte Funktion durch ein Polynom in der Nähe von $x = 0$ „gut" zu approximieren. Wir betrachten im Folgenden das analoge Problem für Funktionen in mehreren Variablen, genauer gesagt für Funktionen $f \in C^{m+1}(U)$, wobei $U \subset \mathbb{R}^n$ eine offene Menge bezeichnet. Gesucht ist ein Polynom p in n Variablen der Ordnung m, d.h. $p(h) = \sum_{|\alpha| \le m} a_\alpha h^\alpha$ für $h = (h_1, \ldots, h_n)$, welches die Funktion f in der Nähe von $x = 0$ „gut" approximiert.

Der folgende Satz von Taylor besitzt äußerlich exakt dieselbe Gestalt wie in der eindimensionalen Situation.

4.1 Theorem. (Satz von Taylor in n Variablen). *Es seien $U \subset \mathbb{R}^n$ eine offene Menge, $a, x \in U$ mit $\overline{ax} \subset U$ und $f \in C^{m+1}(U)$. Dann existiert ein $\xi \in \overline{ax}$ mit*

$$\begin{aligned} f(x) &= \sum_{|\alpha| \le m} \frac{D^\alpha f(a)}{\alpha!}(x-a)^\alpha + \sum_{|\alpha| = m+1} \frac{D^\alpha f(\xi)}{\alpha!}(x-a)^\alpha \\ &=: T_m f(x,a) + R_{m+1} f(x,a). \end{aligned}$$

Beweis. Wir unterteilen den Beweis in zwei Schritte.
Schritt 1: Für $h := (h_1, \ldots, h_n) := x - a \in \mathbb{R}^n$ ist die Funktion

$$F : [0,1] \to \mathbb{R},\ F(t) := f(a + th)$$

nach der Kettenregel $(m+1)$-mal stetig differenzierbar mit

$$F'(t) = \frac{d}{dt} f(a+th) = \sum_{i=1}^n \partial_i f(a+th) \cdot h_i = (\nabla h) f(a+th).$$

Dasselbe Argument angewandt auf $g := (\nabla h) f$ liefert

$$F''(t) = \frac{d}{dt} g(a+th) = (\nabla h) g(a+th) = (\nabla h)^2 f(a+th),$$

und via Induktion erhalten wir $F \in C^{m+1}[0,1]$ mit

$$F^{(\ell)}(t) = (\nabla h)^\ell f(a+th), \quad \ell = 0,\dots,m+1,\ t \in [0,1]. \tag{4.1}$$

Schritt 2: Wir wenden nun den Satz von Taylor in einer Variablen auf die Funktion F an und erhalten wegen (4.1) und Beispiel 3.4

$$\begin{aligned}
f(x) = F(1) &= \sum_{\ell=0}^{m} \frac{F^{(\ell)}(0)}{\ell!} + \frac{F^{(m+1)}(\tau)}{(m+1)!} \\
&= \sum_{\ell=0}^{m} \frac{(\nabla h)^\ell f(a)}{\ell!} + \frac{(\nabla h)^{m+1} f(a+\tau h)}{(m+1)!} \\
&= \sum_{\ell=0}^{m} \sum_{|\alpha|=\ell} \frac{h^\alpha D^\alpha f(a)}{\alpha!} + \sum_{|\alpha|=m+1} \frac{h^\alpha D^\alpha f(a+\tau h)}{\alpha!} \\
&= \sum_{|\alpha|\le m} \frac{D^\alpha f(a)}{\alpha!}(x-a)^\alpha + \sum_{|\alpha|=m+1} \frac{D^\alpha f(\xi)}{\alpha!}(x-a)^\alpha
\end{aligned}$$

für ein geeignetes $\tau \in [0,1]$ und damit $\xi := a + \tau h \in \overline{ax}$.

□

.

4.2 Bemerkungen.
a) Analog zum Fall einer Variablen, nennt man

$$T_m f(x,a) = \sum_{|\alpha|\le m} \frac{D^\alpha f(a)}{\alpha!}(x-a)^\alpha$$

das *Taylorpolynom* von f der *Ordnung* m im *Enwicklungspunkt* a. Ferner heißt

$$R_{m+1} f(x,a) = \sum_{|\alpha|=m+1} \frac{D^\alpha f(\xi)}{\alpha!}(x-a)^\alpha$$

das *Restglied*, hier ausgedrückt in der *Lagrangeschen Form.*
b) Für $m = 0$ ist der Satz von Taylor identisch mit dem Mittelwertsatz.
c) Wir geben noch explizit das Taylorpolynom zweiter Ordnung in n Variablen an. Wegen $Df(a)h = f'(a)h$ und $D^2 f(a)(h,h) = h^T \cdot f''(a) \cdot h$ gilt für das Taylorpolynom zweiter Ordnung

$$T_2 f(x,a) = f(a) + f'(a)(x-a) + \frac{1}{2}(x-a)^T f''(a)(x-a)$$

und explizit

$$T_2 f(x,a) = f(a) + \sum_{i=1}^{n} \partial_i f(a)(x_i - a_i) + \frac{1}{2}\sum_{i,j=1}^{n} \partial_{ij} f(a)(x_i-a_i)(x_j-a_j).$$

d) Für den Spezialfall $n = 2$ und $m = 3$ lautet das Taylorpolynom $T_3 f$ an der Stelle (x, y)

$$\begin{aligned}
T_3 f((x,y),a) &= f(a) + f_x(a)(x-a_1) + f_y(a)(y-a_2) \\
&\quad + \frac{1}{2} f_{xx}(a)(x-a_1)^2 + \frac{1}{2} f_{yy}(a)(y-a_2)^2 + f_{xy}(a)(x-a_1)(y-a_2) \\
&\quad + \frac{1}{6} f_{xxx}(a)(x-a_1)^3 + \frac{1}{6} f_{yyy}(a)(y-a_2)^3 \\
&\quad + \frac{1}{2} f_{xxy}(a)(x-a_1)^2(y-a_2) + \frac{1}{2} f_{xyy}(a)(x-a_1)(y-a_2)^2 .
\end{aligned}$$

e) Ist $U \subset \mathbb{R}^n$ offen, $f \in C^m(U)$ und $a \in U$, so gilt

$$\lim_{x \to a} \frac{f(x) - T_m(x)}{\|x-a\|_2^m} = 0.$$

Um dies einzusehen, wählen wir zunächst δ so klein, dass $U_\delta(a) \subset U$ gilt. Nach dem Satz von Taylor existiert für alle $x \in U_\delta(a)$ ein $\tau \in [0,1]$ mit

$$\begin{aligned}
f(x) - T_m f(x,a) &= (f(x) - T_{m-1} f(x,a)) + (T_{m-1} f(x,a) - T_m f(x,a)) \\
&= \sum_{|\alpha|=m} \frac{D^\alpha f(a + \tau(x-a)) - D^\alpha f(a)}{\alpha!} (x-a)^\alpha .
\end{aligned}$$

Da $\frac{|(x-a)^\alpha|}{\|(x-a)\|_2^m} \leq 1$ für alle α mit $|\alpha| = m$ und da $D^\alpha f$ für solche α stetig ist, folgt

$$0 \leq \frac{f(x) - T_m(x)}{\|x-a\|^m} \leq \sum_{|\alpha|=m} \frac{|D^\alpha f(a + \tau(x-a)) - D^\alpha f(a)|}{\alpha!} \xrightarrow{x \to a} 0.$$

4.3 Beispiel. Im folgenden Beispiel berechnen wir explizit das Taylorpolynom zweiter Ordnung der Funktion

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

im Punkt $(0,0)$. Nach Bemerkung 4.2 c) gilt

$$T_2 f((x,y),(0,0)) = f(0,0) + x f_x(0,0) + y f_y(0,0) + \frac{x^2}{2} f_{xx}(0,0) + \frac{y^2}{2} f_{yy}(0,0) + xy f_{xy}(0,0).$$

Berechnet man die jeweiligen Ableitungen im Punkt $(0,0)$, so ergibt sich

$$T_2 f((x,y),(0,0)) = 1 + \frac{x}{2} + y - \frac{x^2}{2} - \frac{y^2}{8} + \frac{xy}{4}.$$

Ausgehend vom Mittelwertsatz 2.11 in Integralform , d.h.

$$f(x) = f(a) + \int_0^1 Df(a+th)hdt \tag{4.2}$$

mit $h = x - a$, wollen wir noch das Restglied der Taylorentwicklung in Integralform angeben. Der Taylorsche Satz lautet dann wie folgt.

4.4 Korollar. (Satz von Taylor mit Restglied in Integralform). *Es seien $U \subset \mathbb{R}^n$ eine offene Menge, $a, x \subset U$ mit $\overline{ax} \in U$ und $f \in C^{m+1}(U)$. Dann gilt*

$$f(x) \quad = \quad \sum_{|\alpha| \leq m} \frac{D^\alpha f(a)}{\alpha!}(x-a)^\alpha + \int_0^1 \frac{(1-t)^m}{m!}((\nabla h)^{m+1} f)(a+th)dt$$

für $h = x - a$.

Um dies einzusehen, notieren wir zunächst, dass die Behauptung für $m = 0$ genau mit dem obigen Mittelwertsatz in Integralform übereinstimmt.
Für den Fall $m = 1$ definieren wir für $t \in [0, 1]$ zwei Funktion u, v durch $u(t) := Df(a+th)h$ und $v(t) := t-1$. Die Produktregel zusammen mit der obigen Darstellung (4.2) impliziert, dass

$$f(x) = f(a) + Df(a)h + \int_0^1 (1-t)((\nabla h)^2 f)(a+th)dt$$

gilt. Für $m \geq 2$ folgt die Behauptung via eines Induktionsbeweises, welchen wir dem Leser als Übungsaufgabe überlassen.

4.5 Bemerkung. Wir wollen an dieser Stelle noch den Begriff der Tangentialebene diskutieren. Ist $U \subset \mathbb{R}^n$ offen und $f \in C^1(U)$ eine Funktion, so sagen wir dass durch die Gleichung $z = f(x), x \in U$, eine Fläche F in einem $(n+1)$-dimensionalen Raum dargestellt wird. Dabei ist graph $f = \{(x, z) \in \mathbb{R}^{n+1}, z = f(x), x \in U\}$.
Wir haben gesehen, dass das Taylorpolynom $T_1 f(x, a)$ als affine Funktion betrachtet werden kann, welche f in der Nähe von a approximiert. Die durch

$$z = T_1 f(x, a) = f(a) + Df(a)(x-a)$$

dargestellte Hyperebene heißt *Tangentialebene* an die Fläche $z = f(x)$ im Punkt $(a, f(a))$. Der Vektor

$$\nu = (\nabla f(a), -1)$$

heißt *Normale* der Tangentialebene im Punkt $(a, f(a))$. Im Fall $n = 1$ benutzt man die Bezeichnung Kurve anstatt Fläche und Tangente anstatt Tangentialebene. In diesem Fall haben wir die Gleichung der Tangente an die Kurve $z = f(x)$, nämlich

$$z = f(a) + f'(a)(x-a)$$

schon in Kapitel IV kennengelernt. Im Fall $n = 2$ lautet die Gleichung der Tangetialebene im Punkt (a_1, a_2)

$$z = f(a_1, a_2) + f_x(a_1, a_2)(x - a_1) + f_y(a_1, a_2)(y - a_2).$$

5 Lokale Extrema

In diesem Abschnitt untersuchen wir hinreichende Kriterien für lokale Maxima und Minima von Funktionen mehrerer reeller Variablen. Wir beginnen mit der Definition eines lokalen Extremwerts.

5.1 Definition. Es seien $U \subset \mathbb{R}^n$ offen und $f : U \to \mathbb{R}$ eine Funktion. Ein Punkt $a \in U$ heißt *lokales Maximum (Minimum)* von f, falls eine Umgebung $\tilde{U} \subset U$ von a existiert mit

$$f(x) \leq f(a) \quad \text{für alle } x \in \tilde{U} \quad (f(x) \geq f(a) \quad \text{für alle } x \in \tilde{U}).$$

Ein *lokales Extremum* ist ein lokales Maximum oder Minimum.

5.2 Satz. *Es seien $U \subset \mathbb{R}^n$ offen und $f : U \to \mathbb{R}^n$ eine in $a \in U$ partiell differenzierbare Funktion. Besitzt f in a ein lokales Extremum, so gilt*

$$\text{grad} f(a) = 0.$$

Beweis. Wir wählen $\delta > 0$ so klein, dass die Funktionen g_i definiert durch

$$g_i(-\delta, \delta) \to \mathbb{R}, \quad g_i(t) = f(a + te_i), \quad i = 1, \dots, n$$

alle wohldefiniert und differenzierbar in $t = 0$ sind. Da alle Funktionen g_i in $t = 0$ ein lokales Extremum besitzen, gilt nach Satz IV.3.8 aus Analysis I

$$g_i'(0) = \partial_i f(a) = 0, \quad i = 1, \dots, n.$$

□

Ist $f : U \to \mathbb{R}$ partiell differenzierbar in $a \in U$ und gilt $\text{grad} f(a) = 0$, so heißt a *kritischer Punkt* von f.

5.3 Beispiele.
a) Betrachte die Funktion $f : \mathbb{R}^2 \to \mathbb{R}$ gegeben durch $f(x, y) := x^2 + y^2$. Dann gilt

$$\text{grad} f(x) = (2x, 2y) = 0 \Leftrightarrow x = y = 0.$$

Weiter gilt $f(x, y) > 0$ für alle $(x, y) \neq (0, 0)$, welches bedeutet dass f in $(0, 0)$ ein isoliertes Minimum besitzt.

b) Betrachtet man die Funktion $f : \mathbb{R}^2 \to \mathbb{R}$, $f(x,y) := x^2 - y^2$, so gilt

$$\operatorname{grad} f(x,y) = (2x, -2y) = 0 \Leftrightarrow x = y = 0.$$

Ferner gilt $f(x,0) > 0$ für alle $x \neq 0$ und $f(0,y) < 0$ für alle $y \neq 0$. Dies bedeutet, dass f in $(0,0)$ kein lokales Extremum besitzt, sondern einen sogenannten Sattelpunkt.

Wir wollen nun der Frage nachgehen, wie man die oben skizzierten Fälle systematisch unterscheiden kann. Zunächst sei an das hinreichende Kriterium für Extremwerte von Funktionen einer reellen Variablen aus Analysis I erinnert. Es sei $f : \mathbb{R} \to \mathbb{R}$ eine 2-mal differenzierbare Funktion und a kritischer Punkt von f. Dann ist a

- ein lokales Minimum, falls $f''(a) > 0$ und
- ein lokales Maximum, falls $f''(a) < 0$

gilt. Für Funktionen in mehreren Variablen ersetzen wir $f''(a)$ durch die Hesse-Matrix $(\partial_i \partial_j f(a))_{n \times n}$. Diese wurde in 3.1 für eine Funktion $f \in C^2(U)$ im Punkt $a \in U$, $U \subset \mathbb{R}^n$ offen, definiert als

$$H_f(a) = \begin{pmatrix} \partial_1\partial_1 f(a) & \partial_1\partial_2 f(a) & \cdots & \partial_1\partial_n f(a) \\ \partial_2\partial_1 f(a) & \cdots & \cdots & \vdots \\ \vdots & \cdots & \cdots & \vdots \\ \vdots & \cdots & \cdots & \vdots \\ \partial_n\partial_1 f(a) & \cdots & \cdots & \partial_n\partial_n f(a) \end{pmatrix}_{n \times n}$$

5.4 Bemerkungen.
a) Aufgrund des Satzes von Schwarz 3.2 ist $H_f(a)$ eine symmetrische Matrix.
b) Die Polynomformel 3.4 c) impliziert für $h = (h_1, \ldots, h_n) \in \mathbb{R}^n$ und $f \in C^2(U)$

$$\begin{aligned} 2\sum_{|\alpha|=2} \frac{h^\alpha D^\alpha f(a)}{\alpha!} &= (h_1 D_1 + \ldots + h_n D_n)^2 f(a) \\ &= \sum_{i,j=1}^{n} h_i h_j D_i D_j f(a) = (h | H_f(a) h). \end{aligned}$$

Somit gilt nach dem Satz von Taylor und Bemerkung 4.2 c)

$$f(x) = f(a) + (\operatorname{grad} f(a)^T | (x-a)) + \frac{1}{2}((x-a), H_f(a)(x-a)) + r(x), \quad x \in U$$

mit $\lim_{x \to a} \frac{r(x)}{||x-a||^2} = 0$.

Zur Bestimmung von Extremwerten benötigen wir ferner die folgende Begriffe aus der Linearen Algebra.

5.5 Definition. Eine symmetrische Matrix $T \in M_n(\mathbb{R})$ heißt
a) *positiv definit*, falls $(x|Tx) > 0$ für alle $x \in \mathbb{R}^n \backslash \{0\}$,
b) *negativ definit*, falls $(x|Tx) < 0$ für alle $x \in \mathbb{R}^n \backslash \{0\}$,
c) *indefinit*, falls $x, y \in \mathbb{R}^n$ existieren mit $(x|Tx) > 0$ und $(y|Ty) < 0$.

Die Eigenschaft einer Matrix T positiv bzw. negativ definit zu sein, lässt sich, wie im folgenden Satz beschrieben, insbesondere auch durch die Vorzeichen der Eigenwerte von T charakterisieren.

5.6 Satz. *Für eine symmetrische Matrix $T \in M_n(\mathbb{R})$ gilt:*

$$\begin{aligned} a)\ T \text{ ist positiv definit} \quad &\Leftrightarrow \quad \text{alle Eigenwerte von } T \text{ sind } > 0 \\ &\Leftrightarrow \quad \det \begin{pmatrix} t_{11} & \cdots & t_{1k} \\ \vdots & & \vdots \\ t_{k1} & \cdots & t_{kk} \end{pmatrix}_{k \times k} > 0 \text{ für alle } k = 1, \ldots, n. \end{aligned}$$

$$\begin{aligned} b)\ T \text{ ist negativ definit} \quad &\Leftrightarrow \quad \text{alle Eigenwerte von } T \text{ sind } < 0 \\ &\Leftrightarrow \quad (-1)^k \det \begin{pmatrix} t_{11} & \cdots & t_{1k} \\ \vdots & & \vdots \\ t_{k1} & \cdots & t_{kk} \end{pmatrix}_{k \times k} > 0 \text{ für alle } k = 1, \ldots, n. \end{aligned}$$

c)

$$\begin{aligned} T \text{ ist indefinit} \quad &\Leftrightarrow \quad T \text{ besitzt positive und negative Eigenwerte} \\ &\Leftrightarrow \quad \text{obige Unterdeterminaanten sind alle } \neq 0 \text{ und genügen} \\ &\qquad weder der Bedingung unter a) noch der unter b) \end{aligned}$$

Für den Beweis verweisen wir auf die Lineare Algebra.

5.7 Theorem. (Hinreichendes Kriterium für lokale Extrema).
Es sei $U \subset \mathbb{R}^n$ offen, $f \in C^2(U)$ eine Funktion und $a \in U$ ein kritischer Punkt von f. Dann gelten die folgenden Aussagen.

a) *Ist $H_f(a)$ positiv definit, so besitzt f in a ein lokales Minimum.*

b) *Ist $H_f(a)$ negativ definit, so besitzt f in a ein lokales Maximum.*

c) Ist $H_f(a)$ indefinit, so besitzt f in a kein lokales Extremum.

Beweis. Nach Bemerkung 5.4 b) und da grad $f(a) = 0$ ist, gilt für $h := x - a$ mit $x \in U$

$$f(x) = f(a) + \frac{1}{2}(h, H_f(a)h) + r(x)$$

mit $\lim_{x \to a} \frac{r(x)}{\|h\|^2} = 0$. Dies bedeutet, dass für alle $\varepsilon > 0$ ein $\delta > 0$ existiert mit $0 \leq |r(x)| < \varepsilon \|h\|^2$ für alle $h \in U_\delta(0) := \{v \in \mathbb{R}^n : \|v\| \leq \delta\}$.
a) Nach Voraussetzung ist $H_f(a)$ positiv definit. Da die stetige Funktion

$$h \mapsto (h|H_f(a)h)$$

auf der kompakten Einheitssphäre $S := \{v \in \mathbb{R}^n : \|v\| = 1\}$ ein Minimum annimmt, existiert ein $v_0 \in S$ mit

$$(v|H_f(a)v) \geq (v_0|H_f(a)v_0) =: m > 0$$

für alle $v \in S$. Setzt man $v := \frac{h}{\|h\|}$ für $h \in \mathbb{R}^n \backslash \{0\}$, so gilt

$$(h|H_f(a)h) \geq m\|h\|^2 \text{ für alle } h \in \mathbb{R}^n.$$

Wir wählen nun $\delta > 0$ so klein, dass $\frac{m}{4}\|h\|^2 > |r(x)| \geq 0$ für alle $h = x - a \in U_\delta(0)$ gilt. Daher ist

$$\begin{aligned} f(x) &\overset{\text{Taylor}}{=} f(a) + \frac{1}{2}\underbrace{(h|H_f(a)h)}_{\geq m\|h\|^2} + \underbrace{r(x)}_{> -\frac{m}{4}\|h\|^2} > f(a) + \frac{m}{2}\|h\|^2 - \frac{m}{4}\|h\|^2 \\ &= f(a) + \frac{m}{2}\|h\|^2 > f(a), \quad h \in U_\delta(0), \end{aligned}$$

und f besitzt somit in a ein lokales Minimum.
b) Die Aussage b) erhalten wir durch Anwenden von a) auf $-f$.
c) Nach Voraussetzung existieren $v, w \in \mathbb{R}^n$ mit $(v|Hv) > 0$ und $(w|Hw) < 0$. Wir wählen nun $\delta > 0$ so, dass $a + tv \in U$ und $a + tw \in U$ für alle $t \in (-\delta, \delta)$ gilt und setzen

$$\begin{aligned} F_v : (-\delta, \delta) \to \mathbb{R} \quad F_v(t) &:= f(a + tv) \\ F_w : (-\delta, \delta) \to \mathbb{R} \quad F_w(t) &:= f(a + tw). \end{aligned}$$

Die Funktionen F_v und F_w sind zweimal stetig differenzierbar und es gilt

$$F_v'(0) = \nabla f(a) \cdot v = 0, \quad F_w'(0) = \nabla f(a) \cdot w = 0,$$

sowie

$$F_v''(0) = (\nabla v)^2 f(a) = (v|H_f(a)v) > 0, \quad F_w''(0) = (\nabla w)^2 f(a) = (w|H_f(a)w) < 0.$$

Dies bedeutet, dass F_v in $t = 0$ ein lokales Minimum und F_w in $t = 0$ ein lokales Maximum besitzt, welches zeigt, dass f in a kein lokales Extremum besitzen kann.

□

5.8 Beispiele.
Wir betrachten zunächst nochmals die Beispiele aus 5.3.
a) Es sei $f(x,y) = x^2 + y^2$.
Es gilt

$$H_f(0,0) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

und somit ist $H_f(0,0)$ positiv definit und f besitzt daher in $(0,0)$ ein lokales Minimum.
b) Für $f(x,y) = x^2 - y^2$ gilt

$$H_f(0,0) = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Somit ist $H_f(0,0)$ indefinit und f besitzt in $(0,0)$ kein lokales Extremum.
c) Als weiteres Beispiel betrachten wir $f : \mathbb{R}^2 \to \mathbb{R}$, $f(x,y) = x^3 + y^3 - 3xy$. Dann implizieren die Gleichungen

$$f_x(x,y) = 3x^2 - 3y = 0, \quad f_y(x,y) = 3y^2 - 3x = 0,$$

dass $(0,0)$ und $(1,1)$ kritische Punkte von f sind. Die Hesse-Matrix von f lautet

$$H_f(x,y) = \begin{pmatrix} 6x & -3 \\ -3 & 6y \end{pmatrix}$$

und daher gilt

$$H_f(0,0) = \begin{pmatrix} 0 & -3 \\ -3 & 0 \end{pmatrix}, \quad H_f(1,1) = \begin{pmatrix} 6 & -3 \\ -3 & 6 \end{pmatrix}.$$

Die beiden Eigenwerte von $H_f(0,0)$ sind ± 3 und somit ist $H_f(0,0)$ indefinit und f besitzt in $(0,0)$ keins lokales Extremum. Die beiden Eigenwerte von $H_f(1,1)$ sind hingegen strikt positiv, welches bedeutet, dass $H_f(1,1)$ positiv definit ist und f in $(1,1)$ ein lokales Minimum besitzt.

6 Differentation parameterabhägiger Integrale

In der Physik werden häufig Variationsprinzipien angewandt, um zum Beispiel die Bahn, längs deren sich das System bewegt, als Extremale eines bestimmten Variationsproblems zu bestimmen. Für die rigorose Behandlung solcher Variationsprobleme benötigen wir Differenzierbarkeitseigenschaften parameterabhängiger Integrale. Wir beschränken uns in diesem Abschnitt auf Integrale mit kompakten Integrationsbereichen; den allgemeinen Fall nicht kompakter Inegrationsbereiche werden wir erst im Rahmen des Lebesgue-Integrals behandeln.

5.8 Beispiele.
Wir betrachten zunächst nochmals die Beispiele aus 5.3.
a) Es sei $f(x,y) = x^2 + y^2$.
Es gilt

$$H_f(0,0) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$

und somit ist $H_f(0,0)$ positiv definit und f besitzt daher in $(0,0)$ ein lokales Minimum.
b) Für $f(x,y) = x^2 - y^2$ gilt

$$H_f(0,0) = \begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix}.$$

Somit ist $H_f(0,0)$ indefinit und f besitzt in $(0,0)$ kein lokales Extremum.
c) Als weiteres Beispiel betrachten wir $f : \mathbb{R}^2 \to \mathbb{R},\ f(x,y) = x^3 + y^3 - 3xy$. Dann implizieren die Gleichungen

$$f_x(x,y) = 3x^2 - 3y = 0, \quad f_y(x,y) = 3y^2 - 3x = 0,$$

dass $(0,0)$ und $(1,1)$ kritische Punkte von f sind. Die Hesse-Matrix von f lautet

$$H_f(x,y) = \begin{pmatrix} 6x & -3 \\ -3 & 6y \end{pmatrix}$$

und daher gilt

$$H_f(0,0) = \begin{pmatrix} 0 & -3 \\ -3 & 0 \end{pmatrix}, \quad H_f(1,1) = \begin{pmatrix} 6 & -3 \\ -3 & 6 \end{pmatrix}.$$

Die beiden Eigenwerte von $H_f(0,0)$ sind ± 3 und somit ist $H_f(0,0)$ indefinit und f besitzt in $(0,0)$ keins lokales Extremum. Die beiden Eigenwerte von $H_f(1,1)$ sind hingegen strikt positiv, welches bedeutet, dass $H_f(1,1)$ positiv definit ist und f in $(1,1)$ ein lokales Minimum besitzt.

6 Differentation parameterabhägiger Integrale

In der Physik werden häufig Variationsprinzipien angewandt, um zum Beispiel die Bahn, längs deren sich ein System bewegt, als Extremale eines bestimmten Variationsproblems zu bestimmen. Für die rigorose Behandlung solcher Variationsprobleme benötigen wir Differenzierbarkeitseigenschaften parameterabhängiger Integrale. Wir beschränken uns in diesem Abschnitt auf Integrale mit kompakten Integrationsbereichen; den allgemeinen Fall nicht kompakter Inegrationsbereiche werden wir erst im Rahmen des Lebesgue-Integrals behandeln.

Es seien $U \subset \mathbb{R}^n$ offen und $J \subset \mathbb{R}$ ein kompaktes Intervall. Wir betrachten eine Funktion $f : U \times J \to \mathbb{R}$ mit der Eigenschaft, dass für jedes $x \in U$ die Funktion $t \mapsto f(x,t)$ stetig ist und definieren die Funktion $F : U \to \mathbb{R}$ durch

$$F(x) := \int_J f(x,t)dt.$$

Dann gilt der folgende Satz

6.1 Theorem. (Differentation von parameterabhängigen Integralen).
a) Ist f stetig auf $U \times J$, so ist F stetig auf U.
b) Ist f zusätzlich nach x_i stetig partiell differenzierbar, so ist F nach x_i steti partiell differenzbar und man darf „unter dem Integral differenzieren“, d.h. es gilt

$$\frac{\partial F(x)}{\partial x_i} = \int_J \frac{\partial f(x,t)}{\partial x_i} dt.$$

Beweis. a) Wir bemerken zunächst, dass es genügt den Satz für $U \subset \mathbb{R}$ zu beweisen. Es sei $x \in U \in \mathbb{R}$ und $(x_j) \subset U$ eine Folge mit $x_j \to x$. Dann ist f auf der kompakten Menge $J \times \{x, x_1, x_2, \ldots\}$ gleichmäßig stetig, d.h. für jedes $\varepsilon > 0$ existiert ein $\delta > 0$ mit $|f(x_k,t) - f(x,t)| < \varepsilon$ für alle $x \in U$, falls nur $|x_k - x| < \delta$ gilt. Dies bedeutet, dass $f(x_k, \cdot)$ gleichmäßig auf J gegen $f(x, \cdot)$ konvergiert. Nach Satz V.2.13 gilt daher

$$\lim_{k\to\infty} \int_J f(x_k,t)dt = \int_J f(x,t)dt,$$

und somit ist F stetig auf U.
b) Es sei $K \subset U$ eine kompakte Menge mit $x \in K$. Auf dem Kompaktum $K \times J$ ist $\frac{\partial f}{\partial x}$ gleichmäßig stetig, d.h. für jedes $\varepsilon > 0$ existiert ein $\delta > 0$ mit $|\frac{\partial f}{\partial x}(\tilde{x},t) - \frac{\partial f}{\partial x}(x,t)| < \varepsilon$ für alle $|x - \tilde{x}| \le \delta$. Nach dem klassischen Mittelwertsatz IV.2.4 existiert ein $\xi_k \in (x, x_k)$ mit

$$\frac{f(x_k,t) - f(x,t)}{x_k - x} = \frac{\partial f}{\partial x}(\xi_k, t).$$

Wir wählen nun $N \in \mathbb{N}$ so, dass $|x - x_k| < \delta$ gilt für alle $k \ge N$. Damit gilt auch $|x - \xi_k| < \delta$ und

$$\left| \frac{\partial f}{\partial x}(x,t) - \frac{f(x_k,t) - f(x,t)}{x_k - x} \right| = \left| \frac{\partial f}{\partial x}(x,t) - \frac{\partial f}{\partial x}(\xi_k,t) \right| < \varepsilon.$$

Somit konvergiert

$$\frac{f(x_k,\cdot) - f(x,\cdot)}{x_k - x} \to \frac{\partial f}{\partial x}(x,\cdot)$$

gleichmäßig auf J. Nach Satz V.2.13 konvergieren daher die zugehörigen Integrale, d.h. es gilt

$$\lim_{k\to\infty} \frac{F(x_k) - F(x)}{x_k - x} = \lim_{k\to\infty} \int_J \frac{f(x_k, t) - f(x,t)}{x_k - x} dt = \int_J \frac{\partial f}{\partial x}(x,t)dt.$$

Somit ist F partiell nach x differenzierbar und es gilt

$$F'(x) = \int_J \frac{\partial f}{\partial x}(x,t)dt.$$

Schließlich ist F' stetig nach Aussage a), da $\frac{\partial f}{\partial x}$ stetig auf $U \times J$ vorausgesetzt war.

□

Als erste Anwendung dieses Satzes beweisen wir die Vertauschbarkeit für iterierte Integrale.

6.2 Satz. *Ist $f : [a,b] \times [c,d] \to \mathbb{R}$ stetig, so gilt*

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

Beweis. Wir definieren Funktionen $F_1, F_2 : [a,b] \to \mathbb{R}$ durch

$$F_1(\xi) := \int_a^\xi \left(\int_c^d f(x,t)dt \right) dx, \quad F_2(\xi) := \int_c^d \left(\int_a^\xi f(x,t)dx \right) dt.$$

Da der Integrand von F_1 stetig auf $[a,b]$ ist, ist F_1 nach dem Hauptsatz der Differential- und Integralrechnung differenzierbar mit $F'(\xi) = \int_c^d f(\xi,t)dt$. Die Funktion F_2 ist nach dem obigen Theorem 6.1 ebenfalls differenzierbar mit $F_2'(\xi) = \int_c^d f(\xi,t)dt$. Daher gilt $F_1' = F_2'$ und wegen $F_1(a) = F_2(a) = 0$ auch $F_1 = F_2$.

□

Durch wiederholtes Anwenden des obigen Verfahrens kann man das iterierte Integral einer stetigen Funktion auf einem Quader $Q := [a_1, b_1] \times \ldots \times [a_k, b_k] \subset \mathbb{R}^k$ als

$$\int_Q f(x)dx := \int_{a_k}^{b_k} \left(\ldots \int_{a_2}^{b_2} \left(\int_{a_1}^{b_1} f(x_1, \ldots, x_k)dx_1 \right) dx_2 \ldots \right) dx_k$$

definieren.

Im Folgenden betrachten wir die sogenannte Eulersche Differentialgleichung der Variationsrechnung. Hierunter versteht man die folgende Problematik. Zwischen zwei koaxialen Kreislinien soll diejenige Rotationsfläche bestimmt werden, welche kleinsten Flächeninhalt besitzt. Genauer gesagt, suchen wir zu zwei gegebenen Punkten $(a, \alpha$

und (b,β) mit $a<b$ eine stetig differenzierbare Funktion $f:[a,b]\to\mathbb{R}_+$ mit $f(a)=\alpha$ und $f(b)=\beta$ so, dass die durch Rotation ihres Graphen um die x-Achse entstehende Fläche möglichst kleinen Flächeninhalt hat. Den Flächeninhalt einer solchen Fläche werden wir später als

$$F(f)=2\pi\int_a^b f(x)\sqrt{1+f'(x)^2}dx$$

bestimmen.

Ausgehend von diesem Beispiel betrachten wir allgemeiner eine zweimal stetig differenzierbare Funktion

$$L:[a,b]\times\mathbb{R}\times\mathbb{R}\to\mathbb{R},\ (t,y,p)\mapsto L(t,y,p),$$

setzen für $\alpha,\beta\in\mathbb{R}$

$$V:=\{\varphi\in C^2[a,b]:\varphi(a)=\alpha,\varphi(b)=\beta\}$$

und definieren

$$J:V\to\mathbb{R},\ J(\varphi):=\int_a^b L(t,\varphi(t),\varphi'(t))dt.$$

Gesucht wird ein $\varphi\in V$, in dem J ein Extremum annimmt. Im obigen Beispiel wäre also $L(t,x,p)=x\sqrt{1+p^2}$. Das hier formulierte Extremalproblem ist von besonderer Art, da der Definitionsbereich von J eine Teilmenge des unendlich-dimensionalen Vektorraums V ist.
Der folgende Satz gibt eine notwendige Bedingung für die Existenz einer Extremalstelle von J an.

6.3 Satz. (Eulersche Differentialgleichung). *Gilt* $J(\varphi)=\inf_{\psi\in V}J(\psi)$ *für ein* $\varphi\in V$, *so gelten die Eulerschen Differentialgleichungen*

$$\frac{d}{dt}\frac{\partial L}{\partial p}(t,\varphi(t),\varphi'(t))=\frac{\partial L}{\partial y}(t,\varphi(t),\varphi'(t)).$$

Beweis. Es sei $\varphi\in V$ mit $J(\varphi)\le J(\psi)$ für alle $\psi\in V$ und $g\in C^2[a,b]$ eine Funktion mit $g(a)=0=g(b)$. Dann ist $\varphi+\varepsilon g\in V$ für alle $\varepsilon\in\mathbb{R}$ und es gilt

$$J(\varphi)\le J(\varphi+\varepsilon g).$$

Setzt man $F:\mathbb{R}\to\mathbb{R}, F(\varepsilon):=J(\varphi+\varepsilon g)$, so hat F in $\varepsilon=0$ ein Minimum; es gilt also $\frac{dF}{d\varepsilon}(0)=0$. Nach Satz 6.1 dürfen wir unter dem Integral differenzieren und erhalten

$$\begin{aligned}\frac{dF}{d\varepsilon}(\varepsilon) &= \int_a^b \frac{d}{d\varepsilon}L(t,\varphi+\varepsilon g,\varphi'+\varepsilon g')dt\\ &= \int_a^b \frac{\partial L}{\partial y}(t,\varphi+\varepsilon g,\varphi'+\varepsilon g')g+\frac{\partial L}{\partial p}(t,\varphi+\varepsilon g,\varphi'+\varepsilon g')g'dt\end{aligned}$$

Integriert man den zweiten Term auf der rechten Seite partiell, so erhalten wir

$$\int_a^b \frac{\partial L}{\partial p} g' dt = \underbrace{\frac{\partial L}{\partial p} g\Big|_a^b}_{=0} - \int_a^b g(t) \frac{d}{dt}\left(\frac{\partial L}{\partial p}\right) dt$$

und somit gilt

$$0 = \frac{dF}{d\varepsilon}(0) = \int_a^b \left[\frac{\partial L}{\partial y}(t,\varphi,\varphi') - \frac{d}{dt}\left(\frac{\partial L}{\partial p}\right)(t,\varphi,\varphi')\right] g dt$$

für jede Funktion $g \in C^2[a,b]$ mit $g(a) = g(b) = 0$. Die Behauptung folgt nun aus dem folgenden Lemma.

□

6.4 Lemma. *Ist $f : [a,b] \to \mathbb{R}$ eine stetige Funktion und gilt für jede Funktion $g \in C^2[a,b]$ mit $g(a) = 0 = g(b)$*

$$\int_a^b f(t)g(t)dt = 0,$$

so ist $f \equiv 0$ auf $[a,b]$.

Beweis. Da f stetig ist, genügt es zu zeigen, dass $f \equiv 0$ auf (a,b) gilt. Wir nehmen an, dass $f(x) \neq 0$ ist für ein $x \in (a,b)$. OBdA sei $f(x) = \varepsilon > 0$. Die Stetigkeit von f impliziert, dass eine Umgebung $U_\delta(x)$ von x existiert mit $f(t) \geq \varepsilon/2$ für alle $t \in U_\delta(x)$. Wir wählen nun eine Funktion $g \in C^2[a,b]$ mit $g \geq 0$ und $g(x) > 0$ und $g(t) = 0$ für alle $t \in [a,b] \backslash U_\delta(x)$. Daher gilt

$$0 = \int_a^b f(t)g(t)dt = \int_{x-\delta}^{x+\delta} f(t)g(t)dt \geq \frac{\varepsilon}{2} \underbrace{\int_{x-\delta}^{x+\delta} g(t)dt}_{>0} > 0.$$

Widerspruch!

□

6.5 Beispiel.
Wie oben sei $V = \{\varphi \in C^2[a,b] : \varphi(a) = \alpha, \text{ und } \varphi(b) = \beta\}$. Motiviert durch die Bogenlänge von Kurven betrachten wir

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

Dann gilt $L(t,y,p) = \sqrt{1+p^2}$,

$$\frac{\partial L}{\partial y} = 0, \quad \frac{\partial L}{\partial p}(t,y,p) = \frac{p}{\sqrt{1+p^2}}$$

und die Eulersche Differentialgleichung lautet

$$\frac{d}{dt}\,\frac{\varphi'(t)}{\sqrt{1+\varphi'(t)^2}} = \frac{\partial L}{\partial y} = 0.$$

Deshalb gilt

$$\frac{\varphi''}{\sqrt{1+\varphi'^2}} - \varphi'\frac{\varphi'\varphi''}{(1+\varphi'^2)^{\frac{3}{2}}} = 0,$$

und somit $\varphi''(t) = 0$. Damit ergibt sich

$$\varphi(t) = \alpha + \beta t$$

und wir haben gezeigt, dass die Gerade die kürzeste Verbindung zwischen zwei Punkten darstellt.

Im Folgenden verallgemeinern wir die obige Strategie auf die Situation von n-Funktionen. Die Funktion L ist dann von der Form

$$L : [a,b] \times \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}, \quad (t, y_1, \ldots, y_n, p_1, \ldots, p_n) \mapsto L(t, y_1, \ldots, y_n, p_1, \ldots, p_n)$$

und für V gilt

$$V = \{f \in C^2([a,b], \mathbb{R}^n) : f(a) = \alpha, f(y) = \beta\}$$

für $\alpha, \beta \in \mathbb{R}^n$. Definiert man $J : V \to \mathbb{R}$ via

$$J(\varphi) = \int_a^b L(t, \varphi_1(t), \ldots, \varphi_n(t), \varphi_1'(t), \ldots, \varphi_n'(t))dt,$$

so impliziert ein Minimum von J in $\varphi = (\varphi_1, \ldots, \varphi_n) \in V$, dass

$$\frac{d}{dt}L_{p_i}(t, \varphi, \varphi') - L_{y_i}(t, \varphi, \varphi') = 0, \quad i = 1, \ldots, n$$

gilt.
Betrachtet man ein physikalisches System beschrieben durch die Zeitkoordinate t und Ortskoordinaten $\varphi(t) = (\varphi_1(t), \ldots, \varphi_n(t))$, so heißt

a) $L(t, \varphi, \varphi')$ die *Lagrange-Funktion* und

b) es gilt $L = T - U$, wobei $T = T(\varphi, \varphi')$ die *kinetische Energie* und $U = U(\varphi)$ die *potentielle Energie* des Systems beschreibt.

c) Ferner wird $J(\varphi) = \int_a^b L(t, \varphi(t), \varphi'(t))dt$ in der Physik auch *Wirkungsintegral* genannt.

Das *Hamiltonsches Prinzip* der Mechanik besagt, dass zwischen zwei Zeitpunkten t_0, t_1 die Bewegung des Systems so verläuft, dass das Integral

$$J(\varphi) = \int_{t_0}^{t_1} T(\varphi, \varphi') - U(\varphi))dt,$$

minimal wird. Die Eulerschen Differentialgleichungen impliziert dann

$$\frac{d}{dt}\frac{\partial T}{\partial \varphi_i'} - \frac{\partial}{\partial \varphi_i}(T - U) = 0, \quad i = 1, \dots, n.$$

In der Mechanik heißen diese Gleichungen die *Lagrangeschen Bewegungsgleichungen.* Betrachten wir speziell die Bewegung eines Massenpunkts unter dem Einfluss eines nur vom Ort abhängigen Potentials U, so gilt mit $x = (x_1, x_2, x_3)$ und $v = (x_1', x_2', x_3')$

$$T = \frac{m}{2}\sum_{i=1}^{3} v_i^2 = \frac{m}{2}\sum_{i=1}^{3}(x_i')^2(t) \quad \text{und } L(x, v) = \frac{m}{2}(v_1^2 + v_2^2 + v_3^2) - U(x_1, x_2, x_3).$$

Da die Lagrangefunktion L nicht explizit von t abhängt, gilt

$$\frac{\partial L}{\partial x_i} = -\frac{\partial U}{\partial x_i}, \quad \text{und } \frac{\partial L}{\partial v_i} = m v_i$$

und die Eulersche Differentialgleichung lautet

$$\frac{d}{dt}(m x_i'(t)) + \frac{\partial U}{\partial x_i}(x_i(t)) = 0.$$

Daher lauten die Bewegungsgleichung in diesem Fall

$$m x_i'' = -\mathrm{grad}\, U(x), \quad i = 1, 2, 3.$$

6.6 Beispiel. (Rotationsminimalflächen). Wir betrachten nun das eingangs erwähnte Beispiel der Rotationsfläche und definieren hierzu J als

$$J(\varphi) = \int_a^b \varphi(t)\sqrt{1 + \varphi'(t)^2}dt. \tag{6.1}$$

Die Eulersche Differentialgleichung lautet

$$L_{pp}\varphi'' + L_{py}\varphi' + L_{pt} - L_y = 0.$$

Ist L unabhängig von t, so gilt für $E_\varphi := L_p(\varphi, \varphi')\varphi' - L(\varphi, \varphi')$

$$\begin{aligned} \frac{d}{dt}E_\varphi &= (L_{py}\varphi'^2 + L_{pp}\varphi'\varphi'' + L_p\varphi'') - L_y\varphi' - L_p\varphi'' \\ &= \varphi'(L_{py}\varphi' + L_{pp}\varphi'' - L_y) = 0, \end{aligned}$$

und jede Lösung der Eulerschen Differentialgleichung erfüllt in diesem Fall

$$E_\varphi = L_p(\varphi, \varphi')\varphi' - L(\varphi, \varphi') = \text{constant.}$$

In der Physik interpretiert man E_φ als die Energie des Systems. Betrachtet man nun speziell J definiert wie in (6.1), so gilt

$$L(t, x, p) = x\sqrt{1+p^2}$$

und $\frac{\partial L}{\partial p}(t, x, p) = \frac{xp}{\sqrt{1+p^2}}$ und $\frac{\partial L}{\partial y} = \sqrt{1+p^2}$. Die Eulersche Gleichung lautet daher

$$\frac{d}{dt}(\varphi\frac{\varphi'}{\sqrt{1+\varphi'^2}}) = \sqrt{1+\varphi'^2} \tag{6.2}$$

und da L unabhängig von t ist, gilt $L_p(\varphi, \varphi')\varphi' - L(\varphi, \varphi') = c$ für ein $c \in \mathbb{R}$. Damit folgt

$$\frac{xp^2}{\sqrt{1+p^2}} - x\sqrt{1+p^2} = -c$$

und somit

$$\frac{\varphi}{\sqrt{1+\varphi'^2}} = \text{constant.}$$

Damit vereinfacht sich die Eulersche Gleichung zu

$$c\varphi'' = \frac{1}{c}\varphi \Leftrightarrow \varphi'' - \frac{1}{c^2}\varphi = 0$$

und wir erhalten als Lösung von (6.2)

$$\varphi(t) = c\cosh(\frac{1}{c}(t-t_0)).$$

Diese Funktionen heißen auch *Kettenlinien.* Schließlich bestimmen wir noch die Konstanten c und t_0 für den Fall $\alpha = \beta$ und $a = -b$ wie folgt. Setzt man aus Symmetriegründen $t_0 = 0$ so haben wir die Gleichung

$$\frac{\cosh b/c}{b/c} = \frac{\alpha}{b}$$

zu lösen. Es *existiert* dann ein $c\ in\mathbb{R}$, so dass für $\alpha/b = c$ genau eine Lösung dieser Gleichung exisitiert. Zusammengefasst haben wir bewiesen, dass das Problem der Rotationsminimalfläche für dieses c und $\alpha/b = c$ höchstens eine Lösung besitzt.

VIII Umkehrabbildungen und Implizite Funktionen

In diesem Abschnitt untersuchen wir vier Themenkomplexe, die Frage wann eine stetig differenzierbare Funktion eine ebensolche Umkehrfunktion besitzt, das Auflösen von Gleichungen in $\mathbb{R}^n$ und damit zusammenhängend den Satz über implizite Funktionen, Untermannigfaltigkeiten des $\mathbb{R}^n$ und damit zusammenhängend die Bestimmung von Extremwerten unter Nebenbedingungen.

Die Beantwortung der beiden ersten Fragen erweist sich als deutlich schwieriger als in der eindimensionalen Situation, da der Zwischenwertsatz aus der Analysis I kein n-dimensionales Analogon besitzt. Unser zentrales Hilfsmittel zur Herleitung des Satzes über die lokale Umkehrbarkeit einer stetig differenzierbaren Abbildung ist der Banachsche Fixpunktsatz. Wir zeigen mit seiner Hilfe zunächst, dass unter gewissen Bedingungen lokal eine stetige Umkehrung existiert. Die Kettenregel impliziert dann, dass diese wiederum stetig differenzierbar ist.

Der Satz über implizite Funktionen beschäftigt sich dann mit der Frage unter welchen Bedingungen eine Gleichung $f(x, y) = 0$ in der Nähe einer Nullstelle von f eine differenzierbare Auflösung $y = g(x)$ besitzt. Dieser Satz führt aus geometrischer Sicht zum Begriff der Untermannigfaltigkeit in $\mathbb{R}^n$; dies sind diejenigen Teilmengen die lokal wie offene Teilmengen eines $\mathbb{R}^d$ aussehen. Die zunächst geometrisch motivierte Einführung der Begriffe des Tangential- bzw. Normalenraums an eine Untermannigfaltigkeit erlauben dann auch einen eleganten Beweis der sogenannten Multiplikatorenregel von Lagrange, einer notwendigen Bedingung für die Existenz von Extremwerten unter Nebenbedingungen.

1 Der Satz über die Umkehrabbildung

Wir beginnen diesen Abschnitt mit der Definition des Begriffs des Diffeomorphismus.

1.1 Definition. Eine bijektive und stetig differenzierbare Funktion $f : U \to V$ zwischen offenen Mengen $U \subset \mathbb{R}^n$ und $V \subset \mathbb{R}^m$ heißt *Diffeomorphismus*, falls die Umkehrabbildung $f^{-1} : V \to U$ ebenfalls stetig differenzierbar ist.

1.2 Bemerkungen.
a) Ist $I \subset \mathbb{R}$ ein offenes Intervall und $f: I \to J$ eine stetig differenzierbare Funktion mit $f'(x) \neq 0$ für alle $x \in I$, so ist f streng monoton, die Umkehrfunktion existiert im Intervall $J = f(I)$ und sie ist nach dem Satz über die Umkehrfunktion IV.1.8 aus Analysis I auch differenzierbar.
b) Ein zu a) analoges Resultat gilt nicht für Funktionen in $\mathbb{R}^n$ für $n \geq 2$! Als Gegenbeispiel betrachten wir die Polarkoordinatenabbildung

$$f: (0,\infty) \times \mathbb{R} \to \mathbb{R}^2 \backslash \{(0,0)\}, \quad f(r,\varphi) := (r\cos\varphi, r\sin\varphi).$$

Diese ist surjektiv und stetig differenzierbar, $J_f(r,\varphi)$ ist invertierbar für alle $(r,\varphi) \in (0,\infty) \times \mathbb{R}$, aber f ist nicht injektiv. Also ist f auch nicht invertierbar.
c) Es seien $U \subset \mathbb{R}^n$, $V \subset \mathbb{R}^m$ und $f: U \to V$ ein Diffeomorphismus mit der Umkehrabbildung $g := f^{-1}: V \to U$. Nach der Kettenregel gilt

$$\begin{aligned} g'(f(x)) \cdot f'(x) &= (g \circ f)'(x) = (\mathrm{id}_U)'(x) = \mathrm{id}_{\mathbb{R}^n}, \quad x \in U, \\ f'(g(y)) \cdot g'(y) &= (f \circ g)'(y) = (\mathrm{id}_V)'(y) = \mathrm{id}_{\mathbb{R}^m}, \quad y \in V, \end{aligned}$$

und aus der Linearen Algebra folgt, dass $n = m$ gilt und dass für $y = f(x)$ die Abbildungen $g'(y)^{-1} \in \mathcal{L}(\mathbb{R}^n)$ und $f'(x) \in \mathcal{L}(\mathbb{R}^n)$ zueinander inverse Isomorphismen sind.
d) Die Aussagen b) und c) implizieren, dass $Df(x)$ für alle $x \in U$ invertierbar ist, falls $f: U \to V$ ein Diffeomorphismus ist. Die Umkehrung gilt nur, falls man f geeignet einschränkt.

1.3 Theorem. (Satz über die Umkehrabbildung). *Es seien $U \subset \mathbb{R}^n$ eine offene Menge und $f: U \to \mathbb{R}^n$ eine stetig differenzierbare Funktion, für welche $Df(a) \in \mathcal{L}(\mathbb{R}^n)$ für $a \in U$ invertierbar ist. Dann existiert eine offene Umgebung V von $b := f(a)$ und eine Umgebung $\tilde{U} \subset U$ von a, derart dass die Einschränkung $\tilde{f}: \tilde{U} \to V$ von f auf $\tilde{U}$ ein Diffeomorphismus ist. Für die Umkehrabbildung gilt außerdem*

$$D\tilde{f}^{-1}(b) = (Df(a))^{-1}.$$

Für den Beweis benötigen wir die folgenden beiden Hilfssätze.

1.4 Lemma. (Schrankensatz). *Es seien $U \subset \mathbb{R}^n$ offen, $f: U \to \mathbb{R}^m$ eine differenzierbare Funktion und $a, x \in U$ mit $\overline{ax} \in U$. Gilt*

$$\sup_{t\in[0,1]} \|Df(a + t(x-a))\| =: L < \infty,$$

so ist $\|f(x) - f(a)\| \leq L\|x-a\|$.

Für den Beweis verweisen wir auf die Übungen.

1.5 Lemma. (Stetigkeit der Inversion). *Die Menge* $GL_n(\mathbb{R}) := \{A \in M_n(\mathbb{R}) : A$ *in-vertierbar* $\}$ *ist offen in* $\mathbb{R}^{n^2}$ *und die Abbildung*

$$\mathrm{Inv} : GL_n(\mathbb{R}) \to M_n(\mathbb{R}),\ A \mapsto A^{-1}$$

ist stetig.

Beweis. Da $GL_n(\mathbb{R}) = \det^{-1}(\{x \in \mathbb{R} : x \neq 0\})$ gilt, ist die Menge $GL_n(\mathbb{R})$ als Urbild der Menge $\{x \in \mathbb{R} : x \neq 0\}$ unter der stetigen Funktion $\det : M_n(\mathbb{R}) \to \mathbb{R}$ offen. Für die Stetigkeit der Inversenabbildung verweisen wir auf die Übungen.

□

Beweis von Theorem 1.3. Wir unterteilen den relativ umfangreichen Beweis in sechs Teilschritte und beginnen mit der folgenden Vorbemerkung.
Betrachtet man anstelle von f die Funktion $Df(a)^{-1} \circ f : U \to \mathbb{R}^n$ so folgt, dass wir oBdA

$$Df(a) = \mathrm{Id}_{\mathbb{R}_n} \tag{1.1}$$

wählen können. Ferner dürfen wir oBdA ebenfalls

$$a = 0 \quad \text{und} \quad f(a) = 0$$

wählen, indem wir die Funktion $x \mapsto f(x+a) - f(a)$ für $x \in \{z \in \mathbb{R}^n : z + a \in U\}$ betrachten.
Schritt 1: Ziel ist es die Gleichung $y = f(x)$ für „kleine“ $y \in \mathbb{R}^n$ nach x aufzulösen. Hierzu setzen wir für $y \in \mathbb{R}^n$ und $x \in U$

$$\varphi_y(x) := y + x - f(x).$$

Es gilt dann $y = f(x)$ genau dann, wenn $\varphi_y(x) = x$ ist, d.h. genau dann, wenn x ein Fixpunkt der Abbildung φ_y ist.
Schritt 2: Wir wenden auf die obige Gleichung den Banachschen Fixpunktsatz an. Zunächst ist $\varphi_0 : U \to \mathbb{R}^n$ stetig differenzierbar und die Skalierung (1.1) impliziert, dass

$$D\varphi_0(0) = \mathrm{Id}_{\mathbb{R}^n} - \mathrm{Id}_{\mathbb{R}^n} = 0$$

gilt. Daher existiert ein $r > 0$ so, dass $\overline{U}_{2r}(0) = \{z \in \mathbb{R}^n : |z| \leq 2r\} \subset U$ und

$$\|D\varphi_0(x) - \underbrace{D\varphi_0(0)}_{=0}\| = \|D\varphi_0(x)\| \leq \frac{1}{2}, \quad x \in \overline{U}_{2r}(0), \tag{1.2}$$

gilt. Da $D\varphi_y = D\varphi_0$ ist, folgt aus dem Schrankensatz Lemma 1.4

$$\|\varphi_y(x_1) - \varphi_y(x_0)\| \leq \frac{1}{2}\|x_1 - x_0\|, \quad x_0,\ x_1 \in \overline{U}_{2r}(0) \tag{1.3}$$

und somit

$$\|\varphi_y(x)\| \leq \|\varphi_y(x) - \varphi_y(0)\| + \|\underbrace{\varphi_y(0)}_{=y}\| \leq \frac{1}{2}\|x\| + \|y\| < 2r, \tag{1.4}$$

falls $\|x\| \leq 2r$ und $\|y\| < r$ ist. Mit anderen Worten bedeutet dies, dass für $\|y\| < r$ die Abbildung

$$\varphi_y : \overline{U}_{2r}(0) \to \overline{U}_{2r}(0)$$

eine strikte Kontraktion ist. Da $\overline{U}_{2r}(0)$ als abgeschlossene Teilmenge des $\mathbb{R}^n$ vollständig ist, folgt aus dem Banachschen Fixpunktsatz, dass für alle $y \in \mathbb{R}^n$ mit $\|y\| < r$ genau ein Fixpunkt $x \in \overline{U}_{2r}(0)$ von φ_y existiert, für welchen wegen (1.4) sogar $\|x\| < 2r$ gilt. Setzt man

$$V := U_r(0) \text{ und } \tilde{U} := f^{-1}(V) \cap U_{2r}(0),$$

so bedeutet dies, dass die Einschränkung $\tilde{f}$ von f auf $\tilde{U}$ bijektiv ist.
Schritt 3: Wir zeigen als nächstes, dass $g := (\tilde{f})^{-1} : V \to \tilde{U}$ stetig ist. Für jedes $x \in \tilde{U}$ gilt $x = \varphi_0(x) + f(x)$. Damit und wegen (1.3) gilt für alle $y_0, y_1 \in V$ die Abschätzung

$$\|g(y_1) - g(y_0)\| \leq \underbrace{\|\varphi_0(g(y_1)) - \varphi_0(g(y_0))\|}_{\leq \frac{1}{2}\|g(y_1) - g(y_0)\|} + \|f(g(y_1)) - f(g(y_0))\|$$

und somit

$$\|g(y_1) - g(y_0)\| \leq 2\|f(g(y_1)) - f(g(y_0))\| = 2\|y_1 - y_0\|. \tag{1.5}$$

Dies bedeutet, dass g sogar Lipschitz-stetig ist.
Schritt 4: Wir zeigen, dass $Df(x) \in \mathcal{L}(\mathbb{R}^n)$ für alle $x \in \tilde{U}$ invertierbar ist. Zunächst gilt

$$f(x) = x - \varphi_0(x), \quad \text{und} \quad Df(x) = \mathrm{Id}_{\mathbb{R}^n} - D\varphi_0(x), \quad x \in \tilde{U}.$$

Gilt $Df(x)v = 0$ für ein $v \in \mathbb{R}^n$, so folgt $v = [D\varphi_0(x)]v$ und wegen (1.2) gilt

$$\|v\| \leq \|D\varphi_0(x)\| \cdot \|v\| \leq \frac{1}{2}\|v\|,$$

also $v = 0$. Daher ist $Df(x)$ injektiv und die Dimensionsformel aus der Linearen Algebra impliziert, dass $Df(x)$ auch surjektiv ist. Also ist $Df(x)$ invertierbar.
Schritt 5: Wir zeigen, dass g differenzierbar ist. Hierzu sei $y_0 \in V$ und $k \in \mathbb{R}^n$ mit $y_0 + k \in V$. Setzt man $x_0 := g(y_0)$ und $h := g(y_0 + k) - g(y_0)$, so ist $k = f(x_0 + h) - f(x_0)$ und

$$g(y_0 + k) - g(y_0) - [Df(x_0)]^{-1}k = h - [Df(x_0)]^{-1}(f(x_0 + h) - f(x_0))$$

Da f in x_0 differenzierbar ist, gilt

$$f(x_0 + h) - f(x_0) = Df(x_0)h + r(x_0 + h)$$

mit $\lim_{h\to 0} \frac{r(x_0+h)}{\|h\|} = 0$. Daher ist

$$g(y_0 + k) - g(y_0) = [Df(x_0)]^{-1}k - [Df(x_0)]^{-1}r(x_0 + h)$$

und es bleibt zu zeigen, dass

$$\lim_{k\to 0} \frac{[Df(x_0)]^{-1}r(x_0 + h)}{\|k\|} = 0 \tag{1.6}$$

gilt. Aufgrund von Gleichung (1.5) gilt

$$\|h\| = \|(x_0 + h) - x_0\| \leq 2\| \underbrace{f(x_0 + h) - f(x_0)}_{=k} \| = \|2k\|,$$

und für $k \to 0$, also auch $h \to 0$, folgt

$$\frac{\|r(x_0 + h)\|}{\|k\|} \leq \frac{2}{\|h\|}\|r(x_0 + h)\| \overset{h\to 0}{\longrightarrow} 0,$$

da f differenzierbar ist. Da $[Df(x_0)]^{-1} \in \mathcal{L}(\mathbb{R}^n)$ stetig ist, folgt schließlich (1.6).
Schritt 6: Anwenden der Kettenregel analog zu Bemerkung 1.2 c) liefert

$$D_g(y) = [Df(g(y))]^{-1}$$

und Lemma 1.5 impliziert, dass $D_g : V \to \mathcal{L}(\mathbb{R}^n)$ stetig ist. Daher ist $\tilde{f} : \tilde{U} \to V$ ein Diffeomorphismus und der Satz ist vollständig bewiesen.

□

Der obige Satz über die lokale Umkehrbarkeit hat zahlreiche Konsequenzen. Als unmittelbare Folgerung notieren wir die Sätze über offene Abbildungen und die Diffeomorphie.

1.6 Korollar. (Satz von der offenen Abbildung). *Es sei $U \subset \mathbb{R}^n$ offen und $f : U \to \mathbb{R}^n$ eine stetig differenzierbare Funktion, derart dass $Df(x) \in \mathcal{L}(\mathbb{R}^n)$ für alle $x \in U$ invertierbar ist. Dann ist $f(U)$ offen in $\mathbb{R}^n$.*

Beweis. Nach dem Umkehrsatz existiert zu jedem $x \in U$ eine offene Umgebung $U_x \subset U$ von x so, dass $f(U_x) \subset \mathbb{R}^n$ offen ist. Wegen

$$f(U) = \bigcup_{x\in U} f(U_x)$$

und da beliebige Vereinigungen offener Mengen wiederum offen sind, ist auch $f(U)$ offen.

$\square$

Die obige Aussage bedeutet, dass $f(O)$ offen ist für alle offenen Mengen $O \subset U$. Abbildungen mit dieser Eigenschaft heißen auch *offene Abbildungen*.

1.7 Korollar. *Ist in der Situation des Satzes von der offenen Abbildung die Funktion f zusätzlich injektiv, so ist f ein Diffeomorphismus von U auf die offene Menge $f(U) \subset \mathbb{R}^n$.*

Beweis. Die Umkehrabbildung $g : f(U) \to U$ ist stetig, da für jede offene Menge $O \subset U$ das Urbild $g^{-1}(O) = f(O)$ nach dem Satz über die offene Abbildung offen ist. Nach dem Satz über die Umkehrabbildung ist f sogar ein Diffeomorphismus.

$\square$

1.8 Bemerkung. Verwendet man den Satz über die Umkehrabbildung um nichtlineare Gleichungssysteme zu lösen, so erhält man folgende Aussage:
Ist $\det Df(x_0) \neq 0$ für ein $x_0 \in U$, so existieren Umgebungen U von x_0 und V von $f(x_0)$ derart, dass das Gleichungssystem

$$\begin{array}{rcl} f_1(x_1, \dots, x_n) & = & y_1 \\ \vdots & & \vdots \\ f_n(x_1, \dots, x_n) & = & y_n \end{array}$$

für jeden Wert $(y_1, \dots, y_n) \in V$ genau eine Lösung

$$x_1 = x_1(y_1, \dots, y_n), \dots, x_n = x_n(y_1, \dots, y_n)$$

in U besitzt. Ferner besitzen die Funktionen $x_1, \dots, x_n$ dieselbe Regularität wie $f_1, \dots, f_n$.

1.9 Beispiele.
a) *Ebene Polarkoordinaten*
Jeder Punkt $(x, y) \in \mathbb{R}^2$ besitzt eine Darstellung durch Polarkoordinaten, d.h.

$$x = r\cos\varphi,\ y = r\sin\varphi \quad \text{mit } r = (x^2 + y^2)^{1/2}.$$

Für die Abbildung $f : (r, \varphi) \mapsto (x, y)$ gilt

$$\det Df(r, \varphi) = \begin{vmatrix} \cos\varphi & -r\sin\varphi \\ \sin\varphi & r\cos\varphi \end{vmatrix} = r \neq 0, \quad (x, y) \neq (0, 0).$$

Zu jedem Punkt $(x, y) \neq (0, 0)$ existieren unendlich viele Urbilder $(r, \varphi + 2k\pi)$; in Übereinstimmung mit dem Umkehrsatz gibt es jedoch in einer Umgebung von $(x_0, y_0) = (r_0 \cos\varphi_0, r_0 \sin\varphi_0) \neq (0, 0)$ eine unendlich oft differenzierbare Umkehrfunktion gegeben durch

$$r = (x^2 + y^2)^{1/2} \quad \text{und } \varphi = \arctan\frac{y}{x}, \quad x_0 \neq 0$$

wobei jener Funktionszweig des Arcustangens zu wählen ist, für welchen (x_0, y_0) den Wert φ_0 ergibt. Gilt $x_0 = 0$, so wählen wir φ als $\varphi = \operatorname{arccot}\frac{x}{y}$.
Die Polarkoordinatendarstellung bildet also den Streifen

$$S = \{(r, \varphi) \in \mathbb{R}^2 : r > 0, |\varphi| < \pi\}$$

diffeomorph auf die längs der negativen reellen Achse geschlitzten Ebene $\mathbb{R}^2_- := \mathbb{R}^2 \setminus \{(x, y) : x \leq 0, y = 0\}$ ab.
b) *Der komplexe Logarithmus*
Die komplexe Zahl $z \in \mathbb{C}$ besitze die Darstellung $z = x + iy$. Beschränkt man z auf den Streifen $S = \mathbb{R} \times (-\pi, \pi]$ und setzt man $w := e^z$, so folgt $|w| = e^x$ und $\arg w = y$. Dies bedeutet, dass die Exponentialfunktion den Streifen S bijektiv auf $\mathbb{C}\setminus\{0\}$ abbildet. Ihre Umkehrabbildung ist durch $(x, y) \mapsto (\log|w|, \arg w)$ mit $y \in (-\pi, \pi]$ gegeben. Die Periodizität der Exponentialfunktion, d.h. $e^z = e^{z+2\pi i}$, impliziert, dass auch jeder um $2\pi ki$ verschobene Streifen $S_k = 2\pi ki + S$ mit $k \in \mathbb{Z}$ bijektiv auf $\mathbb{C}\setminus\{0\}$ abgebildet wird. Die obige Umkehrformel bleibt gültig mit dem Zusatz, dass jetzt $(2k-1)\pi < y \leq (2k+1)\pi$ gefordert werden muss.
Bezeichnet man für gegebenes $w \neq 0$ jede die Gleichung $e^z = w$ erfüllende komplexe Zahl z als Logarithmus von w, so sehen wir dass es in jedem Streifen S_k genau einen Logarithmus von w gibt. Die verschiedenen Logarithmen unterscheiden sich lediglich um Vielfache von $2\pi i$ und sie sind durch

$$\log w = \log|w| + i \arg w$$

gegeben. Wird das Argument durch $\arg w \in (-\pi, \pi)$ eingeschränkt, so spricht man vom *Hauptzweig* des Logarithmus. In der Sprache des Umkehrsatzes dreht es sich bei $w = e^z$ um die Funktion $f = (f_1, f_2)$ mit

$$f_1(x, y) = e^x \cos y, \quad f_2(x, y) = e^x \sin y$$

und

$$\det\, Df(x, y) = \begin{vmatrix} e^x \cos y & -e^x \sin y \\ e^x \sin y & e^x \cos y \end{vmatrix} = e^{2x} \neq 0.$$

Beschränkt man (x, y) auf den Streifen $S = \mathbb{R} \times (-\pi, \pi)$, so beschreibt $f = (f_1, f_2)$ einen Diffeomorphismus mit dem Bildbereich $\mathbb{R}^2_-$. Die Umkehrabbildung ist dann der Hauptzweig des Logarithmus gegeben durch

$$x = \frac{1}{2}\log(f_1^2 + f_2^2), \quad y = \arg(f_1, f_2) \in (-\pi, \pi).$$

2 Der Satz über implizite Funktionen

Im vorangegangenen Abschnitt beschäftigten wir uns mit der Lösbarkeit von nichtlinearen Gleichungssystemen, bei welchen die Anzahl der Gleichungen mit der Anzahl der Variablen übereinstimmte. Im Folgenden betrachten wir nun die Lösbarkeit von solchen Systemen, bei denen mehr Variablen als Gleichungen vorhanden sind. Genauer gesagt betrachten wir m Gleichungen für $m+k$ Unbekannte, d.h.

$$\begin{array}{rcl} f_1(x_1,\dots,x_k,y_1,\dots,y_m) & = & 0 \\ \vdots & & \vdots \\ f_m(x_1,\dots,x_k,y_1,\dots,y_m) & = & 0 \end{array}$$

und untersuchen die Frage unter welchen Bedingungen sich $y_1,\dots,y_m$ durch $x_1,\dots,x_n$ ausdrücken lassen, oder anders formuliert, ob sich das obige Gleichungssystem nach y auflösen lässt.
Betrachtet man den Spezialfall von *linearen Gleichungssystemen* der Form

$$Ax+By=0$$

mit $A\in M_{mk}(\mathbb{R})$, $B\in M_m(\mathbb{R})$ und $x=(x_1,\dots,x_k)^T$ und $y=(y_1,\dots,y_m)^T$, so ist dies möglich, falls B invertierbar ist; es gilt dann

$$y=-B^{-1}Ax.$$

Im Folgenden wollen wir *lokal* ein analoges Resultat für stetig differenzierbare Funktionen $f=(f_1,\dots,f_m):U\subset\mathbb{R}^k\times\mathbb{R}^m\to\mathbb{R}^m$ herleiten. Hierzu definieren wir

$$\begin{array}{lll} a=(a_1,\dots,a_k), \quad b=(b_1,\dots,b_m), & (a,b) & :=(a_1,\dots,a_k,b_1,\dots,b_m), \\ x=(x_1,\dots,x_k), \quad y=(y_1,\cdots,y_m), & (x,y) & :=(x_1,\dots,x_k,y_1,\dots,y_m), \end{array}$$

$$f(a,b):=f(a_1,\dots,a_k,b_1,\dots,b_m), \quad f(x,y):=f(x_1,\dots,x_k,y_1,\dots,y_m),$$

$$D_xf(a,b):=\begin{pmatrix} \frac{\partial f_1}{\partial x_1}(a,b) & \dots & \frac{\partial f_1}{\partial x_k}(a,b) \\ \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1}(a,b) & \dots & \frac{\partial f_m}{\partial x_k}(a,b) \end{pmatrix}_{m\times k}$$

und

$$D_yf(a,b):=\begin{pmatrix} \frac{\partial f_1}{\partial y_1}(a,b) & \cdots & \frac{\partial f_1}{\partial y_m}(a,b) \\ \vdots & & \vdots \\ \frac{\partial f_m}{\partial y_1}(a,b) & \cdots & \frac{\partial f_m}{\partial y_m}(a,b) \end{pmatrix}_{m\times m}$$

Der folgende Satz über implizite Funktionen bildet den Hauptsatz dieses Abschnitts.

2.1 Theorem. (Satz über implizite Funktionen). *Es seien $U \subset \mathbb{R}^k \times \mathbb{R}^m$ eine offene Menge und $f : U \to \mathbb{R}^m$ eine stetig differenzierbare Funktion. Ferner sei $(a, b) \in U$ mit $f(a, b) = 0$ und*

$$\det(D_y f(a, b)) \neq 0.$$

Dann existiert eine offene Umgebung $W \subset \mathbb{R}^k$ von a und eine eindeutig bestimmte stetig differenzierbare Abbildung $\varphi : W \to \mathbb{R}^m$ mit $\varphi(a) = b$, $(x, \varphi(x)) \in U$ für alle $x \in W$ und

$$f(x, \varphi(x)) = 0 \quad \textit{für alle } x \in W.$$

Ferner gilt

$$D\varphi(x) = -[D_y f(x, \varphi(x))]^{-1} D_x f(x, \varphi(x)), \quad x \in W.$$

Bevor wir den Beweis beginnen, wollen wir uns zunächst die Aussage des Satzes im Spezialfall $m = 1 = k$ veranschaulichen. Hierzu sei $f : \mathbb{R}^2 \to \mathbb{R}$ in Umgebung von $(x_0, y_0) \in \mathbb{R}^2$ stetig differenzierbar und es gelte $f(x_0, y_0) = 0$, sowie

$$D_y f(x_0, y_0) \neq 0.$$

Nach dem Satz über implizite Funktionen existiert dann ein $\delta > 0$ und eine auf dem Intervall $I := (x_0 - \delta, x_0 + \delta)$ eindeutig bestimmte stetig differenzierbare Abbildung φ mit $\varphi(x_0) = y_0$ und

$$f(x, \varphi(x)) = 0 \quad \text{für alle } x \in I.$$

Beweis. Betrachte die Abbildung $F : U \subset \mathbb{R}^k \times \mathbb{R}^m \to \mathbb{R}^k \times \mathbb{R}^m$ definiert durch

$$F(x, y) := (x, f(x, y)).$$

Dann ist F stetig differenzierbar und es gilt

$$\text{(2.1)} \quad DF(x, y)(h, k) = (h, D_x f(x, y) \cdot h + D_y f(x, y) \cdot k), \quad (x, y) \in U, (h, k) \in \mathbb{R}^k \times \mathbb{R}^m.$$

Weiter ist $DF(a, b) \in \mathcal{L}(\mathbb{R}^k \times \mathbb{R}^m)$ invertierbar, denn (2.1) impliziert

$$DF(a, b)(h, k) = 0 \Longrightarrow h = 0 \Longrightarrow \underbrace{D_y f(a, b) k}_{\text{invertierbar n. Vor.}} = 0 \Longrightarrow k = 0.$$

Somit ist $DF(a, b)$ injektiv, also auch bijektiv und daher invertierbar.
Wir können also den Satz 1.3 über die Umkehrabbildung auf F in (a, b) anwenden und

erhalten eine offene Umgebung $\tilde{U} \subset U$ von (a, b) mit der Eigenschaft, dass $\tilde{F} := F|_{\tilde{U}} : \tilde{U} \to F(\tilde{U}) =: V$ eine stetig differenzierbare Umkehrabbildung $G : V \to \tilde{U}$ besitzt.
Da F in den ersten Koordinaten wie die Identität wirkt, gilt dies auch für G, d.h. es existiert eine stetig differenzierbare Abbildung $g : V \to \mathbb{R}^m$ mit

$$G(\xi, \eta) = (\xi, g(\xi, \eta)) \quad \text{für alle } (\xi, \eta) \in V.$$

Für $(x, y) \in \tilde{U}$ gilt somit

$$\begin{aligned} f(x,y) = 0 &\iff F(x,y) = (x,0) \iff (x,y) = G(x,0) = (x, g(x,0)) \\ &\iff y = g(x,0) \end{aligned} \tag{2.2}$$

und insbesondere $b = g(a, 0)$. Setzt man

$$W := \{x \in \mathbb{R}^k : \begin{pmatrix} x \\ 0_{\mathbb{R}^m} \end{pmatrix} \in V\},$$

so ist mit V auch W offen und

$$V \ni F(a, b) = (a, f(a, b)) = (a, 0),$$

d.h. W ist eine Umgebung von a. Wir definieren schließlich die Funktion

$$\varphi : W \to \mathbb{R}^m, \ \varphi(x) := g(x, 0),$$

welche wegen (2.2) die behauptete lokale Auflösung der Gleichung $f(x, y) = 0$ nach y in einer Umgebung des Punkts (a, b) liefert.
Die Ableitung der Funktion φ lässt sich via der Kettenregel wie folgt bestimmen: aus $f(x, \varphi(x)) = 0$ folgt

$$D_x f(x, \varphi(x)) \cdot \mathrm{Id}_{\mathbb{R}^k} + D_y f(x, \varphi(x)) \cdot D\varphi(x) = 0.$$

Für $(x, y) = (a, b)$ gilt wegen $\varphi(a) = b$

$$D\varphi(a) = -[D_y f(a, b)]^{-1} \cdot D_x f(a, b).$$

□

Wir bemerken an dieser Stelle noch, dass sich die Ableitung $D\varphi(a)$ von φ wie in Satz 2.1 angegeben *ohne* explizite Kenntnis von φ bestimmen lässt!

2.2 Beispiele.
a) *Höhenlinien*: Es sei $U \subset \mathbb{R}^2$ eine offene Menge, $f : U \to \mathbb{R}$ eine stetig differenzierbare Funktion und $c \in \mathbb{R}$. Dann bezeichnet

$$N_f(c) = \{(x, y) \in U : f(x, y) = c\}$$

die *Niveaumenge* von f. Die Niveaumenge wird desöfteren auch als *Höhenlinie* bezeichnet, wobei im Allgemeinen die Niveaumenge keine „Linie“ sein muss. Der Satz über implizite Funktionen im Fall $m = k = 1$ angewandt auf die Funktion

$$f_c : U \to \mathbb{R}, \ f_c(x,y) = f(x,y) - c$$

liefert folgendes Resultat. Ist $(a,b) \in U$ mit $f(a,b) = c$ und gilt

$$\operatorname{grad} f_c(a,b) = (f_x(a,b), f_y(a,b)) \neq (0,0), \tag{2.3}$$

so kann man die Gleichung $f(x,y) = c$

1. in einer Umgebung von a nach y auflösen, falls $f_y(a,b) \neq 0$ gilt und

2. in Umgebung von b nach x auflösen, falls $f_x(a,b) \neq 0$ gilt.

Mit anderen Worten bedeutet dies, dass Niveaumengen durch Punkte (a,b), die (2.3) lokal erfüllen, sich als stetig differenzierbare Abbildungen der Form $x \mapsto (x, \varphi(x))$ im Fall 1. bzw. $y \mapsto (\psi(y), y)$ im Fall 2. darstellen lassen. Die Niveaumenge wird in diesem Fall tatsächlich durch *Höhenlinien* beschrieben.
b) *Gleichungssysteme*
Für $k = 1$ und $m = 2$ betrachten wir das Gleichungsystem

$$\begin{array}{lll} f_1(x,y_1,y_2) & = x^3 + y_1^3 + y_2^3 - 7 & = 0 \\ f_2(x,y_1,y_2) & = xy_1 + y_1y_2 + xy_2 + 2 & = 0 \end{array}$$

mit $f_1(2,-1,0) = f_2(2,-1,0) = 0$. Für $f = (f_1, f_2)$ gilt

$$D_y f(2,-1,0) = \begin{pmatrix} 3y_1^2 & 3y_2^2 \\ x + y_2 & x + y_1 \end{pmatrix}_{|(2,-1,0)} = \begin{pmatrix} 3 & o \\ 2 & 1 \end{pmatrix}.$$

Da diese Matrix invertierbar ist, existieren in einer Umgebung W von $a = 2$ zwei stetig differenzierbare Funktionen $\varphi_1, \varphi_2 : W \to \mathbb{R}$ mit

$$(\varphi_1(2), \varphi_2(2)) = (-1, 0)$$

und

$$f_1(x, \varphi_1(x), \varphi_2(x)) = 0 \quad \text{sowie} \quad f_2(x, \varphi_1(x), \varphi_2(x)) = 0.$$

Für die Ableitungen von φ_1 und φ_2 gilt

$$\begin{pmatrix} \varphi_1'(2) \\ \varphi_2'(2) \end{pmatrix} = - \begin{pmatrix} 3 & 0 \\ 2 & 1 \end{pmatrix}^{-1} \begin{pmatrix} \frac{\partial f_1}{\partial x}(2,-1,0) \\ \frac{\partial f_2}{\partial x}(2,-1,0) \end{pmatrix} = \begin{pmatrix} -\frac{1}{3} & 0 \\ \frac{2}{3} & -1 \end{pmatrix} \begin{pmatrix} 3x^2 \\ y_1 + y_2 \end{pmatrix}_{|(2,-1,0)} = \begin{pmatrix} -4 \\ 9 \end{pmatrix}$$

3 Untermannigfaltigkeiten und Extremwerte unter Nebenbedingungen

Der obige Satz über implizite Funktionen führt in geometrischer Sicht zum Begriff der differenzierbaren Mannigfaltigkeit. Letzterer spielt in der modernen Mathematik eine wichtige Rolle und wir wollen hier erste Eigenschaften aufzeigen.

3.1 Definition. Eine Teilmenge M des $\mathbb{R}^n$ heißt d-dimensionale *differenzierbare Untermannigfaltigkeit* des $\mathbb{R}^n$, wenn es zu jedem $a \in M$ eine in $\mathbb{R}^n$ offene Umgebung U von a und einen Diffeomorphismus $\varphi : U \to V$ auf eine offene Teilmenge V des $\mathbb{R}^n$ gibt mit

$$\varphi(U \cap M) = V \cap (\mathbb{R}^d \times \{0\}).$$

Ein- bzw. zweidimensionale Untermannigfaltigkeiten des $\mathbb{R}^n$ heißen auch in $\mathbb{R}^n$ *eingebettete Kurven* bzw. *Flächen.* Differenzierbare Untermannigfaltigkeiten des $\mathbb{R}^n$ der Dimension $n-1$ heißen auch *Hyperflächen.* . Im Folgenden verstehen wir unter einer (Unter)Mannigfaltigkeit eine differenzierbare Untermannigfaltigkeit des $\mathbb{R}^n$.
Beispiele von Mannigfaltigkeiten lassen sich relativ einfach, wie der folgende Satz zeigt, durch Graphen von differenzierbaren Funktionen erhalten.

3.2 Satz. *Ist O offen in $\mathbb{R}^d$ und $f : O \to \mathbb{R}^n$ eine stetig differenzierbare Funktion, so ist graph f eine d-dimensionale Untermannigfaltigkeit des $\mathbb{R}^{d+n}$.*

Beweis. Setzt man $U := O \times \mathbb{R}^n$ und

$$\varphi : U \to \mathbb{R}^{d+n} = \mathbb{R}^d \times \mathbb{R}^n,\ (x,y) \mapsto (x, y - f(x)),$$

so ist φ stetig differenzierbar mit $\mathrm{Im}(\varphi) = U$. Ferner ist $\varphi : U \to U$ bijektiv mit $\varphi^{-1}(x,z) = (x, z + f(x))$, also ein Diffeomorphismus von U auf sich, und es gilt

$$\varphi(U \cap \text{graph } (f)) = O \times \{0\} = U \cap (\mathbb{R}^d \times \{0\}).$$

□

Der folgende Satz vom regulären Wert stellt eines der nützlichsten Mannigfaltigkeitskriterien dar. Wir erinnern zunächst an den folgenden Satz aus der Linearen Algebra: Ist $f : \mathbb{R}^n \to \mathbb{R}^m$ eine reguläre, d.h. surjektive und lineare Abbildung, so ist für jedes $c \in \mathbb{R}^m$ der Lösungsraum der Gleichung $f(x) = c$ ein affiner Unterraum des $\mathbb{R}^n$ der Dimension $n - m$. Im Folgenden verallgemeinern wir dieses Ergebnis auf den Fall einer Gleichung $f(x) = c$ für eine stetig differenzierbare Funktion f und einem sogenannten regulären Wert c.

3.3 Definition. Ein Punkt $x \in U \subset \mathbb{R}^d$ heißt *regulärer Punkt* der stetig differenzierbaren Funktion $f : U \to \mathbb{R}^n$, falls die Ableitung $Df(x) \in \mathcal{L}(\mathbb{R}^d, \mathbb{R}^n)$ surjektiv ist. Ferner heißt ein Punkt $y \in \mathbb{R}^n$ *regulärer Wert* von f, wenn alle $x \in f^{-1}(y)$ reguläre Punkte von f sind.

3.4 Bemerkungen.
a) Gilt $d < n$, so besitzt f keine regulären Punkte.
b) Gilt $d \geq n$, so ist $x \in U$ genau dann ein regulärer Punkt von f, wenn die Ableitung $Df(x)$ den Rang n besitzt.
c) Gilt $n = 1$, so ist $x \in U$ genau dann ein regulärer Punkt von f, wenn $\nabla f(x) \neq 0$ ist.
d) Ein Punkt $y \in \mathbb{R}^n$ ist genau dann ein regulärer Wert von f, wenn die Matrix $Df(x)$ in allen Punkten $x \in f^{-1}(y)$ den Rang n hat. Gilt $n = 1$, so bedeutet dies, dass $f'(x) \neq 0$ ist in allen solchen Punkten.
e) Ist $a \in U$ ein regulärer Punkt einer stetig differenzierbaren Funktion f, so existieren nach dem Satz über implizite Funktionen (nach einer möglichen orthogonalen Transformation des $\mathbb{R}^n$) n Variablen, so dass das Gleichungssystem

$$\begin{aligned} f_1(x_1, \ldots, x_d) &= 0 \\ \vdots \qquad & \ \ \vdots \\ f_n(x_1, \ldots, x_d) &= 0 \end{aligned}$$

in einer Umgebung von a eindeutig nach diesen Variablen als Funktion der übrigen $d - n$ Variablen aufgelöst werden kann.
f) Ist $0 \in \mathrm{Im}(f)$ ein regulärer Wert einer stetig differenzierbaren Funktion $f : U \to \mathbb{R}^n$, so existiert zu jedem $a \in f^{-1}(0)$ eine Umgebung V in $\mathbb{R}^d$, so dass sich $f^{-1}(0) \cap V$ als Graph einer stetig differenzierbaren Funktion in $d - n$ Variablen darstellen lässt.

Wir kommen nun zum angekündigten Satz vom regulären Wert.

3.5 Theorem. *Ist $U \subset \mathbb{R}^d$ eine offene Menge und c ein regulärer Wert einer stetig differenzierbaren Funktion $f : U \to \mathbb{R}^n$, so ist $f^{-1}(c)$ eine $(d-n)$-dimensionale Untermannigfaltigkeit des $\mathbb{R}^n$.*

Nach den obigen Vorbereitungen ist der Beweis einfach und folgt direkt aus Satz 3.2 und Bemerkung 3.4 e).

3.6 Beispiele.
a) Die euklidische $(n-1)-$*Sphäre*

$$S^{n-1} := \{x \in \mathbb{R}^n : \|x\|_2 = 1\}$$

ist eine $(n-1)$-dimensionale Untermannigfaltigkeit des $\mathbb{R}^n$. Um dies einzusehen, betrachten wir die stetig differenzierbare Funktion

$$f : \mathbb{R}^n \to \mathbb{R},\; x \mapsto \|x\|_2^2.$$

Da $\nabla f(x) = 2x$ für alle $x \in \mathbb{R}^n$ gilt, ist 1 ein regulärer Wert von f. Weiter, da $S = f^{-1}(1)$ folgt die Behauptung aus dem Satz vom regulären Wert.
b) Betrachte auf $U = \mathbb{R}^3 \backslash (\{0\} \times \{0\} \times \mathbb{R})$ die stetig differenzierbare Funktion

$$f : U \to \mathbb{R},\; f(x_1, x_2, x_3) := \left(\sqrt{x_1^2 + x_2^2} - 2\right)^2 + x_3^2.$$

Da 1 ein regulärer Wert von f ist, ist $T^2 := f^{-1}(1)$ eine zweidimensionale Untermannigfaltigkeit des $\mathbb{R}^3$; T^2 ist die zweidimensionale *Torusfläche*, welche durch Rotation des in $(x_1 - x_3)$-Ebene liegenden Kreises $(x_1-2)^2 + x_3^2 = 1$ um die x_3-Achse entsteht.
c) Ist A eine reelle, symmetrische $n \times n$-Matrix mit $\det A \neq 0$, so ist die *Quadrik*

$$Q := \{x \in \mathbb{R}^n : (x^T|Ax) = 1\}$$

eine $(n-1)$-dimensionale Untermannigfaltigkeit des $\mathbb{R}^n$. Für den Beweis wählen wir $f : \mathbb{R}^n \to \mathbb{R}$, $f(x) = (x^T|Ax)$ und notieren, dass $Q = f^{-1}(1)$ gilt. Da $Df(x) = 2x^T A \neq 0$ für alle $x \in \mathbb{R}^n$ mit $x \neq 0$ gilt, ist 1 ein regulärer Wert von f und die Behauptung folgt wiederum aus dem Satz vom regulären Wert.

Will man die Konzepte der Differentialrechnung auf Abbildungen zwischen Untermannigfaltigkeiten übertragen, so erweist es sich als sehr nützlich, lineare Strukturen wie den Tangential- und Normalenraum an Mannigfaltigkeiten im euklidischen $\mathbb{R}^n$ einzuführen.
Wir betrachten zunächst den Begriff des Tangentialraums.

3.7 Definition. Es sei M eine Teilmenge des $\mathbb{R}^n$. Ein Vektor $v \in \mathbb{R}^n$ heißt *Tangentialvektor an M im Punkt $a \in M$*, wenn es in M eine stetig differenzierbare Kurve $\gamma : (-\varepsilon, \varepsilon) \to M$, $\varepsilon > 0$, mit $\gamma(0) = a$ und $\gamma'(0) = v$ gibt. Die Menge aller Tangentialvektoren an M in a heißt *Tangentialkegel* von M in a und wird mit T_aM bezeichnet. Ist T_aM ein Vektorraum, so wird er auch Tangential*raum* genannt.

3.8 Satz. *Es sei M eine $d-$dimensionale Untermannigfaltigkeit des $\mathbb{R}^n$. Dann gilt für jedes $a \in M$:*
a) T_aM ist ein Vektorraum der Dimension d.
b) Gilt $M = f^{-1}(c)$ für eine stetig differenzierbare Funktion $f : U \to \mathbb{R}^m$ auf einer offenen Menge $U \subset \mathbb{R}^n$ zu einem regulären Wert c von f, so ist

$$T_aM = \text{kern } Df(a) = \{v \in \mathbb{R}^n : Df(a)v = 0\}.$$

Beweis. a) Die Behauptung ist leicht einzusehen für die d-dimensionale Untermannigfaltigkeit $M = V \cap (\mathbb{R}^d \times \{0\})$, wobei V offen in $\mathbb{R}^n$ ist. In diesem Fall ist

$$T_a(V \cap (\mathbb{R}^d \times \{0\}) = \mathbb{R}^d \times \{0\}.$$

Für den allgemeinen Fall, betrachten wir die Abbildung $\varphi : U \to V$ aus der Definition 3.1 der Untermannigfaltigkeit M, welche jeder Kurve $\gamma : (-\varepsilon, \varepsilon) \to M \cap U$ die Bildkurve $\tilde{\gamma} := \varphi \circ \gamma$ in $V \cap (\mathbb{R}^d \times \{0\})$ zuordnet. Jede Kurve in $V \cap (\mathbb{R}^d \times \{0\})$ ist eine solche Bildkurve und für zwei Kurven γ und $\tilde{\gamma}$ gilt

$$\gamma'(t) = [D\varphi(a)]^{-1}\tilde{\gamma}'(t).$$

Daher folgt $T_a(M \cap U) = [D\varphi(a)]^{-1}(\mathbb{R}^d \times \{0\})$ und wegen $T_aM = T_a(M \cap U)$ folgt die Behauptung a).
b) Für $\gamma : (\varepsilon, \varepsilon) \to M$ gilt $f \circ \gamma = c$ und somit $Df(a)\gamma'(0) = 0$. Daher ist

$$T_aM \subset \text{kern } Df(a).$$

Da $Df(a) : \mathbb{R}^d \to \mathbb{R}^n$ nach Voraussetzung surjektiv ist, gilt nach der Dimensionsformel aus der Linearen Algebra dim kern $Df(a) = n - m =$ dim $M = d$. Daher ist T_aM kein echter Untervektorraum von kern $Df(a)$ und die Behauptung ist bewiesen.

□

Unter einem *Normalenvektor* einer Untermannigfaltigkeit $M \subset \mathbb{R}^n$ im Punkt $a \in M$ versteht man einen Vektor $v \in \mathbb{R}^n$ der senkrecht zum Tangentialraum T_aM steht; der *Normalenraum* N_aM ist das orthogonale Komplement zu T_aM, d.h.

$$N_aM := (T_aM)^\perp.$$

Der obige Satz über den Tangentialraum ergibt folgendes Korollar.

3.9 Korollar. *Es sei $M = f^{-1}(c)$ die Niveaumenge einer stetig differenzieraren Funktion $f = (f_1, \ldots, f_{n-d}) : U \to \mathbb{R}^{n-d}$ zum regulären Wert $c \in \mathbb{R}^{n-d}$. Dann bilden die Gradienten $\nabla f_1(a), \ldots, \nabla f_{n-d}(a)$ in $a \in M$ eine Basis des Normalenraums, d.h. es gilt*

$$N_aM = [\nabla f_1(a), \ldots, \nabla f_{n-d}(a)].$$

Beweis. Die Zeilen der Matrix $Df(a)$ werden durch den obigen Gradienten beschrieben. Nach Satz 3.8 liegt $v \in \mathbb{R}^n$ genau dann in T_aM, wenn $(\nabla f_i(a)|v) = 0$ für $i = 1, \ldots, n-d$ gilt; daher stehen die Vektoren $\nabla f_i(a)$ für $i = 1, \ldots, n-d$ senkrecht auf T_aM. Die Matrix $Df(a)$ hat nach Voraussetzung den Rang $n - d$ und somit sind die Vektoren $\nabla f_i(a)$ für $i = 1, \ldots, n - d$ linear unabhängig und bilden eine Basis für N_aM.

□

3.10 Beispiele.
a) Für die Sphäre S^{n-1} in $\mathbb{R}^n$ gilt $S^{n-1} = f^{-1}(1)$ mit $f : \mathbb{R}^n \to \mathbb{R},\ x \mapsto \|x\|_2$. Wegen $\nabla f(a) = 2a$ gilt

$$N_a S^{n-1} = \{\lambda a : \lambda \in \mathbb{R}\}$$

für $a \in S^{n-1}$.
b) Ein Normalenvektor an die Torusfläche T^2 beschrieben in Beispiel 3.6 b) im Punkt (a_1, a_2, a_3) ist gegeben durch $2(a - h)$ mit

$$h = (\frac{2a_1}{(a_1^2 + a_2^2)^{1/2}}, \frac{2a_2}{(a_1^2 + a_2^2)^{1/2}}, 0).$$

c) Ist $U \subset \mathbb{R}^n$ offen und $f : U \to \mathbb{R}$ eine stetig differenzierbare Funktion, so ist die *Einheitsnormale* $\nu(x)$, d.h. ein Normalenvektor der Länge 1, an die Mannigfaltigkeit $M = \text{graph} f$ im Punkt $a = (x, f(x))$ gegeben durch

$$\nu(x) = \frac{(-\nabla f(x), 1)}{(1 + |\nabla f(x)|^2)^{1/2}}.$$

In vielen Anwendungen ist nicht nur das bloße Extremum einer Funktion $f : \mathbb{R}^n \to \mathbb{R}$ gesucht, sondern es sind Extremwerte unter Nebenbedingungen zu bestimmen. Genauer seien $f : U \subset \mathbb{R}^n \to \mathbb{R}$ und $g = (g_1, \ldots, g_m) : U \subset \mathbb{R}^n \to \mathbb{R}^m$ gegebene Funktionen und $M := \{x \in U : g(x) = 0\}$. Gesucht sind Punkte $x_0 \in M$ mit

$$f(x_0) \le f(x) \quad \text{für alle } x \in M \cap U_{x_0}$$

oder

$$f(x_0) \ge f(x) \quad \text{für alle } x \in M \cap U_{x_0}$$

für eine Umgebung $U_{x_0} \subset U$ von x_0. Ein solcher Punkt heißt *lokales Extremum* von f unter der *Nebenbedingung* $g = 0$.
Die folgende Multiplikatorregel von Lagrange beschreibt ein notwendiges Kriterium für das Vorliegen von solchen Extremwerten unter Nebenbedingungen.

3.11 Satz. (Multiplikatorregel von Lagrange). *Es seien $U \subset \mathbb{R}^n$ offen, $m < n$, $f : U \to \mathbb{R}$ sowie $g = (g_1, \ldots, g_m) : U \to \mathbb{R}^m$ stetig differenzierbare Funktionen. Ferner sei 0 ein regulärer Wert von g und es gelte $M = g^{-1}(0) \ne \emptyset$. Besitzt f in x_0 unter der Nebenbedingung $g = 0$ ein lokales Extremum, so existieren $\lambda_1, \ldots, \lambda_m \in \mathbb{R}$ mit*

$$Df(x_0) = \sum_{j=1}^{m} \lambda_j Dg_j(x_0).$$

Beweis. Nach dem Satz vom regulären Wert ist M eine $(n-m)$-dimensionale Untermannigfaltigkeit des $\mathbb{R}^n$ und es gibt es zu $v \in T_{x_0}M$ eine stetig differenzierbare

Kurve $\gamma : (-\varepsilon, \varepsilon) \to M$ mit $\gamma(0) = x_0$ und $\gamma'(0) = v$. Die durch $F : (-\varepsilon, \varepsilon) \to \mathbb{R}$, $F(t) := f(\gamma(t))$ definierte Funktion hat in $t = 0$ ein lokales Extremum. Daher ist $F'(0) = 0$, welches $(\mathrm{grad} f(x_0)|v) = 0$ und somit $\mathrm{grad} f(x_0) \in N_{x_0}M$ bedeutet. Nach Satz 3.9 gibt es daher eindeutig bestimmte $\lambda_1, \ldots, \lambda_m$ mit der behaupteten Eigenschaft. $\square$

3.12 Bemerkungen.
a) Die Zahlen $\lambda_1, \ldots, \lambda_m$ heißen *Lagrange-Multiplikatoren.*
b) Im speziellen Fall $m = 1$ gilt: seien $U \subset \mathbb{R}^n$ offen, $f, g : U \to \mathbb{R}$ stetig differenzierbare Funktionen. Ist $Dg(x_0) \neq 0$ und besitzt f in x_0 ein lokales Extremum unter der Nebenbedingung $g = 0$, so existiert ein $\lambda \in \mathbb{R}$ mit

$$Df(x_0) = \lambda \cdot Dg(x_0).$$

Um dies einzusehen, nummerieren wir die Koordinaten so, dass

$$Dg(x_0) = \underbrace{(\xi_1, \ldots, \xi_{n-1}, \eta)}_{D_x g(x_0)} \qquad \text{mit } \eta \neq 0$$

gilt und wenden Satz 3.11 an.
c) Für $f, g : U \subset \mathbb{R}^2 \to \mathbb{R}$ und $c \in \mathbb{R}$ setze $N_c := f^{-1}(c)$ und $M := g^{-1}(0)$. Dann gilt $T_{x_0}N = T_{x_0}M$ und besitzt f in x_0 ein Minimum unter der Nebenbedingung $g = 0$ mit $f(x_0) = c$, so ergibt sich das folgende Bild.

In einem Schnittpunkt x von N_c und M kann kein Extremwert von $f_{|M}$ liegen, da durch Verschieben von x längs M sowohl kleinere als auch größere Funktionswerte von f auf M erreicht werden können.

3.13 Beispiel.
Es sei $A \in M_n(\mathbb{R})$ eine symmetrische Matrix,

$$f : \mathbb{R}^n \to \mathbb{R},\ f(x) := (x|Ax)$$

und $S := \{x \in \mathbb{R}^n : \|x\|_2 = 1\}$ die $(n-1)$-Sphäre. Da S kompakt und f stetig aud S ist, nimmt f ein Maximum auf S an. Wir bestimmen die Maximalstelle x_0 mit Hilfe der Multiplikatorenregel von Lagrange. Hierzu beobachten wir, dass $S = g^{-1}(0)$ gilt für

$$g : \mathbb{R}^n \to \mathbb{R},\ g(x) = (x|x) - 1.$$

Wie müssen daher f unter der Nebenbedingung g maximieren. Nach Beispiel 1.3 sind die Funktionen f und g stetig differenzierbar und es gilt

$$Df(x_0) = 2Ax_0,\ Dg(x_0) = 2x_0 \neq 0.$$

Nach der Lagrangeschen Multiplikatorenregel existiert ein $\lambda \in \mathbb{R}$ mit

$$Ax_0 = \lambda x_0,$$

d.h. x_0 ist ein Eigenvektor der Matrix A. Da $\|x_0\| = 1$ gilt, folgt

$$(x_0|Ax_0) = (x_0|\lambda x_0) = \lambda.$$

Wir haben damit den folgenden Satz aus der Linearen Algebra bewiesen.

3.14 Satz. *Ist $A \in M_n(\mathbb{R})$ eine symmetrische Matrix, so ist*

$$M := \max_{\|x\|=1}(x|Ax) \in \mathbb{R}$$

ein Eigenwert von A und jedes $x_0 \in \mathbb{R}$ mit $\|x_0\| = 1$ und $(x_0|Ax_0) = M$ ist ein zu M gehöriger Eigenvektor.

Als weitere Anwendung beweisen wir den den folgenden Spektralsatz für symmetrische Matrizen aus der Linearen Algebra.

3.15 Satz. (Spektralsatz für symmetrische Matrizen). *Ist $A \in M_n(\mathbb{R})$ eine symmetrische Matrix, so existieren*

$$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n, \quad \lambda_j \in \mathbb{R} \quad \textit{und} \quad x_1, \ldots, x_n \in S^{n-1}$$

mit $Ax_j = \lambda x_j$ für alle $j = 1, \ldots, n$. Ferner bilden die Vektoren $x_1, \ldots, x_n$ eine Orthonormalbasis des $\mathbb{R}^n$ und bezüglich dieser Basis besitzt A Diagonalgestalt.

Beweis. Nach dem Satz 3.14 existieren $x_1 \in S$ und $\lambda_1 \in \mathbb{R}$ mit $Ax_1 = \lambda_1 x_1$. Wir konstruieren einen weiteren Vektor $x_2 \in S$ wie folgt: definiert man $g := (g_0, g_1)$ durch

$$g_0(x) := \|x\|^2 - 1, \quad g_1(x) = 2(x_1|x),$$

so ist $g^{-1}(0) = S \cap \{x_1\}^\perp =: K$ eine kompakte Menge. Für die Funktion $f : \mathbb{R}^n \to \mathbb{R}, x \mapsto (x|Ax)$ existiert nach dem Satz vom Maximum III.3.9 ein $x_2 \in K$ mit $f(x) \leq f(x_2)$ für alle $x \in K$. Ferner ist $Dg(x)$ invertierbar für alle $x \in K$. Nach der Multiplikatorenregel von Lagrange, Satz 3.11, existieren daher $\mu_0, \mu_1 \in \mathbb{R}$ mit

$$\nabla f(x_2) = 2Ax_2 \overset{\text{Lagrange}}{=} \mu_0 \nabla g_0(x_2) + \mu_1 \nabla g_1(x_2) = \mu_0 2x_2 + \mu_1 2x_1. \tag{3.1}$$

Aufgrund von Satz 3.14 und da nach Konstruktion $(x_2|x_1) = 0$ gilt, folgt

$$(Ax_2|x_1) = (x_2|Ax_1) \overset{3.14}{=} (x_2|\lambda x_1) = \lambda(x_2|x_1) = 0.$$

Zusammen mit Gleichung (3.1) liefert dies

$$0 = (Ax_2|x_1) = \mu_0 \underbrace{(x_2|x_1)}_{=0} + \mu_1 \underbrace{(x_1|x_1)}_{=1} = \mu_1.$$

Gleichung (3.1) impliziert daher $Ax_2 = \mu_0 x_2$, welches bedeutet, dass μ_0 ein Eigenwert von A zum Eigenvektor x_2 ist. Der Wert von μ_0 berechnet sich schließlich zu

$$\mu_0 = \mu_0(x_2|x_2) = (Ax_2|x_2) = f(x_2).$$

Iterieren wir dieses Verfahren, so folgt die Behauptung.

$\square$

IX Kurven und Vektorfelder

Wir beginnen dieses Kapitel mit dem Begriff einer Kurve in $\mathbb{R}^n$ und lassen uns hierbei von einem Kurvenbegriff, der aus der Physik, genauer aus der Kinematik herrührt, leiten. Er beschreibt die Abstraktion der Bewegung eines Punktes im Raum, die durch die Angabe des Ortes $\gamma(t)$ zum Zeitpunkt t gegeben ist. Dieser Ansatz geht auf den französischen Mathematiker C. JORDAN (1838-1922) zurück. Kurven in diesem Sinn können sehr überraschende Eigenschaften besitzen. Zum Beispiel überdeckt die von G. PEANO (1858-1932) konstruierte Kurve vollständig ein Quadrat.
Eine der ersten Aufgaben der Kurventheorie ist die Rektifikation, also die Bestimmung der Bogenlänge einer Kurve. Eng verbunden mit dieser Problematik sind Funktionen von beschränkter Variation.
Im zweiten Paragraphen diskutieren wir dann Vektorfelder und Kurvenintegrale. Letztere sind Integrale, welche sicht nicht nur über Intervalle, sondern über stetig differenzierbare Bilder von Intervallen, also über Kurven, erstrecken. Diese Erweiterung des Integralbegriffs hat viele wichtige Konsequenzen. Zum Beispiel lassen sich hiermit jene Vektorfelder charakterisieren, welche als Gradienten von Potentialen erhalten werden können. Die grundlegenden Begriffe der Divergenz und Rotation eines Vektorfelds spielen in diesem Zusammenhang ebenfalls ein wichtige Rolle.

1 Kurven

Wir beginnen diesen Abschnitt mit der Definition des Kurvenbegriffs.

1.1 Definition. Eine stetige Abbildung $\gamma : I \to \mathbb{R}^n$ von einem Intervall $I \subset \mathbb{R}$ heißt *Kurve in* $\mathbb{R}^n$. Die Kurve heißt *(stetig) differenzierbar*, falls γ (stetig) differenzierbar ist. Das Bild $\gamma(I)$ wird auch die *Spur* von γ genannt.

Nach der obigen Definition ist eine Kurve also nicht nur eine Punktmenge in $\mathbb{R}^n$, sondern zu ihr gehört auch der durch γ übermittelte Ablaufplan des Durchlaufens der Spur.

1.2 Beispiele.
a) Ist $f : I \subset \mathbb{R} \to \mathbb{R}$ eine stetige Funktion, so ist

$$\gamma : I \to \mathbb{R}^2, \ \gamma(t) := (t, f(t))$$

eine Kurve. Die Spur von γ ist der Graph von f.
Ferner, ist f differenzierbar, so ist auch γ differenzierbar und es gilt $\gamma'(t) = (1, f'(t))$.
b) Für $r > 0$ beschreibt

$$\gamma : [0, 2\pi] \to \mathbb{R}^2,\ \gamma(t) := (r\cos t, r\sin t)$$

eine Kreisbewegung um $0 \in \mathbb{R}^2$ mit Radius r. Da γ differenzierbar ist mit $\gamma'(t) = (-r\sin t, r\cos t)$, ist der Betrag der Geschwindigkeit der Kreisbewegung gleich r.
c) Für $a \in \mathbb{R}^n$ und $v \in \mathbb{R}^m\backslash\{0\}$ beschreibt

$$\gamma : \mathbb{R} \to \mathbb{R}^m,\ \gamma(t) := a + vt$$

eine geradlinige Bewegung in Richtung v mit Geschwindigkeit $\gamma'(t) = v$.
d) Für $r > 0$ und $c \neq 0$ beschreibt

$$\gamma : \mathbb{R} \to \mathbb{R}^3,\ \gamma(t) := (r\cos t, r\sin t, ct)$$

eine *Schraubenlinie*. Die Spur liegt auf dem Zylinder $\{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 = r^2\}$ und $2\pi c$ heißt die *Ganghöhe*.

e) Die *Neilsche Parabel* ist gegeben durch

$$\gamma : [-1, 1] \to \mathbb{R}^2,\ \gamma(t) := (t^2, t^3).$$

Sie war nach dem Kreis historisch gesehen die erste Kurve, für welche man die Bogenlänge berechnen konnte.
f) Die *Zykloide* wird beschrieben durch die Kurve

$$\gamma : \mathbb{R} \to \mathbb{R}^2,\ \gamma(t) = (t - \sin t, 1 - \cos t).$$

Die Zykloide beschreibt die Bewegung eines Randpunktes der Einheitskreisscheibe, welche auf der x-Achse abrollt.

Die Zykloide ist auch aus anderer Sicht interessant. Sie ist die Lösung des sogenannten *Brachystochronenproblems*, des Variationsproblems für die Kurve mit der kürzesten

Laufzeit eines Massenpunkts zwischen zwei festen Punkten unter dem Einfluss der Schwerkraft. Bernoulli, Huygens und Leibniz fanden als Lösung dieses Problems die Zykloide.

1.3 Definition. Ist $\gamma : I \to \mathbb{R}^n$ eine differenzierbare Kurve, so heißt

$$\gamma'(t) = (\gamma_1'(t), \ldots, \gamma_n'(t)) \in \mathbb{R}^n$$

der *Tangentialvektor* der Kurve γ in t.

Der Vektor $\gamma'(t)$ lässt sich als die Geschwindigkeit der Kurve γ im Punkt t interpretieren und besitzt den Betrag

$$\|\gamma'(t)\| = (|\gamma_1'(t)|^2 + \ldots + |\gamma_n'(t)|^2)^{\frac{1}{2}}.$$

Im Folgenden beschäftigen wir uns mit der Bogenlänge einer gegebenen Kurve. Unsere Idee besteht darin, die Bogenlänge durch geeignete Polygonzüge zu approximieren. Wir betrachten daher eine Partition P des Intervalls $[a, b]$, d.h. $a = t_0 < t_1 < \ldots < t_k = b$ und definieren die Länge eines Polygonzugs mit den Ecken $\gamma(t_0), \ldots, \gamma(t_k)$ als

$$L_{P,\gamma} := \sum_{j=1}^{k} \|\gamma(t_j) - \gamma(t_{j-1})\|.$$

1.4 Definition. Eine Kurve $\gamma : [a, b] \to \mathbb{R}^n$ heißt *rektifizierbar* mit Länge L_γ, falls

$$L_\gamma := \sup_P L_{P,\gamma} < \infty$$

gilt. Die Zahl L_γ heißt *Bogenlänge* von γ.

1.5 Satz. *Es sei $\gamma : [a, b] \to \mathbb{R}^n$ eine stetig differenzierbare Kurve. Dann ist γ rektifizierbar und es gilt*

$$L_\gamma = \int_a^b \|\gamma'(t)\|_2 dt = \int_a^b (|\gamma_1'(t)|^2 + \ldots + |\gamma_n'(t)|^2)^{\frac{1}{2}} dt.$$

Insbesondere hat der Graph einer C^1-Funktion $f : [a, b] \to \mathbb{R}$ die Länge

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

Zum Beweis verwenden wir das Integral eines n-Tupels stetiger Funktionen. Setzt man

$$\int_a^b \gamma(t)dt := (\int_a^b \gamma_1(t)dt, \ldots, \int_a^b \gamma_n(t)dt),$$

so gilt

$$\| \int_a^b \gamma(t)dt\| \leq \int_a^b \|\gamma(t)\|dt,$$

denn

$$\begin{aligned} \| \int_a^b \gamma(t)dt\|^2 &= (\int_a^b \gamma(t)dt|v) \quad \text{mit} \quad v = \int_a^b \gamma(t)dt \in \mathbb{R}^n \\ &= \int_a^b (\gamma(t)|v)dt \leq \int_a^b \|\gamma(t)\|dt\|v\|. \end{aligned}$$

Beweis. Die Partition P sei gegeben durch $P : a = t_0 < t_1 < \ldots < t_k = b$. Der Fundamentalsatz der Differential- und Integralrechnung impliziert

$$\begin{aligned} (1.1) \qquad L_{P,\gamma} &= \sum_{j=1}^k \|\gamma(t_j) - \gamma(t_{j-1})\| = \sum_{j=1}^k \| \int_{t_{j-1}}^{t_j} \gamma'(t)dt\| \\ &\leq \sum_{j=1}^k \int_{t_{j-1}}^{t_j} \|\gamma'(t)\|dt = \int_a^b \|\gamma'(t)\|dt =: L. \end{aligned}$$

Daher ist γ rektifizierbar und es gilt $L_\gamma \leq L$.
In einem zweiten Schritt zeigen wir, dass $L_\gamma = L$ gilt. Hierzu genügt es zu zeigen, dass für alle $\varepsilon > 0$ eine Partition $P : a = t_0 < \ldots < t_k = b$ von $[a, b]$ existiert mit

$$L_{P,\gamma} \geq L - \varepsilon.$$

Da γ' auf dem kompakten Intervall $[a, b]$ gleichmäßig stetig ist, existiert ein n-Tupel $\varphi = (\varphi_1, \ldots, \varphi_n)$ von Treppenfunktionen mit

$$\|\gamma'(t) - \varphi(t)\| \leq \frac{\varepsilon}{2(b-a)}, \quad t \in [a, b].$$

Wähle nun eine Partition $P : a = t_0 < t_1 < \ldots < t_k = b$ von $[a, b]$, so dass jedes $\varphi_1, \ldots, \varphi_n$ auf den Intervallen (t_{j-1}, t_j) für $j = 1, \ldots, k$ konstant ist.
Für dieses P gilt die Behauptung, denn zunächst gilt für alle $j = 1, \ldots, k$

$$\begin{aligned} \| \int_{t_{j-1}}^{t_j} \gamma'(t)dt\| &\geq \| \int_{t_{j-1}}^{t_j} \varphi(t)dt\| - \| \int_{t_{j-1}}^{t_j} (\gamma'(t) - \varphi(t))dt\| \\ &\geq \int_{t_{j-1}}^{t_j} \|\varphi(t)\|dt - \int_{t_{j-1}}^{t_j} \|\gamma'(t) - \varphi(t)\|dt \\ &\geq \int_{t_{j-1}}^{t_j} \|\varphi(t)\|dt - \frac{\varepsilon}{2(b-a)}(t_j - t_{j-1}). \end{aligned}$$

Aufsummieren über $j = 1, \ldots, k$ ergibt mit (1.1)

$$\begin{aligned} L_{P,\gamma} &\geq \int_a^b \|\varphi(t)\| dt - \frac{\varepsilon}{2} \\ &\geq \int_a^b \|\gamma'(t)\| dt - \int_a^b \|\varphi(t) - \gamma'(t)\| dt - \frac{\varepsilon}{2} \\ &\geq \int_a^b \|\gamma'(t)\| dt - \frac{\varepsilon}{2} - \frac{\varepsilon}{2} \geq L - \varepsilon. \end{aligned}$$

□

1.6 Beispiele. a) Wir berechnen die Länge des *Zykloidenbogens*

$$\gamma : \mathbb{R} \to \mathbb{R}^2, \ \gamma(t) = (t - \sin t, 1 - \cos t)$$

wie folgt. Da γ differenzierbar ist mit Ableitung $\gamma'(t) = (1 - \cos t, \sin t)$, gilt

$$\|\gamma'(t)\|^2 = 1 - 2\cos t + \underbrace{\cos^2 t + \sin^2 t}_{=1} = 2 - 2\cos t = 4\sin^2(\frac{t}{2}).$$

Daher ist

$$L_{\gamma_{|[0,2\pi]}} = \int_0^{2\pi} 2|\sin(\frac{t}{2})| dt = -4\cos(\frac{t}{2})|_0^{2\pi} = 8.$$

b) Ein Beispiel einer nicht rektifizierbaren Kurve ist durch

$$\gamma : [0,1] \to \mathbb{R}, \quad t \mapsto t^2 \cos\frac{\pi}{t^2}$$

gegeben. Um dies einzusehen, definieren wir die Partition P von $[0,1]$ als $P : 0 = t_0, n^{-1/2}, (n-1)^{-1/2}, \ldots, 2^{-1/2}, 1$ und bemerken, dass $\gamma(n^{-1/2}) = \frac{(-1)^n}{n}$ gilt. Daher gilt $L_{P,\gamma} > 1 + 1/2 + \ldots + 1/n$ und es ist $L_\gamma = \infty$.

Der Begriff der rektifizierbaren Kurve ist sehr eng mit dem Konzept der Funktionen von beschränkter Variation verwandt. Genauer gesagt, definieren wir zu Funktionen $f : I = [a,b] \to \mathbb{R}$ sowie zu Zerlegungen P von I die *Variation* von f durch

$$\mathrm{var}_{P,f} := \sum_{j=1}^n |f(t_j) - f(t_{j-1})|.$$

Das Supremum über alle Zerlegungen, d.h.

$$V_a^b(f) := \sup_P \mathrm{var}_{P,f}$$

heißt *Totalvariation* von f über I. Gilt $V_a^b(f) < \infty$, so heißt f von *beschränkter Variation* auf I. Die Klasse aller dieser Funktionen bezeichnen wir mit $BV(I)$.
Ist $f : [a,b] \to \mathbb{R}$ eine Kurve, so gilt also $L_f = V_a^b(f)$. Im folgenden Lemma stellen wir grundlegende Eigenschaften von Funktionen $f \in BV[a,b]$ bereit.

1.7 Lemma. *Für Funktionen $f \in BV[a,b]$ gelten die folgenden Aussagen:*
a) $BV[a,b] \subset B[a,b]$ und es gilt $|f(a) - f(b)| \leq V_a^b(f)$
b) Der Raum $BV[a,b]$ ist ein Vektorraum und sogar eine Algebra und es gelten die Ungleichungen

$$\begin{aligned} V_a^b(\lambda f + \mu g) &\leq |\lambda| V_a^b(f) + |\mu| V_a^b(g), \quad \lambda, mu \in \mathbb{R}, f, g \in BV[a,b] \\ V_a^b(fg) &\leq \|f\|_\infty V_a^b(g) + \|g\|_\infty V_a^b(f). \end{aligned}$$

c) Für $a < c < b$ gilt

$$V_a^b(f) = V_a^c(f) + V_c^b(f).$$

d) Ist f monoton in $[a,b]$, so gilt $V_a^b(f) = |f(b) - f(a)|$.
e) Ist $f \in C^1([a,b])$, so gilt $V_a^b(f) = \int_a^b |f'(t)| dt$.

Den nicht sehr schwierigen Beweis überlassen wir dem Leser als Übungsaufgabe.

Im folgenden Satz charakterisieren wir Funktionen von beschränkter Variation als diejenigen Funktionen, welche sich als Differenz zweier monotoner Funktionen darstellen lassen .

1.8 Satz. *Eine Funktion $f : [a,b] \to \mathbb{R}$ ist genau dann von beschränkter Variation, wenn $f = g - h$ für zwei in $[a,b]$ monoton wachsende Funktionen g und h gilt.*

Beweis. Für $f \in BV[a,b]$ und $t \in [a,b]$ setze $g(t) := V_a^t(f)$. Für $a \leq c < d \leq b$ gilt nach Lemma 1.7 c)

$$0 \leq V_c^d(f) = V_a^d(f) - V_a^c(f) = g(d) - g(c),$$

welches bedeutet, dass g monoton wachsend ist. Weiter gilt nach Lemma 1.7 a)

$$f(d) - f(c) \leq V_c^d(f) = g(d) - g(c)$$

und somit gilt für $h := g - f$ die Ungleichung $h(c) \leq h(d)$. Also ist auch h monoton wachsend.
Die umgekehrte Richtung folgt direkt aus Lemma 1.7 d) und b).

□

Betrachtet man Kurven $f = (f_1, \ldots, f_n) : I \to \mathbb{R}^n$, so gilt

$$V_a^b(f_i) \leq L_f \leq V_a^b(f_1) + \ldots + V_a^b(f_n).$$

Somit erhalten wir den zuvor beschriebenen Zusammenhang zwischen Funktionen mit beschränkter Variation und rektifizierbaren Kurven.

1.9 Satz. *Eine Kurve $f : I \to \mathbb{R}^n$ ist genau dann rektifizierbar, wenn alle Komponentenfunktionen f_i von beschränkter Variation in I sind.*

2 Vektorfelder und Kurvenintegrale

Ist $U \subset \mathbb{R}^n$ eine offene Menge und $F = (f_1, \ldots, f_n) : U \to \mathbb{R}^n$ eine stetige Funktion, so wird diese Abbildung oft auch als *Vektorfeld* bezeichnet. Wir können uns Vektorfelder veranschaulichen, indem wir uns an jedem Punkt $x \in U$ den Vektor $F(x)$ „angeheftet" denken.

Wichtige Beispiele von Vektorfeldern in der Physik sind sogenannte Kraft- oder Geschwindigkeitsfelder. Im Folgenden wollen wir die folgenden Klassen von Vektorfeldern genauer untersuchen.

2.1 Beispiele.
a) *konstante Vektorfelder*. Diese sind durch $F(x) := y$ für ein $y \in \mathbb{R}^n$ definiert.
b) *Zentralfelder*. Ist $I = [a, b] \subset \mathbb{R}$ ein Intervall, $K := \{x \in \mathbb{R}^n : a < |x| < b\}$ eine Kugelschale in $\mathbb{R}^n$ und $g : I \to \mathbb{R}$ eine stetige Funktion, so heißt $F : K \to \mathbb{R}^n$ definiert durch

$$F(x) := g(\|x\|)x$$

Zentralfeld.
c) *Rotationsfelder*. Ist $I = [a, b]$ ein Intervall, $K := \{x \in \mathbb{R}^2 : a < |x| < b\}$ ein Kreisring in $\mathbb{R}^2$ und $g : I \to \mathbb{R}$ eine stetige Funktion, so heißt $F : K \to \mathbb{R}^n$ definiert durch

$$F(x) := g(\|x\|)(-x_2, x_1)^T$$

Rotationsfeld.
d) Ein Vektorfeld F heißt *Gradientenfeld*, wenn eine stetig differenzierbare Funktion $V : U \to \mathbb{R}$ existiert mit

$$\text{grad } V = F.$$

Von besonderer Wichtigkeit sind die Begriffe der Divergenz und Rotation von Vektorfeldern, welche wir wie folgt definieren.

2.2 Definition. Ist $U \subset \mathbb{R}^n$ offen und $F : U \to \mathbb{R}^n$ ein stetig differenzierbares Vektorfeld, so heißt die Funktion

$$\text{div} F(x) := \frac{\partial F_1}{\partial x_1} + \ldots + \frac{\partial F_n}{\partial x_n}$$

die *Divergenz* von F im Punkt $x \in U$.

Wir überprüfen leicht, dass für C^1-Vektorfelder F und G sowie für skalare Funktionen $h : \mathbb{R}^n \to \mathbb{R}$ die Beziehungen

$$\operatorname{div}(F+G) = \operatorname{div} F + \operatorname{div} G \quad \text{und} \quad \operatorname{div}(h \cdot F) = \nabla h \cdot F + h \cdot \operatorname{div} F$$

gelten.

2.3 Definition. Ist $U \subset \mathbb{R}^3$ und $F : U \to \mathbb{R}^3$ ein stetig partiell differenzierbares Vektorfeld, so definiert man $\operatorname{rot} F : U \to \mathbb{R}^3$, die *Rotation* von F, via

$$\operatorname{rot} F := \left(\frac{\partial F_3}{\partial x_2} - \frac{\partial F_2}{\partial x_3}, \frac{\partial F_1}{\partial x_3} - \frac{\partial F_3}{\partial x_1}, \frac{\partial F_2}{\partial x_1} - \frac{\partial F_1}{\partial x_2} \right).$$

Benützt man das Vektorprodukt aus der Linearen Algebra, so kann man formal auch kurz

$$\operatorname{rot} F = \nabla \times F \quad \text{und} \quad \operatorname{div} F = \nabla \cdot F$$

schreiben. Für stetig differenzierbare Vektorfelder F und G sowie für Funktionen $h \in C^1(\mathbb{R}^3, \mathbb{R})$ gelten dann die folgenden Beziehungen.

a) $\operatorname{rot}(h \cdot F) = h \cdot \operatorname{rot} F - F \times \nabla h$

b) $\operatorname{div}(F \times G) = G \cdot \operatorname{rot} F - F \cdot \operatorname{rot} G$

c) $\operatorname{rot}(\nabla h) = 0$

d) $\operatorname{div}(\operatorname{rot} F) = 0$

e) $\operatorname{rot}(\operatorname{rot} F) = \nabla(\operatorname{div} F) - \Delta F$.

Zur Motivation des Begriffs des *Kurvenintegrals* betrachten wir einen Körper, der eine gegebene glatte Kurve $\gamma : [a,b] \to \mathbb{R}^n$ unter dem Einfluss eines Kraftfelds F durchläuft. Die geleistete Arbeit ist dann gegeben durch das Integral $\int_a^b F(\gamma(t))\gamma'(t)dt$, da für die geleistete Arbeit nur diejenige Komponente der Kraft relevant ist, welche in Richtung der Bewegung zeigt. In Analogie hierzu definieren wir das Kurvenintegral wie folgt.

2.4 Definition. Es sei $\gamma : [a,b] \to \mathbb{R}^n$ eine stückweise stetig differenzierbare Kurve und f eine auf $\gamma([a,b])$ definierte stetige reellwertige Funktion. Dann heißt

$$\int_\gamma f(x)dx := \int_a^b f(\gamma(t))\gamma_j'(t)dt, \quad j = 1, \ldots, n,$$

das *Kurvenintegral von f bezüglich x längs γ*. Ist $f = (f_1, \ldots, f_n) : \gamma([a,b]) \to \mathbb{R}^n$ stetig, so definieren wir das *Kurvenintegral von f längs γ* durch

$$\int_\gamma f(x)dx := \int_\gamma f_1(x)dx_1 + \ldots + \int_\gamma f_n(x)dx_n.$$

2.5 Bemerkung. Das Kurvenintegral ist unabhängig von gewählten Parametrisierung der Kurve.

2.6 Beispiel. Will man die Funktion f gegeben durch $f(x_1,x_2) := (x_2,-x_1)^T$ entlang des Halbkreises, d.h. entlang der Kurve γ gegeben durch

$$\gamma : [-\frac{\pi}{2},\frac{\pi}{2}] \to \mathbb{R}, \quad \gamma(t) = (\cos t, \sin t)$$

integrieren, so gilt mit $x = (x_1,x_2)^T$

$$\begin{aligned}\int_\gamma f(x)dx &= \int_\gamma f_1(x)dx_1 + \int_\gamma f_2(x)dx_2 \\ &= \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \sin t(-\sin t)dt + \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} (-\cos t)\cos t dt = -\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \underbrace{(\sin^2 t + \cos^2 t)}_{=1} dt = -\pi.\end{aligned}$$

2.7 Definition. Es seien $U \subset \mathbb{R}^n$ offen und $F : U \to \mathbb{R}^n$ ein Vektorfeld.
a) Ist F ein Gradientenfeld, d.h. existiert ein Funktion $V \in C^1(U,\mathbb{R})$ mit $\nabla V = F$ in U, so heißt V *Potential* von F.
b) Es sei $\gamma : [a,b] \to U \subset \mathbb{R}^n$ eine stückweise stetig differenzierbare Kurve mit $\gamma(a) = x_0$ und $\gamma(b) = y_0$. Hat für je zwei Punkte $x,y \in U$ das Integral $\int_\gamma F dx$ längs jeder in U von x_0 nach y_0 laufenden Kurve denselben Wert, so heißt das Integral $\int_\gamma F(x)dx$ *wegunabhängig.*

Der folgende Satz charaktierisiert diejenigen Vektorfelder deren Kurvenintegral wegunabhängig ist.

2.8 Satz. *Ist $U \subset \mathbb{R}^n$ eine offene und konvexe Menge und $F : U \to \mathbb{R}^n$ ein stetig differenzierbares Vektorfeld, so ist das Integral $\int_\gamma F(x)dx$ genau dann wegunabhängig, wenn F ein Gradientenfeld ist. In diesem Fall gilt*

$$\int_\gamma F(x)dx = V(\gamma(b)) - V(\gamma(a)),$$

wobei V ein Potential von F und $\gamma : [a,b] \to U \subset \mathbb{R}^n$ eine Kurve bezeichnet.

Beweis. "$\Longleftarrow$": Ist V ein Potential von F, so gilt

$$\int_\gamma F(x)dx = \int_a^b \sum_{j=1}^n F_j(\gamma(t))\gamma_j'(t)dt = \int_a^b \frac{\mathrm{dV}(\gamma(t))}{\mathrm{dt}}dt = V(\gamma(b)) - V(\gamma(a)).$$

Die umgekehrte Richtung "$\Longrightarrow$" überlassen wir dem Leser als Übungsaufgabe.

$\square$

Wir wollen nun der Frage nachgehen wie man auf einfache Art und Weise überprüfen kann, ob F ein Gradientenfeld ist. Eine notwendige Bedingung hierür kann man einfach angeben. Ist nämlich F ein stetig differenzierbares Gradientenfeld, so gilt für das Potential $V \in C^2(U)$ und

$$\frac{\partial^2 V}{\partial x_i \partial x_j} = \frac{\partial^2 V}{\partial x_j \partial x_i},$$

d.h. es gilt

$$\frac{\partial f_i}{\partial x_j} = \frac{\partial f_j}{\partial x_i} \quad \text{in } \Omega$$

für alle $i, j = 1, \ldots, n$.

Wir wollen nun die umgekehrte Frage beantworten, nämlich ob die obige Integrabilitätsbedingung die Wegunabhängigkeit des Integrals impliziert. Es ist nicht überraschend, dass die Antwort auf diese Frage von der Geometrie des Gebietes abhängt. In diesem Zusammhang nennen wir $U \subset \mathbb{R}^n$ ein *Sterngebiet* bezüglich $x_0 \in U$, falls $U \subset \mathbb{R}^n$ offen ist und $\overline{x_0 x} \in U$ für alle $x \in U$ gilt. Insbesondere sind konvexe Mengen Sterngebiete.

2.9 Satz. *Ist $U \subset \mathbb{R}^n$ ein Sterngebiet bzgl. $x_0 \in U$ und $F = (f_1, \ldots, f_n) : U \to \mathbb{R}^n$ ein stetig differenzieres Vektorfeld, welches der Integrabilitätsbedingung*

$$\frac{\partial f_i}{\partial x_j} = \frac{\partial f_j}{\partial x_i}, \quad i, j = 1, \ldots, n,$$

genügt, so ist F ein Gradientenfeld.

Beweis. Es sei oBdA $x_0 = 0$. Setzt man $V(x) = \sum_{i=1}^n \left(\int_0^1 f_i(tx)dt\right)x_i$, so gilt

$$\begin{aligned}
\frac{\partial V}{\partial x_j}(x) &= \sum_{i=1}^n \left(\frac{\partial}{\partial x_j}\int_0^1 f_i(tx)dt\right) x_i + \sum_{i=1}^n \left(\int_0^1 f_i(tx)dt\right)\frac{\partial x_i}{\partial x_j} \\
&= \sum_{i=1}^n \left(\int_0^1 t\frac{\partial}{\partial x_j} f_i(tx)dt\right) x_i + \int_0^1 f_j(tx)dt.
\end{aligned}$$

Ferner ist

$$\frac{d}{dt}(tf_j(tx)) = f_j(tx) + t \underbrace{\frac{d}{dt}f_j(tx)}_{=\sum_{i=1}^n \frac{\partial}{\partial x_i} f_j(tx)x_i} \overset{\text{Vor.}}{=} f_j(tx) + t\sum_{i=1}^n \frac{\partial}{\partial x_j} f_i(tx)x_i$$

und somit gilt

$$\frac{\partial V}{\partial x_j}(x) = \int_0^1 \frac{d}{dt}(tf_j(tx))dt = tf_j(tx) \mid_{t=0}^{t=1} = f_j(x).$$

$\square$

2.10 Bemerkung. Ist $U \subset \mathbb{R}^3$ eine offene Kugel und $F : U \to \mathbb{R}^3$ ein stetig differenzierbares Vektorfeld, so besagt Satz 2.9, dass F genau dann ein Gradientenfeld ist, falls $\operatorname{rot} F = 0$ gilt.