

5. August 2020

A N A L Y S I S II

Sommersemester 2020

Ernst Kuwert

Mathematisches Institut
Universität Freiburg

Inhaltsverzeichnis

1	Metrische Räume	1
2	Partielle Ableitungen	11
3	Die Ableitung	17
4	Erste Anwendungen der Differentialrechnung	27
5	Potenzreihen	43
6	Parameterabhängige Integrale	51
7	Diffeomorphismen	57
8	Implizite Funktionen	63
9	Kurvenintegrale und Gradientenfelder	69

1 Metrische Räume

In Analysis 1, Kapitel 5, haben wir die Euklidische Norm von $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ und den Euklidischen Abstand von $x, y \in \mathbb{R}^n$ definiert:¹

$$(1.1) \quad |x| = \left(\sum_{i=1}^n x_i^2 \right)^{\frac{1}{2}}$$

$$(1.2) \quad d(x, y) = |x - y| = \left(\sum_{i=1}^n (x_i - y_i)^2 \right)^{\frac{1}{2}}.$$

Damit konnten wir die Begriffe Beschränktheit, ε -Umgebung und Konvergenz erklären. Tatsächlich ist der Begriff des Abstands bzw. moderner der Metrik sehr allgemein; zum Beispiel kann auch für Funktionen in vielen Situationen ein Abstand sinnvoll definiert werden. Diese Idee der *Geometrisierung der Analysis* wurde von Hilbert, Fréchet und Hausdorff Anfang des 20. Jahrhunderts entwickelt. Im Folgenden führen wir den Begriff des metrischen Raums (X, d) abstrakt ein, und behandeln den Euklidischen Raum als zentrales Beispiel.

Definition 1.1 (Metrischer Raum) *Ein metrischer Raum ist eine Menge X mit einer Funktion $d : X \times X \rightarrow [0, \infty)$, die für alle $x, y, z \in X$ folgende Eigenschaften hat:*

Positivität: $d(x, y) \geq 0$ mit Gleichheit genau wenn $x = y$,

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

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

Wir nennen $d(x, y)$ auch den Abstand von x und y .

In dieser Definition kann X eine beliebige Menge sein, insbesondere muss X kein Vektorraum sein. Betrachten Sie als Beispiel die Menge X aller Bahnhöfe in Frankreich und

$$(1.3) \quad d(x, y) = \begin{cases} \text{minimale Fahrzeit von } x \text{ nach } y \text{ über Paris} & \text{für } x \neq y, \\ 0 & \text{für } x = y. \end{cases}$$

Viele interessante metrische Räume sind normierte Vektorräume.

Definition 1.2 (Norm) *Eine Norm auf dem reellen (oder komplexen) Vektorraum X ist eine Funktion $\|\cdot\| : X \rightarrow \mathbb{R}$ mit folgenden Eigenschaften:*

Positivität: $\|x\| \geq 0$, mit Gleichheit genau wenn $x = 0$.

Halblinearität: $\|\lambda x\| = |\lambda| \cdot \|x\|$ für alle $\lambda \in \mathbb{R}$, $x \in X$.

Dreiecksungleichung: $\|x + y\| \leq \|x\| + \|y\|$ für alle $x, y \in X$.

Für die Euklidische Norm auf \mathbb{R}^n sind Positivität und Halblinearität offensichtlich. Die Dreiecksungleichung folgt aus der Ungleichung von Cauchy-Schwarz, siehe Analysis 1, Satz 5.5. Andere Normen auf \mathbb{R}^n sind zum Beispiel die 1-Norm und die Maximumsnorm

$$(1.4) \quad \|x\|_1 = \sum_{i=1}^n |x_i| \quad \text{und} \quad \|x\|_\infty = \max_{1 \leq i \leq n} |x_i|.$$

¹Wir schreiben die Euklidische Norm hier mit einfachen Betragsstrichen.

Jeder normierte Vektorraum $(X, \|\cdot\|)$ wird zu einem metrischen Raum, indem wir den Abstand von zwei Punkten x, y erklären durch

$$(1.5) \quad d(x, y) = \|x - y\| \quad \text{für } x, y \in X.$$

Denn offensichtlich gilt $d(x, y) \geq 0$ mit Gleichheit nur für $x = y$, sowie

$$\begin{aligned} d(y, x) &= \|y - x\| = \|(-1)(x - y)\| = |(-1)| \|x - y\| = d(x, y), \\ d(x, z) &= \|x - z\| = \|(x - y) + (y - z)\| \leq \|x - y\| + \|y - z\| = d(x, y) + d(y, z). \end{aligned}$$

Insbesondere ist \mathbb{R}^n ein metrischer Raum mit dem üblichen euklidischen Abstandsbegriff.

Definition 1.3 Sei X ein metrischer Raum. Die offene Kugel um x_0 mit Radius $r > 0$ ist

$$B_r(x_0) = \{x \in X : d(x, x_0) < r\}.$$

Bezüglich der Euklidischen Norm auf \mathbb{R}^n gilt also wie gewohnt

$$B_r(x_0) = \{x \in \mathbb{R}^n : |x - x_0| < r\}.$$

Es ist instruktiv, sich die Kugeln $B_r(x_0)$ für die französische Eisenbahnmeterik aus (1.3) sowie die Kugeln $B_1(0)$ für die Normen $\|\cdot\|_1$ und $\|\cdot\|_\infty$ auf \mathbb{R}^n zu überlegen.

Definition 1.4 (Offene Mengen) Sei X ein metrischer Raum. Eine Menge $\Omega \subset X$ heißt offen, falls zu jedem $x \in \Omega$ ein $\varepsilon > 0$ existiert mit $B_\varepsilon(x) \subset \Omega$.

Beispiel 1.1 Die Kugel $B_r(x_0)$ ist offen in X , vgl. Analysis 1, Beispiel 5.3. Sei nämlich $x \in B_r(x_0)$ gegeben. Dann ist $\varepsilon = r - d(x, x_0) > 0$ und für $y \in B_\varepsilon(x)$ folgt

$$d(y, x_0) \leq d(y, x) + d(x, x_0) < \varepsilon + d(x, x_0) = r,$$

also $B_\varepsilon(x) \subset B_r(x_0)$, was zu zeigen war.

Satz 1.1 (Topologie) Das System der offenen Teilmengen eines metrischen Raums X bildet eine Topologie, das heißt es gelten folgende Eigenschaften:

- (a) \emptyset, X sind offen.
- (b) Der Durchschnitt von endlich vielen offenen Mengen ist offen.
- (c) Die Vereinigung von beliebig vielen offenen Mengen ist offen.

BEWEIS: (vgl. Analysis 1, Satz 5.8) Aussage (a) ist klar. Für (b) sei $x \in \bigcap_{i=1}^N \Omega_i$, wobei $\Omega_1, \dots, \Omega_N$ endlich viele offene Teilmengen von X sind. Dann gibt es $\varepsilon_i > 0$ mit $B_{\varepsilon_i}(x) \subset \Omega_i$. Es folgt $\varepsilon = \min_{1 \leq i \leq N} \varepsilon_i > 0$ sowie $B_\varepsilon(x) \subset B_{\varepsilon_i}(x) \subset \Omega_i$ für jedes i , das heißt $B_\varepsilon(x) \subset \bigcap_{i=1}^N \Omega_i$.

Für (c) sei nun $x \in \bigcup_{\lambda \in \Lambda} \Omega_\lambda$, wobei Λ eine beliebige Indexmenge ist. Dann ist $x \in \Omega_{\lambda_0}$ für (mindestens) ein $\lambda_0 \in \Lambda$. Da Ω_{λ_0} offen, gibt es ein $\varepsilon > 0$ mit $B_\varepsilon(x) \subset \Omega_{\lambda_0}$, also erst recht $B_\varepsilon(x) \subset \bigcup_{\lambda \in \Lambda} \Omega_\lambda$. \square

Ein abzählbarer Schnitt von offenen Mengen muss nicht offen sein. Zum Beispiel sind die Kugeln $B_{\frac{1}{n}}(0)$, $n \in \mathbb{N}$, offen im \mathbb{R}^n , nicht aber der Schnitt

$$\bigcap_{n=1}^{\infty} B_{\frac{1}{n}}(0) = \{0\}.$$

Eine offene Menge $\Omega \subset X$ mit $x \in \Omega$ nennt man auch offene Umgebung von x . Insbesondere wird die offene Kugel $B_{\varepsilon}(x)$ als ε -Umgebung von x bezeichnet.

Anstatt von einer Metrik auf einer Menge X auszugehen, kann man allgemeiner mit einem System von Teilmengen von X starten, das die Eigenschaften (a),(b) und (c) erfüllt, und diese Mengen dann offen nennen. Ein solches System von Mengen heißt *Topologie* (griechisch $\tau\acute{o}\pi\omicron\varsigma$ Ort oder Lage), statt von einem metrischen Raum spricht man dann von einem topologischen Raum. Die Konzepte der Konvergenz und Stetigkeit sind auch dann noch definierbar, aber das wollen wir hier nicht vertiefen.

Lemma 1.1 (Hausdorff-Trennungseigenschaft) *In einem metrischen Raum X gibt es zu zwei Punkten $x, y \in X$ mit $x \neq y$ ein $\varepsilon > 0$ mit $B_{\varepsilon}(x) \cap B_{\varepsilon}(y) = \emptyset$.*

BEWEIS: Sei $z \in B_{\varepsilon}(x) \cap B_{\varepsilon}(y)$. Dann folgt $d(x, y) \leq d(x, z) + d(z, y) < 2\varepsilon$. Also ist die Behauptung richtig für jedes $\varepsilon \leq \frac{1}{2}d(x, y)$. \square

Definition 1.5 (Konvergenz) *Sei X ein metrischer Raum. Die Folge $(x_k)_{k \in \mathbb{N}}$ von Punkten $x_k \in X$ konvergiert gegen $x \in X$, falls gilt:*

Für alle $\varepsilon > 0$ gibt es ein $K \in \mathbb{R}$ mit $x_k \in B_{\varepsilon}(x)$ für alle $k > K$.

Äquivalent dazu ist $d(x_k, x) \rightarrow 0$ mit $k \rightarrow \infty$.

Der Grenzwert ist eindeutig bestimmt, denn wäre $y \neq x$ ebenfalls Grenzwert von (x_k) , so wählen wir $\varepsilon > 0$ wie in Lemma 1.1 und erhalten für k hinreichend groß den Widerspruch

$$x_k \in B_{\varepsilon}(x) \cap B_{\varepsilon}(y) = \emptyset.$$

Definition 1.6 (abgeschlossene Teilmenge) *Eine Teilmenge A eines metrischen Raums X heißt abgeschlossen, wenn folgende Implikation stets gilt:*

$$x_k \in A, \quad x_k \rightarrow x \in X \quad \Rightarrow \quad x \in A.$$

Die Eigenschaften *offen* und *abgeschlossen* sind nicht Gegensätze. Die leere Menge und der ganze Raum X sind sowohl offen als auch abgeschlossen. Es gilt aber folgende Komplementarität.

Satz 1.2 *In einem metrischen Raum X gilt für jede Menge $M \subset X$:*

$$M \text{ offen} \quad \Leftrightarrow \quad X \setminus M \text{ abgeschlossen.}$$

BEWEIS: Im Fall $X = \mathbb{R}^n$ wurde das in Analysis 1, Satz 5.7, gezeigt. Das Argument gilt analog für jeden metrischen Raum X . \square

Folgerung 1.1 Für die abgeschlossenen Teilmengen eines metrischen Raums X gilt:

- a) \emptyset, X sind abgeschlossen.
- b) Die Vereinigung von endlich vielen abgeschlossenen Mengen ist abgeschlossen.
- c) Der Durchschnitt von beliebig vielen abgeschlossenen Mengen ist abgeschlossen.

BEWEIS: Folgt aus Satz 1.1 und Satz 1.2. □

Die Vereinigung von unendlich vielen abgeschlossenen Mengen ist nicht notwendig abgeschlossen, zum Beispiel $\bigcup_{n=1}^{\infty} [\frac{1}{n}, 1] = (0, 1] \subset \mathbb{R}$.

Beispiel 1.2 (induzierte Metrik) Sei (X, d) metrischer Raum. Auf jeder Menge $M \subset X$ haben wir dann die induzierte Metrik

$$(1.6) \quad d^M : M \times M \rightarrow [0, \infty), \quad d^M(x, y) = d(x, y),$$

und (M, d^M) ist selbst ein metrischer Raum. Zum Beispiel ist die Sphäre $\mathbb{S}^{n-1} = \{x \in \mathbb{R}^n : |x| = 1\}$ ein metrischer Raum mit dem Euklidischen Abstand $d^{\mathbb{S}^{n-1}}(x, y) = |x - y|$. Für die Kugeln bezüglich der induzierten Abstandsfunktion gilt allgemein

$$B_r^M(x) = \{y \in M : d^M(y, x) < r\} = \{y \in M : d(y, x) < r\} = M \cap B_r(x).$$

Die offenen Mengen in (M, d^M) sind genau die Mengen $M \cap \tilde{U}$ mit \tilde{U} offen in X . Denn ist $x \in M \cap \tilde{U}$, so gibt es ein $\varepsilon > 0$ mit $B_\varepsilon(x) \subset \tilde{U}$, also $B_\varepsilon^M(x) \subset M \cap \tilde{U}$. Die Mengen des Typs sind also offen. Ist umgekehrt $U \subset (M, d^M)$ eine beliebige offene Menge, so gibt es zu jedem $x \in U$ ein $\varepsilon(x) > 0$ mit $B_{\varepsilon(x)}^M(x) \subset U$. Es folgt

$$U = \bigcup_{x \in U} B_{\varepsilon(x)}^M(x) = \bigcup_{x \in U} M \cap B_{\varepsilon(x)}(x) = M \cap \underbrace{\bigcup_{x \in U} B_{\varepsilon(x)}(x)}_{\text{offen in } X}.$$

Weiter gilt: die abgeschlossenen Mengen in (M, d^M) sind genau von der Form $M \cap \tilde{A}$ mit \tilde{A} abgeschlossen in X . Denn es gilt

$$M \setminus (M \cap \tilde{A}) = M \cap (X \setminus \tilde{A}) = \text{offen in } M.$$

Also $M \cap \tilde{A}$ abgeschlossen in M nach Satz 1.2. Und jede in M abgeschlossene Menge hat diese Form, denn für \tilde{U} offen in X ist

$$A = M \setminus (M \cap \tilde{U}) = M \cap \underbrace{(X \setminus \tilde{U})}_{\text{abg. in } X}.$$

In der eindimensionalen Analysis wurden meist Funktionen auf einem Intervall I mit Randpunkten $a < b$ betrachtet. Im mehrdimensionalen Fall werden wir oft Kugeln $B_r(x)$ oder achsenparallele Quader $I_1 \times \dots \times I_n$ betrachten, bisweilen aber auch kompliziertere Mengen. Dafür sind die folgenden Begriffe nützlich.

Definition 1.7 Sei X ein metrischer Raum und $M \subset X$. Dann definieren wir

$$\begin{aligned} \text{int } M &= \{x \in M : \exists \varepsilon > 0 \text{ mit } B_\varepsilon(x) \subset M\} && \text{(Menge der inneren Punkte von } M), \\ \overline{M} &= \{x \in X : \forall \varepsilon > 0 \text{ ist } B_\varepsilon(x) \cap M \neq \emptyset\} && \text{(Abschluss von } M), \\ \partial M &= \{x \in X : \forall \varepsilon > 0 \text{ sind } B_\varepsilon(x) \cap M, B_\varepsilon(x) \cap (X \setminus M) \neq \emptyset\} && \text{(Rand von } M). \end{aligned}$$

Trivialerweise gilt $\text{int } M \subset M \subset \overline{M}$. Außerdem ist $\text{int } \Omega = \Omega$ für $\Omega \subset X$ offen sowie $\overline{\overline{M}} = \overline{M}$ für $M \subset X$ abgeschlossen.

Beispiel 1.3 Auf dem \mathbb{R}^n mit der euklidischen Abstandsfunktion $d(x, y) = |x - y|$ gilt für die Kugel $B_r(x) = \{y \in \mathbb{R}^n : d(y, x) < r\}$:

$$\overline{B_r(x)} = \{y \in \mathbb{R}^n : d(y, x) \leq r\} \quad \text{und} \quad \partial B_r(x) = \{y \in \mathbb{R}^n : d(y, x) = r\}.$$

Beweis Übungsaufgabe.

Satz 1.3 Sei M Teilmenge des metrischen Raums X .

(a) $\text{int } M$ ist offen, und es gilt die Implikation

$$\Omega \text{ offen, } \Omega \subset M \quad \Rightarrow \quad \Omega \subset \text{int } M.$$

(b) \overline{M} ist abgeschlossen, und es gilt die Implikation

$$A \text{ abgeschlossen, } A \supset M \quad \Rightarrow \quad A \supset \overline{M}.$$

(c) ∂M ist abgeschlossen und es gilt $\partial M = \overline{M} \setminus \text{int } M$.

BEWEIS: Für (a) sei $x \in \text{int } M$, also $B_r(x) \subset M$ für ein $r > 0$. Für $y \in B_r(x)$ gilt dann $B_\varepsilon(y) \subset B_r(x) \subset M$ mit $\varepsilon = r - d(y, x) > 0$, vgl. Beispiel 1.1. Es folgt $B_r(x) \subset \text{int } M$, damit ist $\text{int } M$ offen. Sei nun Ω offen und $\Omega \subset M$. Zu $x \in \Omega$ gibt es ein $\varepsilon > 0$ mit $B_\varepsilon(x) \subset \Omega$, also auch $B_\varepsilon(x) \subset M$, das heißt $x \in \text{int } M$.

Für (b) verwenden wir (a) und Satz 1.2. Nach Definition ist $X \setminus \overline{M} = \text{int}(X \setminus M)$, also ist $\overline{M} = X \setminus \text{int}(X \setminus M)$ abgeschlossen. Ist nun $A \subset X$ eine beliebige abgeschlossene Menge mit $A \supset M$, so ist $X \setminus A$ offen sowie $X \setminus A \subset X \setminus M$, also $X \setminus A \subset \text{int}(X \setminus M)$ nach (a), und somit $A \supset \overline{M}$. Dies beweist (b).

Nach Definition gilt weiter $\partial M = \overline{M} \cap \overline{(X \setminus M)}$, also ist ∂M abgeschlossen nach (b) und Folgerung 1.1. Ferner ist ebenfalls nach Definition $X \setminus \text{int } M = \overline{X \setminus M}$, folglich

$$\partial M = \overline{M} \cap (X \setminus \text{int } M) = \overline{M} \setminus \text{int } M.$$

□

Die Punkte in \overline{M} sind entweder Häufungspunkte oder isolierte Punkte von M .

Definition 1.8 Ein Punkt $x \in X$ heißt

Häufungspunkt von $M \Leftrightarrow$ für jedes $\varepsilon > 0$ ist $M \cap B_\varepsilon(x) \setminus \{x\}$ nichtleer,

isolierter Punkt von $M \Leftrightarrow$ es gibt ein $\varepsilon > 0$ mit $M \cap B_\varepsilon(x) = \{x\}$.

Ist $x \in X$ Häufungspunkt von M , so enthält $B_\varepsilon(x) \cap M \setminus \{x\}$ sogar unendlich viele Punkte. Denn würde die Menge nur aus endlich vielen Punkten y_1, \dots, y_N bestehen, so ist $\delta = \min_{1 \leq i \leq N} d(y_i, x) > 0$ und dann $B_\delta(x) \cap M \setminus \{x\} = \emptyset$, ein Widerspruch. Insbesondere können wir eine Folge $x_k \in M \setminus \{x\}$ bestimmen mit $x_k \rightarrow x$.

Definition 1.9 Eine Teilmenge M eines metrischen Raums X heißt dicht, falls $\overline{M} = X$.

Bekanntes Beispiel sind die rationalen Zahlen \mathbb{Q} im metrischen Raum \mathbb{R} , beziehungsweise die rationalen Punkte \mathbb{Q}^n im \mathbb{R}^n .

Neben der Offenheit und Abgeschlossenheit spielte in Analysis 1 eine dritte Eigenschaft von Mengen eine wesentliche Rolle, nämlich die Kompaktheit.

Definition 1.10 (Folgenkompaktheit) Eine Menge $M \subset (X, d)$ heißt kompakt, wenn jede Folge $(x_k)_{k \in \mathbb{N}}$ in M eine Teilfolge $(x_{k_p})_{p \in \mathbb{N}}$ hat, die gegen ein $x \in M$ konvergiert.

Eine Teilmenge M von (X, d) ist genau dann kompakt, wenn der metrische Raum (M, d^M) kompakt ist. Denn die induzierte Metrik d^M ist die Einschränkung von d auf M , also bedeuten Konvergenz bezüglich d und d^M dasselbe. Für die Kompaktheit von Teilmengen des \mathbb{R}^n hat man das Kriterium von Heine-Borel, siehe Analysis 1, Satz 10.2.

Satz 1.4 (Kompaktheit im \mathbb{R}^n) Eine Menge $M \subset \mathbb{R}^n$ ist genau dann kompakt, wenn sie abgeschlossen und beschränkt ist.

Diese Aussage ist aber in vielen metrischen Räumen falsch, das heißt es kann abgeschlossene und beschränkte Teilmengen geben, die nicht kompakt sind.

Definition 1.11 (Stetigkeit) Seien X, Y metrische Räume, und $M \subset X$. Eine Abbildung $f : M \rightarrow Y$ heißt stetig im Punkt $x_0 \in M$, wenn es zu jedem $\varepsilon > 0$ ein $\delta > 0$ gibt mit

$$d(f(x), f(x_0)) < \varepsilon \quad \text{für alle } x \in M \text{ mit } d(x, x_0) < \delta,$$

oder äquivalent mit $f(M \cap B_\delta(x_0)) \subset B_\varepsilon(f(x_0))$. Die Abbildung f heißt stetig (als Ganzes), wenn f in jedem Punkt $x_0 \in M$ stetig ist.

Wir müssten hier eigentlich $d_X(\cdot, \cdot)$ und $d_Y(\cdot, \cdot)$ schreiben, denn X und Y sind ja verschiedene metrische Räume. Die folgende Eigenschaft impliziert Stetigkeit (aber nicht umgekehrt).

Definition 1.12 (Lipschitz) $f : M \rightarrow Y$ heißt Lipschitzstetig mit Konstante $L \geq 0$, falls

$$d(f(x), f(x')) \leq L d(x, x') \quad \text{für alle } x, x' \in X.$$

Beispiel 1.4 Die Abstandsfunktion von einem Punkt $x_0 \in X$, das heißt

$$f : X \rightarrow \mathbb{R}, f(x) = d(x, x_0),$$

ist Lipschitzstetig mit Konstante $L = 1$, denn aus der Dreiecksungleichung folgt

$$f(x) = d(x, x_0) \leq d(x, x') + d(x', x_0) = d(x, x') + f(x').$$

Durch Vertauschen von x und x' folgt $|f(x) - f(x')| \leq d(x, x')$ wie gewünscht.

Die Stetigkeit von Abbildungen zwischen metrischen Räumen kann alternativ mittels Folgen beschrieben werden. Der Beweis aus Analysis 1, Satz 7.1, gilt ganz analog.

Satz 1.5 (Folgenkriterium der Stetigkeit) Seien X, Y metrische Räume, und $x_0 \in M \subset X$. Für $f : M \rightarrow Y$ sind folgende Aussagen äquivalent:

- (1) f ist stetig in x_0 .
- (2) Für jede Folge $x_k \in M$ mit $x_k \rightarrow x_0$ gilt $f(x_k) \rightarrow f(x_0)$.

Die folgende Charakterisierung der Stetigkeit einer Abbildung verwendet nur die offenen Mengen (und nicht die Metriken). Deshalb kann sie in topologischen Räumen als Definition benutzt werden. Beachte, dass sie nicht in einem Punkt x_0 lokalisiert ist, sondern sich auf die Abbildung als Ganzes bezieht.

Satz 1.6 (Charakterisierung stetiger Abbildungen) Eine Abbildung $f : X \rightarrow Y$ zwischen metrischen Räumen ist genau dann stetig, wenn für jede offene Menge $V \subset Y$ das Urbild $f^{-1}(V)$ offen in X ist.

BEWEIS: Sei f stetig, $V \subset Y$ offen und $x_0 \in f^{-1}(V)$. Dann ist $y_0 = f(x_0) \in V$, also gilt $B_\varepsilon(y_0) \subset V$ für geeignetes $\varepsilon > 0$. Es gibt dann ein $\delta > 0$ mit $f(B_\delta(x_0)) \subset B_\varepsilon(y_0) \subset V$, also $B_\delta(x_0) \subset f^{-1}(V)$ wie verlangt.

Umgekehrt sei $y_0 = f(x_0)$ und $\varepsilon > 0$ gegeben. Nach Voraussetzung ist dann $f^{-1}(B_\varepsilon(y_0))$ offen, das heißt es gibt ein $\delta > 0$ mit $B_\delta(x_0) \subset f^{-1}(B_\varepsilon(y_0))$ beziehungsweise $f(B_\delta(x_0)) \subset B_\varepsilon(y_0)$. \square

Folgende Aussagen über Stetigkeit und kompakte Mengen sind oft nützlich.

Satz 1.7 (Bilder kompakter Mengen) Seien X, Y metrische Räume. Ist $M \subset X$ kompakt und $f : M \rightarrow Y$ stetig, so gilt:

- (1) $f(M)$ ist kompakte Teilmenge von Y .
- (2) Ist f injektiv, so ist $f^{-1} : f(M) \rightarrow M$ stetig.

BEWEIS: Sei y_k eine Folge in $N = f(M)$, also $y_k = f(x_k)$ mit $x_k \in M$. Da M kompakt, gibt es eine Teilfolge mit $x_{k_j} \rightarrow x \in M$. Aus der Stetigkeit von f folgt $y_{k_j} = f(x_{k_j}) \rightarrow f(x) \in N$.

Sei nun f injektiv. Wäre f^{-1} in $y = f(x)$ unstetig, so gibt es eine Folge $y_k = f(x_k)$ mit $y_k \rightarrow y$, aber $d(x_k, x) \geq \varepsilon > 0$ für alle k . Da M kompakt, gibt es eine Teilfolge $x_{k_j} \rightarrow x' \in M$. Es folgt $d(x', x) \geq \varepsilon$. Aber f ist stetig in x' , also $f(x') = \lim_{j \rightarrow \infty} f(x_{k_j}) = y$, im Widerspruch zur Injektivität von f . \square

Satz 1.8 (Extrema) Sei M kompakte Teilmenge des metrischen Raums X . Dann ist jede stetige Funktion $f : M \rightarrow \mathbb{R}$ beschränkt und nimmt ihr Infimum und Supremum an.

BEWEIS: vgl. Analysis 1, Satz 10.1. \square

Beispiel 1.5 Sei X metrischer Raum und $K \subset X$ kompakt. Dann gibt es zu jedem $x_0 \in X$ einen nächsten Punkt $x \in K$, das heißt

$$d(x, x_0) = \inf_{y \in K} d(y, x_0) = \text{dist}(x_0, K).$$

Der Punkt x ist nicht notwendig eindeutig, betrachte etwa $K = \{1, -1\} \subset \mathbb{R}$ und $x_0 = 0$.

Eine weitere nützliche Konsequenz der Kompaktheit ist wie folgt:

Satz 1.9 (Gleichmäßige Stetigkeit) Sei X kompakter metrischer Raum. Dann ist jede stetige Abbildung $f : X \rightarrow Y$ sogar gleichmäßig stetig.

BEWEIS: Nach Definition 13.6 in Analysis 1 ist f gleichmäßig stetig wenn gilt:

$$\lim_{\delta \searrow 0} \omega_f(\delta) = 0 \quad \text{wobei} \quad \omega_f(\delta) = \sup_{d(x,x') < \delta} d(f(x), f(x')).$$

Dazu äquivalent ist folgende Aussage (prüfen!): zu jedem $\varepsilon > 0$ gibt es ein $\delta > 0$, so dass

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

Wäre also f nicht gleichmäßig stetig, so gibt es ein $\varepsilon > 0$ und Folgen $x_n, x'_n \in X$ mit $d(x_n, x'_n) \rightarrow 0$, aber $d(f(x_n), f(x'_n)) \geq \varepsilon$. Da X kompakt, konvergiert die Folge x_n nach evtl. Auswahl einer Teilfolge gegen ein $x \in X$. Wegen $d(x'_n, x) \leq d(x'_n, x_n) + d(x_n, x)$ konvergiert dann auch die Folge x'_n gegen dieses x , und wegen Stetigkeit folgt

$$\varepsilon \leq d(f(x_n), f(x'_n)) \leq d(f(x_n), f(x)) + d(f(x), f(x'_n)) \rightarrow 0,$$

ein Widerspruch. □

Definition 1.13 (Zusammenhang) Ein metrischer Raum X heißt zusammenhängend, wenn folgende Implikation gilt: ist $E \subset X$ nichtleer, abgeschlossen und offen, so ist $E = X$.

Eine Funktion $f : X \rightarrow \mathbb{R}$ heißt lokal konstant, wenn jeder Punkt $x \in X$ eine offene Umgebung U hat mit $f|_U$ konstant. Der Zusammenhang kann damit auch so charakterisiert werden.

Lemma 1.2 Für einen metrischen Raum sind folgende Aussagen äquivalent:

- (a) X ist zusammenhängend.
- (b) Jede lokal konstante Funktion $f : X \rightarrow \mathbb{R}$ ist konstant.

BEWEIS: Für (a) \Rightarrow (b) sei eine lokal konstante Funktion $f : X \rightarrow \mathbb{R}$ gegeben. Wir wählen ein $x_0 \in X$ und zeigen, dass $E = \{x \in X : f(x) = f(x_0)\}$ offen und abgeschlossen ist. Aus (a) folgt dann $E = X$ bzw. f konstant. Ist $x \in E$ und U offene Umgebung mit $f|_U$ konstant, so folgt $f|_U = f(x) = f(x_0)$, also $U \subset E$ und damit ist E offen. Ist andererseits $x_k \in E$ mit $x_k \rightarrow x$, so folgt $f(x) = \lim_{k \rightarrow \infty} f(x_k) = f(x_0)$ wegen f stetig, also ist E abgeschlossen.

Es gelte nun (b), und $E \subset X$ sei nichtleer, offen und abgeschlossen. Betrachte

$$\chi_E : X \rightarrow \mathbb{R}, \quad \chi_E(x) = \begin{cases} 1 & \text{für } x \in E, \\ 0 & \text{für } x \notin E. \end{cases}$$

Da E und $X \setminus E$ offen sind, ist χ_E lokal konstant und damit konstant. Da E nichtleer, folgt $E = X$, das heißt X ist zusammenhängend. □

Der Zusammenhang eines Raums hängt von der Metrik ab. In Aussage (b) des Lemmas steckt das in dem Wort *lokal*. Ist U offene Umgebung von x , so gibt es ein $\varepsilon > 0$ mit $B_\varepsilon(x) \subset U$, siehe Definition 1.4. Also ist $f : X \rightarrow \mathbb{R}$ genau dann lokal konstant, wenn es zu jedem $x \in X$ ein $\varepsilon > 0$ gibt mit $f|_{B_\varepsilon(x)}$ konstant.

Eine Teilmenge X von (Y, d) heißt zusammenhängend, wenn der metrische Raum (X, d^X) zusammenhängend ist. Nach Beispiel 1.2 sind die Kugeln bezüglich der induzierten Metrik d^X genau von der Form $X \cap B_\varepsilon(x)$. Also ist $f : X \rightarrow \mathbb{R}$ genau dann lokal konstant, wenn es zu jedem $x \in X$ ein $\varepsilon > 0$ gibt mit $f|_{X \cap B_\varepsilon(x)}$ konstant.

Satz 1.10 (zusammenhängende Mengen in \mathbb{R}) *Eine Menge $X \subset \mathbb{R}$ ist genau dann zusammenhängend, wenn sie ein (verallgemeinertes) Intervall ist.*

BEWEIS: Sei X ein Intervall, also $(a, b) \subset X$ mit $a = \inf X$, $b = \sup X$, und $f : X \rightarrow \mathbb{R}$ sei lokal konstant. Dann gibt es zu jedem $x \in X$ ein $\varepsilon > 0$ mit f konstant auf $X \cap (x - \varepsilon, x + \varepsilon)$. Ist $x \in (a, b)$, so ist $(x - \varepsilon, x + \varepsilon) \subset (a, b)$ für $\varepsilon > 0$ klein, also folgt $f'(x) = 0$. Im Fall $a \in X$ ist f konstant auf $[a, a + \varepsilon)$, analog wenn $b \in X$. Laut Mittelwertsatz ist nun f konstant, siehe Folgerung 10.1 in Analysis 1. Somit ist X zusammenhängend, siehe Lemma 1.2.

Sei umgekehrt X zusammenhängend. Wir zeigen $(a, b) \subset X$ mit a, b wie oben. Angenommen es gibt $x_0 \in (a, b)$ mit $x_0 \notin X$. Betrachte

$$f : X \rightarrow \mathbb{R}, f(x) = \begin{cases} 0 & \text{für } x < x_0, \\ 1 & \text{für } x > x_0. \end{cases}$$

Die Funktion f ist lokal konstant auf X , also konstant. Das bedeutet $b = \sup X \leq x_0$ oder $a = \inf X \geq x_0$, ein Widerspruch. \square

Lemma 1.3 *Seien X, Y metrische Räume. Ist X zusammenhängend und $f : X \rightarrow Y$ stetig, so ist $f(X)$ auch zusammenhängend.*

BEWEIS: Sei $g : f(X) \rightarrow \mathbb{R}$ lokal konstant, das heißt zu $x \in X$ gibt es eine offene Umgebung $V \subset Y$ von $f(x)$ mit $g|_{f(X) \cap V}$ konstant. Dann ist $g \circ f$ konstant auf $f^{-1}(V) \ni x$. Da X zusammenhängend ist $g \circ f$ konstant, und damit ist $g : f(X) \rightarrow \mathbb{R}$ konstant. \square

Aus dem Lemma und Satz 1.10 ergibt sich der Zwischenwertsatz, in folgender Form.

Folgerung 1.2 *Ist X zusammenhängend und $f : X \rightarrow \mathbb{R}$ stetig, so ist $f(X)$ ein Intervall.*

2 Partielle Ableitungen

Für Funktionen auf dem \mathbb{R}^n gibt es mehrere Ableitungskonzepte. Die partiellen Ableitungen sind am einfachsten, es sind die eindimensionalen Ableitungen in Richtung der Koordinatenachsen. Im Folgenden bezeichnet e_1, \dots, e_n die Standardbasis des \mathbb{R}^n , also $e_j = (0, \dots, 1, \dots, 0)$ mit der 1 an der j -ten Stelle.

Definition 2.1 Sei $\Omega \subset \mathbb{R}^n$ offen und $f : \Omega \rightarrow \mathbb{R}^m$. Die partielle Ableitung von f nach x_j an der Stelle $x \in \Omega$ ist der Grenzwert (falls existent)

$$\partial_j f(x) = \lim_{t \rightarrow 0} \frac{f(x + te_j) - f(x)}{t} = \left. \frac{d}{dt} f(x + te_j) \right|_{t=0}.$$

Andere Bezeichnungen: $\frac{\partial f}{\partial x_j}(x)$ oder $f_{x_j}(x)$.

Halten wir alle x_i mit $i \neq j$ fest, so ergibt sich lokal die Funktion einer Variablen

$$\varphi(s) = f(x_1, \dots, x_{j-1}, s, x_{j+1}, \dots, x_n) \quad \text{für } s \in (x_j - \delta, x_j + \delta).$$

Die Ableitung von φ im Punkt $s = x_j$ ist

$$\varphi'(x_j) = \lim_{t \rightarrow 0} \frac{\varphi(x_j + t) - \varphi(x_j)}{t} = \lim_{t \rightarrow 0} \frac{f(x + te_j) - f(x)}{t} = \partial_j f(x).$$

Bei der Berechnung der partiellen Ableitung $\partial_j f$ können wir also die gewohnte eindimensionale Ableitung nach x_j bilden und dabei die anderen Variablen als Konstanten behandeln.

Die wohlbekanntenen Differentiationsregeln für Funktionen einer Variablen ergeben in diesem Kontext direkt folgende Aussagen.

Satz 2.1 (Ableitungsregeln) Sei $\Omega \subset \mathbb{R}^n$ offen und $x \in \Omega$. Die Existenz der partiellen Ableitungen $\partial_j f(x)$ und $\partial_j g(x)$ sei hier stets vorausgesetzt. Dann gelten folgende Aussagen:

(a) Linearität: für $f, g : \Omega \rightarrow \mathbb{R}^m$ und $\alpha, \beta \in \mathbb{R}$ gilt

$$\partial_j(\alpha f + \beta g)(x) = \alpha \partial_j f(x) + \beta \partial_j g(x).$$

(b) Komponentenweise Differentiation: für $f : \Omega \rightarrow \mathbb{R}^m$ gilt, wenn eine der Seiten existiert,

$$\partial_j f(x) = \sum_{i=1}^m \partial_j f_i(x) e_i.$$

(c) Produktregel: für $f : \Omega \rightarrow \mathbb{R}^m$ und $g : \Omega \rightarrow \mathbb{R}$ gilt

$$\partial_j(fg)(x) = (\partial_j f)(x)g(x) + f(x)(\partial_j g)(x).$$

(d) Quotientenregel: für $f : \Omega \rightarrow \mathbb{R}^m$ und $g : \Omega \rightarrow \mathbb{R}$ mit $g(x) \neq 0$ gilt

$$\partial_j \left(\frac{f}{g} \right) (x) = \frac{(\partial_j f)(x)g(x) - f(x)(\partial_j g)(x)}{g(x)^2}.$$

(e) Kettenregel: sei f reellwertig. Ist I offenes Intervall mit $f(\Omega) \subset I$ und $\varphi : I \rightarrow \mathbb{R}$ differenzierbar, so gilt

$$\partial_j(\varphi \circ f)(x) = \varphi'(f(x))\partial_j f(x).$$

Beispiel 2.1 Wir betrachten die Euklidische Abstandsfunktion vom Nullpunkt

$$r : \mathbb{R}^n \rightarrow \mathbb{R}, r(x) = |x| = \sqrt{x_1^2 + \dots + x_n^2}.$$

In $x \neq 0$ existieren die partiellen Ableitungen, und zwar gilt mit der Kettenregel

$$\partial_j r(x) = \frac{2x_j}{2\sqrt{x_1^2 + \dots + x_n^2}} = \frac{x_j}{r} \quad \text{für } x \in \mathbb{R}^n \setminus \{0\}.$$

Im Punkt $x = 0$ existieren die partiellen Ableitungen nicht, denn $r(0 + te_i) = |t|$ ist in $t = 0$ nicht differenzierbar. Aber für $x \neq 0$ ist die Funktion $\partial_j r$ wieder partiell differenzierbar; wir erhalten mit der Quotientenregel die zweiten partiellen Ableitungen

$$\partial_i(\partial_j r)(x) = \frac{(\partial_i x_j)r - x_j \partial_i r}{r^2} = \frac{1}{r} \left(\delta_{ij} - \frac{x_i x_j}{r^2} \right) \quad \text{für } x \in \mathbb{R}^n \setminus \{0\}.$$

Für $f = \varphi \circ r : \mathbb{R}^n \setminus \{0\} \rightarrow \mathbb{R}$, mit $\varphi : (0, \infty) \rightarrow \mathbb{R}$ zweimal differenzierbar, folgt

$$\begin{aligned} \partial_j f(x) &= \varphi'(r)\partial_j r = \varphi'(r)\frac{x_j}{r}, \\ \partial_i(\partial_j f)(x) &= \varphi''(r)\partial_i r \partial_j r + \varphi'(r)\partial_i(\partial_j r) = \varphi''(r)\frac{x_i x_j}{r^2} + \frac{\varphi'(r)}{r} \left(\delta_{ij} - \frac{x_i x_j}{r^2} \right). \end{aligned}$$

Wir betrachten nun den Laplaceoperator

$$\Delta f = \sum_{i=1}^n \partial_i^2 f.$$

Die Gleichung $\Delta f = 0$ spielt in der komplexen Analysis, der Theorie der Minimalflächen und der Elektrostatik eine zentrale Rolle, ihre Lösungen heißen harmonische Funktionen. Wir rechnen jetzt die rotations-symmetrischen harmonischen Funktionen aus, und zwar erhalten wir die Gleichung

$$0 \stackrel{!}{=} \Delta f(x) = \varphi''(r) + \frac{n-1}{r}\varphi'(r) = r^{1-n}(r^{n-1}\varphi'(r))'.$$

Die Lösungen dieser Gleichung sind, mit Integrationskonstanten $a, b \in \mathbb{R}$,

$$\varphi(r) = \begin{cases} a \frac{r^{2-n}}{2-n} + b & \text{für } n \geq 3 \\ a \log r + b & \text{für } n = 2. \end{cases}$$

Für $n = 3$ ist $f(x) = -\frac{1}{r}$ das Coulombpotential einer Punktladung.

Wir wollen nun zweite und höhere partielle Ableitungen allgemein einführen. Für $f : \Omega \rightarrow \mathbb{R}^m$ sei die Ableitungsfunktion $\partial_j f : \Omega \rightarrow \mathbb{R}^m$ definiert. Wir setzen, falls existent,

$$(2.1) \quad \partial_{ij}^2 f(x) := \partial_i(\partial_j f)(x) \quad (\text{alternative Notation } \frac{\partial^2 f}{\partial x_i \partial x_j}(x) \text{ oder } f_{x_i x_j}(x)).$$

Entsprechend für Ableitungen k -ter Ordnung: seien $i_1, \dots, i_k \in \{1, \dots, n\}$. Ist die Funktion $\partial_{i_2 \dots i_k}^{k-1} f : \Omega \rightarrow \mathbb{R}^m$ bereits definiert so setzen wir, falls existent,

$$(2.2) \quad \partial_{i_1 \dots i_k}^k f(x) = \partial_{i_1}(\partial_{i_2 \dots i_k}^{k-1} f)(x).$$

Für viele Anwendungen ist es wesentlich, dass die partiellen Ableitungen nicht nur existieren sondern zusätzlich stetige Funktionen sind.

Definition 2.2 (C^k -Räume) Sei $\Omega \subset \mathbb{R}^n$ offen und $k \in \mathbb{N}_0 \cup \{\infty\}$. Wir bezeichnen mit $C^k(\Omega, \mathbb{R}^m)$ die Menge aller k -mal stetig differenzierbaren Funktionen auf Ω mit Werten im \mathbb{R}^m , das heißt alle partiellen Ableitungen $\partial_{i_1 \dots i_j}^j f$ der Ordnung $j \leq k$ (bzw. $j < \infty$ im Fall $k = \infty$) sind definiert und stetig auf Ω . Im reellwertigen Fall, also $m = 1$, schreiben wir kürzer $C^k(\Omega) = C^k(\Omega, \mathbb{R})$.

Wir kommen nun zu der Frage, ob die Ableitungen ∂_i und ∂_j vertauschen. Allein aus der Existenz der zweiten partiellen Ableitungen folgt das nicht, wie das folgende Beispiel zeigt:

$$(2.3) \quad f(x, y) = \begin{cases} xy \frac{x^2 - y^2}{x^2 + y^2} & \text{für } (x, y) \neq (0, 0) \\ 0 & \text{für } (x, y) = (0, 0) \end{cases}$$

Für diese Funktion ist $\partial_1 \partial_2 f(0, 0) = 1$, aber $\partial_2 \partial_1 f(0, 0) = -1$. Beide Ableitungen existieren, aber sie sind nicht gleich.

Satz 2.2 (von Schwarz) Sei $\Omega \subset \mathbb{R}^n$ offen. Ist $f \in C^2(\Omega)$, so vertauschen für $1 \leq i, j \leq n$ die Ableitungen nach x_i und x_j :

$$\partial_i \partial_j f = \partial_j \partial_i f \quad \text{auf } \Omega.$$

BEWEIS: Wir sammeln vorab ein paar Fakten zu Differenzenquotienten. Für $g \in C^1(\Omega)$ ist

$$(2.4) \quad \partial_i g(x) = \lim_{s \rightarrow 0} \Delta_i^s g(x) \quad \text{mit } \Delta_i^s g(x) = \frac{g(x + se_i) - g(x)}{s}.$$

Es gilt $\frac{d}{ds} g(x + se_i) = \partial_i g(x + se_i)$, also folgt aus dem Mittelwertsatz eine Darstellung

$$(2.5) \quad \Delta_i^s g(x) = \partial_i g(x + \alpha se_i) \quad \text{für ein } \alpha \in [0, 1].$$

Weiter berechnen wir

$$\begin{aligned} \partial_i(\Delta_j^t g)(x) &= \frac{d}{ds} (\Delta_j^t g)(x + se_i)|_{s=0} \\ &= \frac{d}{ds} \frac{g(x + se_i + te_j) - g(x + se_i)}{t} \Big|_{s=0} \\ &= \frac{\partial_i g(x + te_j) - \partial_i g(x)}{t}, \end{aligned}$$

also gilt die Vertauschungsregel

$$(2.6) \quad \partial_i(\Delta_j^t g)(x) = \Delta_j^t(\partial_i g)(x).$$

Nun betrachte für $s, t \neq 0$ den zweiten Differenzenquotienten $\Delta_i^s(\Delta_j^t f)(x)$. Es gilt

$$\Delta_i^s(\Delta_j^t f)(x) = \frac{(\Delta_j^t f)(x + se_i) - (\Delta_j^t f)(x)}{s} \xrightarrow{t \rightarrow 0} \frac{\partial_j f(x + se_i) - \partial_j f(x)}{s} = \Delta_i^s(\partial_j f)(x),$$

das heißt

$$(2.7) \quad \lim_{t \rightarrow 0} \Delta_i^s(\Delta_j^t f)(x) = \Delta_i^s(\partial_j f)(x).$$

Jetzt wenden wir (2.5) und (2.6) an. Für gewisse $\alpha, \beta \in [0, 1]$ folgt

$$\begin{aligned} \Delta_i^s(\Delta_j^t f)(x) &= \partial_i(\Delta_j^t f)(x + \alpha se_i) \\ &= \Delta_j^t(\partial_i f)(x + \alpha se_i) \\ &= \partial_j(\partial_i f)(x + \alpha se_i + \beta te_j). \end{aligned}$$

Nach Voraussetzung ist $\partial_j(\partial_i f)$ im Punkt x stetig. Somit gibt es zu $\varepsilon > 0$ ein $\delta > 0$ mit

$$(2.8) \quad |\Delta_i^s(\Delta_j^t f)(x) - \partial_j(\partial_i f)(x)| < \varepsilon \quad \text{für } 0 < |s|, |t| < \delta.$$

Hier lassen wir nun $t \rightarrow 0$ gehen, es folgt wegen (2.7)

$$|\Delta_i^s(\partial_j f)(x) - \partial_j(\partial_i f)(x)| \leq \varepsilon \quad \text{für } 0 < |s| < \delta.$$

Nach Definition der partiellen Ableitung, siehe (2.4), ist das die Behauptung. \square

Bemerkung. Der Beweis zeigt etwas mehr: existieren die Ableitungen $\partial_i f$, $\partial_j f$, $\partial_j \partial_i f$ auf einer Umgebung von x und ist $\partial_j \partial_i f$ stetig in x , so existiert auch $\partial_i \partial_j f(x)$ und ist gleich $\partial_j \partial_i f(x)$. Der wesentliche Punkt im Beweis ist die Abschätzung (2.8). Danach gilt

$$\lim_{s, t \rightarrow 0, s, t \neq 0} \Delta_i^s(\Delta_j^t f)(x) = \partial_j(\partial_i f)(x),$$

egal wie s, t gegen Null gehen.

Folgerung 2.1 Für eine Funktion $f \in C^k(\Omega, \mathbb{R}^m)$ vertauschen die partiellen Ableitungen bis zur Ordnung k , das heißt für jede Permutation $\sigma \in S_k$ gilt

$$\partial_{i_{\sigma(1)}} \dots \partial_{i_{\sigma(k)}} f = \partial_{i_1} \dots \partial_{i_k} f.$$

BEWEIS: Nach Satz 2.2 können benachbarte Operatoren ∂_i, ∂_j vertauscht werden. Die symmetrische Gruppe wird durch Vertauschungen erzeugt (siehe Lineare Algebra). \square

Der Begriff der partiellen Ableitung allein ist nicht geeignet, um die mehrdimensionale Differentialrechnung zu entwickeln. Entscheidendes Manko ist, dass aus der Existenz der partiellen Ableitungen $\partial_1 f, \dots, \partial_n f$ in $x \in \Omega$ nicht die Stetigkeit von f im Punkt x folgt.

Beispiel 2.2 Sei $\Omega = \mathbb{R}^2$ und

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

Dann gilt $f(x, 0) = 0 = f(0, y)$, insbesondere $\partial_1 f(0, 0) = 0 = \partial_2 f(0, 0)$. Aber für $c(t) = (t, t)$ gilt $f(c(t)) = 1/2$ für alle $t \neq 0$, das heißt f ist nicht stetig im Nullpunkt.

Also kann die Verkettung $f \circ c$ mit einer Kurve unstetig sein. Aber dann ist $f \circ c$ auch nicht differenzierbar, siehe Analysis 1, Satz 9.1, und eine Kettenregel kann es nicht geben. Die Definition der partiellen Ableitungen macht explizit von den Koordinaten auf \mathbb{R}^n Gebrauch. Es wäre denkbar, dass sich ein besserer Ableitungsbegriff ergibt, wenn alle Richtungen gleichberechtigt betrachtet werden. Dies führt auf den Begriff der Richtungsableitung.

Definition 2.3 (Richtungsableitung) Sei $\Omega \subset \mathbb{R}^n$ offen und $f : \Omega \rightarrow \mathbb{R}^m$. Die Richtungsableitung von f an der Stelle $x \in \Omega$ in Richtung $v \in \mathbb{R}^n$ ist der Grenzwert (falls existent)

$$\partial_v f(x) = \lim_{t \rightarrow 0} \frac{f(x + tv) - f(x)}{t} = \frac{d}{dt} f(x + tv)|_{t=0}.$$

Beispiel 2.3 Die Richtungsableitung von $r(x) = |x|$ in $x \in \mathbb{R}^n \setminus \{0\}$ in Richtung $v \in \mathbb{R}^n$ ist

$$\partial_v r(x) = \frac{d}{dt} \sqrt{|x|^2 + 2t\langle x, v \rangle + t^2|v|^2} |_{t=0} = \left\langle \frac{x}{|x|}, v \right\rangle.$$

Es gibt aber wieder schlechte Nachrichten: selbst wenn in $x \in \Omega$ alle Richtungsableitungen existieren, kann die Funktion im Punkt x trotzdem unstetig sein.

Beispiel 2.4 Betrachte jetzt auf $\Omega = \mathbb{R}^2$ die Funktion

$$f(x, y) = \begin{cases} \frac{2xy^2}{x^2 + y^4} & (x, y) \neq (0, 0) \\ 0 & (x, y) = (0, 0). \end{cases}$$

Dann existieren im Punkt $(0,0)$ alle Richtungsableitungen, denn für $v = (a, b) \neq (0, 0)$ ist

$$\partial_v f(0, 0) = \lim_{t \rightarrow 0} \frac{2ab^2}{a^2 + t^2b^4} = \begin{cases} 2b^2/a & \text{für } a \neq 0 \\ 0 & \text{für } a = 0. \end{cases}$$

Dennoch ist f im Nullpunkt unstetig, denn für $c(t) = (t^2, t)$ gilt $f(c(t)) = 1$ für alle $t \neq 0$.

3 Die Ableitung

Das Konzept der mehrdimensionalen Ableitung beruht auf dem Ansatz, dass eine differenzierbare Funktion mit einer affin-linearen Funktion lokal in erster Ordnung übereinstimmt, siehe Analysis 1, Lemma 9.2. Zur Abgrenzung von den partiellen Ableitungen verwendet man auch den Begriff der totalen Ableitung. Wir betrachten hier Abbildungen zwischen \mathbb{R}^n und \mathbb{R}^m , den Raum der linearen Abbildungen bezeichnen wir mit $L(\mathbb{R}^n, \mathbb{R}^m)$. Die Definition ist aber auch für beliebige normierte Räumen X, Y sinnvoll, nur muss dann $L(X, Y)$ als Raum der stetigen linearen Abbildungen erklärt werden. Man spricht von Differenzierbarkeit im Sinne von Fréchet.

Definition 3.1 (Ableitung) Sei $\Omega \subset \mathbb{R}^n$ offen. $f : \Omega \rightarrow \mathbb{R}^m$ heißt differenzierbar in $x_0 \in \Omega$, falls es ein $A \in L(\mathbb{R}^n, \mathbb{R}^m)$ gibt, so dass gilt:

$$(3.1) \quad \lim_{x \rightarrow x_0} \frac{f(x) - (f(x_0) + A(x - x_0))}{|x - x_0|} = 0.$$

Mit der Substitution $h = x - x_0$ erhalten wir die äquivalente Fassung

$$(3.2) \quad \lim_{h \rightarrow 0} \frac{f(x_0 + h) - (f(x_0) + Ah)}{|h|} = 0.$$

Die Abbildung $A \in L(\mathbb{R}^n, \mathbb{R}^m)$ ist durch (3.1) eindeutig bestimmt, siehe Satz 3.1, und heißt Ableitung von f in x_0 . Notation: $Df(x_0) = A$.

Eine Basis kommt in der Definition nicht explizit vor. Zum Rechnen werden aber in aller Regel die Standardbasen benutzt. Eine lineare Abbildung $A \in L(\mathbb{R}^n, \mathbb{R}^m)$ hat dann eine zugehörige Matrix $a \in \mathbb{R}^{m \times n}$, und zwar gilt

$$Ax = \sum_{i=1}^m \sum_{j=1}^n a_{ij} x_j e_i.$$

Umgekehrt entspricht jeder Matrix $a \in \mathbb{R}^{m \times n}$ durch diese Formel eine lineare Abbildung $A \in L(\mathbb{R}^n, \mathbb{R}^m)$. Oft wird zwischen linearer Abbildung und Matrix gar nicht unterschieden.

Satz 3.1 (Berechnung und Eindeutigkeit der Ableitung) Die Funktion $f : \Omega \rightarrow \mathbb{R}^m$ sei in $x_0 \in \Omega$ differenzierbar. Dann hat f in x_0 die Richtungsableitungen

$$(3.3) \quad \partial_v f(x_0) = Df(x_0)v \quad \text{für alle } v \in \mathbb{R}^n,$$

und $Df(x_0)$ hat bezüglich der Standardbasen die Matrixdarstellung (Jacobimatrix)

$$(3.4) \quad (\partial_j f_i(x_0)) = \begin{pmatrix} \partial_1 f_1(x_0) & \dots & \dots & \partial_n f_1(x_0) \\ \vdots & & & \vdots \\ \vdots & & & \vdots \\ \partial_1 f_m(x_0) & \dots & \dots & \partial_n f_m(x_0) \end{pmatrix} \in \mathbb{R}^{m \times n}.$$

Insbesondere ist die Ableitung durch (3.1) eindeutig bestimmt.

BEWEIS: Für $v = 0$ sind beide Seiten von (3.3) nach Definition gleich Null. Für $v \neq 0$ berechnen wir mit $A = Df(x_0)$,

$$\begin{aligned} \left| \frac{f(x_0 + tv) - f(x_0)}{t} - Av \right| &= \frac{|f(x_0 + tv) - (f(x_0) + A(tv))|}{|t|} \\ &= \frac{|f(x_0 + tv) - (f(x_0) + A(tv))|}{|tv|} |v|. \end{aligned}$$

Für $t \rightarrow 0$ geht die rechte Seite gegen Null nach (3.1), also folgt $\partial_v f(x_0) = DF(x_0)v$. Setzen wir $v = e_j$ ein und berechnen die Ableitung komponentenweise, siehe Satz 2.1, so folgt weiter

$$Df(x_0)e_j = \partial_j f(x_0) = \sum_{i=1}^m \partial_j f_i(x_0)e_i.$$

□

Um die Differenzierbarkeit einer Funktion $f : \Omega \rightarrow \mathbb{R}^m$ im Punkt $x \in \Omega \subset \mathbb{R}^n$ zu zeigen, kann man in zwei Schritten vorgehen. Erstens berechnet man die Jacobimatrix, also die partiellen Ableitungen im Punkt x . Zweitens prüft man, ob die Entwicklung (3.1) gilt, wenn A die Jacobimatrix ist. Nach Satz 3.1 ist das die einzig mögliche Wahl.

Beispiel 3.1 Die komplexe Funktion $f(z) = z^2$ lautet in reellen Koordinaten

$$f : \mathbb{R}^2 \rightarrow \mathbb{R}^2, f(z) = \begin{pmatrix} x^2 - y^2 \\ 2xy \end{pmatrix} \quad \text{wobei } z = \begin{pmatrix} x \\ y \end{pmatrix}.$$

Wir schreiben die Punkte im \mathbb{R}^2 hier als Spaltenvektoren, zwecks Konsistenz mit der Notation der Jacobimatrix. Diese ist

$$A = \begin{pmatrix} 2x & -2y \\ 2y & 2x \end{pmatrix}.$$

Damit berechnen wir für $\zeta = \begin{pmatrix} \xi \\ \eta \end{pmatrix}$ nach (3.1) den Approximationsfehler

$$\begin{aligned} f(z + \zeta) - (f(z) + A\zeta) &= \begin{pmatrix} (x + \xi)^2 - (y + \eta)^2 \\ 2(x + \xi)(y + \eta) \end{pmatrix} \\ &\quad - \left(\begin{pmatrix} x^2 - y^2 \\ 2xy \end{pmatrix} + \begin{pmatrix} 2x & -2y \\ 2y & 2x \end{pmatrix} \begin{pmatrix} \xi \\ \eta \end{pmatrix} \right) \\ &= \begin{pmatrix} \xi^2 - \eta^2 \\ 2\xi\eta \end{pmatrix}. \end{aligned}$$

Mit $|\zeta| = (\xi^2 + \eta^2)^{1/2}$ ist die Norm rechts abgeschätzt durch $C|\zeta|^2$, also folgt

$$\frac{f(z + \zeta) - (f(z) + A\zeta)}{|\zeta|} \rightarrow 0 \quad \text{mit } \zeta \rightarrow 0.$$

Beispiel 3.2 (Lineare Abbildungen) Sei $\Omega \subset \mathbb{R}^n$ offen und $A \in L(\mathbb{R}^n, \mathbb{R}^m)$. Dann ist

$$f : \Omega \rightarrow \mathbb{R}^m, f(x) = Ax \quad \text{für alle } x \in \Omega,$$

in allen $x_0 \in \Omega$ differenzierbar mit Ableitung $Df(x_0) = A$. Dies folgt sofort wegen $f(x_0 + h) = A(x_0 + h) = Ax_0 + Ah = f(x_0) + Ah$.

Beispiel 3.3 (Quadratische Formen) Sei $b : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ symmetrische Bilinearform. Wir betrachten die quadratische Form

$$f : \mathbb{R}^n \rightarrow \mathbb{R}, f(x) = \frac{1}{2} b(x, x).$$

Um die Ableitung im Punkt $x \in \mathbb{R}^n$ zu bestimmen, entwickeln wir

$$f(x+h) = \frac{1}{2} b(x+h, x+h) = \underbrace{f(x) + b(x, h)}_{\text{affinlinear in } h} + \underbrace{\frac{1}{2} b(h, h)}_{\text{quadratisch in } h}.$$

Es folgt $Df(x)h = b(x, h)$, denn der Restterm hat die Abschätzung

$$|b(h, h)| \leq \sum_{i,j=1}^n |b(e_i, e_j)| |h_i| |h_j| \leq C|h|^2 \quad \text{mit } C = \sum_{i,j=1}^n |b(e_i, e_j)|.$$

Beispiel 3.4 (Funktionen einer Variablen) Natürlich muss das Konzept auch in diesem Fall Sinn machen. Die Funktion $f : I = (a, b) \rightarrow \mathbb{R}^m$ habe in $x \in I$ die Ableitung $f'(x) \in \mathbb{R}^m$ im Sinne von Analysis 1. Dann ist f differenzierbar in x im Sinne von Definition 3.1 mit

$$Df(x) : \mathbb{R} \rightarrow \mathbb{R}^m, \quad Df(x)h = f'(x)h.$$

Denn es gilt für $h \neq 0$, siehe auch Lemma 9.2 in Analysis 1,

$$\frac{|f(x+h) - (f(x) + f'(x)h)|}{|h|} = \left| \frac{f(x+h) - f(x)}{h} - f'(x) \right| \rightarrow 0 \quad \text{mit } h \rightarrow 0.$$

Für reelle Funktionen $f : \Omega \rightarrow \mathbb{R}$ ist $Df(x) \in L(\mathbb{R}^n, \mathbb{R})$, also Element des Dualraums von \mathbb{R}^n . Es ist anschaulicher, den zugehörigen Vektor im \mathbb{R}^n zu betrachten.

Definition 3.2 (Gradient) Sei $\Omega \subset \mathbb{R}^n$ offen und $f : \Omega \rightarrow \mathbb{R}$ differenzierbar in $x \in \Omega$. Der Gradient von f im Punkt x ist der Vektor

$$\text{grad } f(x) = \sum_{j=1}^n \partial_j f(x) e_j = \begin{pmatrix} \partial_1 f(x) \\ \vdots \\ \partial_n f(x) \end{pmatrix} \in \mathbb{R}^n.$$

Formal gehen wir vom Zeilenvektor $Df(x)$ zum Spaltenvektor $\text{grad } f(x)$ mit denselben Einträgen über. Eine Charakterisierung ohne Koordinaten ist wie folgt: der Gradient ist der eindeutig bestimmte Vektor im \mathbb{R}^n mit der Eigenschaft

$$(3.5) \quad \langle \text{grad } f(x), v \rangle = Df(x)v \quad \text{für alle } v \in \mathbb{R}^n.$$

Dabei ist $\langle \cdot, \cdot \rangle$ das Standardskalarprodukt. Ist $\text{grad } f(x) = 0$, so heißt x kritischer Punkt von f . Ist x nicht kritisch, so ist die Richtung von $\text{grad } f(x)$ diejenige, in der f am stärksten ansteigt. Denn für $v \in \mathbb{R}^n$ mit $|v| = 1$ folgt aus der Ungleichung von Cauchy-Schwarz

$$(3.6) \quad \partial_v f(x) = \langle \text{grad } f(x), v \rangle \leq |\text{grad } f(x)|, \quad \text{Gleichheit genau wenn } v = \frac{\text{grad } f(x)}{|\text{grad } f(x)|}.$$

Beispiel 3.5 Der Gradient der Funktion $f(x) = \varphi(r)$ mit $r(x) = |x|$ ist nach Beispiel 2.1

$$\text{grad } f(x) = \varphi'(r) \frac{x}{r} \quad \text{für } x \neq 0.$$

Beispiel 3.6 Sei $b : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ symmetrische Bilinearform und $B \in \mathbb{R}^{n \times n}$ die zugehörige Matrix, also $B_{ij} = b(e_i, e_j)$. Es gilt dann, da B symmetrisch,

$$b(x, y) = \langle Bx, y \rangle \quad \text{für alle } x, y \in \mathbb{R}^n.$$

Wir betrachten wieder die quadratische Form

$$f : \mathbb{R}^n \rightarrow \mathbb{R}, \quad f(x) = \frac{1}{2} b(x, x).$$

Nach Beispiel 3.3 gilt für alle $v \in \mathbb{R}^n$

$$\langle \text{grad } f(x), v \rangle = Df(x)v = b(x, v) = \langle Bx, v \rangle.$$

Also ist $\text{grad } f(x) = Bx$.

In Analysis 1 haben wir die Ableitung mit der Existenz der Tangente an den Graphen der Funktion motiviert. Im n -dimensionalen erwarten wir analog die Existenz einer n -dimensionalen Tangentialebene. Eine reellwertige Funktion f auf $\Omega \subset \mathbb{R}^n$ kann immer als Höhenfunktion einer Landschaft über der Grundfläche Ω interpretiert werden. Betrachte dazu den Graph der Funktion

$$G = \{(y, f(y)) : y \in \Omega\} \subset \Omega \times \mathbb{R} \subset \mathbb{R}^{n+1}.$$

Wir wollen zeigen, dass der Graph im Punkt $p = (x, f(x))$ eine Tangentialebene hat, wenn f im Punkt x differenzierbar ist. Betrachte dazu für $\lambda > 0$ die Mengen

$$G_{p,\lambda} = \frac{1}{\lambda}(G - p) = \left\{ \left(\frac{y-x}{\lambda}, \frac{f(y)-f(x)}{\lambda} \right) : y \in \Omega \right\}.$$

Der Graph G wird um $-p$ verschoben, wobei $p = (x, f(x))$ im Nullpunkt landet, dann wird mit dem Faktor $\frac{1}{\lambda}$ gestreckt. Wir wollen die $G_{p,\lambda}$ wieder als Graphen schreiben. Substituieren wir $y = x + \lambda z$, so folgt mit $\Omega_{x,\lambda} = \{z : x + \lambda z \in \Omega\}$

$$G_{p,\lambda} = \{(z, f_{x,\lambda}(z)) : z \in \Omega_{x,\lambda}\} \quad \text{für} \quad f_{x,\lambda}(z) = \frac{f(x + \lambda z) - f(x)}{\lambda}.$$

Da Ω offen, gibt es ein $\varepsilon > 0$ mit $B_\varepsilon(x) \subset \Omega$. Es folgt $B_R(0) \subset \Omega_{x,\lambda}$ für $\lambda < \frac{\varepsilon}{R}$. Für $\lambda > 0$ hinreichend klein ist $f_{x,\lambda}(z)$ also definiert, und es gilt

$$\lim_{\lambda \searrow 0} f_{x,\lambda}(z) = Df(x)z \quad \text{für alle } z \in \mathbb{R}^n.$$

In diesem Sinn konvergieren die reskalierten Graphen $G_{p,\lambda}$ gegen die Menge

$$T_p G = \{(z, Df(x)z) : z \in \mathbb{R}^n\} \subset \mathbb{R}^{n+1}.$$

$T_p G$ ist das Bild der linearen Abbildung $z \mapsto (z, Df(x)z)$, also linearer Unterraum von \mathbb{R}^{n+1} mit Basis $(e_1, \partial_1 f(x)), \dots, (e_n, \partial_n f(x))$. Einheitsnormale von $T_p G$ ist

$$\nu(p) = \frac{(-\text{grad } f(x), 1)}{\sqrt{1 + |\text{grad } f(x)|^2}} \quad \text{für } p = (x, f(x)).$$

Im Beweis der Differentiationsregeln brauchen wir eine Abschätzung für lineare Abbildungen $A \in L(\mathbb{R}^n, \mathbb{R}^m)$ aus Analysis 1, Beispiel 7.10. Und zwar hatten wir mit Cauchy-Schwarz

$$(3.7) \quad |Ax| = \left| \sum_{j=1}^n x_j A e_j \right| \leq \left(\sum_{j=1}^n |A e_j|^2 \right)^{1/2} \left(\sum_{j=1}^n x_j^2 \right)^{1/2} = |A| |x|.$$

Dabei bezeichnet $|A| = \left(\sum_{j=1}^n |A e_j|^2 \right)^{1/2}$ die Euklidische Norm von $A \in L(\mathbb{R}^n, \mathbb{R}^m)$. Es folgt, dass jede lineare Abbildung $A : \mathbb{R}^n \rightarrow \mathbb{R}^m$ Lipschitzstetig ist mit Konstante $|A|$, vgl. Analysis 1, Beispiel 7.10:

$$|Ax - Ay| = |A(x - y)| \leq |A| |x - y|.$$

Die optimale, also kleinstmögliche Norm $\|A\|$ mit einer Abschätzung (3.7) heißt Operatornorm. Für uns reicht die Euklidische Norm aus, die Optimalität spielt keine Rolle. Wird \mathbb{R}^n durch einen unendlichdimensionalen Raum ersetzt, so gilt (3.7) im allgemeinen nicht und lineare Abbildungen sind dann nicht automatisch stetig.

Satz 3.2 (Differenzierbarkeit \Rightarrow Stetigkeit) Sei $\Omega \subset \mathbb{R}^n$ offen. Ist $f : \Omega \rightarrow \mathbb{R}^m$ differenzierbar in x_0 , so ist f stetig in x_0 .

BEWEIS: Wie soeben besprochen, sind affinere Funktionen stetig auf \mathbb{R}^n . Es reicht daher zu zeigen, dass die Funktion $\varphi(x) = f(x) - (f(x_0) + Df(x_0)(x - x_0))$ stetig in x_0 ist. Aber $\varphi(x_0) = 0$, und nach Definition der Differenzierbarkeit gilt

$$\varphi(x) = |x - x_0| \frac{\varphi(x)}{|x - x_0|} \rightarrow 0 \quad \text{mit } x \rightarrow x_0.$$

□

Wir müssen jetzt die Differentiationsregeln erarbeiten.

Satz 3.3 (Kettenregel) Seien $f : U \rightarrow \mathbb{R}^m$ und $g : V \rightarrow \mathbb{R}^p$ mit $U \subset \mathbb{R}^n$, $V \subset \mathbb{R}^m$ offen und $f(U) \subset V$. Sind f in x_0 und g in $f(x_0)$ differenzierbar, so ist auch $g \circ f$ in x_0 differenzierbar und es gilt die Kettenregel

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

Für die zugehörigen Jacobimatrizen bedeutet das mit $y_0 = f(x_0)$

$$\frac{\partial (g \circ f)_i}{\partial x_k}(x_0) = \sum_{j=1}^m \frac{\partial g_i}{\partial y_j}(y_0) \frac{\partial f_j}{\partial x_k}(x_0) \quad \text{für } 1 \leq i \leq p, 1 \leq k \leq n.$$

BEWEIS: Sei $y_0 = f(x_0)$, $Df(x_0) = A$, $Dg(y_0) = B$. Wir definieren für hinreichend kleine $\xi \in \mathbb{R}^n \setminus \{0\}$, $\eta \in \mathbb{R}^m \setminus \{0\}$ die Funktionen

$$\varepsilon_f(\xi) = \frac{f(x_0 + \xi) - (f(x_0) + A\xi)}{|\xi|} \quad \text{und} \quad \varepsilon_g(\eta) = \frac{g(y_0 + \eta) - (g(y_0) + B\eta)}{|\eta|}.$$

Mit $\varepsilon_f(0) = 0$ und $\varepsilon_g(0) = 0$ sind beide Funktionen nach Voraussetzung im Nullpunkt stetig. Offensichtliche Kandidatin für die Ableitung von $g \circ f$ in x_0 ist BA , also berechnen wir

$$\begin{aligned} & \frac{(g \circ f)(x_0 + \xi) - ((g \circ f)(x_0) + BA\xi)}{|\xi|} \\ &= \frac{g(y_0 + A\xi + |\xi|\varepsilon_f(\xi)) - (g(y_0) + BA\xi)}{|\xi|} \\ &= \frac{g(y_0) + B\eta + |\eta|\varepsilon_g(\eta) - (g(y_0) + BA\xi)}{|\xi|} \quad \text{wobei } \eta = A\xi + |\xi|\varepsilon_f(\xi) \\ &= B\varepsilon_f(\xi) + \frac{|\eta|}{|\xi|}\varepsilon_g(\eta). \end{aligned}$$

Wegen $|B\varepsilon_f(\xi)| \leq |B||\varepsilon_f(\xi)|$ und $|\eta| \leq (|A| + |\varepsilon_f(\xi)|)|\xi| \leq C|\xi|$ konvergiert die rechte Seite wie gewünscht gegen Null. \square

Beispiel 3.7 Spezialfall ist die Verkettung $f \circ c$ einer Kurve $c : (a, b) \rightarrow \Omega \subset \mathbb{R}^n$ und einer Funktion $f : \Omega \rightarrow \mathbb{R}$. Ist c differenzierbar in $t \in (a, b)$ und f differenzierbar in $c(t)$, so folgt

$$\frac{d(f \circ c)}{dt}(t) = \sum_{j=1}^n \frac{\partial f}{\partial x_j}(c(t)) \frac{dc_j}{dt}(t),$$

beziehungsweise in vektorieller Form

$$(f \circ c)'(t) = Df(c(t))c'(t) = \langle \text{grad } f(c(t)), c'(t) \rangle.$$

Ist $f \circ c$ konstant, so folgt $\text{grad } f(c(t)) \perp c'(t)$. Anschaulich: der Gradient von f steht senkrecht auf Kurven in der Niveaumenge $\{x \in \Omega : f(x) = \text{const.}\}$, also auf die ganze Niveaumenge. Im Fall $n = 2$ kann man sich die Niveaumenge als Höhenlinie vorstellen.

Wie bei Funktionen einer Variablen kann die Ableitung vektorwertiger Funktionen auf die einzelnen Komponenten zurückgeführt werden.

Satz 3.4 (komponentenweise Differentiation) $f : \Omega \rightarrow \mathbb{R}^m$ ist genau dann in $x_0 \in \Omega$ differenzierbar, wenn alle Komponenten $f_i : \Omega \rightarrow \mathbb{R}$, $i = 1, \dots, m$, in x_0 differenzierbar sind. Ist $P_i : \mathbb{R}^m \rightarrow \mathbb{R}$ Projektion auf die i -te Koordinate, so gilt $Df_i(x_0) = P_i Df(x_0)$.

BEWEIS: Es gilt nach Definition

$$Df(x_0) = A \quad \Leftrightarrow \quad \lim_{x \rightarrow x_0} \frac{f(x) - (f(x_0) + A(x - x_0))}{|x - x_0|} = 0.$$

Die Konvergenz im \mathbb{R}^n ist gleichbedeutend mit der Konvergenz aller Komponenten. Durch Anwendung von P_i ergibt sich daher weiter die äquivalente Formulierung

$$(3.8) \quad \lim_{x \rightarrow x_0} \frac{f_i(x) - (f_i(x_0) + P_i A(x - x_0))}{|x - x_0|} = 0 \quad \text{für alle } i = 1, \dots, m.$$

Aus $Df(x_0) = A$ folgt somit $Df_i(x_0) = P_i A$. Sei umgekehrt $Df_i(x_0) = A_i$ für $i = 1, \dots, m$. Wir definieren $A \in L(\mathbb{R}^n, \mathbb{R}^m)$ durch $Av = \sum_{i=1}^m (A_i v) e_i$. Dann ist $P_i A = A_i$, also gilt (3.8) und somit $Df(x_0) = A$. \square

Wir zeigen schließlich die weiteren klassischen Ableitungsregeln.

Satz 3.5 (Ableitungsregeln) Sei $\Omega \subset \mathbb{R}^n$ offen, und $f, g : \Omega \rightarrow \mathbb{R}$ seien differenzierbar im Punkt $x \in \Omega$. Dann gelten folgende Aussagen:

(a) Linearität: für $\alpha, \beta \in \mathbb{R}$ ist $\alpha f + \beta g$ in x differenzierbar mit Ableitung

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

(b) Produktregel: fg ist in x differenzierbar mit Ableitung

$$D(fg)(x) = Df(x)g(x) + f(x)Dg(x).$$

(c) Quotientenregel: ist $g(x) \neq 0$, so ist f/g auf einer Umgebung von x definiert und

$$D\left(\frac{f}{g}\right)(x) = \frac{Df(x)g(x) - f(x)Dg(x)}{g(x)^2}.$$

BEWEIS: Wir setzen $Df(x) = A$, $Dg(x) = B$, und für $h \neq 0$

$$\varepsilon_f(h) = \frac{f(x+h) - (f(x) + Ah)}{|h|}, \quad \varepsilon_g(h) = \frac{g(x+h) - (g(x) + Bh)}{|h|}.$$

Nach Voraussetzung gilt $\varepsilon_f(h) \rightarrow 0$, $\varepsilon_g(h) \rightarrow 0$ mit $h \rightarrow 0$. Mit der jeweils behaupteten Ableitung ist nun für $h \rightarrow 0$ der Grenzwert in (3.2) nachzuprüfen. Für (a) gilt

$$\frac{(\alpha f + \beta g)(x+h) - ((\alpha f + \beta g)(x) + (\alpha A + \beta B)h)}{|h|} = \alpha \varepsilon_f(h) + \beta \varepsilon_g(h) \rightarrow 0.$$

Für (b) berechnen wir mit etwas mehr Mühe

$$\begin{aligned} & \frac{(fg)(x+h) - ((fg)(x) + (Ag(x) + f(x)B)h)}{|h|} \\ &= \frac{(f(x) + Ah + \varepsilon_f(h)|h|)(g(x) + Bh + \varepsilon_g(h)|h|) - (f(x)g(x) + g(x)Ah + f(x)Bh)}{|h|} \\ &= \frac{1}{|h|} (Ah)(Bh) + \varepsilon_f(h)(g(x) + Bh + \varepsilon_g(h)|h|) + \varepsilon_g(h)(f(x) + Ah). \end{aligned}$$

Wie in (3.7) bemerkt gilt $|Ah| \leq |A||h|$ sowie $|Bh| \leq |B||h|$, also geht die rechte Seite mit $h \rightarrow 0$ gegen Null. In (c) können wir $m = 1$ und $f \equiv 1$ annehmen, denn sonst schreiben wir $f/g = f(1/g)$ und verwenden (b). Es gilt

$$\begin{aligned} & \frac{1}{|h|} \left(\frac{1}{g(x+h)} - \left(\frac{1}{g(x)} - \frac{Bh}{g(x)^2} \right) \right) \\ &= \frac{1}{|h|} \frac{1}{g(x)g(x+h)} \left(g(x) - (g(x) + Bh + \varepsilon_g(h)|h|) + \frac{g(x+h)}{g(x)} Bh \right) \\ &= \frac{1}{g(x)g(x+h)} \left(\left(\frac{g(x+h)}{g(x)} - 1 \right) \frac{Bh}{|h|} - \varepsilon_g(h) \right). \end{aligned}$$

Wegen $g(x) \neq 0$ und $g(x+h) \rightarrow g(x)$ mit $h \rightarrow 0$ nach Satz 3.2 geht die rechte Seite wieder gegen Null mit $h \rightarrow 0$. \square

Die Quotientenregel kann auch eleganter mit der Kettenregel gezeigt werden: man verwendet $1/g = h \circ g$ mit $h(y) = \frac{1}{y}$. Für die Produktregel gibt es ein ähnliches Argument: es ist $fg = h \circ \phi$ mit $\phi(x) = (f(x), g(x)) \in \mathbb{R}^2$ und $h(y_1, y_2) = y_1 y_2$. Nach Satz 3.4 ist ϕ differenzierbar, und h nach Beispiel 3.3.

Wie besprochen kann aus der Existenz der partiellen Ableitungen nicht auf die Differenzierbarkeit geschlossen werden, ja nicht einmal auf die Stetigkeit. Das ist schade, denn die partiellen Ableitungen sind so schön einfach auszurechnen, während die Definition 3.1 eventuell schwierig zu verifizieren ist. Zum Glück können wir aber doch die partiellen Ableitungen einsetzen.

Satz 3.6 (stetig partiell differenzierbar \Rightarrow differenzierbar) Sei $\Omega \subset \mathbb{R}^n$ offen. Die Funktion $f : \Omega \rightarrow \mathbb{R}^m$ sei in Ω nach x_1, \dots, x_n partiell differenzierbar. Sind die Funktionen $\partial_j f : \Omega \rightarrow \mathbb{R}^m$ in $x \in \Omega$ stetig, so ist f in x differenzierbar.

BEWEIS: Wegen Satz 3.4 können wir $m = 1$ annehmen. Mit Satz 3.1 kennen wir bereits die einzig mögliche Kandidatin für die Ableitung, nämlich

$$A : \mathbb{R}^n \rightarrow \mathbb{R}, \quad Ah = \sum_{k=1}^n \partial_k f(x) h_k.$$

Für $h \in \mathbb{R}^n$ hinreichend klein ist $f(x+h)$ mit $f(x) + Ah$ zu vergleichen, dazu wollen wir den Mittelwertsatz verwenden. Da wir nur in Achsenrichtungen differenzieren können, laufen wir längs der Kanten des Quaders, das heißt wir betrachten die Punkte $p_k = x + \sum_{i=1}^k h_i e_i$ mit $k = 0, \dots, n$. Es gilt für geeignete $s_k \in [0, 1]$

$$f(p_k) - f(p_{k-1}) = f(p_{k-1} + h_k e_k) - f(p_{k-1}) = \partial_k f(p_{k-1} + s_k h_k e_k) h_k.$$

Es folgt nun

$$\begin{aligned} \frac{|f(x+h) - (f(x) + Ah)|}{|h|} &= \frac{1}{|h|} \left| \sum_{k=1}^n (f(p_k) - f(p_{k-1}) - \partial_k f(x) h_k) \right| \\ &= \frac{1}{|h|} \left| \sum_{k=1}^n \left(\partial_k f(p_{k-1} + s_k h_k e_k) - \partial_k f(x) \right) h_k \right| \\ &\leq \sum_{k=1}^n \left| \partial_k f \left(x + \sum_{i=1}^{k-1} h_i e_i + s_k h_k e_k \right) - \partial_k f(x) \right|. \end{aligned}$$

Die rechte Seite geht mit $h \rightarrow 0$ gegen Null, da $\partial_k f$ im Punkt x stetig ist. \square

Es gibt differenzierbare Funktionen, die nicht stetig differenzierbar sind. In Analysis 1, Serie 13, Aufgabe 4 hatten wir das Beispiel

$$f : \mathbb{R} \rightarrow \mathbb{R}, \quad f(x) = \begin{cases} x^2 \cos \frac{1}{x} & \text{für } x \neq 0, \\ 0 & \text{für } x = 0. \end{cases}$$

Aber hier gilt: Ausnahmen bestätigen die Regel, in den meisten Fällen ist Satz 3.6 das Mittel der Wahl, um die Differenzierbarkeit einer Funktion zu begründen. Dabei ist hilfreich, dass die Ableitungsregeln auch in der Klasse der C^k -Funktionen gelten.

Folgerung 3.1 Sei $k \in \mathbb{N}_0 \cup \{\infty\}$.

- (a) Mit $f, g \in C^k(\Omega, \mathbb{R}^m)$ gilt $\alpha f + \beta g \in C^k(\Omega, \mathbb{R}^m)$ für $\alpha, \beta \in \mathbb{R}$.
- (b) Aus $f, g \in C^k(\Omega)$ folgt $fg \in C^k(\Omega)$, sowie $f/g \in C^k(\Omega)$ falls $g \neq 0$ auf Ω .
- (c) Sind $f \in C^k(U, \mathbb{R}^m)$, $g \in C^k(V, \mathbb{R}^p)$ mit $U \subset \mathbb{R}^n$, $V \subset \mathbb{R}^m$ offen und $f(U) \subset V$, so ist $g \circ f \in C^k(U, \mathbb{R}^p)$.

BEWEIS: Im Fall $k = 0$ sind die Aussagen wohlbekannt. Die Behauptungen (a) und (b) folgen nun aus den Rechenregeln für die partielle Ableitung, siehe Satz 2.1, mit Induktion über k . Sind zum Beispiel $f, g \in C^k(\Omega)$ für ein $k \geq 1$, so gilt induktiv $\partial_j(fg) = (\partial_j f)g + f(\partial_j g) \in C^{k-1}(\Omega)$, also $fg \in C^k(\Omega)$.

Für $k \geq 1$ sind die Abbildungen f und g aus (c) differenzierbar nach Satz 3.6. Dann ist $g \circ f$ ebenfalls differenzierbar wegen der Kettenregel, Satz 3.3, mit partiellen Ableitungen

$$\frac{\partial(g \circ f)_i}{\partial x_k} = \sum_{j=1}^m \frac{\partial g_i}{\partial y_j} \circ f \frac{\partial f_j}{\partial x_k}.$$

Nun sind $\partial_k f_j \in C^{k-1}(U)$, $\partial_j g_i \in C^{k-1}(V)$ nach Voraussetzung, also $\partial_j g_i \circ f \in C^{k-1}(U)$ nach Induktion. Es folgt $\partial_k(g \circ f)_i \in C^{k-1}(U)$ mit der Produktregel aus (b), also ist $g \circ f$ von der Klasse C^k . □

4 Erste Anwendungen der Differentialrechnung

Ein Grundproblem der Analysis ist es, Eigenschaften einer Funktion aus Informationen über ihre Ableitung zu gewinnen. Für Funktionen einer Variablen, also $f : [a, b] \rightarrow \mathbb{R}$, stehen dazu zwei Argumente zur Verfügung:

a) der Mittelwertsatz (Analysis 1, Kapitel 10):

$$f(b) - f(a) = f'(\tau)(b - a) \quad \text{für ein } \tau \in (a, b);$$

b) der Hauptsatz der Differential- und Integralrechnung (Analysis 1, Kapitel 14):

$$f(b) - f(a) = \int_a^b f'(t) dt.$$

Ein Nachteil des Mittelwertsatzes ist, dass er für vektorwertige Funktionen so nicht gilt. Dies zeigt das Beispiel $f(t) = (\cos t, \sin t)$ auf $[0, 2\pi]$. Ein zweiter Nachteil im Hinblick auf Anwendungen ist mit der Tatsache verbunden, dass die Zwischenstelle im allgemeinen nicht eindeutig ist. Hängt die Funktion f von weiteren Variablen ab, so erlaubt der Satz keine Kontrolle über die Abhängigkeit der Zwischenstelle.

Im folgenden verwenden wir meistens den Hauptsatz, wobei dann $f(t)$ eine C^1 -Funktion sein sollte. Genauer reicht es wenn $f(t)$ stetig auf $[a, b]$ und stückweise C^1 ist. Denn sei $a = t_0 < \dots < t_N = b$ eine Unterteilung, so dass f' auf den offenen Teilintervallen (t_{i-1}, t_i) stetig ist und in den Randpunkten einseitige Grenzwerte $f'_\pm(t_i)$ hat. Dann gilt

$$(4.1) \quad f(b) - f(a) = \sum_{k=1}^N (f(t_k) - f(t_{k-1})) = \sum_{k=1}^N \int_{t_{k-1}}^{t_k} f'(t) dt = \int_a^b f'(t) dt.$$

Wie kann das eindimensionale Argument nun für Funktionen mehrerer Variabler $f : \Omega \rightarrow \mathbb{R}$ mit $\Omega \subset \mathbb{R}^n$ eingesetzt werden? Die einfache Antwort: indem f längs Kurven $\gamma : [a, b] \rightarrow \Omega$, $\gamma = \gamma(t)$, ausgewertet wird.

Lemma 4.1 Sei $\gamma : I = [a, b] \rightarrow \Omega$ stetig und stückweise C^1 . Dann gilt für $f \in C^1(\Omega, \mathbb{R}^m)$

$$(4.2) \quad f(\gamma(b)) - f(\gamma(a)) = \int_a^b Df(\gamma(t))\gamma'(t) dt.$$

BEWEIS: Nach Folgerung 3.1 ist $f \circ \gamma$ stückweise C^1 , also folgt aus (4.1) und der Kettenregel

$$f(\gamma(b)) - f(\gamma(a)) = \int_a^b \frac{d}{dt} f(\gamma(t)) dt = \int_a^b Df(\gamma(t))\gamma'(t) dt.$$

□

Im Beweis trat das Integral einer vektorwertigen Funktion auf. Dieses kann komponentenweise erklärt werden, das heißt für $v \in C^0([a, b], \mathbb{R}^m)$ ist

$$\int_a^b v(t) dt = \int_a^b \left(\sum_{i=1}^m v_i(t) e_i \right) dt = \sum_{i=1}^m \left(\int_a^b v_i(t) dt \right) e_i.$$

Alternativ kann man prüfen, dass die Definition des Integrals in Analysis 1, Kapitel 13, mittels Riemannscher Summen ohne Änderung auch für Funktionen mit Werten im \mathbb{R}^m funktioniert. Man kann sich so oder so davon überzeugen, dass der Hauptsatz der Differential- und Integralrechnung ganz analog für vektorwertige Funktionen gilt.

Satz 4.1 (Konstanzsatz) Für $f : \Omega \rightarrow \mathbb{R}^m$, Ω offen und zusammenhängend, gilt:

$$Df(x) = 0 \quad \text{für alle } x \in \Omega \quad \Rightarrow \quad f \text{ ist konstant.}$$

BEWEIS: Zu $x \in \Omega$ wähle $\varrho > 0$ mit $B_\varrho(x) \subset \Omega$. Zu $y \in B_\varrho(x)$ betrachte $\gamma : [0, 1] \rightarrow B_\varrho(x)$, $\gamma(t) = x + t(y - x)$. Mit Lemma 4.1 folgt

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

Also ist f lokal konstant, und dann nach Lemma 1.2 konstant. \square

Wir wollen nicht nur die Konstanz von Funktionen zeigen, sondern ähnlich wie im Eindimensionalen auch Wachstumsabschätzungen. Mit der bloßen Existenz von Verbindungswegen lässt sich dann nichts anfangen, eine quantitative Kontrolle ist notwendig. Der häufigste und einfachste Fall ist, wenn wir die gerade Strecke nehmen können.

Definition 4.1 Eine Menge $M \subset \mathbb{R}^n$ heißt konvex, falls folgende Implikation gilt:

$$x_0, x_1 \in M \quad \Rightarrow \quad (1 - t)x_0 + tx_1 \in M \quad \text{für alle } t \in [0, 1].$$

Satz 4.2 (Schrankensatz) Sei $\Omega \subset \mathbb{R}^n$ offen und konvex, und $f \in C^1(\Omega, \mathbb{R}^m)$. Es gebe ein $L < \infty$ mit $|Df(x)| \leq L$ für alle $x \in \Omega$. Dann folgt

$$|f(x_1) - f(x_0)| \leq L|x_1 - x_0| \quad \text{für alle } x_0, x_1 \in \Omega.$$

BEWEIS: Für jede stetige Funktion $\varphi : I = [a, b] \rightarrow \mathbb{R}^m$ gilt die Ungleichung

$$(4.3) \quad \left| \int_a^b \varphi \right| \leq \int_a^b |\varphi|.$$

Dies folgt durch Anwendung der Dreiecksungleichung auf die Riemannschen Summen. Sei nun $\gamma(t) = (1 - t)x_0 + tx_1$ für $0 \leq t \leq 1$. Aus (4.2) und (3.7) folgt, da $\gamma'(t) = x_1 - x_0$,

$$|f(x_1) - f(x_0)| = \left| \int_0^1 Df(\gamma(t))(x_1 - x_0) dt \right| \leq \int_0^1 |Df(\gamma(t))(x_1 - x_0)| dt \leq L|x_1 - x_0|.$$

\square

Die folgende lokale Variante des Schrankensatzes ist ebenfalls oft nützlich.

Folgerung 4.1 Sei $\Omega \subset \mathbb{R}^n$ offen und $f \in C^1(\Omega, \mathbb{R}^m)$. Dann gibt es zu jeder kompakten Menge $K \subset \Omega$ eine Konstante $L < \infty$ mit

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

BEWEIS: Angenommen nicht, dann gibt es zu jedem $k \in \mathbb{N}$ Punkte $x_k, y_k \in K$ mit

$$|f(x_k) - f(y_k)| > k|x_k - y_k| \quad \text{für } k = 1, 2, \dots$$

Da f stetig auf der kompakten Menge K ist, gibt es ein $M < \infty$ mit $|f(x)| \leq M$ für alle $x \in K$ nach Satz 1.8. Weiter können wir nach Wahl einer Teilfolge und Umnummerierung annehmen, dass $x_k \rightarrow x \in K$ mit $k \rightarrow \infty$. Aber

$$|x_k - y_k| < \frac{1}{k} |f(x_k) - f(y_k)| \leq \frac{2M}{k} \rightarrow 0 \quad \text{mit } k \rightarrow \infty,$$

also folgt auch $y_k \rightarrow x$ mit $k \rightarrow \infty$. Wähle nun ein $r > 0$ mit $\overline{B_r(x)} \subset \Omega$. Da Df stetig ist, gibt es wieder nach Satz 1.8 ein $L < \infty$ mit

$$|Df(y)| \leq L \quad \text{für alle } y \in \overline{B_r(x)}.$$

Für hinreichend große k gilt $x_k, y_k \in B_r(x)$, also liefert Satz 4.2

$$k|x_k - y_k| < |f(x_k) - f(y_k)| \leq L|x_k - y_k|,$$

ein Widerspruch für k hinreichend groß. □

Als nächstes beschäftigen wir uns mit Extremwerten, und verallgemeinern die notwendigen und hinreichenden Kriterien aus Analysis 1. Dabei spielt die zweite Ableitung eine entscheidende Rolle. Wir behandeln im Anschluss Grundtatsachen über konvexe Funktionen. Als bekannt setzen wir voraus: auf einer kompakten Teilmenge des \mathbb{R}^n nimmt eine stetige Funktion ihre Extremwerte an.

Definition 4.2 Die Funktion $f : M \rightarrow \mathbb{R}$, $M \subset \mathbb{R}^n$, hat in $x \in M$ ein lokales Minimum, falls es ein $\delta > 0$ gibt mit

$$f(y) \geq f(x) \quad \text{für alle } y \in B_\delta(x) \cap M.$$

Ist sogar $f(y) > f(x)$ für $y \in B_\delta(x) \setminus \{x\}$, so heißt das Minimum isoliert. Ein (isoliertes) lokales Maximum ist entsprechend definiert.

Satz 4.3 (notwendige Bedingung für Extrema) Sei $\Omega \subset \mathbb{R}^n$ offen, und $f : \Omega \rightarrow \mathbb{R}$ habe in $x \in \Omega$ ein lokales Extremum. Ist f differenzierbar in x , so folgt $Df(x) = 0$.

BEWEIS: Für $v \in \mathbb{R}^n$ hat die Funktion $t \mapsto f(x + tv)$ ein lokales Extremum bei $t = 0$, also folgt aus der eindimensionalen Version und Satz 3.1

$$0 = \frac{d}{dt} f(x + tv)|_{t=0} = Df(x)v \quad \text{für alle } v \in \mathbb{R}^n.$$

□

Definition 4.3 Ein Punkt $x \in \Omega$ mit $Df(x) = 0$ heißt kritischer Punkt von f .

Kritische Punkte sind damit Kandidaten für Extremalstellen. Es gibt aber kritische Punkte, in denen weder ein Maximum noch ein Minimum vorliegt, das zeigt schon das eindimensionale Beispiel $f(x) = x^3$ im Punkt $x = 0$. Um die Situation genauer zu analysieren brauchen wir die zweite Ableitung.

Definition 4.4 Sei $f \in C^2(\Omega)$ mit $\Omega \subset \mathbb{R}^n$ offen. Die zweite Ableitung von f im Punkt $x \in \Omega$ ist die Bilinearform

$$D^2 f(x) : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}, D^2 f(x)(v, w) = \sum_{i,j=1}^n \partial_{ij}^2 f(x) v_i w_j.$$

Die Matrix $(\partial_{ij}^2 f(x)) \in \mathbb{R}^{n \times n}$ heißt Hessematrix von f an der Stelle x , und als Hesseform bezeichnet man die zugehörige quadratische Form

$$v \mapsto D^2 f(x)(v, v) = \sum_{i,j=1}^n \partial_{ij}^2 f(x) v_i v_j.$$

Die Hessematrix ist symmetrisch, und $D^2 f(x)$ ist symmetrische Bilinearform. Denn nach Schwarz, Satz 2.2, gilt $\partial_{ij}^2 f = \partial_{ji}^2 f$ für $f \in C^2(\Omega)$, und daraus folgt

$$D^2 f(x)(v, w) = \sum_{i,j=1}^n \partial_{ij}^2 f(x) v_i w_j = \sum_{i,j=1}^n \partial_{ji}^2 f(x) w_j v_i = D^2 f(x)(w, v).$$

Als erstes wollen wir die Formel für die zweite Ableitung längs Kurven herleiten.

Lemma 4.2 Sei $\Omega \subset \mathbb{R}^n$ offen, $f \in C^2(\Omega)$ und $\gamma \in C^2(I, \Omega)$. Dann gilt

$$(4.4) \quad (f \circ \gamma)''(t) = D^2 f(\gamma(t))(\gamma'(t), \gamma'(t)) + Df(\gamma(t))\gamma''(t).$$

BEWEIS: Nach Kettenregel ist $(f \circ \gamma)'(t) = \sum_{j=1}^n \partial_j f(\gamma(t))\gamma'_j(t)$, und weiter

$$(f \circ \gamma)''(t) = \sum_{i,j=1}^n \partial_{ij}^2 f(\gamma(t))\gamma'_i(t)\gamma'_j(t) + \sum_{j=1}^n \partial_j f(\gamma(t))\gamma''_j(t).$$

□

Wir benötigen nun eine lokale Entwicklung, die die zweite Ableitung mit einbezieht.

Lemma 4.3 Sei $\Omega \subset \mathbb{R}^n$ offen und $f \in C^2(\Omega)$. Dann gilt

$$\frac{f(x+h) - (f(x) + Df(x)h + \frac{1}{2}D^2 f(x)(h, h))}{|h|^2} \rightarrow 0 \quad \text{mit } h \rightarrow 0.$$

BEWEIS: Setze $\gamma(t) = x + th$, das heißt nach Lemma 4.2 gilt

$$(f \circ \gamma)'(t) = Df(x+th)h \quad \text{und} \quad (f \circ \gamma)''(t) = D^2 f(x+th)(h, h).$$

Wir berechnen mit dem Hauptsatz und partieller Integration

$$\begin{aligned} (f \circ \gamma)(1) &= (f \circ \gamma)(0) + \int_0^1 (f \circ \gamma)'(t) dt \\ &= (f \circ \gamma)(0) + (f \circ \gamma)'(0) + \int_0^1 (1-t)(f \circ \gamma)''(t) dt. \end{aligned}$$

Einsetzen von $\gamma(t) = x + th$ liefert

$$(4.5) \quad f(x+h) = f(x) + Df(x)h + \frac{1}{2}D^2f(x)(h, h) + \int_0^1 (1-t)(D^2f(x+th) - D^2f(x))(h, h) dt.$$

Wir schätzen den Integranden ab. Nach Cauchy-Schwarz gilt für $Q \in \mathbb{R}^{n \times n}$

$$\left| \sum_{i,j=1}^n Q_{ij}h_i h_j \right| = \left| \sum_{i=1}^n (Qh)_i h_i \right| = |\langle Qh, h \rangle| \leq |Qh| |h| \leq |Q| |h|^2.$$

Zu $\varepsilon > 0$ gibt es ein $\delta > 0$ mit $|D^2f(y) - D^2f(x)| < \varepsilon$ für $|y - x| < \delta$. Es folgt

$$|(D^2f(x+th) - D^2f(x))(h, h)| \leq \varepsilon |h|^2 \quad \text{für } |h| < \delta.$$

Damit ist das Lemma bewiesen. □

Als zweites Hilfsmittel brauchen wir folgende Tatsache über quadratische Formen.

Satz 4.4 (Grundzustand) Sei $b : V \times V \rightarrow \mathbb{R}$ symmetrische Bilinearform auf dem n -dimensionalen Euklidischen Vektorraum $(V, \langle \cdot, \cdot \rangle)$, und

$$\lambda = \inf\{b(x, x) : \|x\| = 1\}.$$

Dann gibt es ein $v \in V$ mit $\|v\| = 1$ so dass $b(v, v) = \lambda$.

BEWEIS: Wir zeigen die Aussage erst im Fall $V = \mathbb{R}^n$, mit dem Standardskalarprodukt. Die Funktion $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $f(x) = b(x, x)$, ist stetig auf \mathbb{R}^n , denn es gilt

$$f(x) = \sum_{i,j=1}^n b_{ij}x_i x_j \quad \text{wobei } b_{ij} = b(e_i, e_j).$$

Die Menge $\{x \in \mathbb{R}^n : |x| = 1\}$ ist kompakt, also wird das Infimum durch ein $v \in \mathbb{R}^n$ mit $|v| = 1$ angenommen. Den allgemeinen Fall führen wir auf \mathbb{R}^n zurück, indem wir in V eine Orthonormalbasis $\mathcal{B} = \{v_1, \dots, v_n\}$ wählen. Für $x \in \mathbb{R}^n$ setze $x_{\mathcal{B}} = \sum_{i=1}^n x_i v_i$, und betrachte die symmetrische Bilinearform

$$b_{\mathcal{B}} : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}, \quad b_{\mathcal{B}}(x, y) = b(x_{\mathcal{B}}, y_{\mathcal{B}}).$$

Wegen $\|x_{\mathcal{B}}\| = |x|$ folgt mit Substitution

$$\inf\{b(v, v) : \|v\| = 1\} = \inf\{b_{\mathcal{B}}(x, x) : |x| = 1\}.$$

Das Infimum wird rechts in einem Punkt x angenommen, also links im Punkt $v = x_{\mathcal{B}}$. □

Definition 4.5 Eine symmetrische Bilinearform $b : V \times V \rightarrow \mathbb{R}$ heißt positiv definit (bzw. positiv semidefinit), falls gilt:

$$b(v, v) > 0 \quad (\text{bzw. } b(v, v) \geq 0) \quad \text{für alle } v \in V \setminus \{0\}.$$

Notation: $b > 0$ bzw. $b \geq 0$. Entsprechend für negativ (semi-)definit.

Beachten Sie, dass *definit* die strikte Ungleichung bedeutet, anders als zum Beispiel bei der Monotonie von Funktionen, wo wir zum Ausschluss der Gleichheit den Begriff *streng monoton* verwenden. Wir bemerken auch, dass es sich nur um eine teilweise Ordnung handelt, es muss nicht einer der Fälle $b \geq 0$ oder $b \leq 0$ gelten. Für $b(x, y) = x_1y_1 - x_2y_2$ auf \mathbb{R}^2 gilt zum Beispiel $b(e_1, e_1) > 0$, aber $b(e_2, e_2) < 0$.

Satz 4.5 (Lokale Extrema) Sei $f \in C^2(\Omega)$, $\Omega \subset \mathbb{R}^n$ offen und $x \in \Omega$.

- (a) Wenn f in x ein lokales Minimum hat, so ist $D^2f(x)$ positiv semidefinit.
- (b) Ist $Df(x) = 0$ und $D^2f(x)$ positiv definit, so hat f in x ein isoliertes lokales Minimum.

BEWEIS: In (a) gilt $Df(x) = 0$ nach Satz 4.3. Für $v \in \mathbb{R}^n$ beliebig hat $t \mapsto f(x + tv)$ bei $t = 0$ ein lokales Minimum, also folgt aus dem eindimensionalen Fall und (4.4)

$$0 \leq \frac{d^2}{dt^2} f(x + tv)|_{t=0} = D^2f(x)(v, v).$$

Für (b) verwende (4.5): zu $\varepsilon > 0$ gibt es ein $\delta > 0$ mit

$$f(x + h) - f(x) = \frac{1}{2} D^2f(x)(h, h) + R(h) \quad \text{mit } |R(h)| < \varepsilon|h|^2 \text{ für } |h| < \delta.$$

Nach Voraussetzung ist $D^2f(x)(h, h) > 0$ für $h \neq 0$, wir brauchen aber hier eine quantitative Version. Nach Satz 4.4 gibt es ein $v \in \mathbb{R}^n$, $|v| = 1$, mit

$$\lambda := \inf_{|w|=1} D^2f(x)(w, w) = D^2f(x)(v, v) > 0.$$

Wir wählen $\varepsilon < \frac{\lambda}{2}$. Mit dem zugehörigen $\delta > 0$ gilt

$$f(x + h) - f(x) \geq \frac{\lambda}{2}|h|^2 - \varepsilon|h|^2 = \left(\frac{\lambda}{2} - \varepsilon\right)|h|^2 > 0 \text{ für } 0 < |h| < \delta.$$

□

Um die Funktion in der Nähe eines kritischen Punkts zu verstehen, ist der folgende Satz aus der Linearen Algebra nützlich.

Satz 4.6 (Hauptachsentransformation) Sei $b : V \times V \rightarrow \mathbb{R}$ symmetrische Bilinearform auf dem Euklidischen Vektorraum $(V, \langle \cdot, \cdot \rangle)$. Dann gibt es eine Orthonormalbasis v_1, \dots, v_n und $\lambda_1 \leq \dots \leq \lambda_n$, so dass gilt:

$$b(v_i, v_j) = \lambda_i \delta_{ij} \quad \text{für alle } i, j = 1, \dots, n.$$

BEWEIS: Setze $\lambda = \inf\{b(x, x) : \|x\| = 1\}$ und wähle $v \in V$ mit $\|v\| = 1$ und $b(v, v) = \lambda$, siehe Satz 4.4. Wir behaupten

$$(4.6) \quad b(v, w) = \lambda \langle v, w \rangle \quad \text{für alle } w \in V.$$

Die Gleichung stimmt für $w = v$, und die Menge der $w \in V$ mit (4.6) ist ein Unterraum. Es reicht daher, die Gleichung für $w \in V$ mit $\langle v, w \rangle = 0$ und $\|w\| = 1$ zu zeigen. Dann ist $\|(\cos t)v + (\sin t)w\|^2 = 1$ für alle $t \in \mathbb{R}$, und aus der Minimumeigenschaft folgt wie behauptet

$$b(v, w) = \frac{1}{2} \frac{d}{dt} b((\cos t)v + (\sin t)w, (\cos t)v + (\sin t)w)|_{t=0} = 0 = \lambda \langle v, w \rangle.$$

Jetzt konstruieren wir induktiv $\lambda_1 \leq \dots \leq \lambda_n$ und v_1, \dots, v_n orthonormal mit

$$b(v_i, w) = \lambda_i \langle v_i, w \rangle \quad \text{für alle } w \in V.$$

Mit $w = v_j$ ist das die Behauptung des Satzes. Für $k = 1$ nehmen wir $\lambda_1 = \lambda$ und $v_1 = v$ wie oben. Seien nun v_1, \dots, v_k und $\lambda_1 \leq \dots \leq \lambda_k$ schon bestimmt für $1 \leq k \leq n - 1$. Setze

$$\lambda_{k+1} = \inf\{b(x, x) : x \in V_k, \|x\| = 1\} \quad \text{mit } V_k = \{v_1, \dots, v_k\}^\perp.$$

Es gilt $V_{k+1} \subset V_k$ und damit $\lambda_{k+1} \geq \lambda_k$. Nach Satz 4.4, angewandt im Raum $V = V_k$, gibt es ein $v_{k+1} \in V_k$ mit $\|v_{k+1}\| = 1$ und $b(v_{k+1}, v_{k+1}) = \lambda_{k+1}$. Aus (4.6) folgt

$$b(v_{k+1}, w) = \lambda_{k+1} \langle v_{k+1}, w \rangle \quad \text{für alle } w \in V_k.$$

Aber da b symmetrisch, gilt induktiv für $1 \leq i \leq k$

$$b(v_{k+1}, v_i) = b(v_i, v_{k+1}) = \lambda_i \langle v_i, v_{k+1} \rangle = 0 = \lambda_{k+1} \langle v_{k+1}, v_i \rangle.$$

Es folgt $b(v_{k+1}, w) = \lambda_{k+1} \langle v_{k+1}, w \rangle$ für alle $w \in V$, der Induktionsschluss. \square

Jede lineare Abbildung $B : V \rightarrow V$ induziert die Bilinearform

$$b(v, w) = \langle Bv, w \rangle \quad \text{für } v, w \in V.$$

B heißt symmetrisch, wenn b symmetrisch ist. Wegen $b(v, w) - \lambda \langle v, w \rangle = \langle Bv - \lambda v, w \rangle$ ist die Gleichung (4.6) gleichbedeutend mit

$$Bv = \lambda v,$$

das heißt v ist Eigenvektor von B zum Eigenwert $\lambda \in \mathbb{R}$. Satz 4.6 besagt: symmetrische Endomorphismen sind diagonalisierbar, genauer gibt es eine Orthonormalbasis aus Eigenvektoren.

Diese Konzepte haben eine unendlichdimensionale Verallgemeinerung, den Spektralsatz, der zum Beispiel in der Quantenmechanik von Bedeutung ist. Der zugrundeliegende Vektorraum ist dort ein Raum von Funktionen, welche die möglichen Zustände eines quantenmechanischen Systems beschreiben. Anstelle des Endomorphismus B tritt der Differentialoperator $H = -\Delta + V(x)$, der sogenannte Hamiltonoperator. Die Eigenfunktion zum kleinsten Eigenwert heißt Grundzustand, die weiteren Eigenfunktionen sind die angeregten Zustände. Eine ähnliche Situation hat man bei den Schwingungen einer Saite, mit Grund- und Obertönen.

Definition 4.6 (konvexe Funktion) Sei $K \subset \mathbb{R}^n$ konvex. Dann heißt $f : K \rightarrow \mathbb{R}$ konvex, falls für alle $x, y \in K$ gilt:

$$f((1-t)x + ty) \leq (1-t)f(x) + tf(y) \quad \text{für alle } t \in [0, 1].$$

f heißt strikt konvex, falls die strikte Ungleichung gilt für $x \neq y$ und $t \in (0, 1)$.

Wie man leicht sieht, ist Konvexität von f äquivalent dazu, dass der Epigraph

$$G^+(f) = \{(x, z) \in K \times \mathbb{R} : z \geq f(x)\}$$

eine konvexe Menge im $\mathbb{R}^{n+1} = \mathbb{R}^n \times \mathbb{R}$ ist.

Satz 4.7 (Konvexitätskriterien) Sei $\Omega \subset \mathbb{R}^n$ offen und konvex, und $f \in C^1(\Omega)$. Die folgenden Aussagen sind äquivalent:

- (a) f ist konvex.
- (b) $f(y) \geq f(x) + Df(x)(y - x)$ für alle $x, y \in \Omega$.
- (c) $(Df(y) - Df(x))(y - x) \geq 0$ für alle $x, y \in \Omega$.

Ist sogar $f \in C^2(\Omega)$, so ist außerdem äquivalent:

- (d) $D^2f(x) \geq 0$ für alle $x \in \Omega$.

BEWEIS: Die Aussage wird jeweils auf den eindimensionalen Fall reduziert, indem wir für $x, y \in \Omega$ die Funktion $\varphi(t) = (1 - t)f(x) + tf(y) - f((1 - t)x + ty)$ betrachten. Unter Voraussetzung (a) hat φ in $t = 0$ ein Minimum, daraus folgt (b):

$$0 \leq \varphi'(0) = f(y) - f(x) - Df(x)(y - x).$$

Aussage (c) folgt aus (b) durch Vertauschen von x und y und Addition. Die Implikation (c) \Rightarrow (a) zeigen wir durch Widerspruch. Angenommen $\varphi(t)$ hat in $\tau \in (0, 1)$ ein Minimum $\varphi(\tau) < 0$. Für $t_1 < t_2$ gilt nach (c) mit $x(t) = (1 - t)x + ty$

$$\varphi'(t_1) - \varphi'(t_2) = \frac{1}{t_2 - t_1} (Df(x(t_2)) - Df(x(t_1)))(x(t_2) - x(t_1)) \geq 0.$$

Für $t < \tau$ folgt $\varphi'(t) \geq \varphi'(\tau) = 0$, und hieraus $\varphi(0) \leq \varphi(\tau) < 0$, ein Widerspruch.

Sei nun $f \in C^2(\Omega)$. Nach (4.5) wissen wir

$$f(x + h) = f(x) + Df(x)h + \int_0^1 (1 - t)D^2f(x + th)(h, h) dt.$$

Mit $h = y - x$ folgt die die Implikation (d) \Rightarrow (b). Umgekehrt folgt (d) aus (b) mit Satz 4.5, denn die Funktion $g(y) = f(y) - (f(x) + Df(x)(y - x))$ hat in x ein Minimum. \square

Eine Funktion f mit $f((1 - t)x + ty) \geq (1 - t)f(x) + tf(y)$ für alle $x, y \in \Omega$, $t \in [0, 1]$, heißt konkav und es gelten entsprechende Aussagen mit umgekehrten Ungleichungen.

Das letzte Thema in diesem Kapitel ist die Taylorentwicklung, und zwar erst im Fall einer Variablen und im Anschluss auch im \mathbb{R}^n . Es geht dabei um den Vergleich einer gegebenen Funktion $f(x)$ mit einem Polynom $P(x)$, das in einer Stelle x_0 „von k -ter Ordnung“ mit f übereinstimmt, also $f^{(i)}(x_0) = P^{(i)}(x_0)$ für $i = 0, 1, \dots, k$. $P(x)$ sollte dann $f(x)$ asymptotisch für $x \rightarrow x_0$ gut approximieren, und das will quantifiziert werden. Für $k = 0, 1, 2$, also konstante, lineare sowie quadratische Polynome, haben wir sowas schon gesehen.

Zur Erinnerung: eine Funktion $P : \mathbb{R} \rightarrow \mathbb{R}$ heißt Polynom vom Grad $k \in \mathbb{N}_0$, wenn es $a_0, \dots, a_k \in \mathbb{R}$ gibt mit $a_k \neq 0$, so dass gilt:

$$P(x) = \sum_{j=0}^k a_j x^j \quad \text{für alle } x \in \mathbb{R}.$$

Im Raum aller Funktionen $f : \mathbb{R} \rightarrow \mathbb{R}$ ist die Menge \mathbb{P}_k der Polynome vom Grad $\leq k$ der durch $1, x, \dots, x^k$ erzeugte Unterraum. Es gilt: für jedes $x_0 \in \mathbb{R}$ bilden die Funktionen $1, x - x_0, \dots, (x - x_0)^k$ eine Basis von \mathbb{P}_k . Wegen $\dim \mathbb{P}_k \leq k + 1$ müssen wir nur die lineare Unabhängigkeit zeigen. Dazu verwenden wir die Ableitungsregel

$$(4.7) \quad P(x) = \sum_{j=0}^k a_j (x - x_0)^j \quad \Rightarrow \quad \left(\frac{d}{dx} \right)^i P(x)|_{x=x_0} = i! a_i.$$

Ist $P(x)$ die Nullfunktion, so folgt $a_i = 0$ für $i = 0, \dots, k$ wie behauptet.

Lemma 4.4 Sei $I = (a, b) \subset \mathbb{R}$, $x_0 \in I$ und $k \in \mathbb{N}_0$. Zu $f \in C^k(I)$ gibt es genau ein Polynom $P \in \mathbb{P}_k$ mit $P^{(i)}(x_0) = f^{(i)}(x_0)$ für $i = 0, 1, \dots, k$, und zwar

$$(4.8) \quad P_k(x) = \sum_{j=0}^k \frac{f^{(j)}(x_0)}{j!} (x - x_0)^j.$$

P_k heißt Taylorpolynom der Ordnung k von f mit Entwicklungspunkt x_0 .

BEWEIS: Für das in (4.8) definierte Polynom gilt $P^{(i)}(x_0) = f^{(i)}(x_0)$ für $i = 0, \dots, k$, wie man mit (4.7) sieht. Zur Eindeutigkeit sei $P \in \mathbb{P}_k$ mit $P^{(i)}(x_0) = 0$ für alle $i = 0, \dots, k$. Wie oben gezeigt gilt eine Darstellung $P(x) = \sum_{j=0}^k a_j (x - x_0)^j$ für alle $x \in \mathbb{R}$. Mit (4.7) folgt $a_i = 0$ für alle $i = 0, \dots, k$. \square

Folgerung 4.2 Das k -te Taylorpolynom mit Entwicklungspunkt x_0 eines Polynoms f vom Grad höchstens k ist f selbst.

In der Situation von Lemma 4.4 heißt die Funktion

$$(4.9) \quad R_k : (a, b) \rightarrow \mathbb{R}, \quad R_k(x) = f(x) - P_k(x)$$

das Restglied k -ter Ordnung der Taylorentwicklung in x_0 . Knackpunkt bei der Taylorentwicklung ist die Abschätzung dieses Restglieds und damit eine Aussage darüber, wie gut die Funktion durch das Taylorpolynom approximiert wird. Hierfür gibt es verschiedene mögliche Darstellungen von R_k .

Satz 4.8 (Integraldarstellung des Restglieds) Sei $f \in C^{k+1}(I)$ für ein $k \in \mathbb{N}_0$, und $P_k(x) = \sum_{j=0}^k \frac{f^{(j)}(x_0)}{j!} (x - x_0)^j$ das k -te Taylorpolynom im Punkt $x_0 \in I$. Dann gilt

$$f(x) = P_k(x) + R_k(x) \quad \text{mit} \quad R_k(x) = \frac{1}{k!} \int_{x_0}^x (x - y)^k f^{(k+1)}(y) dy.$$

BEWEIS: Durch Induktion über $k \in \mathbb{N}_0$. Für $k = 0$ folgt aus dem Hauptsatz

$$f(x) = f(x_0) + \int_{x_0}^x f'(y) dy.$$

Für $k \geq 1$ folgt induktiv mit partieller Integration, vgl. Lemma 4.3 für den Fall $k = 1$,

$$\begin{aligned} f(x) &= P_{k-1}(x) + \frac{1}{(k-1)!} \int_{x_0}^x (x-y)^{k-1} f^{(k)}(y) dy \\ &= P_{k-1}(x) + \frac{1}{(k-1)!} \left(\left[-\frac{(x-y)^k}{k} f^{(k)}(y) \right]_{y=x_0}^{y=x} + \int_{x_0}^x \frac{(x-y)^k}{k} f^{(k+1)}(y) dy \right) \\ &= P_k(x) + \frac{1}{k!} \int_{x_0}^x (x-y)^k f^{(k+1)}(y) dy. \end{aligned}$$

□

Die zweite Darstellung des Restglieds ist vielleicht etwas populärer.

Satz 4.9 (Lagrange-Darstellung des Restglieds) Sei $f \in C^{k+1}(I)$ für ein $k \in \mathbb{N}_0$. Dann gibt es zu $x_0, x \in I$ ein ξ zwischen x_0 und x , so dass gilt:

$$(4.10) \quad f(x) = \sum_{j=0}^k \frac{f^{(j)}(x_0)}{j!} (x - x_0)^j + R_k(x) \quad \text{mit} \quad R_k(x) = \frac{f^{(k+1)}(\xi)}{(k+1)!} (x - x_0)^{k+1}.$$

BEWEIS: Wir verwenden den Mittelwertsatz der Integralrechnung mit Gewicht, siehe Folgerung 13.2, Analysis 1: ist $\varphi \in C^0(I)$ mit $\varphi \geq 0$, so gibt es zu $f \in C^0(I)$ ein $\xi \in I$ mit

$$\int_I f \varphi = f(\xi) \int_I \varphi.$$

Sei nun $x > x_0$. Dann können wir $I = [x_0, x]$ und $\varphi(y) = (x - y)^k$ wählen. Es folgt

$$\frac{1}{k!} \int_{x_0}^x (x - y)^k f^{(k+1)}(y) dy = \frac{1}{k!} f^{(k+1)}(\xi) \int_{x_0}^x (x - y)^k dy = \frac{f^{(k+1)}(\xi)}{(k+1)!} (x - x_0)^{k+1},$$

für ein $\xi \in [x_0, x]$. Der Fall $x < x_0$ ist analog, der Satz ist bewiesen. □

Beispiel 4.1 Betrachte für $x \in (-1, 1)$ die Funktion $f(x) = (1 - x)^{-1/2}$, mit Ableitungen

$$f'(x) = \frac{1}{2}(1 - x)^{-3/2} \quad \text{und} \quad f''(x) = \frac{3}{4}(1 - x)^{-5/2}.$$

Es gilt $f(0) = 1$ und $f'(0) = 1/2$, also lautet das Taylorpolynom der Ordnung Eins in $x_0 = 0$

$$P_1(x) = f(0) + f'(0)x = 1 + \frac{1}{2}x,$$

mit der Lagrange-Restglieddarstellung

$$R_1(x) = \frac{f''(\xi)}{2} x^2 = \frac{3}{8}(1 - \xi)^{-5/2} x^2 \quad \text{für ein } \xi \in [0, x].$$

Als Anwendung erhalten wir für die relativistische Energie eines Teilchens mit Ruhemasse m_0 und Geschwindigkeit v , wenn wir $\beta = v/c$ setzen,

$$E = \frac{m_0 c^2}{\sqrt{1 - \beta^2}} = m_0 c^2 \left(1 + \frac{1}{2} \beta^2 + \frac{f''(\xi)}{2} \beta^4 \right) = m_0 c^2 + \frac{1}{2} m_0 v^2 + \Delta E.$$

Dabei ist der erste Term die Ruheenergie und der zweite die klassische kinetische Energie. Für den relativistischen Korrekturterm ergibt sich aus der Restglieddarstellung die Abschätzung

$$\frac{\Delta E}{E_{kin}} = f''(\xi) \beta^2 \leq f''(\beta^2) \beta^2 < 0,008 \quad \text{für } \beta \leq 0,1.$$

Bei Geschwindigkeiten $v \leq \frac{1}{10}c$ beträgt die relativistische Korrektur weniger als ein Prozent der klassischen kinetischen Energie.

Um das asymptotische Verhalten von Funktionen bei Grenzprozessen zu vergleichen, werden oft die Landauschen Symbole \mathcal{O} und o benutzt. Seien f, g zwei Funktionen, die für $0 < |x - x_0| < \delta$ definiert sind, und es sei $g(x) \neq 0$ für x nahe bei x_0 . Dann schreibt man

$$\begin{aligned} f(x) = o(g(x)) \text{ für } x \rightarrow x_0 &\Leftrightarrow \lim_{x \rightarrow x_0} \frac{|f(x)|}{|g(x)|} = 0, \\ f(x) = \mathcal{O}(g(x)) \text{ für } x \rightarrow x_0 &\Leftrightarrow \limsup_{x \rightarrow x_0} \frac{|f(x)|}{|g(x)|} < \infty. \end{aligned}$$

In Worten: die Funktion $f(x)$ ist klein- o von $g(x)$ beziehungsweise groß- \mathcal{O} von $g(x)$ für $x \rightarrow x_0$. Diese Begriffe sind analog für Grenzwerte $|x| \rightarrow \infty$ usw. erklärt. Wir zeigen jetzt, dass sich $f(x)$ für $x \rightarrow x_0$ asymptotisch wie das k -te Taylorpolynom $P_k(x)$ verhält, bis auf einen Term der von der Ordnung k verschwindet:

$$f(x) = P_k(x) + o(|x - x_0|^k) \quad \text{für } x \rightarrow x_0.$$

Satz 4.10 (Approximation durch das Taylorpolynom) Sei $f \in C^k(I)$ für $k \in \mathbb{N}_0$, und P_k das k -te Taylorpolynom von f mit Entwicklungspunkt $x_0 \in I$. Dann ist P_k das eindeutig bestimmte Polynom vom Grad höchstens k mit

$$\lim_{x \rightarrow x_0} \frac{f(x) - P_k(x)}{(x - x_0)^k} = 0.$$

BEWEIS: Nach Satz 4.9 gibt es zu $x \in I$ ein ξ zwischen x_0 und x mit

$$\frac{f(x) - P_k(x)}{(x - x_0)^k} = \frac{f(x) - P_{k-1}(x)}{(x - x_0)^k} - \frac{1}{k!} f^{(k)}(x_0) = \frac{1}{k!} (f^{(k)}(\xi) - f^{(k)}(x_0)).$$

Da $f^{(k)}$ stetig, ist $|f^{(k)}(\xi) - f^{(k)}(x_0)| < \varepsilon$ für $|x - x_0| < \delta$, womit die Konvergenz gegen Null bewiesen ist. Für die Eindeutigkeit ist zu zeigen, dass für $P(x) = \sum_{j=0}^k a_j (x - x_0)^j$ gilt:

$$\lim_{x \rightarrow x_0} \frac{P(x)}{(x - x_0)^k} = 0 \quad \Rightarrow \quad a_0 = \dots = a_k = 0.$$

Sei induktiv schon $a_0 = \dots = a_{j-1} = 0$ gezeigt mit $0 \leq j \leq k$. Dann folgt

$$a_j = \lim_{x \rightarrow x_0} (x - x_0)^{-j} P(x) = \lim_{x \rightarrow x_0} (x - x_0)^{k-j} (x - x_0)^{-k} P(x) = 0.$$

□

Die mehrdimensionale Taylorentwicklung orientiert sich am Fall $n = 1$, nur ist der Notationsaufwand größer. Sei $\Omega \subset \mathbb{R}^n$ offen und konvex. Für $f \in C^k(\Omega)$ definieren wir die k -te Ableitung $D^k f(x)$ im Punkt $x \in \Omega$ als k -Linearform $D^k f(x) : \mathbb{R}^n \times \dots \times \mathbb{R}^n \rightarrow \mathbb{R}$, wobei

$$(4.11) \quad D^k f(x)(v_1, \dots, v_k) = \sum_{i_1, \dots, i_k=1}^n (\partial_{i_1 \dots i_k}^k f)(x)(v_1)_{i_1} \dots (v_k)_{i_k}.$$

Betrachte jetzt für $x_0, x \in \Omega$ die C^k -Funktion, vgl. Folgerung 3.1,

$$\varphi : [0, 1] \rightarrow \mathbb{R}, \quad \varphi(t) = f(x_0 + th) \quad \text{mit } h = x - x_0.$$

Wir zeigen durch Induktion die Formel

$$(4.12) \quad \varphi^{(k)}(t) = D^k f(x_0 + th)(h, \dots, h).$$

Für $k = 1$ gilt das nach Kettenregel und Satz 3.1, denn

$$\varphi'(t) = Df(x_0 + th)h = \sum_{i=1}^n \partial_i f(x_0 + th)h_i.$$

Für $k \geq 2$ ergibt sich induktiv mit Satz 2.2 von Schwarz

$$\begin{aligned} \varphi^{(k)}(t) &= \frac{d}{dt} \sum_{i_1, \dots, i_{k-1}=1}^n (\partial_{i_1 \dots i_{k-1}}^{k-1} f)(x_0 + th) h_{i_1} \dots h_{i_{k-1}} \\ &= \sum_{i_1, \dots, i_{k-1}=1}^n \sum_{i=1}^n (\partial_{i_1 \dots i_{k-1} i}^k f)(x_0 + th) h_{i_1} \dots h_{i_{k-1}} h_i \\ &= D^k f(x_0 + th)(h, \dots, h). \end{aligned}$$

Satz 4.9, angewandt auf die Funktion φ , liefert sofort eine erste Fassung der mehrdimensionalen Taylorentwicklung.

Lemma 4.5 Sei $\Omega \subset \mathbb{R}^n$ offen und konvex, und sei $f \in C^{k+1}(\Omega)$ für ein $k \in \mathbb{N}_0$. Dann gibt es zu $x_0, x \in \Omega$ ein $\xi = (1 - \tau)x_0 + \tau x$, $\tau \in [0, 1]$, so dass mit $h = x - x_0$ gilt:

$$f(x) = \sum_{j=0}^k \frac{D^j f(x_0)(h, \dots, h)}{j!} + \frac{D^{k+1} f(\xi)(h, \dots, h)}{(k+1)!}.$$

BEWEIS: Wir wenden auf die C^{k+1} -Funktion $\varphi(t) = f(x_0 + th)$ die eindimensionale Taylorsche Formel an, mit Entwicklungspunkt $t_0 = 0$. Nach Satz 4.9 gibt es ein $\tau \in [0, 1]$ mit

$$\varphi(1) = \sum_{j=0}^k \frac{\varphi^{(j)}(0)}{j!} + \frac{\varphi^{(k+1)}(\tau)}{(k+1)!}.$$

Einsetzen von (4.12) liefert die Behauptung. □

Die k -te Ableitung $D^k f(x)(h, \dots, h)$ ist eine Summe von n^k Termen, von denen viele aber gleich sind wegen der Vertauschbarkeit der partiellen Ableitungen. Es ist ökonomischer, die Summe danach zu ordnen, wie oft nach den einzelnen Variablen x_1, \dots, x_n differenziert wird. Gleichzeitig führt das wie im Eindimensionalen auf eine Taylordarstellung mit Basispolynomen. Für einen Multiindex $\alpha = (\alpha_1, \dots, \alpha_n) \in \mathbb{N}_0^n$ setzen wir

$$\begin{aligned} |\alpha| &= \alpha_1 + \dots + \alpha_n && \text{Ordnung von } \alpha, \\ \alpha! &= (\alpha_1)! \cdot \dots \cdot (\alpha_n)! && \alpha\text{-Fakultät}, \\ x^\alpha &= x_1^{\alpha_1} \cdot \dots \cdot x_n^{\alpha_n} && \text{Monom mit Exponent } \alpha, \\ D^\alpha &= \partial_1^{\alpha_1} \dots \partial_n^{\alpha_n} && (D^0 = \text{Id}). \end{aligned}$$

Im Operator D^α wird also α_i mal nach x_i differenziert.

Satz 4.11 (Taylorentwicklung im \mathbb{R}^n) Sei $\Omega \subset \mathbb{R}^n$ offen und konvex, und sei $f \in C^{k+1}(\Omega)$ für ein $k \in \mathbb{N}_0$. Dann gibt es zu $x_0, x \in \Omega$ ein $\xi = (1 - \tau)x_0 + \tau x$, $\tau \in [0, 1]$, so dass gilt:

$$f(x) = \sum_{|\alpha| \leq k} \frac{D^\alpha f(x_0)}{\alpha!} (x - x_0)^\alpha + \sum_{|\alpha|=k+1} \frac{D^\alpha f(\xi)}{\alpha!} (x - x_0)^\alpha.$$

BEWEIS: Sei α Multiindex der Ordnung $|\alpha| = k$. Wieviele Tupel (i_1, \dots, i_k) gibt es, in denen jedes $i \in \{1, \dots, n\}$ genau α_i mal vorkommt? Wähle α_1 Stellen für $i = 1$, aus den übrigen α_2 Stellen für $i = 2$, etc. Das ergibt die Zahl

$$\binom{k}{\alpha_1} \cdot \binom{k - \alpha_1}{\alpha_2} \cdot \dots \cdot \binom{k - (\alpha_1 + \dots + \alpha_{n-1})}{\alpha_n} = \frac{k!}{(\alpha_1)! \dots (\alpha_n)!} = \frac{k!}{\alpha!}.$$

Die behauptete Entwicklung folgt nun aus Lemma 4.5 und der Vertauschbarkeit der partiellen Ableitungen, siehe Folgerung 2.1. \square

Sei $I(n, k)$ die Menge aller n -Multiindizes mit $|\alpha| = k$. Es gilt dann

$$I(n + 1, k) = \bigcup_{\ell=0}^k \{(\alpha, \ell) : \alpha \in I(n, k - \ell)\}.$$

Also gilt $\#I(n + 1, k) = \sum_{\ell=0}^k \#I(n, k - \ell)$, und induktiv folgt leicht $\#I(n, k) \leq (k + 1)^{n-1}$. Für große k ist das viel kleiner als die Zahl n^k aus der vorigen Darstellung der Taylorformel.

Eine Funktion $P : \mathbb{R}^n \rightarrow \mathbb{R}$ heißt Polynom vom Grad $k \geq 0$, wenn es $a_\alpha \in \mathbb{R}$, $|\alpha| \leq k$, gibt mit $a_\alpha \neq 0$ für mindestens ein $|\alpha| = k$, so dass gilt:

$$P(x) = \sum_{|\alpha| \leq k} a_\alpha x^\alpha \quad \text{für alle } x \in \mathbb{R}^n.$$

Für $x_0 \in \mathbb{R}^n$ beliebig bilden die Monome $(x - x_0)^\alpha$ mit $0 \leq |\alpha| \leq k$ eine Basis des Raums \mathbb{P}_k der Polynome vom Grad $\leq k$. Dies folgt wie für $n = 1$ aus der Ableitungsregel

$$P(x) = \sum_{|\beta| \leq k} a_\beta (x - x_0)^\beta \quad \Rightarrow \quad D^\alpha P(x_0) = \alpha! a_\alpha \quad \text{für } |\alpha| \leq k.$$

Es folgt analog zu Lemma 4.4: das k -te Taylorpolynom

$$(4.13) \quad P_k(x) = \sum_{|\alpha| \leq k} \frac{D^\alpha f(x_0)}{\alpha!} (x - x_0)^\alpha$$

ist das eindeutige Polynom vom Grad höchstens k mit $D^\alpha P_k(x_0) = D^\alpha f(x_0)$ für $|\alpha| \leq k$.

Folgerung 4.3 Das k -te Taylorpolynom mit Entwicklungspunkt x_0 eines Polynoms f vom Grad höchstens k ist f selbst.

Beispiel 4.2 (Polynomialformel) Die Funktion $f(x) = (x_1 + \dots + x_n)^k$ ist ein Polynom vom Grad k , und es gilt

$$D^\alpha f(0) = \begin{cases} k! & \text{falls } |\alpha| = k, \\ 0 & \text{sonst.} \end{cases}$$

Mit Folgerung 4.3 (oder direkt durch Abzählen) ergibt sich

$$(x_1 + \dots + x_n)^k = \sum_{|\alpha|=k} \frac{k!}{\alpha!} x^\alpha.$$

Auch im Mehrdimensionalen approximiert das k -te Taylorpolynom bis auf Terme höherer Ordnung, nur müssen wir jetzt im Nenner Beträge setzen, da bekanntlich durch Vektoren nicht dividiert werden kann.

Satz 4.12 (Approximation durch das Taylorpolynom im \mathbb{R}^n) Sei $f \in C^k(\Omega)$ für $k \in \mathbb{N}_0$, und P_k das k -te Taylorpolynom von f mit Entwicklungspunkt $x_0 \in \Omega$. Dann ist P_k das eindeutig bestimmte Polynom vom Grad höchstens k mit

$$\lim_{x \rightarrow x_0} \frac{f(x) - P_k(x)}{|x - x_0|^k} = 0.$$

BEWEIS: Nach Satz 4.11, mit k statt $k + 1$, gibt es zu $x \in \Omega$ ein ξ zwischen x_0 und x mit

$$f(x) - P_k(x) = \sum_{|\alpha|=k} \frac{D^\alpha f(\xi) - D^\alpha f(x_0)}{\alpha!} (x - x_0)^\alpha.$$

Da $D^\alpha f$ stetig und $|(x - x_0)^\alpha| \leq |x - x_0|^k$, folgt die Konvergenz gegen Null. Für die Eindeutigkeit zeigen wir für ein beliebiges Polynom $P(x)$ vom Grad $\leq k$ folgende Implikation:

$$\lim_{x \rightarrow x_0} \frac{P(x)}{|x - x_0|^k} = 0 \quad \Rightarrow \quad P(x) = 0 \text{ für alle } x \in \mathbb{R}^n.$$

Angenommen es gibt $x_1 \in \mathbb{R}^n$ mit $P(x_1) \neq 0$, oBdA $x_1 \neq x_0$. Mit $x_t = x_0 + t(x_1 - x_0)$, $t \in \mathbb{R}$, ist $\varphi(t) = P(x_t)$ eindimensionales Polynom vom Grad $\leq k$, und wegen $\lim_{t \rightarrow 0} x_t = x_0$ folgt

$$\frac{\varphi(t)}{|t|^k} = |x_1 - x_0|^k \frac{P(x_t)}{|t(x_1 - x_0)|^k} = |x_1 - x_0|^k \frac{P(x_t)}{|x_t - x_0|^k} \rightarrow 0 \quad \text{mit } t \rightarrow 0.$$

Nach Beweis von Satz 4.10 ist $\varphi(t) \equiv 0$, im Widerspruch zu $\varphi(1) = P(x_1) \neq 0$. □

Beispiel 4.3 Wir berechnen hier mit der Multiindexnotation die Taylorentwicklung erster Ordnung im Punkt $(1, 1)$ für

$$f(x, y) = \frac{x - y}{x + y}.$$

Es ist $f(1, 1) = 0$, und die partiellen Ableitungen der Funktion lauten

$$\begin{aligned} D^{(1,0)} f(x, y) &= \frac{2y}{(x+y)^2} & D^{(0,1)} f(x, y) &= -\frac{2x}{(x+y)^2} \\ D^{(2,0)} f(x, y) &= -\frac{4y}{(x+y)^3} & D^{(1,1)} f(x, y) &= \frac{2(x-y)}{(x+y)^3} & D^{(0,2)} f(x, y) &= \frac{4x}{(x+y)^3}. \end{aligned}$$

Das Taylorpolynom erster Ordnung ist somit

$$\begin{aligned} P_1(x, y) &= f(1, 1) + D^{(1,0)} f(1, 1)((x, y) - (1, 1))^{(1,0)} + D^{(0,1)} f(1, 1)((x, y) - (1, 1))^{(0,1)} \\ &= \frac{1}{2}(x - 1) - \frac{1}{2}(y - 1) = \frac{1}{2}(x - y). \end{aligned}$$

Das Restglied lautet in Lagrangedarstellung mit Zwischenpunkt (ξ, η)

$$\begin{aligned} R_1(x, y) &= \frac{D^{(2,0)}f(\xi, \eta)}{2!0!} ((x, y) - (1, 1))^{(2,0)} + \frac{D^{(1,1)}f(\xi, \eta)}{1!1!} ((x, y) - (1, 1))^{(1,1)} \\ &\quad + \frac{D^{(0,2)}f(\xi, \eta)}{0!2!} ((x, y) - (1, 1))^{(0,2)} \\ &= \frac{2}{(\xi + \eta)^3} (-\eta(x-1)^2 + (\xi - \eta)(x-1)(y-1) + \xi(y-1)^2). \end{aligned}$$

5 Potenzreihen

In Analysis 1 haben wir Konvergenzkriterien für Reihen behandelt. Wir haben auch die Exponentialreihe sowie die Reihendarstellungen der Funktionen Cosinus und Sinus kennengelernt. Die systematische Diskussion von Reihen dieses Typs, den Potenzreihen, musste aus Zeitgründen verschoben werden, wir holen das jetzt nach.

Sei $f \in C^\infty(I)$, wobei I offenes Intervall mit $0 \in I$. Dann können wir für jedes $k \in \mathbb{N}_0$ das Taylorpolynom P_k mit Entwicklungspunkt $x_0 = 0$ bilden, es lautet

$$P_k(x) = \sum_{i=0}^k \frac{f^{(i)}(0)}{i!} x^i.$$

Für festes k beschreibt $P_k(x)$ das Verhalten von $f(x)$ für $x \rightarrow 0$, und zwar umso genauer je größer k gewählt wird. Jetzt ändern wir den Blickwinkel und fragen: konvergiert für festes $x \in I$ die Folge $P_k(x)$ für $k \rightarrow \infty$? Und wenn ja, ist der Grenzwert dann der Funktionswert $f(x)$? Die Reihe

$$P(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(0)}{k!} x^k$$

heißt Taylorreihe von $f(x)$ mit Entwicklungspunkt $x = 0$. Wir betrachten generell Reihen der folgenden Form.

Definition 5.1 (Potenzreihen) *Eine komplexe Potenzreihe ist eine Reihe der Form*

$$P(z) = \sum_{k=0}^{\infty} a_k z^k \quad \text{mit } a_k \in \mathbb{C}.$$

Allgemeiner hat man Potenzreihen mit Entwicklungspunkt $z_0 \in \mathbb{C}$, das heißt Reihen der Form $P(z) = \sum_{k=0}^{\infty} a_k (z - z_0)^k$. Der Einfachheit halber beschränken wir uns auf $z_0 = 0$. Definition 5.1 lässt offen, ob beziehungsweise für welche $z \in \mathbb{C}$ die Reihe tatsächlich konvergiert. Klar ist $P(0) = a_0$, es kann aber sein, dass die Reihe für alle $z \neq 0$ divergiert, dann ist sie natürlich nicht relevant. Ein Beispiel ist $\sum_{k=0}^{\infty} k^k z^k$, wie der Nullfolgentest zeigt. Bei interessanten Reihen (was immer das ist) erwarten wir aber Konvergenz für Punkte z in einem gewissen Gebiet. Auf diesem Gebiet definiert die Reihe dann eine Funktion. Wir beginnen damit, dieses Konvergenzgebiet genauer zu charakterisieren.

Lemma 5.1 (von Abel) *Sei $P(z) = \sum_{k=0}^{\infty} a_k z^k$, $a_k \in \mathbb{C}$, eine Potenzreihe. Es gebe $\varrho > 0$ und $M \in [0, \infty)$ mit $|a_k| \varrho^k \leq M$ für alle $k \in \mathbb{N}_0$. Dann ist $P(z)$ für $|z| < \varrho$ absolut konvergent. Genauer gilt für alle $k \in \mathbb{N}_0$*

$$(5.1) \quad |P(z) - P_k(z)| \leq \sum_{\ell=k+1}^{\infty} |a_\ell| |z|^\ell \leq \frac{M}{1 - \frac{|z|}{\varrho}} \left(\frac{|z|}{\varrho} \right)^{k+1}.$$

Hier ist $P_k(z) = \sum_{\ell=0}^k a_\ell z^\ell$ die k -te Partialsumme.

BEWEIS: Es gilt $|a_k z^k| = |a_k| \varrho^k (|z|/\varrho)^k \leq M(|z|/\varrho)^k$. Für $|z| < \varrho$ folgt die absolute Konvergenz durch Vergleich mit der geometrischen Reihe. Genauer gilt

$$|P(z) - P_k(z)| \leq \sum_{\ell=k+1}^{\infty} |a_\ell| |z|^\ell \leq M \sum_{\ell=k+1}^{\infty} \left(\frac{|z|}{\varrho}\right)^\ell = \frac{M}{1 - \frac{|z|}{\varrho}} \left(\frac{|z|}{\varrho}\right)^{k+1}.$$

□

Das Konvergenzgebiet einer Potenzreihe sieht allgemein wie folgt aus.

Satz 5.1 (vom Konvergenzradius) *Zu jeder Potenzreihe $P(z) = \sum_{k=0}^{\infty} a_k z^k$ gibt es genau ein $R \in [0, \infty]$, den Konvergenzradius, mit folgender Eigenschaft:*

$$P(z) \text{ ist } \begin{cases} \text{absolut konvergent} & \text{für } |z| < R, \\ \text{divergent} & \text{für } |z| > R. \end{cases}$$

BEWEIS: Die Eindeutigkeit von R ist klar. Zur Existenz definieren wir

$$R = \sup\{|z| : P(z) \text{ konvergiert}\} \in [0, \infty].$$

Ist $|z| < R$, so gibt es nach Definition ein $z_0 \in \mathbb{C}$ mit $|z| < |z_0| \leq R$, so dass $P(z_0)$ konvergiert. Nach dem Nullfolgenkriterium, Satz 6.1 in Analysis 1, folgt $a_k z_0^k \rightarrow 0$. Es gibt also ein $M \in [0, \infty)$ mit $|a_k| |z_0|^k \leq M$ für alle k . Nach Lemma 5.1, mit $\varrho = |z_0|$, ist $P(z)$ absolut konvergent. Andererseits ist die Reihe divergent für $|z| > R$ nach Definition von R . □

Beispiel 5.1 Für die Exponentialreihe $\sum_{k=0}^{\infty} z^k/k!$ gilt, außer im trivialen Fall $z = 0$,

$$\frac{|z^{k+1}/(k+1)!|}{|z^k/k!|} = \frac{|z|}{k+1} \rightarrow 0 \quad \text{mit } k \rightarrow \infty.$$

Nach dem Quotientenkriterium, Satz 6.6 (b) in Analysis 1, konvergiert die Reihe absolut für alle $z \in \mathbb{C}$. Der Konvergenzradius der Reihe ist $R = \infty$.

Beispiel 5.2 Die Binomialreihe zum Parameter $\alpha \in \mathbb{C}$ lautet

$$B_\alpha(z) = \sum_{k=0}^{\infty} \binom{\alpha}{k} z^k = 1 + \alpha z + \frac{\alpha(\alpha-1)}{2} z^2 + \dots$$

Für $\alpha \in \mathbb{N}_0$ bricht die Reihe nach $k = \alpha$ ab, und die Binomische Formel liefert $B_\alpha(z) = (1+z)^\alpha$. Im folgenden sei nun $\alpha \notin \mathbb{N}_0$. Für $z \neq 0$ ist dann $a_k = \binom{\alpha}{k} z^k \neq 0$ für alle $k \in \mathbb{N}_0$, und es gilt

$$\left| \frac{a_{k+1}}{a_k} \right| = \frac{|\alpha - k|}{k+1} |z| \rightarrow |z| \quad \text{mit } k \rightarrow \infty.$$

Nach dem Quotientenkriterium konvergiert die Reihe für $|z| < 1$ und divergiert für $|z| > 1$, das heißt der Konvergenzradius ist $R = 1$.

Wir interessieren uns nun dafür, wie die Funktionen $P_k(z)$ gegen die Funktion $P(z)$ konvergieren. Lässt sich zum Beispiel die Stetigkeit der $P_k(z)$ auf die Grenzfunktion $P(z)$ übertragen?

Das ist ein generelles und grundlegendes Problem: sei $f_k : D \rightarrow \mathbb{R}^m$ eine Folge von Funktionen auf $D \subset \mathbb{R}^n$, die punktweise gegen eine Grenzfunktion $f : D \rightarrow \mathbb{R}^m$ konvergiert:

$$f(x) = \lim_{k \rightarrow \infty} f_k(x) \quad \text{für alle } x \in D.$$

Folgt die Stetigkeit von f aus der Stetigkeit der f_k , unter geeigneten Voraussetzungen? Das folgende Beispiel zeigt, dass die punktweise Konvergenz alleine im allgemeinen nicht die Stetigkeit der Grenzfunktion garantiert.

Beispiel 5.3 Betrachte die stetigen Funktionen $f_k : [0, 1] \rightarrow \mathbb{R}$,

$$f_k(x) = \begin{cases} 1 - kx & \text{für } 0 \leq x < \frac{1}{k}, \\ 0 & \text{sonst.} \end{cases}$$

Die Folge konvergiert punktweise gegen $f : [0, 1] \rightarrow \mathbb{R}$ mit

$$f(x) = \begin{cases} 1 & \text{für } x = 0, \\ 0 & \text{sonst.} \end{cases}$$

Die Grenzfunktion ist im Punkt $x = 0$ nicht stetig.

Na gut, das hätten wir eigentlich auch nicht erwartet. Solche Probleme kennen wir schon von der Vertauschbarkeit des Integrals mit der Konvergenz.

Satz 5.2 (Gleichmäßige Konvergenz und Stetigkeit) Seien $f_k : D \rightarrow \mathbb{R}^m$, $k \in \mathbb{N}$, stetige Funktionen auf $D \subset \mathbb{R}^n$, die gleichmäßig gegen $f : D \rightarrow \mathbb{R}^m$ konvergieren, also

$$\|f_k - f\|_D \rightarrow 0 \quad \text{mit } k \rightarrow \infty.$$

Dann ist f ebenfalls stetig auf D .

BEWEIS: Sei $x_0 \in D$ gegeben. Für $x \in D$ beliebig und $k \in \mathbb{N}$ gilt

$$\begin{aligned} |f(x) - f(x_0)| &\leq |f(x) - f_k(x)| + |f_k(x) - f_k(x_0)| + |f_k(x_0) - f(x_0)| \\ &\leq \|f - f_k\|_D + 2\|f - f_k\|_D. \end{aligned}$$

Für $x \rightarrow x_0$ ergibt sich, da f_k stetig ist,

$$\limsup_{x \rightarrow x_0} |f(x) - f(x_0)| \leq 2\|f - f_k\|_D.$$

Mit $k \rightarrow \infty$ folgt $\limsup_{x \rightarrow x_0} |f(x) - f(x_0)| = 0$, die Stetigkeit von f in x_0 . □

Satz 5.3 (Stetigkeit von Potenzreihen) Sei $P(z) = \sum_{k=0}^{\infty} a_k z^k$ eine komplexe Potenzreihe mit Konvergenzradius $R > 0$. Dann ist $P(z)$ stetig auf der Kreisscheibe $B_R(0) = \{z \in \mathbb{C} : |z| < R\}$.

BEWEIS: Wir zeigen, dass die $P_k(z)$ auf jeder Kreisscheibe $B_r(0)$ mit $r < R$ gleichmäßig konvergieren. Die Behauptung folgt dann aus Satz 5.2. Sei $\varrho \in (r, R)$ beliebig gewählt. Da $P(\varrho)$ konvergiert, gibt es ein $M \in [0, \infty)$ mit $|a_k| \varrho^k \leq M$ für alle k . Nach Lemma 5.1 folgt

$$\|P - P_k\|_{B_r(0)} \leq \frac{M}{1 - \frac{r}{\varrho}} \left(\frac{r}{\varrho}\right)^{k+1} \rightarrow 0 \quad \text{mit } k \rightarrow \infty.$$

Damit ist der Satz bewiesen. \square

Nun zur Frage der Differenzierbarkeit der Grenzfunktion. Aus Gründen der Einfachheit beschränken wir uns auf Funktionen $f : I \rightarrow \mathbb{R}^m$ einer Variablen.

Beispiel 5.4 Betrachte für $\varepsilon > 0$ die Funktionen $f_\varepsilon \in C^1(\mathbb{R})$, $f_\varepsilon(x) = (\varepsilon^2 + x^2)^{1/2}$. Es gilt

$$|x| \leq f_\varepsilon(x) \leq \varepsilon + |x| \quad \text{für alle } x \in \mathbb{R},$$

also konvergieren die f_ε gleichmäßig gegen $f(x) = |x|$. In $x = 0$ ist f nicht differenzierbar.

Satz 5.4 (Vertauschung von Konvergenz und Ableitung) Seien $f_k \in C^1(I, \mathbb{R}^m)$, $I = (a, b)$, punktweise konvergent gegen $f : I \rightarrow \mathbb{R}^m$. Konvergiert die Folge f'_k gleichmäßig gegen ein $g : I \rightarrow \mathbb{R}^m$, also

$$\|f'_k - g\|_I \rightarrow 0 \quad \text{mit } k \rightarrow \infty,$$

so folgt $f' = g$, und $f \in C^1(I, \mathbb{R}^m)$.

BEWEIS: Wähle $x_0 \in I$. Dann besagt der Hauptsatz der Differential- und Integralrechnung

$$f_k(x) = f_k(x_0) + \int_{x_0}^x f'_k(y) dy \quad \text{für alle } x \in I.$$

Nach Analysis 1, Satz 13.3, vertauscht das Integral mit $f'_k \rightarrow g$, also folgt mit $k \rightarrow \infty$

$$f(x) = f(x_0) + \int_{x_0}^x g(y) dy \quad \text{für alle } x \in I.$$

Nach Satz 5.2 ist g stetig. Es folgt $f' = g$ aus dem Hauptsatz. \square

Um den Satz auf Potenzreihen anzuwenden, brauchen wir folgende Hilfsaussage.

Lemma 5.2 Sei $P(z) = \sum_{k=0}^{\infty} a_k z^k$ eine Potenzreihe mit Konvergenzradius $R \in [0, \infty]$. Dann hat die formal differenzierte Reihe

$$Q(z) = \sum_{k=1}^{\infty} k a_k z^{k-1} = \sum_{k=0}^{\infty} (k+1) a_{k+1} z^k$$

denselben Konvergenzradius R .

BEWEIS: $Q(z)$ hat höchstens den Konvergenzradius von $P(z)$, denn $|a_k z^k| \leq |k a_k z^{k-1}| \cdot |z|$ für $k \geq 1$. Für $\varrho \in (0, R)$ ist $P(\varrho)$ konvergent, also gilt $|a_k| \varrho^k \leq M$ für ein $M \in [0, \infty)$. Es folgt für alle $z \in B_\varrho(0)$

$$|k a_k z^{k-1}| = k |a_k| \varrho^k \frac{|z|^{k-1}}{\varrho^k} \leq \frac{k M |z|^{k-1}}{\varrho^k}.$$

Die rechte Reihe konvergiert aber nach dem Quotientenkriterium, denn

$$\frac{(k+1)|z|^k}{\varrho^{k+1}} \left(\frac{k|z|^{k-1}}{\varrho^k} \right)^{-1} = \frac{(k+1)|z|}{k\varrho} \rightarrow \frac{|z|}{\varrho} < 1 \quad \text{mit } k \rightarrow \infty.$$

Also ist $Q(z)$ für jedes $z \in B_R(0)$ absolut konvergent. \square

Satz 5.5 (Differenzierbarkeit von Potenzreihen) Die Potenzreihe $P(z) = \sum_{k=0}^{\infty} a_k z^k$ mit $a_k \in \mathbb{C}$ habe den Konvergenzradius $R > 0$. Dann ist die Funktion

$$P : (-R, R) \rightarrow \mathbb{C}, P(x) = \sum_{k=0}^{\infty} a_k x^k$$

stetig differenzierbar, und ihre Ableitung ergibt sich durch gliedweise Differentiation:

$$(5.2) \quad P'(x) = \sum_{k=1}^{\infty} k a_k x^{k-1} = \sum_{k=0}^{\infty} (k+1) a_{k+1} x^k \quad \text{für alle } x \in (-R, R).$$

Hinweis. Durch Induktion folgt $P \in C^\infty((-R, R), \mathbb{C})$.

BEWEIS: Sei $\varrho \in (0, R)$. Nach Beweis von Satz 5.3 konvergieren die $P_k(x)$ auf $(-\varrho, \varrho)$ gleichmäßig gegen $P(x)$. Nach Lemma 5.2 konvergieren die $P'_k(x)$ auf $(-\varrho, \varrho)$ ebenfalls gleichmäßig gegen $Q(x)$, die gliedweise differenzierte Reihe. Satz 5.4 impliziert $P' = Q$ auf $(-\varrho, \varrho)$, also auf ganz $(-R, R)$. \square

Beispiel 5.5 Für die Potenzreihendarstellung des Logarithmus berechnen wir mit der geometrischen Reihe, für $x \in (-1, 1)$,

$$\frac{d}{dx} \log(1+x) = \frac{1}{1+x} = \sum_{k=0}^{\infty} (-1)^k x^k =: Q(x).$$

$Q(x)$ ergibt sich durch gliedweise Differentiation der Reihe $P(x) = \sum_{k=1}^{\infty} (-1)^{k-1} x^k / k$. Nach Lemma 5.2 haben $P(x)$ und $Q(x)$ denselben Konvergenzradius, also $R = 1$, und Satz 5.5 ergibt $P'(x) = Q(x)$ für $x \in (-1, 1)$. Nun ist $\log(1+x) = P(x) = 0$ für $x = 0$, also erhalten wir

$$\log(1+x) = P(x) = x - \frac{x^2}{2} + \frac{x^3}{3} - + \dots \quad \text{für } x \in (-1, 1).$$

Beispiel 5.6 Die Ableitung der Funktion $\tan : (-\frac{\pi}{2}, \frac{\pi}{2}) \rightarrow \mathbb{R}$ ist nach Quotientenregel

$$\tan' = \left(\frac{\sin}{\cos} \right)' = \frac{\cos^2 + \sin^2}{\cos^2} = 1 + \tan^2.$$

Für die Umkehrfunktion \arctan folgt, für $x \in (-1, 1)$,

$$\frac{d}{dx} \arctan x = \frac{1}{\tan'(\arctan x)} = \frac{1}{1+x^2} = \sum_{k=0}^{\infty} (-1)^k x^{2k} =: Q(x).$$

$Q(x)$ ist die gliedweise Ableitung von $P(x) = \sum_{k=0}^{\infty} (-1)^k x^{2k+1} / (2k+1)$. Der gemeinsame Konvergenzradius ist $R = 1$, und aus Satz 5.5 folgt $P'(x) = Q(x)$ für alle $x \in (-1, 1)$. Da $P(0) = \arctan 0 = 0$, sehen wir

$$\arctan x = \sum_{k=0}^{\infty} \frac{(-1)^k}{2k+1} x^{2k+1} = x - \frac{x^3}{3} + \frac{x^5}{5} - + \dots \quad \text{für } x \in (-1, 1).$$

Hat ein Polynom vom Grad höchstens k mehr als k Nullstellen, so ist es bekanntlich das Nullpolynom. Erstaunlicherweise hat man für Potenzreihen eine Aussage, die als verwandt betrachtet werden kann.

Satz 5.6 (Identitätssatz für Potenzreihen) Sei $P(z) = \sum_{k=0}^{\infty} a_k z^k$ eine Potenzreihe mit Konvergenzradius $R > 0$. Ist $0 \in \mathbb{C}$ Häufungspunkt der Menge $\{z \in \mathbb{C} : P(z) = 0\}$, so folgt $a_k = 0$ für alle $k \in \mathbb{N}_0$.

BEWEIS: Für $\varrho \in (0, R)$ ist $P(\varrho)$ konvergent, also $|a_k| \varrho^k \leq M$ für alle k . Nach (5.1) gilt

$$|P(z) - P_k(z)| \leq \frac{M}{1 - \frac{|z|}{\varrho}} \left(\frac{|z|}{\varrho} \right)^{k+1} \quad \text{für } |z| < \varrho.$$

Für $|z| \leq \varrho/2$ ergibt sich

$$|P(z) - P_k(z)| \leq \frac{2M}{\varrho^{k+1}} |z|^{k+1} =: C |z|^{k+1}.$$

Sei nun induktiv schon $a_0 = \dots = a_{k-1} = 0$ gezeigt für ein $k \in \mathbb{N}_0$. Dann folgt

$$|P(z) - a_k z^k| = |P(z) - P_k(z)| \leq C |z|^{k+1} \quad \text{für } |z| \leq \varrho/2.$$

Sei nun $P(z_i) = 0$ für eine Folge $z_i \neq 0$ mit $z_i \rightarrow 0$. Einsetzen von $z = z_i$ ergibt

$$|a_k| |z_i|^k = |P(z_i) - a_k z_i^k| \leq C |z_i|^{k+1},$$

also $|a_k| \leq C |z_i| \rightarrow 0$ mit $i \rightarrow \infty$. Das zeigt $a_k = 0$. □

Folgerung 5.1 (Koeffizientenvergleich) Seien $P(z) = \sum_{k=0}^{\infty} a_k z^k$ und $Q(z) = \sum_{k=0}^{\infty} b_k z^k$ Potenzreihen mit positivem Konvergenzradius. Ist der Nullpunkt Häufungspunkt der Menge $\{z \in \mathbb{C} : P(z) = Q(z)\}$, so folgt $a_k = b_k$ für alle $k \in \mathbb{N}_0$.

BEWEIS: Die Potenzreihe $F(z) = \sum_{k=0}^{\infty} c_k z^k$ mit $c_k = a_k - b_k$ hat positiven Konvergenzradius, und der Nullpunkt ist Häufungspunkt der Menge $\{z \in \mathbb{C} : F(z) = 0\}$. Die Behauptung folgt damit aus Satz 5.6. □

Beispiel 5.7 Die Binomialreihe mit Parameter $\alpha \in \mathbb{R}$ ist definiert durch

$$B_\alpha(z) = \sum_{k=0}^{\infty} \binom{\alpha}{k} z^k.$$

Sie hat Konvergenzradius $R = 1$, siehe Beispiel 5.2. Für $x \in (-1, 1)$ gilt nach Satz 5.5

$$B'_\alpha(x) = \sum_{k=0}^{\infty} (k+1) \binom{\alpha}{k+1} x^k = \sum_{k=0}^{\infty} (k+1) \frac{\alpha - k}{k+1} \binom{\alpha}{k} x^k = \alpha B_\alpha(x) - x B'_\alpha(x),$$

das heißt $B'_\alpha(x) = \frac{\alpha}{1+x} B_\alpha(x)$. Es folgt mit $f(x) = (1+x)^{-\alpha}$

$$(f B_\alpha)'(x) = f'(x) B_\alpha(x) + f(x) B'_\alpha(x) = f(x) B_\alpha(x) \left(-\frac{\alpha}{1+x} + \frac{\alpha}{1+x} \right) = 0.$$

Wegen $B_\alpha(0) = 1 = f(0)$ ergibt sich die folgende Darstellung (Newton 1665)

$$(1+x)^\alpha = B_\alpha(x) = \sum_{k=0}^{\infty} \binom{\alpha}{k} x^k \quad \text{für alle } x \in (-1, 1).$$

In der Physik wird oft die Näherung $(1+x)^\alpha = 1 + \alpha x$ benutzt, für $|x| \ll 1$.

Bisher haben wir uns immer in der offenen Kreisscheibe bewegt, auf der die Potenzreihe lokal gleichmäßig konvergiert. Zum Schluss des Kapitels betrachten wir die Situation am Rand.

Satz 5.7 (Abelscher Grenzwertsatz) *Ist die Potenzreihe $P(x) = \sum_{k=0}^{\infty} a_k x^k$ für $x = 1$ konvergent, so gilt*

$$\lim_{x \nearrow 1} P(x) = P(1).$$

BEWEIS: Nach Voraussetzung hat P Konvergenzradius $R \geq 1$. Berechne für $0 \leq x < 1$

$$\begin{aligned} P_\ell(1) - P_\ell(x) &= \sum_{k=0}^{\ell} a_k (1 - x^k) \\ &= (1-x) \sum_{k=0}^{\ell} a_k \sum_{j=0}^{k-1} x^j \\ &= (1-x) \sum_{\substack{j,k=0 \\ j < k}}^{\ell} a_k x^j \\ &= (1-x) \sum_{j=0}^{\ell-1} x^j \sum_{k=j+1}^{\ell} a_k \\ &= (1-x) \sum_{j=0}^{\ell-1} x^j (P_\ell(1) - P_j(1)). \end{aligned}$$

Zu $\varepsilon > 0$ gibt es ein $N \in \mathbb{N}$ mit $|P_\ell(1) - P_j(1)| < \varepsilon$ für $j, \ell \geq N$, außerdem gilt $|P_j(1)| \leq C$ für alle j . Damit schätzen wir ab

$$|P_\ell(1) - P_\ell(x)| \leq 2C(1-x) \sum_{j=0}^{N-1} x^j + \varepsilon(1-x) \sum_{j=N}^{\ell-1} x^j \leq 2C(1-x)N + \varepsilon,$$

wobei zuletzt die geometrische Reihe benutzt wurde. Mit $\ell \rightarrow \infty$ und $x \nearrow 1$ folgt

$$\limsup_{x \nearrow 1} |P(1) - P(x)| \leq \varepsilon.$$

□

Beispiel 5.8 Nach dem Leibnizkriterium konvergiert die Reihe $x - \frac{1}{2}x^2 + \frac{1}{3}x^3 \pm \dots$ auch für $x = 1$, also folgt aus Satz 5.7

$$\log 2 = \lim_{x \nearrow 1} \log(1+x) = 1 - \frac{1}{2} + \frac{1}{3} - + \dots \quad (\text{Mercator 1668}).$$

Ebenso konvergiert die Reihe des arctan auch für $x = 1$, und es ergibt sich die Darstellung

$$\frac{\pi}{4} = \arctan 1 = \lim_{x \nearrow 1} \arctan x = 1 - \frac{1}{3} + \frac{1}{5} - + \dots \quad (\text{Gregory 1671, Leibniz 1674}).$$

Zum Schluss kommen wir nochmal auf die Eingangsfrage zurück, nämlich ob die Taylorreihe einer Funktion $f \in C^\infty(I)$, $I = (a, b)$, gegen die Funktion f konvergiert.

Definition 5.2 (analytische Funktion) Sei I offenes Intervall und $f : I \rightarrow \mathbb{R}$. Dann heißt f *analytisch*, wenn jedes $x_0 \in I$ eine Umgebung $(x_0 - \delta, x_0 + \delta)$ hat, auf der f durch eine konvergente Potenzreihe mit Entwicklungspunkt x_0 dargestellt wird:

$$f(x) = \sum_{k=0}^{\infty} a_k (x - x_0)^k \quad \text{für alle } x \in (x_0 - \delta, x_0 + \delta).$$

Man bezeichnet den entsprechenden Raum von Funktionen mit $C^\omega(I)$.

Die darstellende Potenzreihe kann nur die Taylorreihe sein. Denn f ist C^∞ auf $(x_0 - \delta, x_0 + \delta)$ nach Satz 5.5, und gliedweise Differentiation ergibt

$$f^{(i)}(x_0) = \sum_{k=0}^{\infty} \delta_{ik} k! a_k = i! a_i.$$

Eine Funktion ist also genau dann analytisch, wenn sie C^∞ ist und für jedes $x_0 \in I$ gilt: die Taylorreihe mit Entwicklungspunkt x_0 konvergiert punktweise gegen f nahe bei x_0 . Für die analytischen Funktionen gilt also eine Art Fernwirkung: hat man alle Ableitungen in einem Punkt $x_0 \in I$, so kennt man schon die ganze Funktion $f : I \rightarrow \mathbb{R}$. Hat die Menge der Nullstellen von f einen Häufungspunkt $x_0 \in I$, so ist f schon die Nullfunktion. Das folgt sofort aus Satz 5.6, angewandt im Punkt x_0 . Eine analytische Funktion kann für jedes $x_0 \in I$ auf die Kreisscheibe $B_\delta(x_0) \subset \mathbb{C}$ fortgesetzt werden. Dazu muss man nur in die Potenzreihe Zahlen $z \in B_\delta(x_0)$ einsetzen. In der komplexen Analysis betrachtet man komplex differenzierbare Funktionen. Es stellt sich heraus, dass diese immer analytisch sind, daraus ergibt sich die Bedeutung des Begriffs. Im allgemeinen sind aber Funktionen $f \in C^\infty(I)$ nicht analytisch, wie das folgende Beispiel zeigt.

Beispiel 5.9 Betrachte $f : \mathbb{R} \rightarrow \mathbb{R}$ mit

$$f(x) = \begin{cases} e^{-\frac{1}{x}} & \text{für } x > 0 \\ 0 & \text{für } x \leq 0 \end{cases}$$

Es gilt $f \in C^\infty(\mathbb{R})$ nach Übungsaufgabe, und zwar $f^{(k)}(0) = 0$ für alle $k \in \mathbb{N}_0$. Also sind die Koeffizienten der Taylorreihe mit Entwicklungspunkt $x_0 = 0$ alle Null und damit auch alle Partialsummen, die Reihe konvergiert somit gegen die Nullfunktion und nicht gegen f . In der Tat ist $x_0 = 0$ Häufungspunkt der Nullstellenmenge von f .

6 Parameterabhängige Integrale

In diesem Abschnitt behandeln wir Integrale, deren Integranden von zusätzlichen Parametern abhängen. Dies ist eine typische Problematik in zahlreichen Anwendungen. Sei $\Omega \subset \mathbb{R}^n$ offen und $I = [a, b]$ ein kompaktes Intervall. Für eine gegebene Funktion $f : \Omega \times I \rightarrow \mathbb{R}$, $f = f(x, y)$, betrachten wir die neue Funktion

$$(6.1) \quad \phi : \Omega \rightarrow \mathbb{R}, \quad \phi(x) = \int_I f(x, y) dy.$$

Diese Funktion wird als parameterabhängiges Integral bezeichnet, wobei die Parameter hier die Punkte $x = (x_1, \dots, x_n) \in \Omega$ sind. Damit ϕ wohldefiniert ist, müssen die Integrale existieren, also sollte für jedes $x \in \Omega$ die Funktion $f(x, \cdot) : I \rightarrow \mathbb{R}$, $y \mapsto f(x, y)$, Riemann-integrierbar sein. Wir interessieren uns für die Stetigkeit und Ableitung der Funktion $\phi(x)$. Ein nützlicher Begriff ist dabei die Oszillationsfunktion einer Funktion $f : D \rightarrow \mathbb{R}$:

$$\text{osc}(f, \delta) = \sup\{|f(x) - f(x')| : x, x' \in D, |x - x'| < \delta\}.$$

Die Funktion f ist genau dann gleichmäßig stetig, wenn $\lim_{\delta \rightarrow 0} \text{osc}(f, \delta) = 0$.

Satz 6.1 (Stetigkeit von Parameterintegralen) Sei $f : \Omega \times I \rightarrow \mathbb{R}$, $f = f(x, y)$, wobei $\Omega \subset \mathbb{R}^n$ offen und $I = [a, b]$ kompakt. Ist $f \in C^0(\Omega \times I)$, so ist die Funktion

$$\phi : \Omega \rightarrow \mathbb{R}, \quad \phi(x) = \int_I f(x, y) dy,$$

wohldefiniert und stetig.

BEWEIS: Die Funktion $\phi(x)$ ist wohldefiniert, denn für $x \in \Omega$ ist $f(x, \cdot) \in C^0(I)$ und damit Riemann-integrierbar (siehe Satz 13.5). Wir berechnen für $x, x' \in \Omega$

$$|\phi(x') - \phi(x)| \leq \int_I |f(x', y) - f(x, y)| dy \leq |I| \sup_{y \in I} |f(x', y) - f(x, y)|.$$

Wähle $R > 0$ mit $\overline{B_R(x)} \subset \Omega$. Dann ist $f(x, y)$ gleichmäßig stetig auf $\overline{B_R(x)} \times I$, vgl. Satz 13.4, insbesondere folgt für $|x' - x| < \delta \leq R$

$$\sup_{y \in I} |f(x', y) - f(x, y)| \leq \text{osc}(f|_{\overline{B_R(x)} \times I}, \delta) \rightarrow 0 \quad \text{mit } \delta \rightarrow 0.$$

□

Wir gehen direkt weiter zur Differenzierbarkeit und Berechnung der Ableitung.

Satz 6.2 (Differentiation unter dem Integral) Sei $f : \Omega \times I \rightarrow \mathbb{R}$, $f = f(x, y)$, wobei $\Omega \subset \mathbb{R}^n$ offen und $I = [a, b]$ kompakt. Es gelte:

- (a) $f(x, \cdot)$ ist Riemann-integrierbar für jedes $x \in \Omega$.
- (b) Die partielle Ableitung $\frac{\partial f}{\partial x_j}$ existiert und ist stetig auf $\Omega \times I$.

Dann ist $\phi : \Omega \rightarrow \mathbb{R}$, $\phi(x) = \int_I f(x, y) dy$, nach x_j partiell differenzierbar, und zwar gilt

$$\frac{\partial \phi}{\partial x_j}(x) = \int_I \frac{\partial f}{\partial x_j}(x, y) dy \quad \text{für alle } x \in \Omega.$$

Sind f und $\frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n}$ in $C^0(\Omega \times I)$, so ist $\phi \in C^1(\Omega)$.

BEWEIS: Nach dem Hauptsatz der Differential- und Integralrechnung gilt

$$\frac{f(x + he_j, y) - f(x, y)}{h} = \frac{1}{h} \int_0^1 \frac{d}{ds} f(x + she_j, y) ds = \int_0^1 \frac{\partial f}{\partial x_j}(x + she_j, y) ds.$$

Sei wieder $\overline{B_R(x)} \subset \Omega$ und $|h| < \delta \leq R$, dann schätzen wir wie folgt ab:

$$\begin{aligned} \left| \frac{\phi(x + he_j) - \phi(x)}{h} - \int_I \frac{\partial f}{\partial x_j}(x, y) dy \right| &= \left| \int_I \left(\frac{f(x + he_j, y) - f(x, y)}{h} - \frac{\partial f}{\partial x_j}(x, y) \right) dy \right| \\ &= \left| \int_I \int_0^1 \left(\frac{\partial f}{\partial x_j}(x + she_j, y) - \frac{\partial f}{\partial x_j}(x, y) \right) ds dy \right| \\ &\leq \int_I \int_0^1 \left| \frac{\partial f}{\partial x_j}(x + she_j, y) - \frac{\partial f}{\partial x_j}(x, y) \right| ds dy \\ &\leq |I| \operatorname{osc} \left(\frac{\partial f}{\partial x_j} \Big|_{\overline{B_R(x)} \times I}, \delta \right) \rightarrow 0 \quad \text{mit } \delta \rightarrow 0. \end{aligned}$$

Damit ist die Vertauschung der Ableitung mit dem Integral gerechtfertigt. Satz 6.1 impliziert weiter $\frac{\partial \phi}{\partial x_j} \in C^0(\Omega)$. Sind nun f und alle $\frac{\partial f}{\partial x_j}$ in $C^0(\Omega \times I)$, so folgt $\phi, \frac{\partial \phi}{\partial x_j} \in C^0(\Omega)$ aus Satz 6.1 bzw. wie gerade gezeigt. \square

Beispiel 6.1 Wir berechnen hier das Integral der Gaußschen Dichtefunktion (das früher auf 10-Mark-Scheinen zu finden war)

$$\int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi}.$$

Der Beweis ist trickreich, ich wäre wohl selbst nicht drauf gekommen. Setze

$$F : [0, \infty) \rightarrow \mathbb{R}, F(x) = \left(\int_0^x e^{-\xi^2} d\xi \right)^2,$$

und berechne mit Hauptsatz und anschließender Substitution $\xi = xy$, also $d\xi = xdy$,

$$F'(x) = 2e^{-x^2} \int_0^x e^{-\xi^2} d\xi = \int_0^1 2xe^{-(1+y^2)x^2} dy = \int_0^1 \frac{\partial f}{\partial x}(x, y) dy,$$

wobei $f(x, y) = -e^{-(1+y^2)x^2}/(1+y^2)$. Da f auf $(0, \infty) \times [0, 1]$ glatt ist, können wir nach Satz 6.2 den Operator $\frac{\partial}{\partial x}$ herausziehen, und mit $\phi(x) = \int_0^1 f(x, y) dy$ folgt

$$\phi'(x) = \int_0^1 \frac{\partial f}{\partial x}(x, y) dy = F'(x).$$

Nun gilt $F(0) - \phi(0) = \int_0^1 (1+y^2)^{-1} dy = \arctan 1 = \pi/4$, also $F(x) = \phi(x) + \pi/4$ für alle $x \in [0, \infty)$. Aber $|\phi(x)| \leq e^{-x^2} \rightarrow 0$ mit $x \rightarrow \infty$, und so

$$\int_0^{\infty} e^{-x^2} dx = \lim_{x \rightarrow \infty} \sqrt{F(x)} = \frac{\sqrt{\pi}}{2}.$$

In dieser Vorlesung werden wir aus Zeitgründen kein mehrdimensionales Integral behandeln, dies soll in Analysis 3 ausführliches Thema sein. Immerhin können wir als nützliche Anwendung hier die Vertauschbarkeit der Integrationsreihenfolge in Mehrfachintegralen folgern.

Satz 6.3 (Kleiner Fubini) Seien $I = [a, b]$, $J = [\alpha, \beta]$ kompakte Intervalle. Dann gilt

$$\int_{\alpha}^{\beta} \left(\int_a^b f(x, y) dx \right) dy = \int_a^b \left(\int_{\alpha}^{\beta} f(x, y) dy \right) dx \quad \text{für } f \in C^0(I \times J).$$

BEWEIS: Wir betrachten die Funktionen $\phi, \psi : [a, b] \rightarrow \mathbb{R}$ mit

$$\phi(x) = \int_{\alpha}^{\beta} \left(\int_a^x f(\xi, y) d\xi \right) dy \quad \text{und} \quad \psi(x) = \int_a^x \left(\int_{\alpha}^{\beta} f(\xi, y) dy \right) d\xi.$$

Nach Satz 6.1 sind $y \mapsto \int_a^x f(\xi, y) d\xi$ sowie $\xi \mapsto \int_{\alpha}^{\beta} f(\xi, y) dy$ stetig, und damit beide Seiten wohldefiniert mit $\phi(a) = \psi(a) = 0$. Wir zeigen $\phi'(x) = \psi'(x)$ für alle $x \in I$, woraus die Behauptung $\phi(b) = \psi(b)$ folgt. Der Hauptsatz der Differential- und Integralrechnung liefert

$$\psi'(x) = \int_{\alpha}^{\beta} f(x, y) dy.$$

Weiter hat die Funktion $F(x, y) = \int_a^x f(\xi, y) d\xi$ die partielle Ableitung $\frac{\partial F}{\partial x} = f \in C^0(I \times J)$, und aus Satz 6.2 folgt

$$\phi'(x) = \int_{\alpha}^{\beta} \frac{\partial F}{\partial x}(x, y) dy = \int_{\alpha}^{\beta} f(x, y) dy.$$

□

Alternativ kann der kleine Fubini auch durch Approximation mit Riemannschen Summen in beiden Variablen bewiesen werden.

Wir kommen jetzt zu einer Anwendung in der Variationsrechnung, und zwar betrachten wir Integrale des folgenden Typs:

$$\mathcal{F} : C^1(I, \mathbb{R}^n) \rightarrow \mathbb{R}, \quad \mathcal{F}(u) = \int_a^b f(t, u(t), u'(t)) dt.$$

Abstrakt ist \mathcal{F} eine reelle Funktion auf dem Raum $C^1(I, \mathbb{R}^n)$. Da es sich aber nicht um eine Funktion von endlich vielen reellen Variablen handelt, wie wir sie bisher hatten, wird meistens die Bezeichnung *Funktional* oder *Variationsintegral* benutzt. Das Funktional \mathcal{F} ist dabei definiert durch die *Lagrangefunktion*

$$f : I \times \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}, \quad f = f(t, x, v).$$

Hier ein paar Beispiele.

Beispiel 6.2 Die Bogenlänge von $u \in C^1(I, \mathbb{R}^n)$ ist das Funktional

$$\mathcal{F}(u) = \int_a^b |u'(t)| dt, \quad \text{also } f(t, x, v) = |v|.$$

Die Formel kann wie folgt motiviert werden: für eine Zerlegung $a = t_0 < \dots < t_N$ ergibt sich als Näherung der Bogenlänge

$$\sum_{i=1}^N |u(t_i) - u(t_{i-1})| \approx \sum_{i=1}^N |u'(t_i)| \Delta t_i.$$

Rechts steht aber die Riemannsche Summe für die Funktion $|u'(t)|$.

Beispiel 6.3 Soll der Kalorienbedarf beim Querfeldeinlauf ermittelt werden, so spielt nicht nur die Länge der Strecke eine Rolle, sondern auch die wechselnde Qualität des Bodens. Dies könnte durch eine Gewichtsfunktion als Faktor beschrieben werden:

$$\mathcal{F}(u) = \int_a^b \omega(u(t)) |u'(t)| dt, \quad \text{also } f(t, x, v) = \omega(x)|v|.$$

Beispiel 6.4 Hier beschreibt $u \in C^1(I, \mathbb{R}^n)$ die Bahn eines Teilchens der Masse m in einem Kraftfeld mit Potential $V : \mathbb{R}^n \rightarrow \mathbb{R}$. Es ist dann $\frac{m}{2}|u'(t)|^2$ die kinetische und $V(u(t))$ die potentielle Energie des Teilchens zur Zeit t . Das Wirkungsintegral bildet die Differenz aus kinetischer und potentieller Energie, integriert auf I :

$$\mathcal{F}(u) = \int_a^b \left(\frac{m}{2}|u'(t)|^2 - V(u(t)) \right) dt, \quad \text{also } f(t, x, v) = \frac{m}{2}|v|^2 - V(x).$$

Funktionen, für die ein Funktional einen kleinsten oder größten Wert annimmt, sind natürlich von zentralem Interesse. Wir werden zeigen, dass eine extremale Funktion eine gewisse Differentialgleichung erfüllt, die Euler-Lagrange Gleichung. Unser Ansatz besteht darin, Variationen $u(\varepsilon, t)$ der extremalen Funktion $u(t)$ zu betrachten, die von einem Parameter abhängen:

$$u : (-\varepsilon_0, \varepsilon_0) \times I \rightarrow \mathbb{R}^n, \quad u = u(\varepsilon, t), \quad \text{wobei } u(0, \cdot) = u.$$

Ableitung der Variation nach dem Parameter ergibt das zugehörige Vektorfeld

$$\varphi : I \rightarrow \mathbb{R}^n, \quad \varphi(t) = \frac{\partial u}{\partial \varepsilon}(0, t).$$

Lemma 6.1 (Erste Variation) Sei $f = f(t, x, v)$ Lagrangefunktion mit f und $D_v f$ stetig differenzierbar auf $I \times \mathbb{R}^n \times \mathbb{R}^n$. Für $u \in C^2((-\varepsilon_0, \varepsilon_0) \times I, \mathbb{R}^n)$ betrachte

$$\phi(\varepsilon) = \mathcal{F}(u(\varepsilon, \cdot)) = \int_a^b f(t, u(\varepsilon, t), \frac{\partial u}{\partial t}(\varepsilon, t)) dt.$$

Dann gilt mit $\varphi = \frac{\partial u}{\partial \varepsilon}(0, \cdot) : I \rightarrow \mathbb{R}^n$ die Formel

$$(6.2) \quad \frac{d\phi}{d\varepsilon}(0) = \int_a^b \langle L_f(u), \varphi \rangle dt + \left[\langle D_v f(t, u, u'), \varphi \rangle \right]_{t=a}^{t=b}.$$

Dabei ist $\langle \cdot, \cdot \rangle$ das Standardskalarprodukt im \mathbb{R}^n und $L_f(u) : I \rightarrow \mathbb{R}^n$ ist gegeben durch

$$L_f(u) = D_x f(t, u, u') - \frac{d}{dt} [D_v f(t, u, u')].$$

BEWEIS: Die Verkettung $(\varepsilon, t) \mapsto f(t, u(\varepsilon, t), \frac{\partial u}{\partial t}(\varepsilon, t))$ ist von der Klasse C^1 . Deshalb kann nach Satz 6.2 unter dem Integralzeichen differenziert werden. Es folgt mit der Kettenregel

$$\phi'(0) = \int_a^b \sum_{i=1}^n \left(\frac{\partial f}{\partial x_i}(t, u, u') \frac{\partial u_i}{\partial \varepsilon}(0, t) + \frac{\partial f}{\partial v_i}(t, u, u') \frac{\partial^2 u_i}{\partial \varepsilon \partial t}(0, t) \right) dt.$$

Mit $\varphi(t) = \frac{\partial u}{\partial \varepsilon}(0, t)$ folgt, indem wir hinten die Ableitungen vertauschen,

$$(6.3) \quad \phi'(0) = \int_a^b \sum_{i=1}^n \left(\frac{\partial f}{\partial x_i}(t, u, u') \varphi_i(t) + \frac{\partial f}{\partial v_i}(t, u, u') \varphi_i'(t) \right) dt.$$

Schließlich mit partieller Integration im hinteren Term

$$\phi'(0) = \int_a^b \sum_{i=1}^n \underbrace{\left(\frac{\partial f}{\partial x_i}(t, u, u') - \frac{d}{dt} \left[\frac{\partial f}{\partial v_i}(t, u, u') \right] \right)}_{= L_f(u)_i} \varphi_i dt + \left[\sum_{i=1}^n \frac{\partial f}{\partial v_i}(t, u, u') \varphi_i \right]_{t=a}^{t=b}.$$

Die Behauptung des Lemmas folgt, indem wir die Summen als Skalarprodukte schreiben. \square

Um eine optimale Funktion zu charakterisieren, müssen wir sie mit hinreichend vielen Variationen vergleichen. Das folgende Lemma gibt an, wieviele wir tatsächlich brauchen.

Lemma 6.2 Sei $I = (a, b)$. Für $f \in C^0(I, \mathbb{R}^n)$ gelte

$$(6.4) \quad \int_a^b \langle f, \varphi \rangle = 0 \quad \text{für alle } \varphi \in C_c^\infty(I, \mathbb{R}^n).$$

Dann ist f die Nullfunktion.

BEWEIS: Wir zeigen das Lemma erst für $n = 1$. Es gibt eine Funktion $\eta \in C^\infty(\mathbb{R})$ mit

$$\eta(s) = 0 \quad \text{für } |s| \geq 1, \quad \eta \geq 0 \quad \text{und} \quad \int_{\mathbb{R}} \eta = 1.$$

Ein konkretes Beispiel ist, bei passender Wahl von $a > 0$, die Funktion

$$\eta(s) = \begin{cases} a \exp \frac{1}{s^2-1} & \text{für } |s| < 1 \\ 0 & \text{sonst.} \end{cases}$$

Angenommen es ist $\varepsilon := f(t_0) > 0$ für ein $t_0 \in I$. Dann gibt es ein $\delta > 0$ mit $f(t) \geq \varepsilon/2$ für $t \in [t_0 - \delta, t_0 + \delta] \subset I$. Betrachte die reskalierte Funktion

$$\eta_{t_0, \delta}(t) = \frac{1}{\delta} \eta \left(\frac{t - t_0}{\delta} \right).$$

Dann gilt $\eta_{t_0, \delta}(t) = 0$ für $|t - t_0| \geq \delta$, sowie $\int_{\mathbb{R}} \eta_{t_0, \delta} = 1$. Nach Voraussetzung

$$0 = \int_a^b f(t) \eta_{t_0, \delta}(t) dt = \int_{t_0 - \delta}^{t_0 + \delta} f(t) \eta_{t_0, \delta}(t) dt \geq \frac{\varepsilon}{2} \int_a^b \eta_{t_0, \delta}(t) dt = \frac{\varepsilon}{2},$$

ein Widerspruch. Im Fall $f : I \rightarrow \mathbb{R}^n$ liefert die Voraussetzung, für alle $\varphi \in C_c^\infty(I)$,

$$0 = \int_a^b \langle f, \varphi e_i \rangle = \int_a^b f_i \varphi.$$

Aus obigem folgt $f_i = 0$ für $i = 1, \dots, n$, das Lemma ist bewiesen. \square

Satz 6.4 (Euler-Lagrange-Gleichungen) Sei $\mathcal{F}(u) = \int_a^b f(t, u, u')$ ein Variationsintegral mit $f \in C^2(I \times \mathbb{R}^n \times \mathbb{R}^n)$, $f = f(t, x, v)$. Sei $u \in C^2(I, \mathbb{R}^n)$ stationärer Punkt, d. h.

$$\frac{d}{d\varepsilon} \mathcal{F}(u + \varepsilon\varphi)|_{\varepsilon=0} = 0 \quad \text{für alle } \varphi \in C_c^\infty(I, \mathbb{R}^n).$$

Dann gelten die Euler-Lagrange-Gleichungen

$$L_f(u) = 0 \quad \Leftrightarrow \quad \frac{\partial f}{\partial x_i}(t, u, u') - \frac{d}{dt} \frac{\partial f}{\partial v_i}(t, u, u') = 0 \quad \text{für } i = 1, \dots, n.$$

Bemerkung. Es handelt sich um ein System von n Differentialgleichungen zweiter Ordnung, wie man durch Ausdifferenzieren des zweiten Terms sieht.

BEWEIS: Die Randterme in (6.1) verschwinden, da $\varphi(a) = \varphi(b) = 0$ nach Voraussetzung. Die Aussage folgt dann aus den Lemmas 6.1 und 6.2. \square

Beispiel 6.5 (Bogenlänge) Wir betrachten die Bogenlänge aus Beispiel 6.2

$$\mathcal{F}(u) = \int_a^b |u'(t)| dt \quad \text{für } u \in C^2([a, b], \mathbb{R}^n).$$

Die Lagrangefunktion und ihre Ableitung sind

$$f(t, x, v) = |v|, \quad D_v f(t, x, v) = \frac{v}{|v|} \quad \text{falls } v \neq 0.$$

Die Euler-Lagrange-Gleichungen lauten somit

$$L_f(u) = -\frac{d}{dt} \frac{u'}{|u'|} = 0.$$

Die Gleichung sagt aus, dass der Einheitstangentenvektor $\frac{u'}{|u'|}$ konstant ist. Es ist nicht schwer zu sehen, dass $u(t)$ dann die Strecke von $u(a)$ nach $u(b)$ durchläuft. Allerdings wird zur Herleitung der Euler-Lagrange Gleichungen gebraucht, dass $u'(t) \neq 0$ für alle $t \in [a, b]$.

Beispiel 6.6 Bewegung eines Massenpunkts in einem konservativen Kraftfeld:

$$\mathcal{F}(u) = \int_a^b \left(\frac{m}{2} |u'|^2 - V(u(t)) \right) dt, \quad f(t, x, v) = \frac{m}{2} |v|^2 - V(x).$$

Das zugehörige Kraftfeld ist gegeben durch $F(x) = -\text{grad } V(x)$; das Minuszeichen ist in der Physik üblich. Dann ergibt sich

$$\frac{\partial f}{\partial x_i}(t, x, v) = F_i(x), \quad \frac{\partial f}{\partial v_i}(t, x, v) = mv_i.$$

Die Euler-Lagrange-Gleichungen lauten somit $F(u) - mu'' = 0$, es sind die Newtonschen Bewegungsgleichungen.

Viele interessante Parameterintegrale sind uneigentliche Integrale, zum Beispiel bei der Definition der Gammafunktion oder der Fouriertransformation. Aus Zeitgründen können wir darauf jetzt nicht eingehen, werden aber Parameterintegrale nochmals innerhalb der Theorie des Lebesgue-Integrals im dritten Semester aufgreifen.

7 Diffeomorphismen

Thema dieses und des folgenden Abschnitts ist die lokale Lösbarkeit nichtlinearer Gleichungen. Sei $\Omega \subset \mathbb{R}^n$ offen und $f \in C^1(\Omega, \mathbb{R}^m)$, und es sei schon eine Lösung von $f(x_0) = y_0$ gegeben. Dann stellen wir uns folgende Fragen:

- (1) Hat die Gleichung $f(x) = y$ zu jedem y nahe bei y_0 eine Lösung x nahe bei x_0 ?
- (2) Ist x_0 die einzige Lösung der Gleichung $f(x) = y_0$ in einer Umgebung von x_0 ?
- (3) Falls nicht, wie sieht die Lösungsmenge $f^{-1}\{y_0\}$ nahe bei x_0 aus?

Betrachten wir erst den affin-linearen Fall $f(x) = Ax + b$ für $A \in L(\mathbb{R}^n, \mathbb{R}^m)$. Wegen $f(x_0) = y_0$ ist dann $b = y_0 - Ax_0$, das heißt f hat die Form $f(x) = y_0 + A(x - x_0)$. Also gilt

$$f(x) = y \quad \Leftrightarrow \quad A(x - x_0) = y - y_0.$$

In diesem Fall liefert die Lineare Algebra folgende, sogar globale Antworten:

- (1) Es gibt eine Lösung für alle $y \in \mathbb{R}^m \quad \Leftrightarrow \quad \text{rang } A = m.$
- (2) x_0 ist einzige Lösung von $f(x) = y_0 \quad \Leftrightarrow \quad \ker A = \{0\} \quad \Leftrightarrow \quad \text{rang } A = n.$
- (3) $f^{-1}\{y_0\} = x_0 + \ker A$ ist affiner Unterraum der Dimension $n - \text{rang } A.$

Sei nun $f \in C^1(\Omega, \mathbb{R}^m)$ mit $f(x_0) = y_0$. Dann ist f differenzierbar in x_0 , das heißt

$$f(x) = f(x_0) + Df(x_0)(x - x_0) + R_f(x) \quad \text{wobei } R_f(x_0) = 0, DR_f(x_0) = 0.$$

Setzen wir $A = Df(x_0)$, so ergibt sich die Formulierung

$$f(x) = y \quad \Leftrightarrow \quad A(x - x_0) + R_f(x) = y - y_0.$$

Wir wollen dies als Störung der linearen Gleichung auffassen und hoffen, dass sich die Aussagen in einer lokalen Version geeignet übertragen lassen. In diesem Abschnitt geht es um den Fall $n = m$, das heißt es gibt genauso viele Unbekannte wie Gleichungen. Im darauffolgenden Abschnitt über implizite Funktionen behandeln wir den Fall $n \geq m$.

Definition 7.1 Eine Abbildung $f : U \rightarrow V$ zwischen offenen Mengen $U, V \subset \mathbb{R}^n$ heißt *Diffeomorphismus der Klasse C^r* , wobei $r \in \mathbb{N} \cup \{\infty\}$, falls f bijektiv ist und sowohl f als auch f^{-1} sind r -mal stetig differenzierbar.

Beispiel 7.1 Sei $f \in C^1(I)$, $I = (a, b)$, mit $f' > 0$ auf ganz I , also f streng monoton wachsend. Nach Analysis 1, Satz 9.4, ist dann $J := f(I)$ ein offenes Intervall, und die Umkehrfunktion $g : J \rightarrow I$ ist differenzierbar mit Ableitung

$$g' = \frac{1}{f' \circ g} \in C^0(J).$$

Also ist f ein C^1 -Diffeomorphismus auf $J = f(I)$. Im Fall $f' < 0$ auf I folgt das natürlich analog. Umgekehrt: ist $f : I \rightarrow J$ ein C^1 -Diffeomorphismus zwischen offenen Intervallen, mit Umkehrfunktion $g : J \rightarrow I$, so ergibt die Kettenregel

$$g(f(x)) = x \quad \Rightarrow \quad g'(f(x))f'(x) = 1 \quad \Rightarrow \quad f'(x) \neq 0.$$

Nach dem Zwischenwertsatz ist entweder $f' > 0$ oder $f' < 0$ auf I . Zum Beispiel ist die Abbildung $f : (-1, 1) \rightarrow (-1, 1)$, $f(x) = x^3$, zwar bijektiv, genauer streng monoton wachsend, und von der Klasse C^1 , aber wegen $f'(0) = 0$ kann sie kein C^1 -Diffeomorphismus sein. In der Tat, die Umkehrabbildung ist im Punkt $y = 0$ nicht differenzierbar:

$$g : (-1, 1) \rightarrow (-1, 1), g(y) = \begin{cases} \sqrt[3]{y} & \text{für } y \geq 0 \\ -\sqrt[3]{-y} & \text{für } y < 0 \end{cases}$$

Beispiel 7.2 (Polarkoordinaten) Seien $U = \{(r, \theta) \in \mathbb{R}^2 : r > 0, 0 < \theta < 2\pi\}$ und $V = \mathbb{R}^2 \setminus \{(x, 0) : x \geq 0\}$. Wir betrachten die Polarkoordinatenabbildung

$$f \in C^\infty(U, V), f(r, \theta) = (r \cos \theta, r \sin \theta).$$

Die Umkehrabbildung $g : V \rightarrow U$ lautet mit $r = \sqrt{x^2 + y^2}$

$$g(x, y) = \begin{cases} \left(r, \arccos \frac{x}{r}\right) & \text{für } y > 0 \\ \left(r, \frac{\pi}{2} + \arccos \frac{y}{r}\right) & \text{für } x < 0, \\ \left(r, \pi + \arccos\left(-\frac{x}{r}\right)\right) & \text{für } y < 0. \end{cases}$$

Die Darstellungen sind jeweils in C^∞ , also ist f ein C^∞ -Diffeomorphismus.

Beispiel 7.3 (Inversion) Die Inversion an der Sphäre $\mathbb{S}^{n-1} = \{x \in \mathbb{R}^n : |x| = 1\}$ ist

$$f : \mathbb{R}^n \setminus \{0\} \rightarrow \mathbb{R}^n \setminus \{0\}, f(x) = \frac{x}{|x|^2}.$$

Es gilt $f \in C^\infty(\mathbb{R}^n \setminus \{0\}, \mathbb{R}^n)$ und $f^{-1} = f$, also ist f ein C^∞ -Diffeomorphismus. Die beschränkte Menge $B_1(0)$ geht unter f in die unbeschränkte Menge $\mathbb{R}^n \setminus \overline{B_1(0)}$.

Lemma 7.1 (Ableitung der Umkehrfunktion) Sei $f : U \rightarrow V$ bijektiv mit Umkehrabbildung $g : V \rightarrow U$, wobei $U \subset \mathbb{R}^n$ und $V \subset \mathbb{R}^m$ offen. Ist f in x_0 und g in $y_0 = f(x_0)$ differenzierbar, so ist die lineare Abbildung $Df(x_0) : \mathbb{R}^n \rightarrow \mathbb{R}^m$ invertierbar. Insbesondere muss $m = n$ sein. Weiter gilt

$$Dg(y_0) = Df(x_0)^{-1} \quad \text{mit } y_0 = f(x_0).$$

BEWEIS: Aus $g(f(x)) = x$ und $f(g(y)) = y$ folgt jeweils mit der Kettenregel

$$Dg(y_0)Df(x_0) = \text{Id}_{\mathbb{R}^n} \quad \text{und} \quad Df(x_0)Dg(y_0) = \text{Id}_{\mathbb{R}^m}.$$

Also ist $Df(x_0)$ injektiv und surjektiv, sprich invertierbar, und es folgt $m = n$. □

Das Lemma besagt, dass ein Diffeomorphismus zwischen offenen Mengen $U \subset \mathbb{R}^n$ und $V \subset \mathbb{R}^m$ nur möglich ist für $m = n$. Dies wird als Invarianz der Dimension bezeichnet. Nach einem Satz von Brouwer (1910) bleibt die Dimension auch unter Homeomorphismen erhalten, das heißt f und f^{-1} sind nur stetig. Peano hatte zuvor surjektive stetige Abbildungen von einem Intervall auf die Fläche eines Quadrats konstruiert, daher stellte sich die Frage nach der Invarianz der Dimension. Die Peanokurven sind aber keine Homeomorphismen, sie sind nicht injektiv. Der Satz von Brouwer wird mit dem Konzept des Abbildungsgrads bewiesen, das in der nichtlinearen Funktionalanalysis oder der Algebraischen Topologie eingeführt wird.

Man bezeichnet $\det Df(x_0)$ als Jacobideterminante von f im Punkt x_0 . In der Situation von Lemma 7.1 folgt aus dem Determinantenmultiplikationssatz

$$(7.1) \quad \det Dg(y_0) \det Df(x_0) = 1 \quad \text{für } y_0 = f(x_0).$$

Lemma 7.2 (Höhere Ableitungen der Umkehrfunktion) *Seien $U, V \subset \mathbb{R}^n$ offen, und $f : U \rightarrow V$ bijektiv. Ist $f \in C^r(U, V)$ für ein $r \in \mathbb{N} \cup \{\infty\}$, und ist die Umkehrabbildung $g : V \rightarrow U$ differenzierbar, so folgt $g \in C^r(V, U)$.*

BEWEIS: Nach Lemma 7.1 ist $Df(x)$ invertierbar und es gilt $Dg = (Df)^{-1} \circ g$, also nach der Cramerschen Regel

$$(7.2) \quad \frac{\partial g_i}{\partial y_j} = (-1)^{i+j} \frac{M_{ji}(Df)}{\det Df} \circ g.$$

Dabei bezeichnet $M_{ji}(Df)$ die Determinante der Matrix, die aus Df durch Streichen der j -ten Zeile und i -ten Spalte entsteht. Wir zeigen die Behauptung durch Induktion über $r \in \mathbb{N}$. Da g nach Voraussetzung differenzierbar und somit stetig ist, vgl. Satz 3.2, ist für $f \in C^1$ die rechte Seite in (7.2) stetig als Produkt, Quotient und Verkettung stetiger Funktionen, und damit $g \in C^1$. Ist $f \in C^r$ und induktiv schon $g \in C^{r-1}$, so ist die rechte Seite von der Klasse C^{r-1} als Produkt, Quotient und Verkettung von C^{r-1} -Funktionen, siehe Folgerung 3.1, und damit $g \in C^r$, was zu zeigen war. \square

Nach diesen Vorüberlegungen wollen wir die Frage der Existenz einer Lösung angehen. Für eine allgemeine nichtlineare Gleichung kann nicht erwartet werden, dass die Lösung durch eine explizite Formel geliefert wird. Vielmehr brauchen wir einen abstrakten Existenzsatz. Dazu die folgenden Definitionen.

Definition 7.2 *Eine Folge $x_k, k \in \mathbb{N}$, in einem metrischen Raum (X, d) heißt Cauchyfolge, wenn es zu jedem $\varepsilon > 0$ ein $K \in \mathbb{R}$ gibt mit*

$$d(x_k, x_l) < \varepsilon \quad \text{für alle } k, l > K.$$

Ein metrischer Raum heißt vollständig, wenn jede Cauchyfolge x_k in X konvergiert, das heißt es gibt ein $x \in X$ mit $d(x, x_k) \rightarrow 0$ mit $k \rightarrow \infty$.

Natürlich ist \mathbb{R}^n mit der Euklidischen Abstandsfunktion ein vollständiger metrischer Raum. Aber jede abgeschlossene Teilmenge $A \subset \mathbb{R}^n$ ist mit dem Euklidischen Abstand auch ein vollständiger metrischer Raum, denn eine Cauchyfolge $x_k \in A$ ist auch Cauchyfolge in \mathbb{R}^n und konvergiert damit gegen ein $x \in \mathbb{R}^n$, und es gilt $x \in A$ wegen A abgeschlossen.

Satz 7.1 (Fixpunktsatz von Banach) *Sei (X, d) ein vollständiger metrischer Raum, und $F : X \rightarrow X$ eine Kontraktion, das heißt es gibt ein $\theta \in [0, 1)$ mit*

$$(7.3) \quad d(F(x), F(y)) \leq \theta d(x, y) \quad \text{für alle } x, y \in X.$$

Dann gibt es genau ein $x \in X$ mit $F(x) = x$.

BEWEIS: Die Eindeutigkeit ist klar, denn aus $F(x) = x$ und $F(y) = y$ folgt

$$d(x, y) = d(F(x), F(y)) \leq \theta d(x, y) \quad \Rightarrow \quad d(x, y) = 0, \text{ also } x = y.$$

Um den Fixpunkt zu konstruieren, betrachten wir die rekursiv definierte Folge $x_{n+1} = F(x_n)$ mit beliebigem Startwert $x_0 \in X$. Es folgt aus (7.3) für $n \geq 1$

$$(7.4) \quad d(x_{n+1}, x_n) = d(F(x_n), F(x_{n-1})) \leq \theta d(x_n, x_{n-1}).$$

Wir können uns einen müder werdenden Frosch vorstellen, dessen Sprünge jedes Mal um ein Faktor $\theta \in [0, 1)$ kürzer werden. Wie weit kann der Frosch insgesamt kommen? Es folgt per Induktion aus (7.4)

$$(7.5) \quad d(x_{n+1}, x_n) \leq \theta^n d(x_1, x_0) \quad \text{für } n \in \mathbb{N}_0,$$

und hieraus weiter mit der Dreiecksungleichung und der geometrischen Reihe

$$d(x_n, x_0) \leq \sum_{j=0}^{n-1} d(x_{j+1}, x_j) \leq \sum_{j=0}^{n-1} \theta^j d(x_1, x_0) \leq \frac{1}{1-\theta} d(x_1, x_0).$$

Indem wir x_n statt x_0 als Startwert auffassen, haben wir für $m > n$

$$(7.6) \quad d(x_m, x_n) \leq \frac{1}{1-\theta} d(x_{n+1}, x_n) \leq \frac{\theta^n}{1-\theta} d(x_1, x_0).$$

Also ist $(x_n)_{n \in \mathbb{N}_0}$ eine Cauchyfolge, und konvergiert nach Voraussetzung gegen ein $x \in X$. Da F nach Voraussetzung Lipschitzstetig ist (mit Konstante θ), folgt

$$F(x) = \lim_{n \rightarrow \infty} F(x_n) = \lim_{n \rightarrow \infty} x_{n+1} = x,$$

und die Existenz des Fixpunkts ist gezeigt. □

Aus Sicht der Numerik ist eine Abschätzung von Interesse, wie weit die Iteration im n -ten Schritt noch vom gesuchten Fixpunkt entfernt ist. Mit $m \rightarrow \infty$ folgt aus (7.6)

$$d(x, x_n) \leq \frac{\theta^n}{1-\theta} d(x_1, x_0).$$

Das folgende zentrale Resultat wird auch als Umkehrsatz bezeichnet.

Satz 7.2 (über inverse Funktionen) Sei $\Omega \subset \mathbb{R}^n$ offen und $f \in C^1(\Omega, \mathbb{R}^n)$. Ist $Df(x_0) \in L(\mathbb{R}^n, \mathbb{R}^n)$ invertierbar, so gibt es eine offene Umgebung U von x_0 , so dass gilt:

- (a) $V = f(U)$ ist offene Umgebung von $y_0 = f(x_0)$
- (b) $f|_U : U \rightarrow V$ ist Diffeomorphismus der Klasse C^1 .

Zusatz. Ist $f \in C^r(\Omega, \mathbb{R}^n)$ für ein $r \in \mathbb{N} \cup \{\infty\}$, so ist $g = (f|_U)^{-1} \in C^r(V, \mathbb{R}^n)$.

BEWEIS: **Schritt 1** Formulierung als Fixpunktproblem

Mit $y_0 = f(x_0)$, $A := Df(x_0)$ und $R_f(x) := f(x) - (f(x_0) + A(x - x_0))$ hatten wir

$$f(x) = y \quad \Leftrightarrow \quad A(x - x_0) + R_f(x) = y - y_0 \quad \Leftrightarrow \quad x = x_0 + A^{-1}(y - y_0 - R_f(x)).$$

Für $y \in \mathbb{R}^n$ definieren wir also $\phi_y : \Omega \rightarrow \mathbb{R}^n$, $\phi_y(x) = x_0 + A^{-1}(y - y_0 - R_f(x))$, und erhalten

$$(7.7) \quad f(x) = y \quad \Leftrightarrow \quad \phi_y(x) = x.$$

Schritt 2 *Konstruktion der Lösung*

Wir bestimmen $\delta_0 > 0$, so dass für jedes $\delta \in (0, \delta_0]$ die Abbildung $\phi_y : \overline{B_\delta(x_0)} \rightarrow \overline{B_\delta(x_0)}$ definiert und kontrahierend ist, sofern $y \in B_\varepsilon(y_0)$ mit $\varepsilon = \varepsilon(\delta) > 0$. Setze $\Lambda = |A^{-1}| \in (0, \infty)$. $DR_f(x) = Df(x) - A$ ist stetig mit $DR_f(x_0) = 0$, folglich gibt es $\delta_0 > 0$ mit

$$\overline{B_{\delta_0}(x_0)} \subset \Omega \quad \text{und} \quad \|DR_f(x)\| \leq \frac{1}{2\Lambda} \quad \text{für } |x - x_0| \leq \delta_0.$$

Aus dem Schrankensatz, siehe Satz 4.2, folgt

$$(7.8) \quad x_{1,2} \in \overline{B_{\delta_0}(x_0)} \quad \Rightarrow \quad |R_f(x_1) - R_f(x_2)| \leq \frac{1}{2\Lambda} |x_1 - x_2|.$$

Wir berechnen nun

$$|\phi_y(x_1) - \phi_y(x_2)| = |A^{-1}R_f(x_1) - A^{-1}R_f(x_2)| \leq \Lambda |R_f(x_1) - R_f(x_2)|.$$

Also folgt aus (7.8) die Kontraktionseigenschaft

$$(7.9) \quad x_{1,2} \in \overline{B_{\delta_0}(x_0)} \quad \Rightarrow \quad |\phi_y(x_1) - \phi_y(x_2)| \leq \frac{1}{2} |x_1 - x_2|.$$

Wir müssen sicherstellen, dass ϕ_y eine Selbstabbildung ist. Dazu schätzen wir ab

$$\begin{aligned} |\phi_y(x) - x_0| &= |A^{-1}(y - y_0 - R_f(x))| \\ &\leq \Lambda(|y - y_0| + |R_f(x) - R_f(x_0)|) \quad (\text{da } R_f(x_0) = 0) \\ &\leq \Lambda|y - y_0| + \frac{1}{2}|x - x_0| \quad \text{für } x \in \overline{B_{\delta_0}(x_0)} \text{ nach (7.8)}. \end{aligned}$$

Also folgt für $\delta \in (0, \delta_0]$, wenn wir $\varepsilon = \delta/(2\Lambda) > 0$ wählen,

$$(7.10) \quad x \in \overline{B_\delta(x_0)}, y \in B_\varepsilon(y_0) \quad \Rightarrow \quad |\phi_y(x) - x_0| < \Lambda\varepsilon + \frac{1}{2}\delta = \delta.$$

Wegen (7.10) und (7.9) ist $\phi_y : \overline{B_\delta(x_0)} \rightarrow \overline{B_\delta(x_0)}$ eine Kontraktion mit Konstante $\theta = 1/2$. Nach dem Banachschen Fixpunktsatz gibt es zu jedem $y \in B_\varepsilon(y_0)$ genau ein $x \in \overline{B_\delta(x_0)}$ mit $\phi_y(x) = x$, das heißt $f(x) = y$ nach (7.7). Es ist sogar $x \in B_\delta(x_0)$, denn nach (7.10) gilt $|x - x_0| = |\phi_y(x) - x_0| < \delta$. Die Mengen $V = B_\varepsilon(y_0)$ und $U = f^{-1}(V) \cap B_\delta(x_0)$ sind offen, vgl. Satz 1.6 für die Offenheit von U . Also gilt Behauptung (a), und $f|_U : U \rightarrow V$ bijektiv.

Schritt 3 *Differenzierbarkeit der inversen Abbildung*

Sei $g : V \rightarrow U$ die Umkehrabbildung von $f|_U : U \rightarrow V$. Dann gilt

$$(7.11) \quad |g(y) - x_0| = |\phi_y(g(y)) - x_0| \leq \Lambda|y - y_0| + \frac{1}{2}|g(y) - x_0| \quad \Rightarrow \quad |g(y) - x_0| \leq 2\Lambda|y - y_0|.$$

Insbesondere ist g stetig in y_0 mit $g(y_0) = x_0$. Wir zeigen nun $Dg(y_0) = A^{-1}$. Für $y \neq y_0$ ist $g(y) \neq x_0$ und es gilt die Abschätzung

$$\begin{aligned} \frac{|g(y) - (g(y_0) + A^{-1}(y - y_0))|}{|y - y_0|} &= \frac{|\phi_y(g(y)) - x_0 - A^{-1}(y - y_0)|}{|y - y_0|} \\ &= \frac{|A^{-1}R_f(g(y))|}{|y - y_0|} \\ &\leq \Lambda \frac{|R_f(g(y))|}{|g(y) - x_0|} \frac{|g(y) - x_0|}{|y - y_0|}. \end{aligned}$$

Mit $y \rightarrow y_0$ geht die rechte Seite gegen Null, denn es ist $|g(y) - x_0|/|y - y_0| \leq 2\Lambda$ nach (7.11) und $|R_f(x)|/|x - x_0| \rightarrow 0$ mit $x = g(y) \rightarrow x_0$. Dies zeigt $Dg(y_0) = A^{-1}$.

Um die Differenzierbarkeit von g auf V zu bekommen, wählen wir $\delta \in (0, \delta_0]$ so klein, dass $\det Df(x) \neq 0$ für alle $x \in B_\delta(x_0)$. Sei $y \in V$ beliebig gegeben. Die Voraussetzungen des Satzes gelten dann für $f|_U \in C^1(U, \mathbb{R}^n)$ und $x = g(y) \in U$. Wie bewiesen gibt es also $\tilde{U} \subset U$ mit folgenden Eigenschaften: $\tilde{V} = f(\tilde{U})$ ist offene Umgebung von $y = f(x)$, $f|_{\tilde{U}} : \tilde{U} \rightarrow \tilde{V}$ ist bijektiv, und $\tilde{g} = (f|_{\tilde{U}})^{-1}$ ist differenzierbar in y . Aber $g|_{\tilde{V}} = \tilde{g}$ da $f|_U$ injektiv. Somit ist g differenzierbar in $y \in V$.

Lemma 7.2 liefert schließlich $g \in C^1(V, U)$. Ist $f \in C^r(\Omega, \mathbb{R}^n)$ für ein $r \in \mathbb{N} \cup \{\infty\}$, so ist $g \in C^r(V, U)$, ebenfalls nach Lemma 7.2. \square

Als unmittelbare Konsequenz des Satzes halten wir fest:

Folgerung 7.1 Sei $\Omega \subset \mathbb{R}^n$ offen und $f \in C^1(\Omega, \mathbb{R}^n)$. Ist $Df(x)$ invertierbar für alle $x \in \Omega$, so ist $f(\Omega) \subset \mathbb{R}^n$ offen.

BEWEIS: Nach Satz 7.2 hat jeder Punkt $y \in f(\Omega)$ eine offene Umgebung $V \subset f(\Omega)$. \square

Beispiel 7.4 Wie wir in Beispiel 7.1 gesehen haben, bildet eine eindimensionale Funktion $f : (a, b) \rightarrow \mathbb{R}$ mit $f' \neq 0$ das gesamte Definitionsintervall diffeomorph auf das Bildintervall ab, das heißt es gilt eine globale Version des Umkehrsatzes. Das folgende Beispiel zeigt, dass eine entsprechende Aussage für Funktionen mehrerer Variabler im allgemeinen nicht wahr ist. In reellen Koordinaten $z = x + iy$ lautet die komplexe Exponentialfunktion

$$\exp : \mathbb{R}^2 \rightarrow \mathbb{R}^2, \exp(x, y) = (e^x \cos y, e^x \sin y).$$

Es gilt $\exp(\mathbb{R}^2) = \mathbb{R}^2 \setminus \{0\}$. Die Jacobideterminante von \exp ist nirgends Null, genauer gilt

$$D \exp(x, y) = \begin{pmatrix} e^x \cos y & -e^x \sin y \\ e^x \sin y & e^x \cos y \end{pmatrix} \Rightarrow \det D \exp(x, y) = e^{2x} \neq 0.$$

Die Abbildung ist jedoch nicht injektiv, denn es ist $\exp(x, y + 2k\pi) = \exp(x, y)$ für alle $k \in \mathbb{Z}$.

8 Implizite Funktionen

Thema dieses Kapitels ist die lokale Lösung nichtlinearer Gleichungen im unterbestimmten Fall, das heißt es gibt mehr Unbekannte als Bedingungen. Dazu nehmen wir an, dass die Variablen in zwei Gruppen eingeteilt sind:

$$f \in C^1(\Omega, \mathbb{R}^k), f = f(x, y), \quad \text{wobei } (x, y) \in \Omega \subset \mathbb{R}^m \times \mathbb{R}^k.$$

Die Frage aus dem letzten Kapitel lautet präziser:

- Sei $f(x_0, y_0) = z_0$ gegeben. Wie sieht die Lösungsmenge der Gleichung $f(x, y) = z_0$ nahe bei (x_0, y_0) aus?
- Können wir die Gleichung nach y eindeutig auflösen, sprich die Lösungsmenge lokal als Graph $y = g(x)$ darstellen?

Die Lösungen einer Gleichung $f(x, y) = z_0$ kann im allgemeinen nicht explizit durch Umformungen berechnet werden; deshalb wird $y = g(x)$ als implizit gegebene Funktion bezeichnet.

Beispiel 8.1 Betrachte die Gleichung

$$f(x, y) = x^2 + y^2 = 1 \quad \text{für } (x, y) \in \mathbb{R} \times \mathbb{R} = \mathbb{R}^2.$$

Die Lösungsmenge ist der Einheitskreis \mathbb{S}^1 . Hier ist die Auflösung der Gleichung explizit möglich: ist $(x_0, y_0) \in \mathbb{S}^1$ mit $y_0 > 0$, so kann \mathbb{S}^1 in einer Umgebung als Graph $y = \sqrt{1 - x^2}$ dargestellt werden. Analog im Fall $y_0 < 0$, mit Graphenfunktion $y = -\sqrt{1 - x^2}$. Dagegen hat der Punkt $(1, 0)$ keine Umgebung, in der die Gleichung eindeutig nach y aufgelöst werden kann, für $x < 1$ gibt es nahebei die beiden Lösungen $(x, \pm\sqrt{1 - x^2})$, und für $x > 1$ gar keine.

Für eine reelle Funktion $f = f(x, y)$ von zwei Variablen kann die Lösungsmenge der Gleichung $f(x, y) = z_0$ als Höhenlinie interpretiert werden. Allerdings ist die Bezeichnung salopp, es kann Singularitäten geben, in denen die Menge nicht lokal wie eine Linie aussieht. Ein Beispiel ist die Gleichung $xy = 0$, die im Nullpunkt nicht regulär ist. Es kann vorkommen, dass alle Lösungen singuläre Punkte sind, etwa bei der Gleichung $x^2 + y^2 = 0$. Wir betrachten nun den linearen Fall, um ein Kriterium für die lokale Lösbarkeit von Gleichungen zu erhalten.

Beispiel 8.2 Betrachte eine lineare Funktion von zwei Variablen, also

$$f : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}, f(x, y) = ax + by \quad \text{mit } a, b \in \mathbb{R}.$$

Die Gleichung $f(x, y) = z_0$ ist genau dann nach y auflösbar wenn $b \neq 0$, die Funktion lautet

$$y = \frac{1}{b}(z_0 - ax), \quad x \in \mathbb{R}.$$

In höheren Dimensionen ist die Sache analog. Für $f : \mathbb{R}^m \times \mathbb{R}^k \rightarrow \mathbb{R}^k$ linear unterteilen wir die $k \times (m + k)$ -Matrix in eine $k \times m$ -Matrix A und eine $k \times k$ -Matrix B , d. h.

$$f(x, y) = Ax + By \quad \text{mit } A \in L(\mathbb{R}^m, \mathbb{R}^k), \quad B \in L(\mathbb{R}^k, \mathbb{R}^k).$$

Die Gleichung $Ax + By = z_0$ hat zu festem $x \in \mathbb{R}^m$ eine eindeutige Auflösung nach y dann und nur dann, wenn B invertierbar ist. Ist das der Fall, so lautet die Auflösung

$$y = B^{-1}(z_0 - Ax), \quad x \in \mathbb{R}^m.$$

Allgemein schreiben wir die Jacobimatrix von $f = f(x, y)$ in der Form

$$Df(x, y) = (D_x f(x, y), D_y f(x, y)) \in (\mathbb{R}^{k \times m}, \mathbb{R}^{k \times k}).$$

Wenn wir nach $y = g(x)$ auflösen wollen, so sollte nach Beispiel 8.2 die Ableitung $D_y f(x_0, y_0) \in \mathbb{R}^{k \times k}$ invertierbar sein. In den Anwendungen ist die Einteilung in die beiden Variablengruppen nicht immer vorgegeben, das heißt es könnte nach verschiedenen Gruppen von je k Variablen aufgelöst werden. So kann der Einheitskreis in einer Umgebung von $(1, 0)$ zwar nicht als Graph $y = g(x)$ geschrieben werden, wohl aber als Graph $x = g(y)$, und außer in den vier Punkten $\pm e_1, \pm e_2$ könnte sowohl nach x als auch nach y aufgelöst werden.

Merkregel. Die Ableitung nach den Variablen, nach denen aufgelöst werden soll, muss invertierbar sein. Im Fall $k = 1$ bedeutet das einfach

$$\frac{\partial f}{\partial y}(x_0, y_0) \neq 0.$$

Satz 8.1 (über implizite Funktionen) Sei $\Omega \subset \mathbb{R}^m \times \mathbb{R}^k$ offen und $f \in C^1(\Omega, \mathbb{R}^k)$. Ist $f(x_0, y_0) = z_0$ und $D_y f(x_0, y_0) \in L(\mathbb{R}^k, \mathbb{R}^k)$ invertierbar, so gibt es offene Umgebungen $U \subset \mathbb{R}^m$ von x_0 und $V \subset \mathbb{R}^k$ von y_0 , sowie eine Funktion $g \in C^1(U, V)$, mit

$$(8.1) \quad \{(x, y) \in U \times V : f(x, y) = z_0\} = \{(x, g(x)) : x \in U\}.$$

Es folgt $g(x_0) = y_0$, und die Funktion g hat die Ableitung

$$(8.2) \quad Dg(x_0) = -(D_y f(x_0, y_0))^{-1} D_x f(x_0, y_0).$$

Zusatz. Für jedes $r \in \mathbb{N} \cup \{\infty\}$ gilt die Implikation

$$f \in C^r(\Omega, \mathbb{R}^k) \Rightarrow g \in C^r(U, \mathbb{R}^k).$$

BEWEIS: Wir verwenden einen Trick, um den Satz über inverse Funktionen anwenden zu können, und zwar betrachten wir $F \in C^1(\Omega, \mathbb{R}^m \times \mathbb{R}^k)$, $F(x, y) = (x, f(x, y))$. Es gilt

$$DF = \begin{pmatrix} E_m & 0 \\ D_x f & D_y f \end{pmatrix} \in \begin{pmatrix} \mathbb{R}^{m \times m} & \mathbb{R}^{m \times k} \\ \mathbb{R}^{k \times m} & \mathbb{R}^{k \times k} \end{pmatrix}.$$

Es folgt $\det DF(x_0, y_0) = \det D_y f(x_0, y_0) \neq 0$ nach Voraussetzung². Nach dem Umkehrsatz gibt es offene Umgebungen $U_0 \times V$ von (x_0, y_0) , W von (x_0, z_0) , so dass $F : U_0 \times V \rightarrow W$ ein Diffeomorphismus ist. Wir bezeichnen die Umkehrabbildung mit $G \in C^1(W, U_0 \times V)$. Ist $(x, z) \in W$, also $(x, z) = (x, f(x, y))$ mit $(x, y) \in U_0 \times V$ nach Konstruktion, so folgt

$$G(x, z) = G(x, f(x, y)) = G(F(x, y)) = (x, y).$$

Also gilt $G(x, z) = (x, g_0(x, z))$ mit $g_0 \in C^1(W, \mathbb{R}^k)$. Sei U die Menge der $x \in U_0$ mit $(x, z_0) \in W$. Da W offen in $\mathbb{R}^m \times \mathbb{R}^k$, ist U offen in \mathbb{R}^m . Außerdem ist $x_0 \in U$ wegen $(x_0, z_0) \in W$. Für $(x, y) \in U \times V$ berechnen wir

$$\begin{aligned} f(x, y) = z_0 &\Leftrightarrow F(x, y) = (x, z_0) \\ &\Leftrightarrow (x, y) = G(x, z_0) \quad (\text{da } (x, z_0) \in W) \\ &\Leftrightarrow y = g_0(x, z_0). \end{aligned}$$

² $\varphi(D) = \det \begin{pmatrix} E_m & 0 \\ C & D \end{pmatrix}$ ist multilinear and alternierend in den Spalten von D , mit $\varphi(E_k) = 1$

Also gilt der Satz mit $g(x) = g_0(x, z_0)$. Die Formel für die Ableitung folgt aus der Kettenregel:

$$f(x, g(x)) = z_0 \quad \Rightarrow \quad D_x f(x_0, y_0) + D_y f(x_0, y_0) Dg(x_0) = 0.$$

□

Beispiel 8.3 Die Nullstellen einer quadratischen Gleichung hängen von den Koeffizienten ab. Betrachte

$$f : \mathbb{R}^2 \times \mathbb{R} \rightarrow \mathbb{R}, \quad f(p, q, \lambda) = \lambda^2 + p\lambda + q = \left(\lambda + \frac{p}{2}\right)^2 - \left(\frac{p^2}{4} - q\right).$$

Setze $N = \{(p, q, \lambda) \in \mathbb{R}^2 \times \mathbb{R} : f(p, q, \lambda) = 0\}$. Für $(p_0, q_0, \lambda_0) \in N$ berechnen wir

$$\frac{\partial f}{\partial \lambda}(p_0, q_0, \lambda_0) = 2\left(\lambda_0 + \frac{p_0}{2}\right) \quad \text{wobei } 0 \leq \left(\lambda_0 + \frac{p_0}{2}\right)^2 = \frac{p_0^2}{4} - q_0.$$

Nach Satz 8.1 gibt es im Fall $\frac{p_0^2}{4} - q_0 > 0$ lokal eine eindeutige Auflösung $\lambda = \lambda(p, q)$. Das sehen wir natürlich auch direkt mit der p - q -Formel, die Auflösung ist

$$\lambda(p, q) = \begin{cases} -\frac{p}{2} + \sqrt{\frac{p^2}{4} - q} & \text{falls } \lambda_0 > -\frac{p_0}{2}, \\ -\frac{p}{2} - \sqrt{\frac{p^2}{4} - q} & \text{falls } \lambda_0 < -\frac{p_0}{2}. \end{cases}$$

Im Fall $\frac{p_0^2}{4} - q_0 = 0$ ist die Bedingung von Satz 8.1 nicht erfüllt. Und tatsächlich gibt es für $\frac{p^2}{4} < q$ keine Lösung, für $\frac{p^2}{4} > q$ nahebei die zwei Lösungen aus der p - q -Formel.

Beispiel 8.4 Betrachte jetzt $f : \mathbb{R}^n \times \mathbb{R} \rightarrow \mathbb{R}$, $f(b, \lambda) = \lambda^n + b_{n-1}\lambda^{n-1} + \dots + b_0$. Sei λ_0 eine einfache Nullstelle von $f(a, \lambda)$ für $a \in \mathbb{R}^n$ fest, das heißt es gilt

$$f(a, \lambda) = (\lambda - \lambda_0) q(\lambda) \quad \text{für ein Polynom } q(\lambda) \text{ mit } q(\lambda_0) \neq 0.$$

Es folgt $\frac{\partial f}{\partial \lambda}(a, \lambda_0) = q(\lambda_0) \neq 0$. Nach dem Satz über implizite Funktionen existiert eine Umgebung $U \times V$ von (a, λ_0) , so dass zu jedem $b \in U$ genau eine Nullstelle $\lambda(b) \in V$ von $f(b, \cdot)$ existiert. Diese hängt unendlich oft differenzierbar von b ab, und es gilt für $0 \leq i \leq n-1$

$$\frac{\partial \lambda}{\partial b_i}(a) = -\left(\frac{\partial f}{\partial \lambda}(a, \lambda_0)\right)^{-1} \frac{\partial f}{\partial b_i}(a, \lambda_0) = -\frac{\lambda_0^i}{n\lambda_0^{n-1} + (n-1)a_{n-1}\lambda_0^{n-2} + \dots + a_1}.$$

Wir kommen nun zurück zur Interpretation als Höhenlinie bzw. allgemeiner Niveaumenge. Wir hatten im zweidimensionalen Fall bereits heuristisch die Unterscheidung zwischen regulären und singulären Punkten gemacht. In regulären Punkten sieht eine Niveaumenge lokal wie ein Unterraum aus, insbesondere hat die Menge in dem Punkt einen Tangentialraum. Diese Konzepte sollen nun definiert werden. Den Begriff des Tangentialraums hatten wir im Fall von Graphen bereits durch Blow-up eingeführt, siehe Kapitel 3.

Definition 8.1 Sei $1 \leq m \leq n$. Eine Menge $M \subset \mathbb{R}^n$ heisst m -dimensionale Untermannigfaltigkeit des \mathbb{R}^n der Klasse C^r , wobei $r \in \mathbb{N} \cup \{\infty\}$, falls gilt: zu jedem $p \in M$ gibt es eine offene Umgebung $\Omega \subset \mathbb{R}^n$ und einen C^r -Diffeomorphismus $\phi : \Omega \rightarrow \phi(\Omega)$ mit

$$\phi(M \cap \Omega) = (\mathbb{R}^m \times \{0\}) \cap \phi(\Omega).$$

Wir nennen den Diffeomorphismus ϕ eine (lokale) Plättung von M . Im Einzelfall kann der Nachweis, dass eine gegebene Menge $M \subset \mathbb{R}^n$ eine Untermannigfaltigkeit ist, anhand der Definition mühevoll sein. Für Mengen, die als Niveaumengen einer Funktion gegeben sind, liefert jedoch der Satz über implizite Funktionen folgendes Kriterium.

Satz 8.2 (Untermannigfaltigkeitskriterien) Sei $M \subset \mathbb{R}^n$ und $m + k = n$. Dann sind folgende Aussagen äquivalent:

- (1) M ist eine m -dimensionale Untermannigfaltigkeit der Klasse C^r .
- (2) Niveaumengenkriterium: Zu jedem $p \in M$ gibt es eine offene Umgebung $\Omega \subset \mathbb{R}^n$ und eine Funktion $f \in C^r(\Omega, \mathbb{R}^k)$, so dass $M \cap \Omega = f^{-1}(0)$ und $\text{rang } Df(p) = k$.
- (3) Graphenkriterium: Zu $p \in M$ gibt es eine offene Umgebung $U \times V \subset \mathbb{R}^m \times \mathbb{R}^k$ und $g \in C^r(U, V)$, so dass nach geeigneter Permutation der Koordinaten gilt:

$$M \cap (U \times V) = \{(x, g(x)) : x \in U\}.$$

BEWEIS: Wir zeigen (1) \Rightarrow (2) \Rightarrow (3) \Rightarrow (1).

Nach (1) gibt es zu jedem $p \in M$ eine C^r -Plättung $\phi : \Omega \rightarrow \phi(\Omega)$ mit $p \in \Omega$. Definiere $f = \pi_2 \circ \phi$ wobei $\pi_2 : \mathbb{R}^m \times \mathbb{R}^k \rightarrow \mathbb{R}^k$, $\pi_2(x, y) = y$. Dann folgt für $q \in \Omega$ beliebig

$$f(q) = 0 \quad \Leftrightarrow \quad \phi(q) \in \mathbb{R}^m \times \{0\} \quad \Leftrightarrow \quad q \in \phi^{-1}(\mathbb{R}^m \times \{0\}) = M \cap \Omega.$$

π_2 ist linear, insbesondere C^∞ , also ist nach Kettenregel $f \in C^r(\Omega, \mathbb{R}^k)$ mit Ableitung $Df(p) = D\pi_2(f(p))D\phi(p) = \pi_2 D\phi(p)$. Aber $D\phi(p)$ ist invertierbar nach Lemma 7.1, es folgt $\text{rang } Df(p) = \text{rang } \pi_2 = k$.

Ist (2) erfüllt, so ist nach evtl. Permutation der Koordinaten $D_y f(p)$ invertierbar, wobei $(x, y) \in \mathbb{R}^n = \mathbb{R}^m \times \mathbb{R}^k$, und (3) folgt aus dem Satz über implizite Funktionen.

Sei (3) gegeben, also $M \cap (U \times V) = \{(x, g(x)) : x \in U\}$ mit $g \in C^r(U, V)$, eventuell nach Permutation der Koordinaten. Die Abbildung $\phi : U \times \mathbb{R}^k \rightarrow U \times \mathbb{R}^k$, $\phi(x, y) = (x, y - g(x))$, ist bijektiv mit $\phi^{-1}(x, z) = (x, z + g(x))$. Also ist die Einschränkung $\phi : U \times V \rightarrow \phi(U \times V)$ ein C^r -Diffeomorphismus, und es gilt

$$\phi(M \cap (U \times V)) = \phi(\{(x, g(x)) : x \in U\}) = U \times \{0\}.$$

Dies zeigt (1), womit der Satz insgesamt bewiesen ist. □

Beispiel 8.5 Die Sphäre $\mathbb{S}^m = \{x \in \mathbb{R}^{m+1} : |x| = 1\}$ ist eine m -dimensionale Untermannigfaltigkeit im \mathbb{R}^{m+1} der Klasse C^∞ . Denn es gilt

$$\mathbb{S}^m = f^{-1}(\{0\}) \quad \text{für } f : \mathbb{R}^{m+1} \rightarrow \mathbb{R}, \quad f(x) = |x|^2 - 1.$$

Da $Df(x) \neq 0$ für alle $x \in \mathbb{S}^m$, ist das Niveaumengenkriterium aus Satz 8.2 anwendbar.

Definition 8.2 $v \in \mathbb{R}^n$ heisst Tangentialvektor von $M \subset \mathbb{R}^n$ im Punkt $p \in M$, falls es eine Abbildung $\gamma : (-\varepsilon, \varepsilon) \rightarrow M$ gibt mit $\gamma(0) = p$, $\gamma'(0) = v$. Die Menge der Tangentialvektoren von M im Punkt p wird mit $T_p M$ bezeichnet.

Je nach Menge M kann $T_p M$ nur aus dem Nullvektor bestehen, betrachte etwa $M = \{(x, y) \in \mathbb{R}^2 : xy = 0, x, y \geq 0\}$ im Punkt $p = (0, 0)$. Unser Interesse gilt aber dem Fall, wenn M eine Untermannigfaltigkeit ist.

Folgerung 8.1 Sei $M \subset \mathbb{R}^n$ eine m -dim. C^1 -Untermannigfaltigkeit, und $n = m + k$. Ist $p \in M \cap \Omega = f^{-1}(\{0\})$ für eine Funktion $f \in C^1(\Omega, \mathbb{R}^k)$ mit $\text{rang } Df(p) = k$, so gilt

$$T_p M = \ker Df(p).$$

Insbesondere ist $T_p M$ ein m -dimensionaler Unterraum des \mathbb{R}^n .

BEWEIS: Für $\gamma : (-\varepsilon, \varepsilon) \rightarrow M$ ist $f(\gamma(t)) = 0$. Mit $\gamma(0) = p$, $\gamma'(0) = v$ folgt mit Kettenregel

$$0 = \frac{d}{dt} f(\gamma(t))|_{t=0} = Df(p)v, \quad \text{also} \quad T_p M \subset \ker Df(p).$$

Nach Satz 8.2 gibt es andererseits, nach eventueller Permutation der Koordinaten, offene Mengen $U \subset \mathbb{R}^m$, $V \subset \mathbb{R}^k$ mit $p \in U \times V$, sowie $g \in C^1(U, V)$ mit

$$M \cap (U \times V) = \{(x, g(x)) : x \in U\}.$$

Die Graphenabbildung $G \in C^1(U, \mathbb{R}^n)$, $G(x) = (x, g(x))$, bildet nach M ab. Mit $p = (x_0, g(x_0))$ für $x_0 \in U$ geeignet folgt für alle $\xi \in \mathbb{R}^m$

$$DG(x_0)\xi = \frac{d}{dt} G(x_0 + t\xi)|_{t=0} \in T_p M, \quad \text{also} \quad \text{Bild } DG(x_0) \subset T_p M.$$

$DG(x_0)$ ist injektiv, denn $DG(x_0)\xi = (\xi, Dg(x_0)\xi)$, also ist $\dim \text{Bild } DG(x_0) = m$. Andererseits liefern Dimensionsformel und Voraussetzung

$$\dim \ker Df(p) = \dim \mathbb{R}^n - \dim \text{Bild } Df(p) = n - k = m.$$

Zusammen ergibt sich $\text{Bild } DG(x_0) = T_p M = \ker Df(p)$. □

Die Folgerung zeigt: für eine m -dimensionale Untermannigfaltigkeit ist der Tangentialraum ein Vektorraum der Dimension m . Damit ist die Dimension einer Untermannigfaltigkeit wohldefiniert, es kann nicht Plättungen zu verschiedenen m geben. Wir kommen nun zur Multiplikatorenregel von Lagrange.

Satz 8.3 (Extrema mit Nebenbedingungen) Sei $f \in C^1(\Omega, \mathbb{R}^k)$, $\Omega \subset \mathbb{R}^n$ offen. Die Funktion $\varphi \in C^1(\Omega)$ habe in p ein Minimum unter der Nebenbedingung $f(q) = z_0$, das heißt

$$f(p) = z_0 \quad \text{und} \quad \varphi(p) \leq \varphi(q) \quad \text{für alle } q \in \Omega \text{ mit } f(q) = z_0.$$

Ist dann $\text{rang } Df(p) = k$, so gibt es $\lambda_1, \dots, \lambda_k \in \mathbb{R}$ mit $\text{grad } \varphi(p) = \sum_{i=1}^k \lambda_i \text{grad } f_i(p)$.

BEWEIS: Nach Verkleinerung von Ω ist $\text{rang } Df = k$ auf ganz Ω , und $M = f^{-1}(\{z_0\})$ ist m -dimensionale Untermannigfaltigkeit mit $m = n - k$, vgl. Satz 8.2. Ist $\gamma : (-\varepsilon, \varepsilon) \rightarrow M$ mit $\gamma(0) = p$ und $\gamma'(0) = v$, so hat $\varphi \circ \gamma$ in $t = 0$ ein lokales Minimum und folglich

$$0 = \frac{d}{dt} \varphi(\gamma(t))|_{t=0} = \langle \text{grad } \varphi(p), v \rangle, \quad \text{also} \quad \text{grad } \varphi(p) \in (T_p M)^\perp.$$

Analog folgt $\text{grad } f_i(p) \in (T_p M)^\perp$, denn die f_i sind auf M konstant. Wir behaupten, dass die $\text{grad } f_i(p)$, $i = 1, \dots, k$, den Raum $(T_p M)^\perp$ erzeugen. Nach Folgerung 8.1 hat $(T_p M)^\perp$ die Dimension k . Nach Voraussetzung ist aber $\text{rang } Df(p) = k$, das heißt der von den Zeilen aufgespannte Raum hat Dimension k . Dies zeigt unsere Behauptung, und der Satz ist bewiesen. □

Beispiel 8.6 Für $B \in \mathbb{R}^{n \times n}$ symmetrisch betrachten wir das Minimierungsproblem

$$\langle Bx, x \rangle \longrightarrow \min. \quad \text{unter Nebenbedingung } |x|^2 = 1.$$

Wir setzen $\varphi(x) = \langle Bx, x \rangle$ und $f(x) = |x|^2$. Da $f^{-1}\{1\} = \mathbb{S}^{n-1}$ kompakt ist und $\varphi(x)$ stetig, wird das Infimum in einem $v \in \mathbb{S}^{n-1}$ angenommen. Mit Satz 8.3 gibt es ein $\lambda \in \mathbb{R}$ mit

$$\text{grad } \varphi(v) = \lambda \text{ grad } f(v), \quad \text{also } Bv = \lambda v.$$

Somit hat jede symmetrische Matrix B mindestens einen Eigenvektor. Dies wurde in Satz 4.6 schon mit einem direkten Argument gezeigt.

Wir haben die Sätze über inverse und implizite Funktionen im Endlichdimensionalen formuliert, um das Wesentliche ohne zuviel Abstraktion darzustellen. An der Verallgemeinerung auf Abbildungen zwischen Banachräumen besteht aber großes Interesse: in den Anwendungen ist die Gleichung $f(x) = y$ zum Beispiel eine nichtlineare Differentialgleichung, die durch eine gesuchte Funktion x in einem geeigneten Funktionenraum X gelöst werden soll. Eine Inspektion des Beweises des Umkehrsatzes ergibt, dass die Konstruktion der inversen Abbildung einschließlich ihrer Differenzierbarkeit ohne Änderungen auch dann richtig ist, wenn f eine offene Teilmenge des Banachraums X in den Banachraum Y abbildet. Allerdings muss der Begriff der linearen Abbildung wie folgt ergänzt werden:

$$L(X, Y) = \{A : X \rightarrow Y \mid A \text{ linear}, \|A\| < \infty\} \quad \text{mit } \|A\| = \sup_{\|x\|=1} \|Ax\|.$$

$\|A\|$ heißt Operatornorm von A , eine lineare Abbildung mit $\|A\| < \infty$ wird auch als beschränkter Operator bezeichnet. Es ist leicht zu sehen, dass die Bedingung $\|A\| < \infty$ äquivalent zur Stetigkeit von A ist. Im Fall $X = \mathbb{R}^n$ ist automatisch $\|A\| < \infty$, siehe Beispiel 7.10 in Analysis 1, für $\dim X = \infty$ muss das extra verlangt werden. Zum Beispiel wird in der Definition der Differenzierbarkeit $Df(x_0) \in L(X, Y)$ gefordert. Von den Koordinaten des \mathbb{R}^n wurde beim Beweis des Umkehrsatzes nur explizit Gebrauch gemacht, um die höhere Differenzierbarkeit der Inversen zu etablieren. Hier gibt es aber als Alternative die Neumannsche Reihe, das heißt die geometrische Reihe

$$(\text{Id} - A)^{-1} = \sum_{k=0}^{\infty} A^k \quad \text{für } A \in L(X, X) \text{ mit } \|A\| < 1.$$

Zusammenfassend gelten Versionen der Sätze über inverse und implizite Funktionen auch für Abbildungen zwischen Banachräumen.

9 Kurvenintegrale und Gradientenfelder

Beim einer Radtour vom Mathematischen Institut auf den Schauinsland wird Arbeit gegen die Gravitationskraft verrichtet. Es ist dabei egal, welcher Weg gewählt wird: die Gravitationskraft zeigt konstant nach unten, daher entspricht die Arbeit einfach dem Zugewinn an Höhe bzw. Lageenergie. Felder mit einem solchen Erhaltungsgesetz heißen konservativ. Wir wollen notwendige und hinreichende Bedingungen angeben, die konservative Felder charakterisieren.

Die Regel zur Berechnung der Arbeit lautet *Kraft längs Weg*. Wir betrachten die Klasse $PC^1([a, b], \mathbb{R}^n)$ der stückweise C^1 -Wege $\gamma : [a, b] \rightarrow \mathbb{R}^n$, das heißt γ ist stetig und es gibt eine Unterteilung $a = t_0 < \dots < t_N = b$ mit $\gamma|_{[t_{k-1}, t_k]} \in C^1([t_{k-1}, t_k], \mathbb{R}^n)$.

Definition 9.1 (Kurvenintegral) Sei $F \in C^0(\Omega, \mathbb{R}^n)$, $\Omega \subset \mathbb{R}^n$ offen. Das Kurvenintegral von F längs $\gamma \in PC^1([a, b], \Omega)$ ist

$$\int_{\gamma} F \cdot dx := \int_a^b \langle F(\gamma(t)), \gamma'(t) \rangle dt.$$

In der Physik steht $F(x)$ für ein Kraftfeld und dx wird als vektorielles Wegelement bezeichnet. Die Notation ist jedoch rein symbolisch, zur Berechnung des Kurvenintegrals ist nur das Riemannintegral auf der rechten Seite relevant. Dabei ist die Merkregel nützlich, dass $x = \gamma(t)$ und $dx = \gamma'(t) dt$ ersetzt wird.

Lemma 9.1 Das Kurvenintegral hat folgende Eigenschaften:

(a) *Linearität:* sind $F_{1,2} \in C^0(\Omega, \mathbb{R}^2)$ und $\lambda_{1,2} \in \mathbb{R}$, so gilt für $\gamma \in PC^1([a, b], \Omega)$

$$\int_{\gamma} (\lambda_1 F_1 + \lambda_2 F_2) \cdot dx = \lambda_1 \int_{\gamma} F_1 \cdot dx + \lambda_2 \int_{\gamma} F_2 \cdot dx.$$

(b) *Additivität bei Zerlegungen:* ist $\gamma \in PC^1([a, b], \mathbb{R}^n)$ und $a = t_0 < \dots < t_N = b$ eine beliebige Zerlegung von $[a, b]$, so folgt mit $\gamma_i = \gamma|_{[t_{i-1}, t_i]}$

$$\int_{\gamma} F \cdot dx = \sum_{i=1}^N \int_{\gamma_i} F \cdot dx.$$

(c) *Invarianz bei Umparametrisierungen:* sei $\gamma \in PC^1(I_1, \mathbb{R}^2)$ und $\varphi \in C^1(I_2, I_1)$ sei diffeomorph. Dann gilt, je nach Vorzeichen von φ' ,

$$\int_{\gamma \circ \varphi} F \cdot dx = \pm \int_{\gamma} F \cdot dx.$$

BEWEIS: (a) und (b) folgen aus der Definition und den Eigenschaften des Riemannintegrals. Für (c) sei $I_1 = [a_1, b_1]$ und $I_2 = [a_2, b_2]$. Mit der Substitution $\varphi(t) = s$ ergibt sich

$$\begin{aligned} \int_{\gamma \circ \varphi} F \cdot dx &= \int_{a_2}^{b_2} \langle (F \circ \gamma \circ \varphi)(t), (\gamma \circ \varphi)'(t) \rangle dt \\ &= \int_{a_2}^{b_2} \langle F \circ \gamma(\varphi(t)), \gamma'(\varphi(t)) \rangle \varphi'(t) dt \\ &= \int_{\varphi(a_2)}^{\varphi(b_2)} \langle (F \circ \gamma)(s), \gamma'(s) \rangle ds. \end{aligned}$$

Ist $\varphi' > 0$ so gilt $\varphi(a_2) = a_1$ und $\varphi(b_2) = b_1$, wir bekommen das Pluszeichen. Ist $\varphi' < 0$ so sind die Grenzen vertauscht und es gilt das Minuszeichen. \square

Definition 9.2 (konservatives Feld) Ein Vektorfeld $F \in C^0(\Omega, \mathbb{R}^n)$, $\Omega \subset \mathbb{R}^n$ offen, heißt konservativ oder Gradientenfeld, wenn es eine Funktion $\varphi \in C^1(\Omega)$ gibt mit

$$F = \text{grad } \varphi \quad \Leftrightarrow \quad F_i = \partial_i \varphi \text{ für } i = 1, \dots, n.$$

Die Funktion φ heißt Stammfunktion oder Potential von F .

Hinweis. In der Physik ist die Wahl $F = -\text{grad } \varphi$ üblich.

Beispiel 9.1 (Gravitationsfelder) In folgenden Beispielen gilt $F = -\text{grad } \varphi$:

$$\begin{aligned} F : \mathbb{R}^3 &\rightarrow \mathbb{R}^3 & F(x) &= -C e_3 & \varphi(x) &= C x_3 \\ F : \mathbb{R}^3 \setminus \{0\} &\rightarrow \mathbb{R}^3 & F(x) &= C \frac{x}{|x|^3} & \varphi(x) &= \frac{C}{|x|}. \end{aligned}$$

Das erste Feld beschreibt approximativ die Gravitation nahe der Erdoberfläche, das zweite ist das Gravitationsfeld eines beliebigen, rotationssymmetrischen Körpers, nach dem Newtonschen Gravitationsgesetz.

Lemma 9.2 (Eindeutigkeit der Stammfunktion) Ist $\Omega \subset \mathbb{R}^n$ offen und zusammenhängend, so ist eine Stammfunktion von $F \in C^0(\Omega, \mathbb{R}^n)$ eindeutig bestimmt (wenn existent), bis auf eine additive Konstante.

BEWEIS: Sind $\varphi_1, \varphi_2 \in C^1(\Omega)$ Stammfunktionen von F , so folgt

$$\text{grad}(\varphi_2 - \varphi_1) = \text{grad } \varphi_2 - \text{grad } \varphi_1 = F - F = 0.$$

Also ist $\varphi_2 - \varphi_1$ konstant nach Satz 4.1, das heißt $\varphi_2 = \varphi_1 + c$. \square

Wir zeigen jetzt, dass Existenz einer Stammfunktion und Wegunabhängigkeit äquivalent sind.

Satz 9.1 (Wegunabhängigkeit des Kurvenintegrals) Sei $\Omega \subset \mathbb{R}^n$ offen und zusammenhängend. Für ein Vektorfeld $F \in C^0(\Omega, \mathbb{R}^n)$ sind folgende Aussagen äquivalent:

- (a) F ist ein Gradientenfeld.
- (b) Für jede geschlossene PC^1 -Kurve in Ω ist $\int_{\gamma} F \cdot dx = 0$.
- (c) Für zwei PC^1 -Kurven in Ω mit gleichen Anfangs- und Endpunkten ist

$$\int_{\gamma_0} F \cdot dx = \int_{\gamma_1} F \cdot dx.$$

BEWEIS: Ist $F = \text{grad } \varphi$ in Ω , so gilt nach Kettenregel für alle $\gamma \in PC^1([a, b], \Omega)$

$$(9.1) \quad \int_{\gamma} F \cdot dx = \int_a^b \langle \text{grad } \varphi(\gamma(t)), \gamma'(t) \rangle dt = \int_a^b (\varphi \circ \gamma)'(t) dt = [\varphi(x)]_{x=\gamma(a)}^{x=\gamma(b)}.$$

Das Integral ist also gleich für zwei Wege mit gleichem Anfangs- und Endpunkt; für geschlossene Wege ist es Null.

Für (b) \Rightarrow (c) seien $\gamma_i \in PC^1([a_i, b_i], \Omega)$, $i = 1, 2$, mit gleichem Anfangs- und Endpunkt. Dann ist

$$\gamma(t) = \begin{cases} \gamma_1(t) & a_1 \leq t \leq b_1 \\ \gamma_2(b_1 + b_2 - t) & b_1 \leq t \leq b_1 + b_2 - a_2 \end{cases}$$

geschlossen und stückweise C^1 , und aus (b) ergibt sich mit Lemma 9.1

$$0 = \int_{\gamma} F \cdot dx = \int_{\gamma_1} F \cdot dx - \int_{\gamma_2} F \cdot dx.$$

Für (c) \Rightarrow (a) sei $x_0 \in \Omega$ fest. Zu $x \in \Omega$ wählen wir $\gamma_x \in PC^1([0, 1], \Omega)$ mit $\gamma_x(0) = x_0$ und $\gamma_x(1) = x$, siehe Aufgabe 2, Serie 2, für die Existenz von γ . Wäre φ Stammfunktion von F mit $\varphi(x_0) = 0$, so folgt aus (9.1)

$$(9.2) \quad \varphi(x) = \int_{\gamma_x} F \cdot dx.$$

Umgekehrt definieren wir $\varphi : \Omega \rightarrow \mathbb{R}$ durch (9.2) und zeigen, dass dies eine Stammfunktion liefert. Zu $x \in \Omega$ sei $\varepsilon > 0$ mit $B_\varepsilon(x) \subset \Omega$. Wir erhalten eine PC^1 -Kurve von x_0 nach $x + he_j$, $h \in (-\varepsilon, \varepsilon)$, indem wir γ_x zusammensetzen mit

$$c : [0, 1] \rightarrow B_\varepsilon(x), \quad c(t) = x + the_j.$$

Nach Voraussetzung (c) und Lemma 9.1 gilt für $h \neq 0$

$$\frac{\varphi(x + he_j) - \varphi(x)}{h} = \frac{1}{h} \int_c F \cdot dx = \int_0^1 \langle F(x + the_j), e_j \rangle dt \rightarrow F_j(x) \quad \text{mit } h \rightarrow 0.$$

Also gilt $\partial_j \varphi = F_j$ für $j = 1, \dots, n$. □

Die folgende Bedingung ist offensichtlich notwendig für die Existenz einer Stammfunktion.

Satz 9.2 (Rotationsfreiheit von Gradientenfeldern) Sei $\Omega \subset \mathbb{R}^n$ offen. Ist $F \in C^1(\Omega, \mathbb{R}^n)$ ein Gradientenfeld, so gilt für alle $i, j = 1, \dots, n$

$$\partial_i F_j = \partial_j F_i \quad \text{in } \Omega.$$

BEWEIS: Ist $F = \text{grad } \varphi$, so folgt $\varphi \in C^2(\Omega)$ und mit Schwarz, Satz 2.2, gilt

$$\partial_i F_j = \partial_i \partial_j \varphi = \partial_j \partial_i \varphi = \partial_j F_i.$$

□

Für $n = 3$ lässt sich die Bedingung schreiben als $\text{rot } F = 0$, wobei

$$\text{rot } F = (\partial_2 F_3 - \partial_3 F_2, \partial_3 F_1 - \partial_1 F_3, \partial_1 F_2 - \partial_2 F_1).$$

Beispiel 9.2 $F : \mathbb{R}^2 \rightarrow \mathbb{R}^2$, $F(x, y) = (-y, x)$, hat auf keiner offenen Teilmenge eine Stammfunktion, denn es gilt $\partial_1 F_2 = 1$, dagegen $\partial_2 F_1 = -1$.

Das folgende Beispiel ist interessant.

Beispiel 9.3 (Winkelvektorfeld) Wir betrachten

$$W : \mathbb{R}^2 \setminus \{0\} \rightarrow \mathbb{R}^2, W(x, y) = \left(-\frac{y}{x^2 + y^2}, \frac{x}{x^2 + y^2} \right)$$

Die notwendige Bedingung aus Satz 9.2 ist erfüllt, es gilt

$$\partial_1 W_2 = \frac{y^2 - x^2}{(x^2 + y^2)^2} = \partial_2 W_1.$$

Für $y > 0$ hat $W(x, y)$ als Stammfunktion den Winkel mit der x -Achse

$$\varphi : \{(x, y) \in \mathbb{R}^2 : y > 0\} \rightarrow (0, \pi), \varphi(x, y) = \arccos \frac{x}{\sqrt{x^2 + y^2}}.$$

Um das nachzurechnen, verwende $\varphi(r \cos \theta, r \sin \theta) = \theta$ und die Kettenregel:

$$\begin{aligned} \langle \text{grad } \varphi(r \cos \theta, r \sin \theta), (\cos \theta, \sin \theta) \rangle &= \frac{\partial}{\partial r} \varphi(r \cos \theta, r \sin \theta) = 0, \\ \langle \text{grad } \varphi(r \cos \theta, r \sin \theta), (-r \sin \theta, r \cos \theta) \rangle &= \frac{\partial}{\partial \theta} \varphi(r \cos \theta, r \sin \theta) = 1. \end{aligned}$$

Wegen $W(r \cos \theta, r \sin \theta) = \frac{1}{r}(-\sin \theta, \cos \theta)$ gilt ebenfalls

$$\begin{aligned} \langle W(r \cos \theta, r \sin \theta), (\cos \theta, \sin \theta) \rangle &= 0, \\ \langle W(r \cos \theta, r \sin \theta), (-r \sin \theta, r \cos \theta) \rangle &= 1. \end{aligned}$$

Betrachte nun eine Kurve $\gamma : I = [a, b] \rightarrow \mathbb{R}^2 \setminus \{0\}$ in Polardarstellung, also $\gamma(t) = r(t)(\cos \theta(t), \sin \theta(t))$ mit $r, \theta \in C^1(I)$. Wir berechnen

$$(9.3) \quad \int_{\gamma} W \cdot dx = \int_a^b \left\langle \frac{1}{r} \begin{pmatrix} -\sin \theta \\ \cos \theta \end{pmatrix}, r' \begin{pmatrix} \cos \theta \\ \sin \theta \end{pmatrix} + r \theta' \begin{pmatrix} -\sin \theta \\ \cos \theta \end{pmatrix} \right\rangle dt = \theta(b) - \theta(a).$$

Speziell sei $\gamma(t) = (\cos nt, \sin nt)$, $t \in [0, 2\pi]$, mit $n \in \mathbb{Z}$. Dann ist γ geschlossen und es gilt

$$(9.4) \quad \int_{\gamma} W \cdot dx = 2\pi n \quad (\neq 0 \text{ für } n \neq 0).$$

Mit Satz 9.1 folgt, dass W keine Stammfunktion auf ganz $\mathbb{R}^2 \setminus \{0\}$ hat.

Das Beispiel zeigt, dass die Rotationsfreiheit eines Vektorfelds, siehe Satz 9.2, nicht hinreichend ist für die Wegunabhängigkeit des Kurvenintegrals. Es stellt sich die Frage, ob das Kurvenintegral zumindest gleich bleibt, wenn eine Kurve stetig deformiert wird.

Definition 9.3 (Homotopie) Eine Homotopie in Ω zwischen $\gamma_0, \gamma_1 \in C^0([a, b], \Omega)$ ist eine Abbildung $\gamma \in C^0([a, b] \times [0, 1], \Omega)$ mit $\gamma(\cdot, 0) = \gamma_0$ und $\gamma(\cdot, 1) = \gamma_1$. Speziell:

- Homotopie mit festen Endpunkten: $\gamma(a, t), \gamma(b, t)$ konstant für $t \in [0, 1]$ (falls γ_0, γ_1 mit gleichem Anfangs- und Endpunkt)
- geschlossene Homotopie: $\gamma(a, t) = \gamma(b, t)$ für alle $t \in [0, 1]$ (falls γ_0, γ_1 geschlossen).

Im folgenden Lemma berechnen wir die Ableitung des Kurvenintegrals längs Homotopien, sofern diese ausreichend differenzierbar sind.

Lemma 9.3 (Homotopieformel) Sei $\Omega \subset \mathbb{R}^n$ offen und $F \in C^1(\Omega, \mathbb{R}^n)$. Sei $\gamma \in C^2([a, b] \times [0, 1], \Omega)$ eine Homotopie, entweder mit festen Endpunkten oder geschlossen. Dann gilt mit $\gamma_t = \gamma(\cdot, t)$ für $t \in [0, 1]$

$$(9.5) \quad \int_{\gamma_1} F \cdot dx - \int_{\gamma_0} F \cdot dx = - \int_0^1 \int_a^b \sum_{i,j=1}^n (\partial_i F_j - \partial_j F_i) \circ \gamma \frac{\partial \gamma^i}{\partial s} \frac{\partial \gamma^j}{\partial t} ds dt.$$

Ist F rotationsfrei, so sind die Kurvenintegrale über γ_0, γ_1 gleich.

BEWEIS: Durch Differentiation unter dem Integral, Satz 6.2, und partielle Integration bezüglich $s \in [a, b]$ ergibt sich, zunächst für eine beliebige C^2 -Homotopie,

$$\begin{aligned} \frac{d}{dt} \int_{\gamma_t} F \cdot dx &= \frac{d}{dt} \int_a^b \langle F(\gamma(s, t)), \frac{\partial \gamma}{\partial s}(s, t) \rangle ds \\ &= \int_a^b \langle DF \circ \gamma \frac{\partial \gamma}{\partial t}, \frac{\partial \gamma}{\partial s} \rangle ds + \int_a^b \langle F \circ \gamma, \frac{\partial^2 \gamma}{\partial t \partial s} \rangle ds \\ &= \int_a^b \langle DF \circ \gamma \frac{\partial \gamma}{\partial t}, \frac{\partial \gamma}{\partial s} \rangle ds + \langle F \circ \gamma, \frac{\partial \gamma}{\partial t} \rangle \Big|_{s=a}^{s=b} - \int_a^b \langle DF \circ \gamma \frac{\partial \gamma}{\partial s}, \frac{\partial \gamma}{\partial t} \rangle ds. \end{aligned}$$

Der Randterm verschwindet in beiden Fällen:

$$\begin{aligned} \text{feste Endpunkte} &\Rightarrow \frac{\partial \gamma}{\partial t}(a, t) = 0, \quad \frac{\partial \gamma}{\partial t}(b, t) = 0, \\ \text{geschlossen} &\Rightarrow \gamma(a, t) = \gamma(b, t), \quad \frac{\partial \gamma}{\partial t}(a, t) = \frac{\partial \gamma}{\partial t}(b, t). \end{aligned}$$

Damit ist die Formel bewiesen. □

Wir können an dieser Stelle als Anwendung den Fundamentalsatz der Algebra, Satz 5.11 aus Analysis 1, beweisen. Es stellt sich heraus, dass es für die Existenz einer Nullstelle einen geometrischen Grund gibt.

Satz 9.3 (Fundamentalsatz der Algebra) Jedes komplexe Polynom vom Grad $n \geq 1$ hat mindestens eine Nullstelle $z_0 \in \mathbb{C}$.

BEWEIS: Das Winkelvektorfeld $W : \mathbb{R}^2 \setminus \{0\} \rightarrow \mathbb{R}^2$ aus Beispiel 9.3 erfüllt $\partial_1 W_2 = \partial_2 W_1$. Sei $p(z) = z^n + a_{n-1} z^{n-1} + \dots + a_0$ mit $a_i \in \mathbb{C}$ und $n \geq 1$. Schreibe $p(z) = p_n(z) + q(z)$ mit $p_n(z) = z^n$. Betrachte nun die Homotopie

$$\gamma : [0, 2\pi] \times [0, 1] \rightarrow \mathbb{R}^2, \quad \gamma(\theta, t) = p_n(Re^{i\theta}) + t q(Re^{i\theta}).$$

Wir haben $\gamma_0(\theta) = (Re^{i\theta})^n$ und $\gamma_1(\theta) = p(Re^{i\theta})$. Die Homotopie geht a priori nach \mathbb{R}^2 , aber $q(z)$ hat Grad höchstens $n-1$, daher ist $|q(Re^{i\theta})| \leq \frac{1}{2}R^n$ für $R > 0$ hinreichend groß. Es folgt

$$|\gamma(\theta, t)| \geq |p_n(Re^{i\theta})| - |q(Re^{i\theta})| \geq R^n - \frac{1}{2}R^n > 0.$$

Mit Lemma 9.3 und Beispiel 9.3 gilt

$$\int_{\gamma_1} W \cdot dx = \int_{\gamma_0} W \cdot dx = 2\pi n.$$

Nun betrachten wir die zweite, ebenfalls glatte Homotopie

$$\tilde{\gamma} : [0, 2\pi] \times [0, R] \rightarrow \mathbb{R}^2, \tilde{\gamma}(\theta, \varrho) = p(\varrho e^{i\theta}).$$

Hätte $p(z)$ keine Nullstelle in \mathbb{C} , so wäre dies eine Homotopie in $\mathbb{R}^2 \setminus \{0\}$ zwischen $\tilde{\gamma}_R(\theta) = p(Re^{i\theta}) = \gamma_1(\theta)$ und der konstanten Kurve $\tilde{\gamma}_0(\theta) = p(0)$. Wieder mit Lemma 9.3 folgt

$$\int_{\gamma_1} W \cdot dx = \int_{\tilde{\gamma}_0} W \cdot dx = 0,$$

ein Widerspruch. □

Wir kommen nun auf das Problem der Wegunabhängigkeit, bzw. äquivalent der Existenz einer Stammfunktion, zurück. Beispiel 9.3 weist darauf hin, dass es neben der Rotationsfreiheit des Vektorfeldes auch auf das Gebiet Ω ankommt.

Definition 9.4 Eine Menge $\Omega \subset \mathbb{R}^n$ heißt einfach zusammenhängend, wenn jede geschlossene Kurve $\gamma \in C^0([a, b], \Omega)$ in Ω geschlossen homotop zu einer konstanten Kurve ist.

Beispiel 9.4 Eine Menge $\Omega \subset \mathbb{R}^n$ heißt sternförmig, wenn es ein $x_0 \in \Omega$ gibt mit

$$(1-t)x + tx_0 \in \Omega \quad \text{für alle } x \in \Omega, t \in [0, 1].$$

Eine sternförmige Menge ist einfach zusammenhängend, denn jede geschlossene Kurve $\gamma_0 \in C^0([a, b], \Omega)$ ist homotop zur konstanten Kurve in x_0 , durch die Homotopie

$$\gamma : [a, b] \times [0, 1] \rightarrow \Omega, \gamma(s, t) = (1-t)\gamma_0(s) + tx_0.$$

Satz 9.4 (Stammfunktion) Sei $\Omega \subset \mathbb{R}^n$ offen und $F \in C^1(\Omega, \mathbb{R}^n)$. Es gelte

- (a) Für alle $i, j = 1, \dots, n$ ist $\partial_i F_j = \partial_j F_i$ auf Ω .
- (b) Ω ist einfach zusammenhängend.

Dann gibt es auf Ω eine Stammfunktion zu F .

Der Beweis ist im Prinzip klar: nach Satz 9.1 reicht es zu zeigen, dass das Kurvenintegral längs jeder geschlossenen PC^1 -Kurve γ gleich Null ist. Nach Voraussetzung ist aber γ homotop zu einer konstanten Kurve. Wegen F rotationsfrei ist das Kurvenintegral längs der Homotopie konstant nach Lemma 9.3, und damit gleich Null wie verlangt.

Es gibt eine technische Komplikation. Die Homotopie von γ zur konstanten Kurve muss nach Definition 9.3 nur stetig sein. Das Kurvenintegral längs $\gamma_t = \gamma(\cdot, t)$ ist damit nicht definiert, außer für $t = 0, 1$. Erst recht kann die Homotopieformel nicht angewandt werden. Um das zu umgehen, ersetzen wir die γ_t durch stückweise lineare Kurven und verwenden affin-lineare Homotopien. Für diese ist die Homotopieformel gültig.

Lemma 9.4 (affine Homotopie) Sei $\Omega \subset \mathbb{R}^n$ offen und $F \in C^1(\Omega, \mathbb{R}^n)$ mit $\partial_i F_j = \partial_j F_i$ auf Ω für $1 \leq i, j \leq n$. Betrachte eine affine Homotopie

$$\gamma : [a, b] \times [0, 1] \rightarrow \Omega, \gamma(s, t) = (1-t)\gamma_0(s) + t\gamma_1(s).$$

Dabei seien $\gamma_{0,1} \in PC^1([a, b], \Omega)$ mit Randbedingungen

$$\gamma_0(a) = \gamma_1(a), \gamma_0(b) = \gamma_1(b) \quad \text{oder} \quad \gamma_0(a) = \gamma_0(b), \gamma_1(a) = \gamma_1(b).$$

Dann gilt

$$\int_{\gamma_1} F \cdot dx = \int_{\gamma_0} F \cdot dx.$$

BEWEIS: Seien zunächst $\gamma_{0,1}$ von der Klasse C^1 . Wir haben

$$\frac{\partial^2 \gamma}{\partial t \partial s}(s, t) = \gamma_1'(s) - \gamma_0'(s) = \frac{\partial^2 \gamma}{\partial s \partial t}(s, t).$$

Dies reicht für die Rechnung aus Lemma 9.3 aus. Da F rotationsfrei ist, folgt

$$\frac{\partial}{\partial t} \int_{\gamma_t} F \cdot dx = \langle F \circ \gamma, \frac{\partial \gamma}{\partial t} \rangle \Big|_{s=a}^{s=b}.$$

Seien nun $\gamma_{0,1}$ nur in PC^1 . Dann gibt es eine Unterteilung $a = s_0 < \dots, s_N = b$, so dass $\gamma_{0,1} \in C^1$ auf jedem Teilintervall $[s_{k-1}, s_k]$. Wir wenden die Rechnung auf $[s_{k-1}, s_k]$ an und addieren. Dabei heben sich alle Randwerte weg, außer für $s = a, b$. Diese sind aber Null wegen der Randbedingungen. \square

Satz 9.5 (Homotopieinvarianz des Kurvenintegrals) Sei $\Omega \subset \mathbb{R}^n$ offen und $F \in C^1(\Omega, \mathbb{R}^n)$ mit $\partial_i F_j = \partial_j F_i$ auf Ω für $1 \leq i, j \leq n$. Sind dann $\gamma_0, \gamma_1 \in PC^1([a, b], \Omega)$ homotop in Ω mit festen Endpunkten (oder geschlossen homotop), so gilt

$$\int_{\gamma_0} F \cdot dx = \int_{\gamma_1} F \cdot dx.$$

BEWEIS: Sei $\gamma \in C^0([a, b] \times [0, 1], \Omega)$ die Homotopie, also $\gamma(\cdot, 0) = \gamma_0$ und $\gamma(\cdot, 1) = \gamma_1$. Aus Kompaktheitsgründen gibt es ein $\varepsilon > 0$ mit

$$\text{dist}(\gamma([a, b] \times [0, 1]), \mathbb{R}^n \setminus \Omega) > 2\varepsilon.$$

Da γ auf $[a, b] \times [0, 1]$ gleichmäßig stetig ist, gibt es weiter ein $\delta > 0$ mit

$$|\gamma(s, t) - \gamma(s', t')| < \varepsilon \quad \text{für } |s - s'|, |t - t'| < \delta.$$

Wir ersetzen jetzt $\gamma(\cdot, t)$ durch stückweise lineare Kurven. Seien $a = s_0 < \dots < s_N = b$ äquidistant gewählt. Wir definieren $\tilde{\gamma} : [a, b] \times [0, 1] \rightarrow \mathbb{R}^n$ durch

$$\tilde{\gamma}(s, t) = \frac{s_k - s}{s_k - s_{k-1}} \gamma(s_{k-1}, t) + \frac{s - s_{k-1}}{s_k - s_{k-1}} \gamma(s_k, t) \quad \text{für } s \in [s_{k-1}, s_k].$$

Es gilt $\tilde{\gamma}(a, t) = \gamma(a, t)$ und $\tilde{\gamma}(b, t) = \gamma(b, t)$ für alle $t \in [0, 1]$. Wir behaupten, dass die affine Homotopie zwischen γ und $\tilde{\gamma}$ in Ω liegt. Für $s \in [s_{k-1}, s_k]$ haben wir

$$|\tilde{\gamma}(s, t) - \gamma(s, t)| \leq \frac{s_k - s}{s_k - s_{k-1}} |\gamma(s_{k-1}, t) - \gamma(s, t)| + \frac{s - s_{k-1}}{s_k - s_{k-1}} |\gamma(s_k, t) - \gamma(s, t)| < \varepsilon.$$

Für $\lambda \in [0, 1]$ folgt daraus die Abschätzung

$$|(1 - \lambda)\gamma(s, t) + \lambda\tilde{\gamma}(s, t) - \gamma(s, t)| = \lambda|\tilde{\gamma}(s, t) - \gamma(s, t)| < \varepsilon,$$

und die Behauptung folgt. Als zweites zeigen wir, dass für $|t - t'| < \delta$ die affine Homotopie zwischen $\tilde{\gamma}(\cdot, t)$ und $\tilde{\gamma}(\cdot, t')$ ebenfalls in Ω liegt. Und zwar gilt für $s \in [s_{k-1}, s_k]$

$$|\tilde{\gamma}(s, t) - \tilde{\gamma}(s, t')| \leq \frac{s_k - s}{s_k - s_{k-1}} |\gamma(s_{k-1}, t) - \gamma(s_{k-1}, t')| + \frac{s - s_{k-1}}{s_k - s_{k-1}} |\gamma(s_k, t) - \gamma(s_k, t')| < \varepsilon.$$

Daraus ergibt sich für $\mu \in [0, 1]$

$$|(1 - \mu)\tilde{\gamma}(s, t) + \mu\tilde{\gamma}(s, t') - \gamma(s, t)| \leq |\tilde{\gamma}(s, t) - \gamma(s, t)| + \mu |\tilde{\gamma}(s, t) - \tilde{\gamma}(s, t')| < 2\varepsilon.$$

Sei nun $N \in \mathbb{N}$ mit $1/N < \delta$ gewählt. Dann folgt mit Lemma 9.4

$$\int_{\gamma_0} F \cdot dx = \int_{\tilde{\gamma}_0} F \cdot dx \quad \text{und} \quad \int_{\gamma_1} F \cdot dx = \int_{\tilde{\gamma}_1} F \cdot dx, \quad \text{sowie}$$

$$\int_{\tilde{\gamma}_{t_j}} F \cdot dx = \int_{\tilde{\gamma}_{t_{j-1}}} F \cdot dx \quad \text{für } t_j = \frac{j}{N} \text{ mit } j = 1, \dots, N.$$

Damit ist der Satz bewiesen. □

Folgende Tabelle fasst unsere Ergebnisse zum Kurvenintegral zusammen:

F Gradientenfeld	Satz 9.1 \Leftrightarrow	$\int F \cdot dx$ wegunabhängig
\Downarrow Satz 9.2	\Uparrow 1-fach zshg. Satz 9.4 \Uparrow	\Downarrow (klar)
F rotationsfrei	Satz 9.5 \Leftrightarrow	$\int F \cdot dx$ homotopieinvariant

Die Implikation von rechts nach links in der unteren Zeile folgt leicht aus der Homotopieformel, Lemma 9.3, und dem Fundamentallema der Variationsrechnung. Abstrakt aber nicht exakt kann die Sache so gesehen werden: für ein gegebenes Vektorfeld F ist das Kurvenintegral ein Funktional \mathcal{F} auf dem Raum X der geschlossenen Kurven in Ω . Nach Lemma 9.3 ist die Ableitung dieses Funktionals gleich

$$D\mathcal{F}(\gamma)\phi = - \int_a^b \langle (DF - DF^T) \circ \gamma \frac{\partial \gamma}{\partial s}, \phi \rangle ds.$$

F rotationsfrei bedeutet, dass diese Ableitung gleich Null ist. In Konsequenz ist \mathcal{F} konstant auf den (Weg-)Komponenten von X , den Homotopieklassen. Rigoros wird das in Satz 9.5 bewiesen. Ist Ω einfach zusammenhängend, das heißt die Homotopieklasse der konstanten Kurven ist die einzige Komponente, so ist \mathcal{F} identisch Null. Dann existiert eine Stammfunktion, wie in Satz 9.1 gezeigt.