

# Kapitel 6

## Differentialrechnung für Funktionen in mehreren Variablen

### 1 Topologische Grundbegriffe

den Begriff der Stetigkeit auf Funktionen, die auf Teilmengen des  $\mathbb{R}^n$  gegeben sind. Vorab führen wir offene und abgeschlossene Mengen ein. Eine Funktion einer Variablen ist üblicherweise auf einem Intervall definiert, entsprechend könnten wir Funktionen von mehreren Variablen nur auf Rechtecken oder Quadern betrachten. Das wäre aber zu einschränkend, eine kurze Diskussion der topologischen Grundbegriffe ist notwendig.

Bereits bekannt ist die Euklidische Norm auf dem  $\mathbb{R}^n$ , also

$$|x| = \left( \sum_{i=1}^n x_i^2 \right)^{\frac{1}{2}} \quad \text{für } x = (x_1, \dots, x_n) \in \mathbb{R}^n. \quad (1.1)$$

Es gilt  $|x| \geq 0$ ,  $|\lambda x| = |\lambda| |x|$  für  $\lambda \in \mathbb{R}$ , und die Dreiecksungleichung

$$|x + y| \leq |x| + |y| \quad \text{für alle } x, y \in \mathbb{R}^n.$$

Dies hatten wir mit der Ungleichung von Cauchy-Schwarz hergeleitet. Die offene Kugel mit Mittelpunkt  $x \in \mathbb{R}^n$  und Radius  $r > 0$  ist die Menge

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

**Definition 1.1 (offene Menge)** Eine Menge  $D \subset \mathbb{R}^n$  heißt offen, wenn es zu jedem  $x \in D$  ein  $\varepsilon > 0$  gibt mit  $B_\varepsilon(x) \subset D$ .

**Beispiel 1.1** Die Kugel  $B_r(x_0)$  ist offen im Sinn der Definition 1.1. Denn aus der Dreiecksungleichung folgt (Bild)

$$|x - x_0| < r \quad \Rightarrow \quad B_\varepsilon(x) \subset B_r(x_0) \quad \text{für } \varepsilon = r - |x - x_0| > 0.$$

Die leere Menge und  $\mathbb{R}^n$  sind offen, und jede Vereinigung von offenen Mengen ist offen. Bei Durchschnitten muss man aufpassen, im allgemeinen sind nur endliche Durchschnitte wieder offen.

**Definition 1.2 (Konvergenz)** Eine Folge  $x_k \in \mathbb{R}^n$  konvergiert gegen  $x \in \mathbb{R}^n$ , falls gilt:

$$|x_k - x| \rightarrow 0 \quad \text{mit } k \rightarrow \infty.$$

Mit anderen Worten: zu jedem  $\varepsilon > 0$  gibt es ein  $K \in \mathbb{N}$  mit  $x_k \in B_\varepsilon(x)$  für alle  $k \geq K$ .

Die folgende Tatsache wird sehr oft benutzt: eine Folge  $x_k \in \mathbb{R}^n$  konvergiert genau dann gegen  $x \in \mathbb{R}^n$ , wenn sie komponentenweise konvergiert:

$$\lim_{k \rightarrow \infty} x_k = x \quad \Leftrightarrow \quad \lim_{k \rightarrow \infty} (x_k)^i = x^i \quad \text{für } i = 1, \dots, n.$$

Dies ergibt sich aus den Ungleichungen

$$|(x_k)^i - x^i| \leq |x_k - x| \leq \sqrt{n} \max_{1 \leq i \leq n} |(x_k)^i - x^i|.$$

Ich schreibe die Koordinaten hier oben nur um Verwechslung mit dem Folgenindex zu vermeiden.

**Definition 1.3 (abgeschlossene Menge)** Eine Menge  $D \subset \mathbb{R}^n$  heißt abgeschlossen, wenn folgende Implikation gilt:

$$x_k \in D, \quad x_k \rightarrow x \quad \Rightarrow \quad x \in D.$$

Es ist naheliegend, die Begriffe *offen* und *abgeschlossen* als Gegensätze anzusehen. Das ist nicht ganz richtig, vielmehr gilt:

- Es gibt viele Mengen  $D \subset \mathbb{R}^n$ , die weder offen noch abgeschlossen sind, zum Beispiel halboffene Intervalle in  $\mathbb{R}$ .
- Die leere Menge und der ganze  $\mathbb{R}^n$  sind sowohl offen als auch abgeschlossen.
- Eine Menge  $D \subset \mathbb{R}^n$  ist genau dann offen, wenn  $\mathbb{R}^n \setminus D$  abgeschlossen ist.

**Definition 1.4 (Rand und Abschluss einer Menge)** Ein Punkt  $x \in \mathbb{R}^n$  heißt *Randpunkt* von  $D \subset \mathbb{R}^n$ , wenn es in jeder Kugel  $B_\varepsilon(x)$  einen Punkt aus  $D$  und einen Punkt nicht aus  $D$  gibt. Die Menge der Randpunkte wird mit  $\partial D$  bezeichnet, der Abschluss von  $D$  ist die Menge  $\overline{D} = D \cup \partial D$ .

**Beispiel 1.2** Für  $B_r(x_0) = \{x \in \mathbb{R}^n : |x - x_0| < r\}$  behaupten wir

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

Denn ist  $|x - x_0| = r$ , so liegen in  $B_\varepsilon(x)$  Punkte aus  $B_r(x_0)$  (zum Beispiel  $x_0 + t(x - x_0)$  mit  $t \nearrow 1$ ), sowie Punkte nicht aus  $B_r(x_0)$  (zum Beispiel  $x$  selbst). Somit ist  $x$  ein Randpunkt. Für  $|x - x_0| < r$  hatten wir dagegen ein  $\varepsilon > 0$  gefunden mit  $B_\varepsilon(x) \subset B_r(x_0)$ , das heißt  $x$  ist kein Randpunkt. Entsprechend gilt

$$|x - x_0| > r \quad \Rightarrow \quad B_\varepsilon(x) \subset \mathbb{R}^n \setminus B_r(x_0) \quad \text{für } \varepsilon = |x - x_0| - r > 0.$$

Auch diese Punkte sind also keine Randpunkte, die Behauptung ist gezeigt. Es folgt weiter

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

**Definition 1.5 (Grenzwert und Stetigkeit)** Sei  $f : D \rightarrow \mathbb{R}$  wobei  $D \subset \mathbb{R}^n$ .

(1)  $f(x)$  konvergiert für  $x \rightarrow x_0$  gegen  $a \in \mathbb{R}$ , falls es zu jedem  $\varepsilon > 0$  ein  $\delta > 0$  gibt, so dass

$$|f(x) - a| < \varepsilon \quad \text{für alle } x \in D \text{ mit } 0 < |x - x_0| < \delta.$$

Notation:  $\lim_{x \rightarrow x_0} f(x) = a$ .

(2)  $f(x)$  heißt stetig in  $x_0 \in D$ , falls  $\lim_{x \rightarrow x_0} f(x) = f(x_0)$ .

(3)  $f(x)$  heißt stetig auf  $D$ , wenn  $f(x)$  in allen Punkten  $x \in D$  stetig ist.

Diese Definitionen sind analog zum eindimensionalen Fall, insbesondere wiederholen wir die dort gegebenen Hinweise (siehe Kapitel 2.5):

- Für die Existenz des Grenzwerts ist es egal, ob  $f(x)$  im Punkt  $x_0$  definiert ist bzw. welchen Funktionswert die Funktion dort hat.
- Die Definition des Grenzwerts ist nicht sinnvoll, wenn  $x_0$  isolierter Punkt von  $D$  ist, das heißt es gibt ein  $\varepsilon > 0$ , so dass  $x_0$  einziger Punkt aus  $D$  in  $B_\varepsilon(x_0)$ .
- Summe, Produkt und Quotient (außerhalb der Nullstellen des Nenners) von stetigen Funktionen sind wieder stetig. Ebenfalls stetig ist die Verkettung mit einer stetigen Funktion auf  $\mathbb{R}$ .

**Beispiel 1.3** Für jedes  $i = 1, \dots, n$  ist die Koordinatenfunktion  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $f(x) = x_i$ , stetig auf  $\mathbb{R}^n$ . Denn es gilt

$$|f(y) - f(x)| = |y_i - x_i| < \varepsilon \quad \text{für } |y - x| < \varepsilon.$$

Damit sind auch Polynomfunktionen stetig, also Funktionen der Form

$$P(x) = \sum_{i_1, \dots, i_n=1}^N a_{i_1 \dots i_n} x_1^{i_1} \cdot \dots \cdot x_n^{i_n}.$$

**Beispiel 1.4** Sei  $a \in \mathbb{R}^n$  ein fester Punkt. Dann ist die Abstandsfunktion  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $f(x) = |x - a|$ , stetig. Denn für  $x, y \in \mathbb{R}^n$  gilt mit der Dreiecksungleichung

$$f(y) = |y - a| \leq |y - x| + |x - a| = f(x) + |y - x|,$$

also  $f(y) - f(x) \leq |y - x|$ . Vertauschen wir  $x, y$ , so folgt

$$|f(y) - f(x)| \leq |y - x| < \varepsilon \quad \text{für } |y - x| < \varepsilon.$$

**Definition 1.6 (Beschränktheit)** Eine Menge  $D \subset \mathbb{R}^n$  heißt beschränkt, wenn es ein  $K \geq 0$  gibt mit  $|x| \leq K$  für alle  $x \in D$ .

**Satz 1.1 (Existenz von Extremstellen)** Sei  $D \subset \mathbb{R}^n$  abgeschlossen und beschränkt, und  $f : D \rightarrow \mathbb{R}$  sei stetig. Dann gibt es  $x_0, x_1 \in D$  mit

$$f(x_0) \leq f(x) \leq f(x_1) \quad \text{für alle } x \in D.$$

Insbesondere ist die Funktion  $f$  beschränkt.

BEWEIS: (Skizze) Um die Minimalstelle zu finden, wählen wir Punkte  $x_k \in D$  mit

$$\lim_{k \rightarrow \infty} f(x_k) = \inf_{x \in D} f(x).$$

Eine solche Folge nennt man eine Minimalfolge. Angenommen die  $x_k$  konvergieren gegen ein  $x_0 \in \mathbb{R}^n$ . Da  $D$  abgeschlossen ist, folgt dann  $x_0 \in D$  und weiter wegen der Stetigkeit von  $f(x)$

$$f(x_0) = \lim_{k \rightarrow \infty} f(x_k) = \inf_{x \in D} f(x).$$

Somit ist die gesuchte Minimalstelle gefunden. Im allgemeinen muss die Folge  $x_k$  nicht konvergieren, zum Beispiel könnte  $f(x)$  mehrere Minimalstellen haben. Aber es reicht, wenn wir eine konvergente Teilfolge  $x_{k_1}, x_{k_2}, \dots$  mit  $k_1 < k_2 < \dots$  finden, denn diese ist auch eine Minimalfolge, und das Argument von oben ist gültig.

Eine beschränkte Folge im  $\mathbb{R}^n$  hat nun immer eine solche konvergente Teilfolge. Im Fall  $n = 1$  sei dazu  $I_0$  ein Intervall, das alle Folgenglieder enthält. Dann wählen wir eine Teilfolge, die ganz in einer der Hälften von  $I_0$  liegt. Sei  $I_1$  diese Hälfte, wähle dann eine Teilfolge, die ganz in einer Hälfte von  $I_1$  liegt. Wir erhalten sukzessive Teilfolgen, so dass die  $m$ -te Teilfolge ganz in  $I_m$  liegt, wobei  $I_0 \supset I_1 \supset \dots$  und  $|I_m| = 2^{-m}|I_0| \rightarrow 0$ . Allerdings muss keine dieser Teilfolgen konvergieren. Der Trick ist folgendes Diagonalargument von Bolzano und Weierstrass: betrachte die Folge, deren  $m$ -tes Glied das  $m$ -te Glied der  $m$ -ten Teilfolge ist. Dies ist die gewünschte konvergente Teilfolge, denn für jedes  $m$  liegt die Folge ab Nummer  $m$  ganz in  $I_m$ . In Dimension  $n \geq 1$  wendet man dieses Argument auf jede der Koordinaten an.  $\square$

**Beispiel 1.5** Sei  $K \subset \mathbb{R}^n$  kompakt. Dann gibt es zu jedem  $a \in \mathbb{R}^n$  einen nächsten Punkt  $x_0 \in K$ , das heißt es gilt

$$|x_0 - a| \leq |x - a| \quad \text{für alle } x \in K.$$

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

**Beispiel 1.6** Betrachte die quadratische Form

$$q : \mathbb{R}^n \rightarrow \mathbb{R}, q(x) = \sum_{i,j=1}^n a_{ij} x_i x_j.$$

Wie oben festgestellt ist  $q(x)$  als Polynomfunktion stetig auf  $\mathbb{R}^n$ . Die Standardsphäre  $\mathbb{S}^{n-1} = \{x \in \mathbb{R}^n : |x| = 1\}$  ist beschränkt (wähle  $K = 1$ ) und abgeschlossen: die Funktion  $f(x) = |x|$  ist stetig, also gilt

$$x_k \in \mathbb{S}^{n-1}, x_k \rightarrow x \quad \Rightarrow \quad |x| = \lim_{k \rightarrow \infty} |x_k| = 1.$$

Somit ist  $x \in \mathbb{S}^{n-1}$ . Aus Satz 1.1 folgt, dass  $q(x)$  eine Minimalstelle auf  $\mathbb{S}^{n-1}$  hat. Dies hatten wir bei der Hauptachsentransformation, siehe Satz 7.3, benutzt.

## 2 Die partiellen Ableitungen

Wir kommen nun zur Differentiation von Funktionen im  $\mathbb{R}^n$ . Um für diese Ableitungen zu definieren, ist die einfachste und vielfach beste Idee, alle Variablen bis auf  $x_j$  als konstant aufzufassen und die resultierende Funktion der einen Variablen  $x_j$  wie üblich zu differenzieren. Auf diese Weise ergeben sich für  $j = 1, \dots, n$  die  $n$  partiellen Ableitungen der Funktion.

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

$$\begin{aligned}\frac{\partial f}{\partial x_j}(x) &= \lim_{t \rightarrow 0} \frac{f(x_1, \dots, x_j + t, \dots, x_n) - f(x_1, \dots, x_j, \dots, x_n)}{t} \\ &= \lim_{t \rightarrow 0} \frac{f(x + te_j) - f(x)}{t}.\end{aligned}$$

Dies ist genau die Ableitung der Funktion  $x_j \mapsto f(x_1, \dots, x_j, \dots, x_n)$ . Ebenfalls übliche Bezeichnungen sind  $\partial_j f(x)$  und  $f_{x_j}(x)$ .

Das Symbol  $\partial$  (statt  $d$ ) weist darauf hin, dass es sich um eine partielle Ableitung nach einer der Variablen handelt. Der folgende Satz formuliert Aussagen der eindimensionalen Analysis im jetzigen Kontext.

**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 vorausgesetzt. Dann gelten folgende Aussagen:

(a) Linearität: für  $f, g : \Omega \rightarrow \mathbb{R}$  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) Produktregel: für  $f, g : \Omega \rightarrow \mathbb{R}$  gilt

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

(d) Quotientenregel: für  $f, g : \Omega \rightarrow \mathbb{R}$  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: ist  $f : \Omega \rightarrow I \subset \mathbb{R}$  und  $\varphi : I \rightarrow \mathbb{R}$  differenzierbar in  $f(x)$ , so folgt

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

**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

$$\partial_j r(x) = \frac{2x_j}{2\sqrt{x_1^2 + \dots + x_n^2}} = \frac{x_j}{r} \quad \text{für } r = |x|.$$

Im Nullpunkt sind die partiellen Ableitungen dagegen nicht definiert, denn  $r(0 + te_i) = |t|$  ist in  $t = 0$  nicht differenzierbar. Die Funktion  $\partial_j r$  ist in  $x \neq 0$  ihrerseits partiell differenzierbar, und wir erhalten die zweiten partiellen Ableitungen

$$\partial_i(\partial_j r)(x) = \frac{(\partial_i x_j)r - x_j \partial_i r}{r^2} = \frac{1}{r} \left( \delta_{ij} - \frac{x_i x_j}{r^2} \right).$$

Ist  $\varphi : (0, \infty) \rightarrow \mathbb{R}$  zweimal differenzierbar, so berechnen wir weiter für  $f = \varphi \circ r : \mathbb{R}^n \rightarrow \mathbb{R}$

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

Der Operator  $\Delta = \sum_{i=1}^n \partial_i^2$  heißt Laplaceoperator, die Lösungen der Gleichung  $\Delta f = 0$  heißen harmonische Funktionen. Wir können jetzt die rotationssymmetrischen harmonischen Funktionen ausrechnen, und zwar erhalten wir

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

Diese Gleichung hat die Lösungen, 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  $f$  das Newtonsche Gravitationspotential bzw. Coulombpotential.

Im Beispiel traten zweite partielle Ableitungen auf. Ist die Ableitungsfunktion  $\partial_j f : \Omega \rightarrow \mathbb{R}$  definiert und wiederum nach  $x_i$  partiell differenzierbar, so setzen wir

$$\partial_i \partial_j f(x) := \partial_i(\partial_j f)(x) \quad (\text{alternative Notation } \frac{\partial^2 f}{\partial x_i \partial x_j}(x)). \quad (2.2)$$

Entsprechend werden partielle Ableitungen beliebiger Ordnung gebildet. Die folgenden Klassen von Funktionen spielen in der Analysis eine wichtige Rolle. Wir werden sehen, dass die Eigenschaft der Stetigkeit der partiellen Ableitungen in vielen Anwendungen wesentlich ist.

**Definition 2.2 ( $C^k$ -Räume)** Sei  $\Omega \subset \mathbb{R}^n$  offen und  $k \in \mathbb{N}_0 \cup \{\infty\}$ . Wir bezeichnen mit  $C^k(\Omega)$  die Menge aller  $k$ -mal stetig differenzierbaren Funktionen auf  $\Omega$ , das heißt alle partiellen Ableitungen  $\partial_{i_1} \dots \partial_{i_j} f$  der Ordnung  $0 \leq j \leq k$  (bzw.  $j < \infty$  im Fall  $k = \infty$ ) sind definiert und stetig auf  $\Omega$ .

Wir wollen nun zeigen, dass die Operatoren  $\partial_i$  und  $\partial_j$  auf  $C^2$ -Funktionen vertauschen. Beispiele zeigen, dass dies nicht allein aus der Existenz der Ableitungen  $\partial_i \partial_j f$  und  $\partial_j \partial_i f$  folgt, die Stetigkeit dieser Ableitungen ist wesentlich.

**Satz 2.2 (Symmetrie der zweiten Ableitung)** Für  $f \in C^2(\Omega)$  gilt

$$\partial_i \partial_j f = \partial_j \partial_i f \quad \text{auf } \Omega \quad \text{für alle } 1 \leq i, j \leq n.$$

BEWEIS: Es reicht, eine Funktion von zwei Variablen zu betrachten. Bilde für  $f = f(x, y)$  nacheinander die Differenzenquotienten nach  $y$  und  $x$ :

$$\begin{aligned}\delta_2^h f(x, y) &= \frac{f(x, y+h) - f(x, y)}{h}, \\ \delta_1^h(\delta_2^h f)(x, y) &= \frac{\delta_2^h f(x+h, y) - \delta_2^h f(x, y)}{h} \\ &= \frac{f(x+h, y+h) - f(x+h, y) - f(x, y+h) + f(x, y)}{h^2}.\end{aligned}$$

Das Ergebnis ist symmetrisch bzgl.  $x$  und  $y$ , wir haben also

$$\delta_1^h(\delta_2^h f)(x, y) = \delta_2^h(\delta_1^h f)(x, y).$$

Mit dem Mittelwertsatz können wir einen Differenzenquotienten durch die Ableitung an einer Zwischenstelle ersetzen. Wir machen das links erst bezüglich  $x$  und dann bzgl.  $y$ , und erhalten  $\xi, \eta$  mit  $|\xi - x|, |\eta - y| \leq |h|$ , so dass gilt:

$$\begin{aligned}\delta_1^h(\delta_2^h f)(x, y) &= \partial_1(\delta_2^h f)(\xi, y) \\ &= \frac{1}{h}(\partial_1 f(\xi, y+h) - \partial_1 f(\xi, y)) \\ &= \partial_2(\partial_1 f)(\xi, \eta).\end{aligned}$$

Analog können wir rechts umformen, es folgt mit anderen Zwischenstellen  $\xi', \eta'$

$$\partial_2(\partial_1 f)(\xi, \eta) = \partial_1(\partial_2 f)(\xi', \eta').$$

Mit  $h \rightarrow 0$  folgt wegen der Stetigkeit  $\partial_2(\partial_1 f)(x, y) = \partial_1(\partial_2 f)(x, y)$ . □

Durch Induktion ergibt sich allgemeiner die

**Folgerung 2.1** Für  $f \in C^k(\Omega)$  vertauschen die partiellen Ableitungen bis zur Ordnung  $k$ .

Eine geringfügige Verallgemeinerung der partiellen Ableitung ist die Richtungsableitung. Sie hat den Vorteil, dass sie nicht von der Wahl des Koordinatensystems explizit abhängt.

**Definition 2.3 (Richtungsableitung)**  $\Omega \subset \mathbb{R}^n$  offen und  $f : \Omega \rightarrow \mathbb{R}$ . Für  $x \in \Omega$  und  $v \in \mathbb{R}^n$  heißt der Grenzwert (falls existent)

$$\partial_v f(x) = \lim_{t \rightarrow 0} \frac{f(x+tv) - f(x)}{t}$$

Richtungsableitung von  $f$  an der Stelle  $x$  in Richtung  $v$ . Dies ist die gewöhnliche Ableitung der Funktion  $t \mapsto f(x+tv)$  an der Stelle  $t=0$ .

**Beispiel 2.2** Die Richtungsableitung von  $r(x) = |x|$  in  $x \in \mathbb{R}^n \setminus \{0\}$  in Richtung  $v \in \mathbb{R}^n$  ist

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

Aus Gründen der Übersichtlichkeit haben wir die partiellen Ableitungen erst nur für skalare Funktionen, also Funktionen mit Werten in  $\mathbb{R}$ , betrachtet. Die Verallgemeinerung auf vektorwertige Funktionen ist aber einfach, die Ableitung wird komponentenweise gebildet. Wir betrachten das mal im Spezialfall der parametrisierten Kurven.

Eine stetige Abbildung  $\gamma : I \rightarrow \mathbb{R}^m$ ,  $I$  Intervall, nennt man eine parametrisierte Kurve. Es ist hilfreich, sich  $\gamma(t)$  als Ortsvektor eines Punkts vorzustellen, der sich abhängig von der Zeit  $t \in I$  bewegt. Der Geschwindigkeitsvektor (oder die Ableitung) von  $\gamma(t)$  ist (falls existent)

$$\gamma'(t) = \begin{pmatrix} \gamma'_1(t) \\ \vdots \\ \gamma'_m(t) \end{pmatrix}.$$

Die Bogenlänge von  $\gamma \in C^1(I, \mathbb{R}^m)$  ist definiert als das Integral

$$L(\gamma) = \int_a^b |\gamma'(t)| dt. \quad (2.3)$$

Zur Motivation dieser Definition betrachte eine Unterteilung  $a = t_0 < \dots < t_N = b$ . Dann sollte die Länge von  $\gamma$  in etwa die Länge des zugehörigen Polygonzugs sein (Bild):

$$L(\gamma) \approx \sum_{k=1}^N |\gamma(t_k) - \gamma(t_{k-1})|.$$

Da  $\gamma'(t)$  auf jedem Teilintervall nahezu konstant ist, gilt (*Weg = Geschwindigkeit  $\times$  Zeit*)

$$|\gamma(t_k) - \gamma(t_{k-1})| \approx |\gamma'(t_{k-1})|(t_k - t_{k-1}) \approx \int_{t_{k-1}}^{t_k} |\gamma'(t)| dt.$$

Summieren wir für  $k = 1, \dots, N$ , so ergibt sich (näherungsweise) die Formel (2.3).

**Beispiel 2.3** Die Schraubenlinie ist die parametrisierte Kurve

$$\gamma : \mathbb{R} \rightarrow \mathbb{R}^3, \gamma(t) = \begin{pmatrix} r \cos t \\ r \sin t \\ ht \end{pmatrix}.$$

Die Bahn der Kurve liegt auf dem Zylinder  $\{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 = r^2\}$ . Tangentenvektor sowie Beschleunigungsvektor lauten

$$\gamma'(t) = \begin{pmatrix} -r \sin t \\ r \cos t \\ h \end{pmatrix} \quad \text{und} \quad \gamma''(t) = \begin{pmatrix} -r \cos t \\ -r \sin t \\ 0 \end{pmatrix}.$$

Bei einem Umlauf wächst die dritte Komponente von  $\gamma(t)$  um den Wert  $2\pi h$ , die sogenannte Ganghöhe der Schraubenlinie. Die Bogenlänge eines Umlaufs ist

$$L(\gamma|_{[0, 2\pi]}) = \int_0^{2\pi} |\gamma'(t)| dt = \int_0^{2\pi} \sqrt{r^2 + h^2} dt = 2\pi \sqrt{r^2 + h^2}.$$

Bei Abrollen des Zylinders wird aus der Schraubenlinie eine Gerade, das betrachtete Stück entspricht der Hypotenuse im rechtwinkligen Dreieck mit Katheten  $2\pi r$  und  $2\pi h$ .

Für eine vektorwertige Funktion von mehreren Variablen werden die partiellen Ableitungen ebenfalls komponentenweise gebildet. Mit  $f : \Omega \rightarrow \mathbb{R}^m$ ,  $\Omega \subset \mathbb{R}^n$ , gilt also

$$\partial_j f(x) = \begin{pmatrix} \partial_j f_1(x) \\ \vdots \\ \partial_j f_m(x) \end{pmatrix} \quad \text{für } j = 1, \dots, n.$$

Im Fall  $n = 2, m = 3$  kann man sich vorstellen, dass  $f$  eine Fläche im  $\mathbb{R}^3$  parametrisiert.

**Beispiel 2.4** Sei  $\gamma : I \rightarrow \mathbb{R}^2$ ,  $\gamma(s) = (x(s), r(s))$ , eine Kurve mit  $r(s) > 0$  für alle  $s \in I$ . Durch Rotation um die  $x$ -Achse erhalten wir dann die Rotationsfläche

$$f : I \times \mathbb{R} \rightarrow \mathbb{R}^3, f(s, \theta) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \theta & -\sin \theta \\ 0 & \sin \theta & \cos \theta \end{pmatrix} \cdot \begin{pmatrix} x(s) \\ r(s) \\ 0 \end{pmatrix} = \begin{pmatrix} x(s) \\ r(s) \cos \theta \\ r(s) \sin \theta \end{pmatrix}.$$

Die partiellen Ableitungen von  $f(s, \theta)$  lauten

$$\frac{\partial f}{\partial s}(s, \theta) = \begin{pmatrix} x'(s) \\ r'(s) \cos \theta \\ r'(s) \sin \theta \end{pmatrix} \quad \text{und} \quad \frac{\partial f}{\partial \theta}(s, \theta) = \begin{pmatrix} 0 \\ -r(s) \sin \theta \\ r(s) \cos \theta \end{pmatrix}.$$

Die Kurve  $s \mapsto f(s, 0) = (\gamma(s), 0)$  hat den Geschwindigkeitsvektor  $\frac{\partial f}{\partial s}(s, 0) = (\gamma'(s), 0)$ . Die Kurve  $s \mapsto f(s, \theta)$  ergibt sich aus  $f(s, 0)$  durch Drehung um den Winkel  $\theta$ , dementsprechend ist  $\frac{\partial f}{\partial s}(s, \theta)$  der gedrehte Geschwindigkeitsvektor. Die Kurve  $\theta \mapsto f(s, \theta)$  beschreibt einen Kreis mit Radius  $r(s)$  in der Ebene  $x = x(s)$ . Der Vektor  $\frac{\partial f}{\partial \theta}(s, \theta)$  ist der Geschwindigkeitsvektor dieses Kreises.

### 3 Die Jacobimatrix

**Definition 3.1 (Jacobimatrix)** Sei  $f \in C^1(\Omega, \mathbb{R}^m)$  wobei  $\Omega \subset \mathbb{R}^n$ . Die Jacobimatrix von  $f$  im Punkt  $x \in \Omega$  ist

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

In der  $i$ -ten Zeile stehen die Ableitungen der  $i$ -ten Komponente von  $f$ , in der  $j$ -ten Spalte stehen die Ableitungen von  $f$  nach der Variablen  $x_j$ .

**Beispiel 3.1** Zweidimensionale Polarkoordinaten sind definiert durch

$$\phi : (0, \infty) \times [0, 2\pi) \rightarrow \mathbb{R}^2 \setminus \{0\}, \phi(r, \theta) = (r \cos \theta, r \sin \theta).$$

Die Jacobimatrix dieser Abbildung lautet

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

**Beispiel 3.2** Betrachte für  $A \in \mathbb{R}^{m \times n}$  und  $b \in \mathbb{R}^m$  die affine Funktion  $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ ,  $f(x) = Ax + b$ . Dann gilt  $f_i(x) = \sum_{j=1}^n a_{ij}x_j + b_i$  und somit

$$\frac{\partial f_i}{\partial x_j}(x) = a_{ij} \quad \text{bzw.} \quad Df(x) = A \quad \text{für alle } x \in \mathbb{R}^n.$$

**Satz 3.1 (von der linearen Approximation)** Ist  $f \in C^1(\Omega, \mathbb{R}^m)$ , so gilt im Punkt  $x_0$  mit  $A = Df(x_0)$  die Approximationseigenschaft

$$\lim_{x \rightarrow x_0} \frac{f(x) - (f(x_0) + A(x - x_0))}{|x - x_0|} = 0. \quad (3.5)$$

BEWEIS: Die Konvergenz ist gleichbedeutend mit Konvergenz der Komponenten, daher können wir  $m = 1$  annehmen. Sei zuerst  $x_0 = 0$  und  $f(0) = 0$ ,  $Df(0) = 0$ . Dann gilt mit dem Mittelwertsatz

$$\begin{aligned} f(x) &= \sum_{j=1}^n (f(x_1, \dots, x_{j-1}, x_j, 0, \dots, 0) - f(x_1, \dots, x_{j-1}, 0, \dots, 0)) \\ &= \sum_{j=1}^n (\partial_j f)(x_1, \dots, x_{j-1}, \xi_j, 0, \dots, 0) x_j. \end{aligned}$$

Also gilt die Abschätzung

$$\frac{|f(x)|}{|x|} \leq \sum_{j=1}^n |(\partial_j f)(x_1, \dots, x_{j-1}, \xi_j, 0, \dots, 0)| \rightarrow 0 \quad \text{mit } x \rightarrow 0.$$

Sind  $x_0$ ,  $f(x_0)$  und  $A = Df(x_0)$  beliebig, so betrachte die  $C^1$ -Funktion

$$g(y) = f(x_0 + y) - (f(x_0) + Ay) = f(x_0 + y) - f(x_0) - \sum_{k=1}^n \partial_k f(x_0) y_k.$$

Es gilt dann  $g(0) = 0$  sowie  $Dg(0) = 0$ , und somit

$$\frac{f(x) - (f(x_0) + A(x - x_0))}{|x - x_0|} = \frac{g(x - x_0)}{|x - x_0|} \rightarrow 0 \quad \text{mit } x \rightarrow x_0.$$

□

Der Satz besagt, dass das Verhalten von  $f(x)$  für  $x \rightarrow x_0$  in erster Ordnung durch die affine Funktion  $f(x_0) + A(x - x_0)$  bestimmt wird. Um zu sagen, dass eine Funktion  $f(x)$  mit  $x \rightarrow x_0$  schneller verschwindet bzw. schwächer wächst als eine Funktion  $g(x)$ , wird oft das Landausche  $o$ -Symbol benutzt:

$$f(x) = o(g(x)) \quad \text{für } x \rightarrow x_0 \quad \Leftrightarrow \quad \lim_{x \rightarrow x_0} \frac{f(x)}{g(x)} \rightarrow 0.$$

Die Aussage des Satzes lautet damit

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

Das ist ziemlich griffig, aber man muss die Bedeutung des  $o$ -Symbols im Kopf behalten. Eine erste Konsequenz des Satzes ist, dass wir die Richtungsableitung mit den partiellen Ableitungen ausrechnen können.

**Folgerung 3.1** Die Richtungsableitung von  $f \in C^1(\Omega)$  in Richtung  $v \in \mathbb{R}^n$  ist

$$\partial_v f(x) = Df(x)v = \sum_{j=1}^n \partial_j f(x_0)v_j.$$

BEWEIS: Im Fall  $v = 0$  stimmt die Aussage. Für  $v \neq 0$  rechnen wir

$$\frac{f(x+tv) - f(x)}{t} - Df(x)v = \frac{f(x+tv) - (f(x) - Df(x)(tv))}{t} \rightarrow 0 \quad \text{mit } t \rightarrow 0.$$

□

**Definition 3.2 (Gradient)** Der Gradient einer Funktion  $f \in C^1(\Omega)$  im Punkt  $x \in \Omega$  ist

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

Ist  $\text{grad } f(x) = 0$ , so heißt  $x$  kritischer Punkt von  $f$ . Ist  $x$  nicht kritisch, so ist die Richtung von  $\text{grad } f(x)$  genau die, in der  $f(x)$  am stärksten ansteigt. Denn für  $v \in \mathbb{R}^n$  mit  $|v| = 1$  gilt nach Folgerung 3.1 und Cauchy-Schwarz

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

Gleichheit tritt genau dann ein, wenn der Einheitsvektor  $v$  in Richtung von  $\text{grad } f(x)$  weist.

Beachten Sie, dass der Gradient nur für skalare Funktionen definiert wird. Für diese ist nach unserer Definition die Jacobimatrix  $Df(x) \in \mathbb{R}^{1 \times n}$ , also der Zeilenvektor mit Einträgen  $\partial_j f(x)$ . Der Gradient ist de facto dasselbe, nur haben wir ihn als Spaltenvektor geschrieben, das heißt als Element von  $\mathbb{R}^{n \times 1}$ . In den Standardkoordinaten ist der Unterschied rein formal, bei Wechseln des Koordinatensystems spielt er aber eine Rolle, weil sich die Jacobimatrix anders transformiert als der Gradient. Das werden wir sehen.

**Beispiel 3.3** 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 3.4** Sei  $q(x) = \langle Ax, x \rangle = \sum_{i,k=1}^n a_{ik}x_i x_k$  quadratische Form der symmetrischen Matrix  $A \in \mathbb{R}^{n \times n}$ . Dann gilt

$$\partial_j q(x) = \sum_{i,k=1}^n a_{ik}(\delta_{ij}x_k + x_i\delta_{jk}) = \sum_{k=1}^n a_{jk}x_k + \sum_{i=1}^n a_{ij}x_i = 2 \sum_{k=1}^n a_{jk}x_k.$$

Das bedeutet  $\text{grad } q(x) = 2Ax$ .

Wir kommen jetzt zur mehrdimensionalen Kettenregel. Wir betrachten erst den wichtigen Spezialfall, wenn eine Funktion längs einer Kurve ausgewertet wird.

**Satz 3.2 (Kettenregel längs Kurven)** Sei  $f \in C^1(\Omega)$  und  $x = x(t)$ ,  $t \in (a, b)$ , sei eine  $C^1$ -Kurve in  $\Omega$  mit  $x(t_0) = x_0$ . Dann gilt

$$\frac{d}{dt}f(x(t))|_{t=t_0} = Df(x_0)x'(t_0) = \sum_{j=1}^n (\partial_j f)(x_0)x'_j(t_0).$$

BEWEIS: Nach Satz 3.1 gilt

$$\varepsilon(x) := \frac{f(x) - (f(x_0) + Df(x_0)(x - x_0))}{|x - x_0|} \rightarrow 0 \quad \text{mit } x \rightarrow x_0.$$

Wir berechnen nun

$$\begin{aligned} \frac{f(x(t)) - f(x_0)}{t - t_0} - Df(x_0)x'(t_0) &= \frac{f(x(t)) - (f(x_0) + Df(x_0)(x(t) - x(t_0)))}{t - t_0} \\ &\quad + Df(x_0)\left(\frac{x(t) - x(t_0)}{t - t_0} - x'(t_0)\right) \\ &= \frac{\varepsilon(x(t))|x(t) - x_0|}{t - t_0} + Df(x_0)\left(\frac{x(t) - x(t_0)}{t - t_0} - x'(t_0)\right). \end{aligned}$$

Da  $\lim_{t \rightarrow t_0} \frac{x(t) - x(t_0)}{t - t_0} = x'(t_0)$ , geht die rechte Seite gegen Null und der Satz ist bewiesen.  $\square$   
Ist  $f(x(t))$  konstant, so folgt

$$0 = \frac{d}{dt}f(x(t)) = Df(x(t))x'(t) = \langle \text{grad } f(x(t)), x'(t) \rangle.$$

Die Mengen  $\{x \in \Omega : f(x) = \text{const.}\}$  sind die Niveaulinien ( $n = 2$ ) oder Niveauflächen ( $n = 3$ ) der Funktion  $f$ . Die Aussage bedeutet, der Gradient von  $f$  steht senkrecht auf diese Linien oder Flächen.

Wir kommen nun zur allgemeinen Fassung der Kettenregel. Sind  $f(x) = Ax$  und  $g(y) = By$  lineare Abbildungen mit  $A \in \mathbb{R}^{m \times n}$  und  $B \in \mathbb{R}^{p \times m}$ , so gilt  $g(f(x)) = BAx$ , und diese Funktion hat die Jacobimatrix  $BA \in \mathbb{R}^{p \times n}$ , siehe Beispiel 3.2. Daraus ergibt sich, dass allgemein die Jacobimatrix einer Verkettung das Produkt der Jacobimatrizen an den jeweiligen Stellen sein muss.

**Satz 3.3 (Kettenregel)** Seien  $f \in C^1(U, \mathbb{R}^m)$  und  $g \in C^1(V, \mathbb{R}^p)$  mit  $U \subset \mathbb{R}^n$ ,  $V \subset \mathbb{R}^m$ . Dann ist  $g \circ f \in C^1(U, \mathbb{R}^p)$  und es gilt die Kettenregel

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

Mit andern Worten gilt 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: Mit  $y(t) = f(x_0 + te_k)$  gilt  $y(0) = f(x_0) = y_0$ , und es folgt aus der Kettenregel längs Kurven, Satz 3.2,

$$\frac{\partial(g \circ f)_i}{\partial x_k}(x_0) = \frac{d}{dt}g_i(y(t))|_{t=0} = \sum_{j=1}^m \frac{\partial g_i}{\partial y_j}(y_0)y'_j(0) = \sum_{j=1}^m \frac{\partial g_i}{\partial y_j}(y_0) \frac{\partial f_j}{\partial x_k}(x_0).$$

□

Oft wird die Kettenregel so beschrieben: um die Ableitung der  $i$ -ten Komponente von  $g \circ f$  nach  $x_k$  zu berechnen, wird  $g$  nach allen Variablen  $y_j = f_j(x)$  abgeleitet, und diese werden nach  $x_k$  nachdifferenziert. Die Interpretation der Ableitung als Matrix bzw. lineare Abbildung ermöglicht oft ein gutes geometrisches Verständnis. Allerdings ist beim Umgang mit Zeilen- und Spaltenvektoren Vorsicht geboten, zum Beispiel kann bei der Produktregel Verwirrung entstehen, wenn eine der beteiligten Funktionen vektorwertig ist. Erst recht wird die Notation kompliziert, wenn zweite oder höhere Ableitungen zu bilden sind. Im Zweifelsfall sollte man auf die partiellen Ableitungen zurückgreifen.

## 4 Erste Anwendungen der Ableitung

Um aus Eigenschaften der Ableitung auf das Verhalten einer Funktion zu schließen, haben wir bei Funktionen einer Variablen zwei Hilfsmittel, den Mittelwertsatz und den Hauptsatz der Differential- und Integralrechnung. Diese können wir auch für Funktionen mehrerer Variabler nutzen, indem wir die Funktionen längs Kurven  $x = x(t)$  auswerten.

**Lemma 4.1** Sei  $f \in C^1(\Omega)$ , wobei  $\Omega \subset \mathbb{R}^n$ , und  $x_0, x_1 \in \Omega$ . Ist  $x : [0, 1] \rightarrow \Omega$  eine  $C^1$ -Verbindungskurve von  $x_0$  nach  $x_1$ , das heißt  $x(0) = x_0$  und  $x(1) = x_1$ , so gilt

$$f(x_1) - f(x_0) = \int_a^b Df(x(t))x'(t) dt. \quad (4.1)$$

BEWEIS: Die Kettenregel für Kurven, Satz 3.2, besagt

$$\frac{d}{dt}f(x(t)) = Df(x(t))x'(t).$$

Durch Integration folgt die Behauptung. □

**Definition 4.1** Eine offene Menge  $\Omega \subset \mathbb{R}^n$  heißt *wegweise zusammenhängend*, wenn es zu je zwei Punkten  $x_0, x_1 \in \Omega$  eine  $C^1$ -Verbindungskurve von  $x_0$  nach  $x_1$  in  $\Omega$  gibt.

**Satz 4.1 (Konstanzsatz)** Sei  $\Omega \subset \mathbb{R}^n$  offen und *wegweise zusammenhängend*. Dann gilt:

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

BEWEIS: Wähle zu  $x_0, x_1$  eine  $C^1$ -Verbindungskurve  $x : [0, 1] \rightarrow \Omega$ . Aus Lemma 4.1 folgt

$$f(x_1) - f(x_0) = \int_a^b Df(x(t))x'(t) dt = 0.$$

Da  $x_0, x_1$  beliebig, ist die Funktion konstant. □

Wählen wir als Verbindungskurven Strecken, so können wir auch quantitative Aussagen machen. Allerdings liegt die Verbindungsstrecke zwischen zwei Punkten  $x_0, x_1 \in \Omega$  nicht unbedingt in  $\Omega$ , das müssen wir extra garantieren.

**Definition 4.2** Eine offene Menge  $\Omega \subset \mathbb{R}^n$  heißt *konvex*, falls gilt:

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

**Satz 4.2 (Schrankensatz)** Sei  $\Omega \subset \mathbb{R}^n$  offen und konvex, und  $f \in C^1(\Omega)$ . Es gelte  $|Df(x)| \leq L < \infty$  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  $x(t) = (1-t)x_0 + tx_1$  gilt  $x'(t) = x_1 - x_0$ , also folgt aus Lemma 4.1

$$\begin{aligned} |f(x_1) - f(x_0)| &= \left| \int_0^1 Df(x(t))(x_1 - x_0) dt \right| \\ &\leq \int_0^1 |Df(x(t))(x_1 - x_0)| dt \\ &\leq \int_0^1 |Df(x(t))| |x_1 - x_0| dt \quad (\text{Cauchy-Schwarz}) \\ &\leq L|x_1 - x_0|. \end{aligned}$$

□

Als nächstes diskutieren wir lokale Extrema von Funktionen mehrerer Variabler.

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

Eine quadratische Form  $q(x) = \langle Ax, x \rangle$  mit  $A \in \mathbb{R}^{n \times n}$  symmetrisch hat den Gradienten  $\text{grad } q(x) = 2Ax$ , also ist der Nullpunkt kritischer Punkt von  $q$ . In der linearen Algebra hatten wir gezeigt, dass  $q(x)$  bezüglich einer geeigneten Orthonormalbasis  $v_1, \dots, v_n$  eine Summe von Quadraten ist:

$$q(x) = \sum_{i=1}^n \lambda_i \xi_i^2 \quad \text{für } x = \sum_{i=1}^n \xi_i v_i.$$

Der Nullpunkt ist genau dann ein Minimum von  $q(x)$ , wenn die Eigenwerte  $\lambda_i$  von  $A$  alle nichtnegativ sind. Es handelt sich um ein striktes Minimum, wenn die  $\lambda_i$  alle positiv sind. Für Maxima gilt genau das Entsprechende, nur mit nichtnegativen bzw. positiven Eigenwerten. Für quadratische Funktionen einer Variablen ist ein kritischer Punkt automatisch ein Extremum. Das gilt nicht bei Funktionen mehrerer Variabler, zum Beispiel hat die quadratische Form  $q(x_1, x_2) = x_1^2 - x_2^2$  im Nullpunkt weder ein Minimum noch ein Maximum, auch nicht lokal.

**Definition 4.4** Eine symmetrische Matrix  $A \in \mathbb{R}^{n \times n}$  heißt positiv definit (bzw. positiv semidefinit), falls gilt

$$\langle Ax, x \rangle > 0 \quad (\text{bzw. } \langle Ax, x \rangle \geq 0) \quad \text{für alle } x \in \mathbb{R}^n, x \neq 0.$$

Entsprechend für negativ (semi-)definit.

Bei dem Begriff der *Monotonie* lassen wir den Gleichheitsfall zu, und sprechen ausdrücklich von *streng monoton*, wenn wir die Gleichheit ausschließen wollen. Hier ist der Sprachgebrauch anders, *definit* bedeutet automatisch die strikte Ungleichung. In manchen Büchern werden die Symbole  $A > 0$  bzw.  $A \geq 0$  benutzt. Das Beispiel oben zeigt allerdings, dass für eine symmetrische Matrix weder  $A \geq 0$  noch  $A \leq 0$  gelten muss. Matrizen mit sowohl positiven als auch negativen Eigenwerten heißen indefinit.

**Definition 4.5** Sei  $f \in C^2(\Omega)$  mit  $\Omega \subset \mathbb{R}^n$  offen. Die zweite Ableitung oder Hesse-Matrix von  $f$  im Punkt  $x \in \Omega$  ist

$$D^2 f(x) = (\partial_{ij}^2 f(x)) \in \mathbb{R}^{n \times n}.$$

Die Symmetrie der Hesse-Matrix wurde in Satz 2.2 gezeigt.

**Satz 4.3 (notwendige Bedingungen für Extrema)** Sei  $\Omega \subset \mathbb{R}^n$  offen, und  $f \in C^2(\Omega)$  habe in  $x \in \Omega$  ein Minimum (Maximum). Dann gilt:

- (1)  $Df(x) = 0$ , also ist  $x$  kritischer Punkt.
- (1)  $D^2 f(x)$  ist positiv semidefinit (negativ semidefinit).

BEWEIS: Für  $v \in \mathbb{R}^n$  berechnen wir mit der Kettenregel

$$\begin{aligned} \frac{d}{dt} f(x + tv) &= \sum_{j=1}^n \partial_j f(x + tv) v_j = Df(x + tv)v \\ \frac{d^2}{dt^2} f(x + tv)|_{t=0} &= \sum_{i,j=1}^n \partial_{ij}^2 f(x) v_i v_j = \langle D^2 f(x)v, v \rangle. \end{aligned}$$

Die Funktion  $t \mapsto f(x + tv)$  hat bei  $t = 0$  ein Minimum, also folgt

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

Dies bedeutet  $Df(x) = 0$ . Weiter folgt aus dem eindimensionalen Fall

$$0 \leq \frac{d^2}{dt^2} f(x + tv)|_{t=0} = \langle D^2 f(x)v, v \rangle \quad \text{für alle } v \in \mathbb{R}^n.$$

Somit ist  $D^2 f(x)$  positiv semidefinit. □

Sei nun  $Df(x) = 0$ , also  $x \in \Omega$  kritischer Punkt der Funktion  $f$ . Was sind *hinreichende Bedingungen* dafür, dass  $f$  am Punkt  $x \in \Omega$  zum Beispiel ein Minimum hat?

**Satz 4.4 (Hinreichende Bedingung für lokales Minimum)** Sei  $x \in \Omega$  kritischer Punkt der Funktion  $f \in C^2(\Omega)$ . Ist  $D^2 f(x)$  positiv definit, so hat  $f$  an der Stelle  $x$  ein lokales, isoliertes Minimum.

Um diese Aussage zu zeigen, brauchen wir eine lokale Entwicklung von  $f$ .

**Satz 4.5 (Taylor-Approximation)** Für  $f \in C^2(\Omega)$  sei  $q(x)$  das Taylorpolynom zweiter Ordnung im Entwicklungspunkt  $x_0 \in \Omega$ , das heißt

$$q(x) = f(x_0) + Df(x_0)(x - x_0) + \frac{1}{2} D^2 f(x_0)(x - x_0, x - x_0).$$

Dann folgt

$$\lim_{x \rightarrow x_0} \frac{f(x) - q(x)}{|x - x_0|^2} = 0.$$

BEWEIS: Wir betrachten die eindimensionale Funktion

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

Die Ableitungen der Funktion  $\varphi$  lauten

$$\varphi'(t) = Df(x_0 + tv)v \quad \text{und} \quad \varphi''(t) = D^2f(x_0 + tv)(v, v).$$

Nach der Taylorschen Formel, siehe Satz 4.1, gibt es ein  $\tau \in [0, 1]$  mit

$$\varphi(1) = \varphi(0) + \varphi'(0) + \frac{1}{2}\varphi''(\tau),$$

beziehungsweise mit  $\xi = (1 - \tau)x_0 + \tau x$

$$\begin{aligned} f(x) &= f(x_0) + Df(x_0)(x - x_0) + \frac{1}{2}D^2f(\xi)(x - x_0, x - x_0) \\ &= q(x) + \frac{1}{2}(D^2f(\xi) - D^2f(x_0))(x - x_0, x - x_0). \end{aligned}$$

Es folgt

$$\frac{f(x) - q(x)}{|x - x_0|^2} = \frac{1}{2}(D^2f(\xi) - D^2f(x_0))\left(\frac{x - x_0}{|x - x_0|}, \frac{x - x_0}{|x - x_0|}\right).$$

Mit  $x \rightarrow x_0$  geht  $D^2f(\xi)$  gegen  $D^2f(x_0)$ , also geht die rechte Seite gegen Null.  $\square$

In der Landau-Notation bedeutet das  $f(x) - q(x) = o(|x - x_0|^2)$ , oder

$$f(x) = f(x_0) + Df(x_0)(x - x_0) + \frac{1}{2}D^2f(x_0)(x - x_0, x - x_0) + o(|x - x_0|^2) \quad \text{für } x \rightarrow x_0.$$

Allgemeiner kann für eine  $C^k$ -Funktion eine Taylorentwicklung der Ordnung  $k$  hergeleitet werden, allerdings erfordert das einen gewissen Notationsaufwand.

Wir zeigen jetzt das hinreichende Kriterium von Satz 4.4. Sei  $Df(x_0) = 0$ , also  $x_0$  kritischer Punkt, und  $D^2f(x_0)$  sei positiv definit. Wir wenden die Hauptachsentransformation auf die symmetrische Matrix  $D^2f(x_0)$  bzw. die zugehörige quadratische Form an. Genauer wurde in Satz 7.3 gezeigt: es gibt einen Vektor  $v_1 \in \mathbb{R}^n$  mit  $|v_1| = 1$ , so dass gilt:

$$D^2f(x_0)(v_1, v_1) = \inf\{D^2f(x_0)(v, v) : |v| = 1\} =: \lambda_1.$$

Da  $D^2f(x_0)$  positiv definit, ist  $\lambda_1 > 0$  und damit gilt für alle  $v \neq 0$

$$D^2f(x_0)(v, v) = D^2f(x_0)\left(\frac{v}{|v|}, \frac{v}{|v|}\right) |v|^2 \geq \lambda_1 |v|^2.$$

Es folgt wegen  $Df(x_0) = 0$

$$\begin{aligned} f(x) - f(x_0) &= \frac{1}{2}D^2f(x_0)(x - x_0, x - x_0) + f(x) - q(x) \\ &\geq \left(\frac{\lambda_1}{2} + \frac{f(x) - q(x)}{|x - x_0|^2}\right) |x - x_0|^2 \\ &\geq \frac{\lambda_1}{4} |x - x_0|^2 > 0, \end{aligned}$$

für  $x \neq x_0$  hinreichend nahe bei  $x_0$ .

## 5 Parameterabhängige Integrale

In diesem Abschnitt geht es um Integrale, die von einem Parameter abhängen, und die Differentiation unter dem Integralzeichen. Solche Integrale treten in vielen Anwendungen auf, als Beispiel sei die Fouriertransformierte genannt:

$$\hat{f}(p) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(x) e^{-ipx} dx.$$

Der Ehrlichkeit halber sei gesagt, dass wir dieses Beispiel hier nicht behandeln werden: wir betrachten nur Integrale auf einem endlichen Intervall, um nicht mit zusätzlichen Konvergenzproblemen kämpfen zu müssen. Die Differentiation unter dem Integralzeichen ist auch bei vielen uneigentlichen Integralen möglich, für hinreichende Kriterien verweisen wir auf die Literatur. Unsere Version mit endlichen Intervallen hat bereits interessante Anwendungen, unter anderem werden wir die Euler-Lagrange Gleichung für eindimensionale Variationsintegrale herleiten.

Sei  $D \subset \mathbb{R}$  ein offenes (Parameter-)Intervall und  $I = [a, b]$  ein kompaktes Intervall. Für eine gegebene Funktion  $f : D \times I \rightarrow \mathbb{R}$ ,  $f = f(x, y)$ , betrachten wir die neue Funktion

$$\phi : D \rightarrow \mathbb{R}, \phi(x) = \int_I f(x, y) dy. \quad (5.2)$$

Diese Funktion nennt man ein parameterabhängiges Integral. Damit  $\phi$  wohldefiniert ist, müssen die Integrale existieren. Das ist garantiert, wenn für alle  $x \in D$  die Funktion  $f(x, \cdot) : I \rightarrow \mathbb{R}$ ,  $y \mapsto f(x, y)$ , stetig ist. Wir interessieren uns für die Stetigkeit und die Ableitung der Funktion  $\phi(x)$ . Hierfür brauchen wir folgende Aussage.

**Satz 5.1 (gleichmäßige Stetigkeit)** Sei  $f : K \rightarrow \mathbb{R}$  stetig, wobei  $K \subset \mathbb{R}^n$  abgeschlossen und beschränkt. Dann ist  $f$  gleichmäßig stetig, das heißt es gilt

$$\text{osc}_f(\delta) = \sup_{x, x' \in K, |x-x'| < \delta} |f(x) - f(x')| \rightarrow 0 \quad \text{mit } \delta \searrow 0.$$

Die Funktion  $\text{osc}_f(\delta)$  heißt Oszillationsfunktion von  $f$ .

Dies kann durch Widerspruch gezeigt werden. Da die Information eher technischer Natur ist, wollen wir hier darauf verzichten. Stattdessen kommen wir gleich zum ersten Resultat.

**Satz 5.2 (Stetigkeit von Parameterintegralen)** Sei  $f \in C^0(D \times I)$ ,  $f = f(x, y)$ , wobei  $D \subset \mathbb{R}^n$  offen und  $I = [a, b]$ . Dann ist die Funktion

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

stetig.

BEWEIS: Sei erst  $D$  abgeschlossen und beschränkt. Dann gilt für  $x, x' \in D$  mit  $|x - x'| < \delta$

$$|\phi(x') - \phi(x)| = \left| \int_I (f(x', y) - f(x, y)) dy \right| \leq (b - a) \text{osc}_f(\delta).$$

Die rechte Seite geht mit  $\delta \searrow 0$  gegen Null. Ist  $D$  offen, so können wir  $D$  durch abgeschlossene und beschränkte Intervalle ausschöpfen, also ist  $\phi(x)$  stetig auf  $D$ .  $\square$

Wir gehen direkt weiter zur Differenzierbarkeit und Berechnung der Ableitung.

**Satz 5.3 (Differentiation unter dem Integral)** Sei  $D \subset \mathbb{R}$  offen und  $I = [a, b]$ . Für  $f \in C^0(D \times I)$  setze  $\phi : D \rightarrow \mathbb{R}$ ,  $\phi(x) = \int_I f(x, y) dy$ . Ist  $\frac{\partial f}{\partial x} \in C^0(D \times I)$ , so folgt

$$\frac{d\phi}{dx}(x) = \int_I \frac{\partial f}{\partial x}(x, y) dy,$$

und es gilt  $\phi \in C^1(D)$ .

BEWEIS: Wir nehmen wieder  $D$  abgeschlossen und beschränkt an. Nach dem Hauptsatz der Differential- und Integralrechnung gilt

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

Berechne nun

$$\begin{aligned} \left| \frac{\phi(x+h) - \phi(x)}{h} - \int_I \frac{\partial f}{\partial x}(x, y) dy \right| &= \left| \int_I \left( \frac{f(x+h, y) - f(x, y)}{h} - \frac{\partial f}{\partial x}(x, y) \right) dy \right| \\ &= \left| \int_I \int_0^1 \left( \frac{\partial f}{\partial x}(x+th, y) - \frac{\partial f}{\partial x}(x, y) \right) dt dy \right| \\ &\leq \int_I \int_0^1 \left| \frac{\partial f}{\partial x}(x+th, y) - \frac{\partial f}{\partial x}(x, y) \right| dt dy \\ &\leq (b-a) \operatorname{osc}_{\frac{\partial f}{\partial x}}(|h|). \end{aligned}$$

Mit  $h \rightarrow 0$  geht die rechte Seite gegen Null, also ist die Formel gezeigt. Die Stetigkeit der Ableitung folgt aus Satz 5.2.  $\square$

Als Warnung betrachten wir für  $f : [a, b] \rightarrow \mathbb{R}$  stetig das Integral

$$\phi(x) = \frac{1}{2} \int_a^b \operatorname{sign}(x-y) f(y) dy.$$

Es ist  $\frac{\partial}{\partial x} \operatorname{sign}(x-y) = \operatorname{sign}'(x-y) = 0$ , außer wenn  $x = y$ . Würden wir diese Ausnahmestelle ignorieren und naiv unter dem Integral differenzieren, so wäre

$$\phi'(x) = \frac{1}{2} \int_a^b \operatorname{sign}'(x-y) f(y) dy = 0 \quad \text{für } x \in (a, b).$$

Das ist aber falsch, durch Aufspalten des Integrals sieht man

$$\phi'(x) = \frac{1}{2} \frac{d}{dx} \left( \int_a^x f(y) dy - \int_x^b f(y) dy \right) = f(x).$$

Formal gilt also die Gleichung

$$f(x) = \int_a^b \operatorname{sign}'(x-y) f(y) dy \quad \text{für } x \in (a, b).$$

Das wird oft in der Form  $\operatorname{sign}'(x-y) = \delta(x-y)$  notiert. Dabei soll  $\delta : \mathbb{R} \rightarrow \mathbb{R}$  eine Funktion sein mit  $\delta(x) = 0$  für  $x \neq 0$ , andererseits soll aber gelten

$$\int_{-\infty}^{\infty} \delta(x) f(x) dx = f(0) \quad \text{sofern } f(x) = 0 \text{ für } |x| \text{ groß.}$$

Eine solche Funktion  $\delta(x)$  kann es aber nicht geben. Um der Notation Sinn zu verleihen, muss man den Raum der Funktionen verlassen und Distributionen einführen. Die Notation kann nützlich sein, wenn man diesen Hintergrund versteht.

Eine nützliche Anwendung der Parameterableitung ist die Vertauschbarkeit der Integrationsreihenfolge in iterierten Integralen.

**Satz 5.4 (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 5.2 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 5.3 folgt

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

□

Jetzt kommen wir zur Variationsrechnung. Auf der Menge aller Kurven  $x : I = [a, b] \rightarrow \Omega$ ,  $x = x(t)$ , betrachten wir das Funktional

$$\mathcal{F} : C^1(I, \Omega) \rightarrow \mathbb{R}, \quad \mathcal{F}(x) = \int_I f(t, x(t), x'(t)) dt.$$

$\mathcal{F}$  ist keine Funktion von endlich vielen Variablen, sondern ist auf dem Raum  $C^1(I, \Omega)$  definiert, darum spricht man auch von einem *Funktional* statt einer Funktion. Die Bezeichnung *Variationsintegral* ist ebenfalls üblich. Die Funktion

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

heißt Lagrangefunktion. Durch sie ist das Funktional  $\mathcal{F}$  definiert.

**Beispiel 5.1** Die Bogenlänge von  $x \in C^1(I, \Omega)$  ist

$$\mathcal{F}(x) = \int_I |x'(t)| dt.$$

Für die Bogenlänge gilt also  $f(t, x, v) = |v|$ .

**Beispiel 5.2** Ist der Boden von unterschiedlicher Qualität, so hängt die Laufzeit nicht nur von der Länge des Wegs ab. Dies modelliert das Integral

$$\mathcal{F}(c) = \int_I \omega(x(t)) |x'(t)| dt.$$

Hier ist also  $f(t, x, v) = \omega(x)|v|$ . Dasselbe Integral tritt auch bei der Lichtausbreitung in Medien mit nichtkonstanter optischer Dichte auf.

**Beispiel 5.3** Es sei  $x = x(t)$  die Bahn eines Teilchens der Masse  $m$  in einem Kraftfeld mit Potential  $V : \Omega \rightarrow \mathbb{R}$ . Zur Zeit  $t$  hat das Teilchen dann die kinetische Energie  $\frac{m}{2}|x'(t)|^2$  und die potentielle Energie  $V(x(t))$ . Das sogenannte Wirkungsintegral ist

$$\mathcal{F}(x) = \int_I \left( \frac{m}{2} |x'(t)|^2 - V(x(t)) \right) dt.$$

Die zugehörige Lagrangefunktion ist  $f(t, x, v) = \frac{m}{2}|v|^2 - V(x)$ .

Besonderes Interesse gilt natürlich den optimalen Kurven oder Bahnen, die das Funktional  $\mathcal{F}(x)$  minimieren oder maximieren gegenüber allen Vergleichskurven. Unser Ziel ist es, diese Kurven zu charakterisieren. Wie bei der Diskussion der Extrema zeigen wir allerdings nur eine notwendige Bedingung. Dazu betrachten wir eine Schar (oder: *Variation*) von Kurven

$$x \in C^2((-\varepsilon_0, \varepsilon_0) \times I, \mathbb{R}^n), \quad x = x(\varepsilon, t) \quad \text{mit } x(0, t) = x(t). \quad (5.3)$$

Weiter definieren wir das Vektorfeld der Variation durch

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

**Lemma 5.1 (Erste Variation)** Sei  $f \in C^2(I \times \Omega \times \mathbb{R}^n)$ ,  $f = f(t, x, v)$ , wobei  $I = [a, b]$  und  $\Omega \subset \mathbb{R}^n$  offen. Definiere für  $x = x(\varepsilon, t)$  wie in (5.3) die Funktion

$$\phi(\varepsilon) = \mathcal{F}(x(\varepsilon, \cdot)) = \int_I f(t, x(\varepsilon, t), \frac{\partial x}{\partial t}(\varepsilon, t)) dt.$$

Dann gilt die Formel

$$\frac{d\phi}{d\varepsilon} \Big|_{\varepsilon=0} = \int_I \sum_{i=1}^n L_f(x)_i \varphi_i dt + \left[ \sum_{i=1}^n \frac{\partial f}{\partial v_i}(t, x, x') \varphi_i \right]_{t=a}^{t=b}. \quad (5.4)$$

Dabei ist  $L_f(x) : I \rightarrow \mathbb{R}^n$  gegeben durch

$$L_f(x)_i = \frac{\partial f}{\partial x_i}(t, x, x') - \frac{d}{dt} \left[ \frac{\partial f}{\partial v_i}(t, x, x') \right].$$

BEWEIS: Differentiation unter dem Integral und die Kettenregel liefern

$$\frac{d\phi}{d\varepsilon} \Big|_{\varepsilon=0} = \int_I \sum_{i=1}^n \left( \frac{\partial f}{\partial x_i}(t, x, x') \frac{\partial x_i}{\partial \varepsilon}(0, t) + \frac{\partial f}{\partial v_i}(t, x, x') \frac{\partial^2 x_i}{\partial \varepsilon \partial t}(0, t) \right) dt.$$

Nun gilt

$$\frac{\partial x}{\partial \varepsilon}(0, t) = \varphi(t) \quad \text{und} \quad \frac{\partial^2 x}{\partial \varepsilon \partial t}(0, t) = \frac{\partial^2 x}{\partial t \partial \varepsilon}(0, t) = \varphi'(t).$$

Durch partielle Integration des hinteren Terms folgt nun

$$\frac{d\phi}{d\varepsilon}\Big|_{\varepsilon=0} = \int_I \sum_{i=1}^n \left( \frac{\partial f}{\partial x_i}(t, x, x') - \frac{d}{dt} \left[ \frac{\partial f}{\partial v_i}(t, x, x') \right] \right) \varphi_i + \left[ \sum_{i=1}^n \frac{\partial f}{\partial v_i}(t, x, x') \varphi_i \right]_{t=a}^{t=b}.$$

Damit ist das Lemma bewiesen.  $\square$

Im folgenden Satz besagt die Voraussetzung, dass die Richtungsableitung des Funktionals  $\mathcal{F}(x)$  an der Stelle  $x$  in Richtung von  $\varphi$  gleich Null ist, für alle  $\varphi$  wie beschrieben. Mit andern Worten können wir  $x$  als kritischen Punkt des Funktionals  $\mathcal{F}(x)$  betrachten.

**Satz 5.5 (Euler-Lagrange-Gleichung)** Sei  $\mathcal{F}(x) = \int_a^b f(t, x, x')$  wie in Lemma 5.1, und

$$\frac{d}{d\varepsilon} \mathcal{F}(x + \varepsilon\varphi)\Big|_{\varepsilon=0} = 0 \quad \text{für alle } \varphi \in C^2(I, \mathbb{R}^n) \text{ mit } \varphi(a) = \varphi(b) = 0.$$

Dann ist  $x(t)$  eine Lösung der Euler-Lagrange-Gleichungen

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

BEWEIS: Wähle  $\varphi = \eta e_i$  mit  $\eta \in C^2(I)$  reellwertig,  $\eta(a) = \eta(b) = 0$ . Die Randterme in Lemma 5.1 verschwinden, da  $\varphi(a) = \varphi(b) = 0$ . Also gilt für alle diese  $\eta$

$$\int_I L_f(x)_i \eta dt = 0 \quad \text{für alle } \eta \in C^2(I) \text{ mit } \eta(a) = \eta(b) = 0.$$

Hieraus folgt  $L_f(x)_i = 0$ .  $\square$

Der Beweis für den letzten Schluss ist nicht schwierig, wir lassen ihn dennoch aus. Stattdessen interpretieren wir die Aussage mit dem Integral-Skalarprodukt

$$\int_I L_f(x)_i \eta dt = \langle L_f(x)_i, \eta \rangle.$$

Wir haben die Information, dass  $L_f(x)_i$  auf alle  $\eta$  senkrecht steht. Der Schluss besagt, dass dies nur die Nullfunktion sein kann.

**Beispiel 5.4** Für die Bogenlänge gilt  $f(t, x, v) = |v|$ , also  $\frac{\partial f}{\partial x_i} = 0$  und  $\frac{\partial f}{\partial v_i} = \frac{v_i}{|v|}$  für  $v \neq 0$ . Damit ist die Euler-Lagrange-Gleichung

$$L_f(x) = -\frac{d}{dt} \frac{x'}{|x|} = 0.$$

Wir müssen dabei  $x'(t) \neq 0$  für alle  $t \in I$  voraussetzen. Die Gleichung sagt aus, dass der Einheitsstangentenvektor von  $x(t)$  konstant ist. Dies bedeutet, dass  $x(t)$  eine gerade Strecke parametrisiert.

**Beispiel 5.5** Für das Wirkungsintegral  $f(t, x, v) = \frac{m}{2}|v|^2 - V(x)$  haben wir  $\frac{\partial f}{\partial v_i} = mv_i$  und

$$\frac{\partial f}{\partial x_i}(t, x, v) = -\frac{\partial V}{\partial x_i}(x) =: F_i(x).$$

Das Vektorfeld  $F(x) = -\text{grad } V(x)$  ist das Kraftfeld des Potentials, das Minuszeichen ist in der Physik üblich. Die Euler-Lagrange Gleichung lautet nun

$$L_f(x) = F(x) - \frac{d}{dt}(mx') = F(x) - mx'' = 0.$$

Dies sind die Newtonschen Bewegungsgleichungen.

## 6 Kurvenintegrale

**Definition 6.1** Auf der offenen Menge  $\Omega \subset \mathbb{R}^n$  sei ein stetiges Vektorfeld  $F : \Omega \rightarrow \mathbb{R}^n$  gegeben. Das Kurvenintegral von  $F$  längs einer  $C^1$ -Kurve  $\gamma : [a, b] \rightarrow \Omega$  ist

$$\int_{\gamma} F \cdot \vec{ds} := \int_a^b F(\gamma(t)) \cdot \gamma'(t) dt.$$

In der Physik ist zum Beispiel  $F$  das Gravitationsfeld, und das Kurvenintegral ergibt die Arbeit, die beim Transport einer Masse längs einer Kurve innerhalb des Feldes verrichtet wird. Wir schreiben hier  $F \cdot \gamma'$  für das Standard-Skalarprodukt, das ist die übliche Notation. Der Ausdruck  $\vec{ds}$  wird als vektorielles Längenelement bezeichnet, jedoch ist das nur eine formale Notation; das Integral ist durch die rechte Seite definiert. Immerhin ist die Merkregel nützlich, dass  $\vec{ds}$  durch  $\gamma'(t) dt$  zu ersetzen ist.

**Beispiel 6.1 (Gravitationsfeld)** Das Gravitationsfeld der Sonne ist gegeben durch

$$F : \mathbb{R}^3 \setminus \{0\} \rightarrow \mathbb{R}^3, F(x) = -\frac{Cx}{|x|^3} \quad \text{mit } C > 0.$$

Längs einer Bahn  $x(t)$ ,  $t \in [a, b]$ , wird folgende Arbeit verrichtet:

$$E(\gamma) = -\int_a^b \frac{Cx(t)}{|x(t)|^3} \cdot x'(t) dt = \int_a^b \frac{d}{dt} \frac{C}{|x(t)|} = \left( \frac{C}{|x(b)|} - \frac{C}{|x(a)|} \right).$$

**Beispiel 6.2 (Winkelvektorfeld I)** Betrachte das Vektorfeld

$$W : \mathbb{R}^2 \setminus \{0\} \rightarrow \mathbb{R}^2, W(x, y) = \left( -\frac{y}{x^2 + y^2}, \frac{x}{x^2 + y^2} \right).$$

Ist  $\gamma(t) = r(t)(\cos \theta(t), \sin \theta(t))$ ,  $t \in [a, b]$ , mit  $C^1$ -Funktionen  $r(t) > 0$  und  $\theta(t)$ , so gilt

$$\int_{\gamma} W \cdot \vec{ds} = \int_a^b \frac{1}{r} \begin{pmatrix} -\sin \theta \\ \cos \theta \end{pmatrix} \cdot \left( r' \begin{pmatrix} \cos \theta \\ \sin \theta \end{pmatrix} + r\theta' \begin{pmatrix} -\sin \theta \\ \cos \theta \end{pmatrix} \right) dt = \theta(b) - \theta(a). \quad (6.5)$$

Also liefert das Kurvenintegral den umlaufenen Winkel.

Wir wollen einige Eigenschaften des Kurvenintegrals zusammenstellen. Direkt aus der Definition folgt die Linearität

$$\int_{\gamma} (\lambda_1 F_1 + \lambda_2 F_2) \cdot \vec{ds} = \lambda_1 \int_{\gamma} F_1 \cdot \vec{ds} + \lambda_2 \int_{\gamma} F_2 \cdot \vec{ds}. \quad (6.6)$$

Als zweites haben wir die Additivität unter Zerlegungen: ist  $a = t_0 < t_1 < \dots < t_N = b$ , so gilt für  $\gamma_i = \gamma|_{[t_{i-1}, t_i]}$

$$\int_{\gamma} F \cdot \vec{ds} = \sum_{i=1}^N \int_{\gamma_i} F \cdot \vec{ds}. \quad (6.7)$$

Insbesondere ist das Kurvenintegral auch für stückweise  $C^1$ -Kurven definiert, zum Beispiel wenn  $C^1$ -Kurven aneinander gehängt werden.

Die dritte Eigenschaft ist die Invarianz unter Umparametrisierungen. Sei  $\gamma : [a, b] \rightarrow \Omega$ ,  $\gamma = \gamma(t)$ , eine gegebene  $C^1$ -Kurve. Durch die Substitution  $t = \tau(u)$  mit  $\tau : [\alpha, \beta] \rightarrow [a, b]$  bijektiv erhalten wir die Kurve

$$\tilde{\gamma} : [\alpha, \beta] \rightarrow \Omega, \quad \tilde{\gamma}(u) = \gamma(\tau(u)).$$

Wir berechnen nun mit der Kettenregel und Substitution  $\tau(u) = t$

$$\begin{aligned} \int_{\tilde{\gamma}} F \cdot \vec{ds} &= \int_{\alpha}^{\beta} F(\tilde{\gamma}(u)) \cdot \tilde{\gamma}'(u) du \\ &= \int_{\alpha}^{\beta} F(\gamma(\tau(u))) \cdot \gamma'(\tau(u)) \tau'(u) du \\ &= \int_{\tau(\alpha)}^{\tau(\beta)} F(\gamma(t)) \cdot \gamma'(t) dt. \end{aligned}$$

$\tau$  heißt orientierungserhaltend, wenn  $\tau(\alpha) = a$  und  $\tau(\beta) = b$ , andernfalls heißt  $\tau$  orientierungsumkehrend. Es gilt also

$$\int_{\gamma \circ \tau} F \cdot \vec{ds} = \pm \int_{\gamma} F \cdot \vec{ds}, \quad \text{falls } \tau \text{ orientierungserhaltend (-umkehrend)}. \quad (6.8)$$

Die Physiker nennen das Gravitationsfeld konservativ, weil der Energieerhaltungssatz gilt. Die verrichtete Arbeit zum Beispiel bei Transport einer Masse von der Mensa zum Schauinsland entspricht genau der zugewonnenen Lageenergie, und diese kommt beim Herunterrollen auch wieder heraus, theoretisch jedenfalls. Der Begriff des konservativen Feldes ist auch in der Mathematik interessant.

**Definition 6.2** Sei  $\Omega \subset \mathbb{R}^n$  offen. Ein Vektorfeld  $F \in C^0(\Omega, \mathbb{R}^n)$  heißt Gradientenfeld (bzw. konservativ), wenn es eine Funktion  $\varphi \in C^1(\Omega)$  gibt mit

$$\text{grad } \varphi = F.$$

Diese Funktion heißt Potentialfunktion von  $F$ .

Ist  $\Omega \subset \mathbb{R}^n$  wegwiese zusammenhängend, so unterscheiden sich zwei Potentialfunktionen  $\varphi_{1,2}$  von  $F$  nur um eine additive Konstante, denn es gilt

$$\text{grad}(\varphi_1 - \varphi_2) = F - F = 0.$$

Die Kenntnis einer Potentialfunktion ist sehr nützlich, denn damit kann das Kurvenintegral mühelos berechnet werden. Ist  $\varphi \in C^1(\Omega)$  Potentialfunktion von  $F(x)$ , so gilt für eine  $C^1$ -Kurve  $\gamma : [a, b] \rightarrow \Omega$  nach der Kettenregel

$$\begin{aligned} \int_{\gamma} F \cdot \vec{ds} &= \int_a^b F(\gamma(t)) \cdot \gamma'(t) dt = \int_a^b \text{grad} \varphi(\gamma(t)) \cdot \gamma'(t) dt \\ &= \int_a^b \frac{d}{dt} \varphi(\gamma(t)) dt = \varphi(\gamma(b)) - \varphi(\gamma(a)). \end{aligned} \quad (6.9)$$

Für alle Kurven mit demselben Anfangs- und Endpunkt liefert das Kurvenintegral also denselben Wert. Die Frage ist nun, wann diese günstige Situation vorliegt, mit anderen Worten:

*Welche Vektorfelder  $F : \Omega \rightarrow \mathbb{R}^n$  haben eine Potentialfunktion?*

Für  $n = 1$  ist eine Potentialfunktion nichts anderes als eine Stammfunktion, also  $\varphi'(x) = F(x)$ . Nach dem Hauptsatz der Differential- und Integralrechnung können wir eine solche Stammfunktion immer finden. Nicht so für  $n \geq 2$ !

**Satz 6.1 (Rotationsfreiheit)** *Für ein Gradientenfeld  $F \in C^1(\Omega, \mathbb{R}^n)$  gilt*

$$\partial_i F_j = \partial_j F_i \quad \text{für alle } i, j = 1, \dots, n.$$

BEWEIS: Ist  $F = \text{grad} \varphi$ , so folgt  $\varphi \in C^2(\Omega)$  und wegen der Vertauschbarkeit der partiellen Ableitungen, Satz 2.2, gilt

$$\partial_i F_j = \partial_i \partial_j \varphi = \partial_j \partial_i \varphi = \partial_j F_i.$$

□

*Bemerkung.* Für  $n = 3$  lässt sich diese Bedingung schreiben als

$$\text{rot } F = (\partial_2 F_3 - \partial_3 F_2, \partial_3 F_1 - \partial_1 F_3, \partial_1 F_2 - \partial_2 F_1) = 0.$$

**Beispiel 6.3**  $F : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ ,  $F(x, y) = (-y, x)$ , hat keine Stammfunktion, denn es ist

$$\partial_1 F_2 = 1, \text{ aber } \partial_2 F_1 = -1.$$

**Beispiel 6.4 (Winkelvektorfeld II)** Für das Vektorfeld

$$W : \mathbb{R}^2 \setminus \{0\} \rightarrow \mathbb{R}^2, W(x, y) = \left( -\frac{y}{x^2 + y^2}, \frac{x}{x^2 + y^2} \right)$$

berechnen wir

$$\partial_1 W_2 = \frac{y^2 - x^2}{(x^2 + y^2)^2} = \partial_2 W_1,$$

das heißt die notwendige Bedingung aus Satz 6.1 ist erfüllt. Nach Beispiel 6.2 gilt jedoch für die Kurven  $\gamma_k : [0, 2\pi] \rightarrow \mathbb{R}^2 \setminus \{0\}$ ,  $\gamma_k(t) = r(\cos kt, \sin kt)$ ,

$$\int_{\gamma_k} W \cdot \vec{ds} = 2\pi k \quad (\neq 0 \text{ für } k \in \mathbb{Z} \setminus \{0\}).$$

Hätte  $W$  eine Potentialfunktion  $\varphi$  auf  $\mathbb{R}^2 \setminus \{0\}$ , so müsste gelten

$$\int_{\gamma_k} W \cdot \vec{ds} = \varphi(\gamma(2\pi k)) - \varphi(\gamma(0)) = 0.$$

Das Beispiel zeigt, dass die Rotationsfreiheit aus Satz 6.1 nicht immer ausreicht, um die Existenz eines Potentials zu garantieren. Wir werden sehen, dass es zusätzlich auf die Geometrie des Gebiets  $\Omega$  ankommt. Zunächst untersuchen wir, wie sich das Integral längs einer Schar von Kurven verhält. Statt *Schar* ist der modernere Ausdruck *Homotopie* üblich.

**Satz 6.2 (Homotopieformel)** Sei  $F \in C^1(\Omega, \mathbb{R}^n)$  und  $\gamma \in C^2([a, b] \times [0, 1], \Omega)$ ,  $\gamma = \gamma(s, t)$ . Dann gilt

$$\begin{aligned} \int_{\gamma(\cdot, 1)} F \cdot \vec{ds} - \int_{\gamma(\cdot, 0)} F \cdot \vec{ds} &= \int_{\gamma(b, \cdot)} F \cdot \vec{ds} - \int_{\gamma(a, \cdot)} F \cdot \vec{ds} \\ &\quad - \int_0^1 \int_a^b \sum_{i,j=1}^n (\partial_i F_j - \partial_j F_i)(\gamma) \frac{\partial \gamma^i}{\partial s} \frac{\partial \gamma^j}{\partial t} ds dt. \end{aligned}$$

BEWEIS: Das Kurvenintegral ist ein Variationsintegral mit der Euler-Lagrange Funktion

$$f(x, v) = F(x) \cdot v = \sum_{i=1}^n F_i(x) v_i.$$

Wir berechnen den zugehörigen Euler-Lagrange Operator, wobei  $' = \frac{\partial}{\partial s}$ .

$$\begin{aligned} L_f(\gamma)_j &= \frac{\partial f}{\partial x_j}(\gamma, \gamma') - \frac{d}{ds} \left[ \frac{\partial f}{\partial v_j}(\gamma, \gamma') \right] \\ &= \sum_{i=1}^n \frac{\partial F_i}{\partial x_j}(\gamma) \gamma'_i - \frac{d}{ds} F_j(\gamma) \\ &= \sum_{i=1}^n \left( \frac{\partial F_i}{\partial x_j}(\gamma) - \frac{\partial F_j}{\partial x_i}(\gamma) \right) \gamma'_i. \end{aligned}$$

Einsetzen des Ergebnisses in die Formel für die erste Variation ergibt

$$\begin{aligned} \frac{d}{dt} \int_{\gamma(\cdot, t)} F \cdot \vec{ds} &= \left[ \sum_{i=1}^n F_i(\gamma) \frac{\partial \gamma_i}{\partial t} \right]_{s=a}^{s=b} \\ &\quad + \int_a^b \sum_{i,j=1}^n \left( \frac{\partial F_i}{\partial x_j}(\gamma) - \frac{\partial F_j}{\partial x_i}(\gamma) \right) \frac{\partial \gamma_i}{\partial s} \frac{\partial \gamma_j}{\partial t} ds. \end{aligned}$$

Die Behauptung folgt nun durch Integration bezüglich  $t \in [0, 1]$ . □

**Folgerung 6.1 (Homotopieinvarianz)** Sei  $\partial_i F_j = \partial_j F_i$  für  $i, j = 1, \dots, n$  in Satz 6.2, und  $\gamma(a, t) = p$  und  $\gamma(b, t) = q$  für alle  $t \in [0, 1]$  (fester Anfangs- und Endpunkt). Dann folgt

$$\int_{\gamma(\cdot, 1)} F \cdot \vec{ds} = \int_{\gamma(\cdot, 0)} F \cdot \vec{ds}.$$

BEWEIS: Aufgrund der Voraussetzungen ist die rechte Seite in Satz 6.2 gleich Null.  $\square$

Wir kommen nun zurück auf die Geometrie des Gebiets.  $\Omega \subset \mathbb{R}^n$  heißt *einfach zusammenhängend*, wenn  $\Omega$  zusammenhängend ist und wenn gilt: sind  $\gamma_{0,1} : [a, b] \rightarrow \Omega$  zwei stetige Kurven von  $p$  nach  $q$ , so gibt es eine stetige Schar von Kurven  $\gamma(\cdot, t) : [a, b] \rightarrow \Omega$ ,  $t \in [0, 1]$ , von  $p$  nach  $q$  mit  $\gamma(\cdot, 0) = \gamma_0$  und  $\gamma(\cdot, 1) = \gamma_1$ . Es ist einfach zu sehen, dass ein konvexes Gebiet einfach zusammenhängend ist.

**Lemma 6.1** Sei  $\Omega \subset \mathbb{R}^n$  einfach zusammenhängend, und  $F \in C^1(\Omega, \mathbb{R}^n)$  erfülle  $\partial_i F_j = \partial_j F_i$ . Sind  $\gamma_0, \gamma_1 : [a, b] \rightarrow \Omega$  Kurven mit dem gleichem Anfangs- und Endpunkt, so gilt

$$\int_{\gamma_1} F \cdot \vec{ds} = \int_{\gamma_0} F \cdot \vec{ds}.$$

BEWEIS: Nach Voraussetzung gibt es eine Schar von Kurven  $\gamma(\cdot, t)$  wie oben beschrieben. Aus Folgerung 6.1 folgt dann die Behauptung. Es gibt dabei ein technisches Problem mit der Regularität der Homotopie, wir brauchen  $C^2$ . Aber das kann geregelt werden.  $\square$

**Satz 6.3 (Existenz eines Potentials)** Sei  $\Omega \subset \mathbb{R}^n$  einfach zusammenhängendes Gebiet. Erfüllt  $F \in C^1(\Omega, \mathbb{R}^n)$  die Integrabilitätsbedingung

$$\partial_i F_j = \partial_j F_i \quad \text{für alle } i, j = 1, \dots, n,$$

so hat  $F$  eine Potentialfunktion  $\varphi \in C^2(\Omega)$ .

BEWEIS: Sei  $x_0 \in \Omega$  fest. Zu  $x \in \Omega$  wählen wir eine Kurve  $\gamma_x \in C^2([0, 1], \Omega)$  mit  $\gamma_x(0) = x_0$  und  $\gamma_x(1) = x$ , und setzen

$$\varphi(x) = \int_{\gamma_x} F \cdot \vec{ds}.$$

Nach Lemma 6.1 hängt der Wert  $\varphi(x)$  nicht von der Wahl der Verbindungskurve  $\gamma_x$  ab, die Funktion  $\varphi$  ist definiert. Sei  $x \in \Omega$  und  $h \in \mathbb{R}$  klein. Wir erhalten eine Kurve von  $x_0$  nach  $x + he_j$ , indem wir  $\gamma_x$  mit der Kurve  $c(t) = x + the_j$ ,  $t \in [0, 1]$ , zusammensetzen. Es folgt

$$\frac{\varphi(x + he_j) - \varphi(x)}{h} = \int_c F \cdot \vec{ds} = \int_0^1 F(x + the_j) \cdot e_j dt \rightarrow F_j(x) \quad \text{mit } h \rightarrow 0.$$

Also gilt  $\partial_j \varphi = F_j$  für  $j = 1, \dots, n$ .  $\square$

**Beispiel 6.5 (Winkelvektorfeld III)** Nach Beispiel 6.2 erfüllt das Vektorfeld

$$W : \mathbb{R}^2 \setminus \{0\} \rightarrow \mathbb{R}^2, W(x, y) = \left( -\frac{y}{x^2 + y^2}, \frac{x}{x^2 + y^2} \right)$$

die notwendige Bedingung  $\partial_1 W_2 = \partial_2 W_1$ . In Beispiel 6.4 haben wir gesehen, dass  $W$  aber keine Potentialfunktion auf  $\mathbb{R}^2 \setminus \{0\}$  hat. Somit ist  $\mathbb{R}^2 \setminus \{0\}$  nicht einfach zusammenhängend. Aber zum Beispiel  $H = \{(x, y) \in \mathbb{R}^2 : y > 0\}$  ist konvex und damit einfach zusammenhängend, auf  $H$  muss es eine Potentialfunktion geben. Tatsächlich können wir sie angeben, und zwar

$$\varphi : H \rightarrow \mathbb{R}, \varphi(x, y) = \arccos \frac{x}{x^2 + y^2}.$$

Beim Kurvenintegral werden Vektorfelder integriert, aber natürlich kann man auch skalare Funktionen längs Kurven integrieren. Wir stellen uns vor, dass  $\gamma : [a, b] \rightarrow \Gamma \subset \mathbb{R}^3$  eine Parametrisierung des Kurvenstücks  $\Gamma \subset \mathbb{R}^3$  ist. Für eine gegebene Funktion  $\varrho : \Gamma \rightarrow \mathbb{R}$  setzen wir dann

$$\int_{\Gamma} \varrho ds := \int_a^b \varrho(\gamma(t)) |\gamma'(t)| dt. \quad (6.10)$$

Für  $\varrho = 1$  erhalten wir das Bogenlängenintegral, und wir interpretieren  $ds = |\gamma'(t)| dt$  als Bogenlängenelement von  $\gamma$ . Die Funktion  $\varrho(x)$  könnte zum Beispiel eine Massendichte oder Ladungsdichte (pro Längeneinheit) sein, das Integral ist dann die Gesamtmasse oder -ladung von  $\Gamma$ . Ist  $\tilde{\gamma}(u) = \gamma(\tau(u))$  eine orientierungserhaltende Umparametrisierung,  $\tau : [\alpha, \beta] \rightarrow [a, b]$ , so ist  $\tau'(u) > 0$  und es folgt

$$\begin{aligned} \int_{\tilde{\gamma}} \varrho ds &= \int_{\alpha}^{\beta} \varrho(\tilde{\gamma}(u)) |\tilde{\gamma}'(u)| du \\ &= \int_{\alpha}^{\beta} \varrho(\gamma(\tau(u))) |\gamma'(\tau(u))| |\tau'(u)| du \\ &= \int_a^b \varrho(\gamma(t)) |\gamma'(t)| dt. \end{aligned}$$

Ist  $t = \tau(u)$  orientierungsumkehrend, so drehen sich die Integrationsgrenzen um, andererseits ist  $\tau'(u) < 0$  und somit  $|\tau'(u)| du = -\tau'(u) du = -dt$ . Im Effekt bleibt das Integral also auch dann gleich, anders als beim Kurvenintegral eines Vektorfelds.

Sei nun auf  $\Gamma$  eine Orientierung und damit eine Einheitstangente  $\vec{T} : \Gamma \rightarrow \mathbb{R}^3$ ,  $|\vec{T}(x)| = 1$ , gegeben. Sei  $\gamma : [a, b] \rightarrow \Gamma$  eine Parametrisierung mit der gegebenen Orientierung. Wir können dann das Kurvenintegral des Vektorfelds  $F$  wie folgt umschreiben:

$$\int_{\gamma} F \cdot \vec{ds} = \int_a^b F(\gamma(t)) \cdot \frac{\gamma'(t)}{|\gamma'(t)|} |\gamma'(t)| dt = \int_a^b F(\gamma(t)) \cdot \vec{T}(\gamma(t)) |\gamma'(t)| dt = \int_{\Gamma} F \cdot \vec{T} ds.$$

Wir sehen deutlicher, dass im Kurvenintegral eines Vektorfelds  $F$  die tangentielle Komponente von  $F$  integriert wird.

Mit diesen Vorbereitungen wollen wir nun die Homotopieformel aus Satz 6.2 in etwas anderem Licht betrachten, und zwar im dreidimensionalen Fall. Wir stellen uns vor, dass die Abbildung  $\gamma(s, t)$  ein Flächenstück parametrisiert:

$$\gamma : [a, b] \times [0, 1] \rightarrow \Sigma \subset \mathbb{R}^3, \quad \gamma = \gamma(s, t).$$

Der Rand  $\partial\Sigma$  von  $\Sigma$  wird durch die vier Kurven  $\gamma(\cdot, 0)$ ,  $\gamma(b, \cdot)$ ,  $\gamma(\cdot, 1)$  und  $\gamma(a, \cdot)$  parametrisiert. Durchlaufen wir den Rand des Rechtecks entgegen dem Uhrzeigersinn, so werden die ersten beiden Kurven orientierungstreu durchlaufen, bei den letzten beiden dreht sich die Orientierung um. Wir erhalten folgende Darstellung:

$$\oint_{\partial\Sigma} F \cdot \vec{T} ds = \int_{\gamma(\cdot, 0)} F \cdot \vec{ds} + \int_{\gamma(b, \cdot)} F \cdot \vec{ds} - \int_{\gamma(\cdot, 1)} F \cdot \vec{ds} - \int_{\gamma(a, \cdot)} F \cdot \vec{ds}. \quad (6.11)$$

Der Kreis beim Integralzeichen symbolisiert, dass über eine geschlossene Kurve integriert wird. Nun kommen wir zum Flächenintegral. Der Flächeninhalt eines Parallelogramms, das

von zwei Vektoren  $a, b \in \mathbb{R}^3$  aufgespannt wird, ist  $|a \times b|$ , wobei  $\times$  das Kreuzprodukt bedeutet. Die Abbildung  $\gamma(s, t)$  bildet ein Quadrat mit Ecken  $(s, t)$  und  $(s + \Delta s, t + \Delta t)$  näherungsweise auf ein Parallelogramm ab, das von den Vektoren  $(\partial_s \gamma) \Delta s$  sowie  $(\partial_t \gamma) \Delta t$  aufgespannt wird. Das Flächenelement sollte daher wie folgt gegeben sein:

$$dA = |\partial_s \gamma \times \partial_t \gamma| ds dt.$$

Das Integral einer Funktion  $\varphi : \Sigma \rightarrow \mathbb{R}$  über die Fläche wird daher wie folgt definiert:

$$\int_{\Sigma} \varphi dA = \int_0^1 \int_a^b \varphi(\gamma(s, t)) |\partial_s \gamma \times \partial_t \gamma| ds dt.$$

Wieder könnte  $\varphi$  eine Massen- oder Ladungsdichte sein, jetzt aber pro Flächeneinheit. Auf die Frage, ob die Definition unabhängig ist von der Parametrisierung, kommen wir später zurück. Die Einheitsnormale der Fläche lautet nun

$$\vec{N} = \frac{\partial_s \gamma \times \partial_t \gamma}{|\partial_s \gamma \times \partial_t \gamma|}.$$

In der folgenden Rechnung verwenden wir den  $\varepsilon_{ijk}$ -Tensor aus Aufgabe 4, Serie1.

$$\begin{aligned} \text{rot } F \cdot (\partial_s \gamma \times \partial_t \gamma) &= \sum_{i,j,k=1}^3 \varepsilon_{ijk} \partial_i F_j e_k \cdot \sum_{l,m,n=1}^3 \varepsilon_{lmn} \partial_s \gamma^l \partial_t \gamma^m e_n \\ &= \sum_{i,j,k,l,m=1}^3 \varepsilon_{ijk} \varepsilon_{lmk} \partial_i F_j \partial_s \gamma^l \partial_t \gamma^m \\ &= \sum_{i,j,l,m=1}^3 (\delta_{il} \delta_{jm} - \delta_{im} \delta_{jl}) \partial_i F_j \partial_s \gamma^l \partial_t \gamma^m \\ &= \sum_{i,j=1}^3 \partial_i F_j \partial_s \gamma^i \partial_t \gamma^j - \sum_{i,j=1}^3 \partial_i F_j \partial_s \gamma^j \partial_t \gamma^i \\ &= \sum_{i,j=1}^3 (\partial_i F_j - \partial_j F_i) \partial_s \gamma^i \partial_t \gamma^j. \end{aligned}$$

Wir erhalten

$$\begin{aligned} \int_{\Sigma} \text{rot } F \cdot \vec{N} dA &= \int_0^1 \int_a^b \text{rot } F \cdot (\partial_s \gamma \times \partial_t \gamma) ds dt \\ &= \int_0^1 \int_a^b \sum_{i,j=1}^3 (\partial_i F_j - \partial_j F_i) \partial_s \gamma^i \partial_t \gamma^j ds dt. \end{aligned} \tag{6.12}$$

Setzen wir (6.11) und (6.12) in die Homotopieformel aus Satz 6.2 ein, so folgt insgesamt

**Satz 6.4 (von Stokes)** *Sei  $\Sigma \subset \mathbb{R}^3$  ein Flächenstück mit Normale  $\vec{N}$ . Sei  $\vec{T}$  die Einheits-tangente auf  $\partial\Sigma$ , so dass  $\vec{N} \times \vec{T}$  ins Innere des Flächenstücks weist. Dann gilt für jedes  $C^1$ -Vektorfeld  $F$  auf  $\mathbb{R}^3$*

$$\int_{\Sigma} \text{rot } F \cdot \vec{N} dA = \oint_{\partial\Sigma} F \cdot \vec{T} ds.$$

Sei  $\Sigma \subset \mathbb{R}^2$  und  $\vec{N} = e_3$  als Normale gewählt. Die Wahl der Tangente  $\vec{T}$  besagt dann, dass bei Durchlaufen des Randes das Gebiet auf der linken Seite liegt, das heißt  $\partial\Sigma$  wird gegen den Uhrzeigersinn durchlaufen. Das war auch unsere Wahl für  $[a, b] \times [0, 1]$ . Wir haben den Satz bewiesen für Flächenstücke, die auf einem Rechteck parametrisiert sind. Allgemeinere Flächen kann man unter Umständen durch Zerlegen behandeln. Für geschlossene Flächen sollten sich die Randterme gegenseitig wegheben, und es sollte folgen

$$\oint_{\Sigma} \operatorname{rot} F \cdot \vec{N} \, dA = 0.$$

Unsere Herleitung des Satzes von Stokes ist insofern rigoros, als die Integrale über die Parametrisierung genau definiert sind und damit die Formel hergeleitet wird. Dagegen haben wir den Begriff des Flächenstücks  $\Sigma$  und seines Randes  $\partial\Sigma$  nicht präzisiert.

## 7 Lokale Auflösung von Gleichungen

Für die Lösbarkeit von linearen Gleichungen haben wir in den linearen Algebra Kriterien entwickelt. Eine nichtlineare Gleichung

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

kann nur in Ausnahmefällen *explizit*, also durch algebraische Umformungen, gelöst werden. Hier stellen wir uns vor, dass  $x$  die Rolle eines Parameters spielt und die Lösung als Funktion  $y = u(x)$  dieses Parameters gefunden werden soll. Da keine Formel für die Lösung angegeben wird, spricht man von einer implizit gegebenen Funktion oder kurz einer impliziten Funktion. Wir betrachten genauer die Situation, dass schon eine konkrete Lösung  $g(x_0, y_0) = 0$  bekannt ist, und wollen die Lösungsmenge nahe bei  $(x_0, y_0)$  analysieren.

**Beispiel 7.1** Betrachte die quadratische Gleichung

$$g(x, y) = y^2 - xy + 1 = 0.$$

Für welche Parameter  $x$  ist die Gleichung lösbar, und wie hängt die Lösung von  $x$  ab? In diesem Fall können wir das voll beantworten:

- Für  $x \in (-2, 2)$  gibt es keine Lösung.
- Für  $x = \pm 2$  gibt es genau eine Lösung, nämlich  $y = \pm 1$ .
- Für  $|x| > 2$  gibt es genau zwei Lösungen  $y = u_{\pm}(x)$ , mit  $u_{\pm}(x) = \frac{x \pm \sqrt{x^2 - 4}}{2}$ .

Nehmen wir etwa die Lösung  $(x_0, y_0) = (3, \frac{3+\sqrt{5}}{2})$ . Für  $|x - x_0| < \delta$  gibt es genau eine Lösung mit  $|y - y_0| < \varepsilon$ , nämlich  $y = u_+(x)$ . Wir dürfen  $\delta > 0$  nicht zu groß wählen, denn für  $x \in (-2, 2)$  gibt es gar keine Lösung. Und wir dürfen  $\varepsilon > 0$  nicht zu groß wählen, sonst haben wir auch den Punkt  $(3, \frac{3-\sqrt{5}}{2})$  und die Eindeutigkeit ist verletzt. Ungünstig wird die Sache für  $(x_0, y_0) = (2, 1)$ . Die Existenz gilt für kein Intervall  $(x_0 - \delta, x_0 + \delta)$ , und die Eindeutigkeit ist in  $(y_0 - \delta, y_0 + \delta)$  ebenfalls nie erfüllt. Beachten Sie

$$\frac{\partial g}{\partial y}(x, y) = 2y - x.$$

Auf der Menge  $\{(x, y) : g(x, y) = 0\}$  ist  $\frac{\partial g}{\partial y}$  nur in den Punkten  $(2, 1)$  und in  $(-2, -1)$  gleich Null. Es wäre zu vermuten, dass diese partielle Ableitung eine Rolle spielt.

**Satz 7.1 (implizite Funktionen)** Sei  $g = g(x, y)$  stetig differenzierbar auf der offenen Menge  $D \subset \mathbb{R}^2$ , und sei  $(x_0, y_0) \in D$  mit  $g(x_0, y_0) = 0$ . Ist  $\frac{\partial g}{\partial y}(x_0, y_0) \neq 0$ , so gibt es  $\delta, \varepsilon > 0$  mit folgender Eigenschaft:

- (1) Zu jedem  $x \in (x_0 - \delta, x_0 + \delta)$  gibt es genau ein  $y \in (y_0 - \varepsilon, y_0 + \varepsilon)$  mit  $g(x, y) = 0$ .  
Damit ist die implizite Funktion  $y = u(x)$  definiert.
- (2)  $u : (x_0 - \delta, x_0 + \delta) \rightarrow (y_0 - \varepsilon, y_0 + \varepsilon)$  ist in  $x_0$  differenzierbar mit Ableitung

$$u'(x_0) = - \frac{\frac{\partial g}{\partial x}(x_0, y_0)}{\frac{\partial g}{\partial y}(x_0, y_0)}.$$

BEWEIS: Wir nehmen  $\frac{\partial g}{\partial y}(x_0, y_0) > 0$  an. Da  $\frac{\partial g}{\partial y}$  stetig, gibt es ein  $\varepsilon > 0$  mit

$$\frac{\partial g}{\partial y}(x, y) > 0 \quad \text{auf } [x_0 - \varepsilon, x_0 + \varepsilon] \times [y_0 - \varepsilon, y_0 + \varepsilon].$$

Es folgt  $g(x_0, y_0 - \varepsilon) < 0 < g(x_0, y_0 + \varepsilon)$ . Da  $g$  stetig, gibt es ein  $\delta \in (0, \varepsilon]$  mit

$$\begin{aligned} g(x, y_0 - \varepsilon) &< 0 & \text{für } x \in [x_0 - \delta, x_0 + \delta] \\ g(x, y_0 + \varepsilon) &> 0 & \text{für } x \in [x_0 - \delta, x_0 + \delta]. \end{aligned}$$

Für  $x \in (x_0 - \delta, x_0 + \delta)$  ist nun  $g(x, \cdot)$  stetig auf  $[y_0 - \varepsilon, y_0 + \varepsilon]$  mit  $g(x, y_0 - \varepsilon) < 0 < g(x, y_0 + \varepsilon)$ . Nach dem Zwischenwertsatz gibt es ein  $y \in (y_0 - \varepsilon, y_0 + \varepsilon)$  mit  $g(x, y) = 0$ . Da  $g(x, \cdot)$  streng monoton wachsend auf  $[y_0 - \varepsilon, y_0 + \varepsilon]$ , ist  $u(x) := y$  eindeutig bestimmt.

Wäre  $u(x)$  nicht stetig in  $x_0$ , so gibt es  $x_n \rightarrow x_0$  mit  $u(x_n) \rightarrow y_1 \neq y_0$ , also

$$g(x_0, y_1) = \lim_{n \rightarrow \infty} g(x_n, u(x_n)) = 0.$$

Das widerspricht aber der Eindeutigkeit, also gilt  $u(x) \rightarrow u(x_0)$ . Um  $u'(x_0)$  zu bestimmen, verwenden wir den Mittelwertsatz: mit geeigneten Zwischenstellen  $\xi, \eta$  gilt

$$\begin{aligned} 0 &= g(x, u(x)) - g(x_0, u(x_0)) \\ &= g(x, u(x)) - g(x_0, u(x)) + g(x_0, u(x)) - g(x_0, u(x_0)) \\ &= \frac{\partial g}{\partial x}(\xi, u(x))(x - x_0) + \frac{\partial g}{\partial y}(x_0, \eta)(u(x) - u(x_0)), \end{aligned}$$

beziehungsweise nach Umformung

$$\frac{u(x) - u(x_0)}{x - x_0} = - \frac{\frac{\partial g}{\partial x}(\xi, g(x))}{\frac{\partial g}{\partial y}(x_0, \eta)}.$$

Mit  $x \rightarrow x_0$  geht  $\xi \rightarrow x_0$  sowie  $u(x), \eta \rightarrow u(x_0)$ , und die Behauptung folgt. □

Die Formel für  $u'(x_0)$  kann immer mit der Kettenregel wieder hergeleitet werden:

$$g(x, u(x)) = 0 \quad \Rightarrow \quad 0 = \frac{d}{dx} g(x, u(x)) = \frac{\partial g}{\partial x}(x, u(x)) + \frac{\partial g}{\partial y}(x, u(x))u'(x).$$

Durch Auflösen folgt dann für alle  $x \in (x_0 - \delta, x_0 + \delta)$  die Ableitung der impliziten Funktion

$$u'(x) = \frac{\frac{\partial g}{\partial x}(x, u(x))}{\frac{\partial g}{\partial y}(x, u(x))}. \quad (7.13)$$

Dies gilt tatsächlich auf dem ganzen Intervall, denn der Satz ist auch im Punkt  $(x, u(x))$  statt  $(x_0, y_0)$  anwendbar.

**Beispiel 7.2** Betrachte die Gleichung

$$g(x, y) = y^n - xy + 1 = 0 \quad \text{mit } n \in \mathbb{N} \text{ ungerade.}$$

Hierfür gibt es keine allgemeine Lösungsformel. Es gilt  $g(0, -1) = 0$ ,

$$\frac{\partial g}{\partial x}(0, -1) = 1 \quad \text{und} \quad \frac{\partial g}{\partial y}(0, -1) = n.$$

Nach dem impliziten Funktionensatz gibt es für  $x$  nahe bei Null eine eindeutige Lösung  $y = u(x)$  nahe bei  $-1$ , und es gilt

$$u'(0) = -\frac{\frac{\partial g}{\partial x}(0, -1)}{\frac{\partial g}{\partial y}(0, -1)} = -\frac{1}{n}.$$

Wir haben also die Näherung  $u(x) \approx -1 - \frac{x}{n}$  für die Nullstelle.

Ein Beweis der höherdimensionalen Version des Satzes über implizite Funktionen würde hier zu weit führen, aber wir wollen diese zumindest formulieren. Im Anschluss werden wir eine geometrische Interpretation des Satzes geben. Betrachte also die  $k$  Gleichungen

$$g(x, y) = 0 \quad \text{mit } (x, y) \in D \subset \mathbb{R}^m \times \mathbb{R}^k \text{ und } g : D \rightarrow \mathbb{R}^k.$$

Wir wollen nach  $y$  auflösen und setzen voraus, dass die Zahl der Gleichungen mit der Zahl der Unbekannten übereinstimmt, was ja bestimmt sinnvoll ist. Bezeichne mit  $D_x g(x_0, y_0) \in \mathbb{R}^{k \times m}$  beziehungsweise  $D_y g(x_0, y_0) \in \mathbb{R}^{k \times k}$  die Jacobimatrizen bezüglich der Variablen  $x \in \mathbb{R}^m$  bzw.  $y \in \mathbb{R}^k$ . Es sei eine Lösung  $(x_0, y_0)$  der Gleichung gegeben. Entwickeln wir  $f(x, y)$  bei  $(x_0, y_0)$  und ignorieren Terme höherer Ordnung, so erhalten wir das lineare Gleichungssystem

$$0 = g(x, y) \approx A(x - x_0) + B(y - y_0) \quad \text{mit } A = D_x g(x_0, y_0), B = D_y g(x_0, y_0).$$

Dieses lineare System ist eindeutig nach  $y$  auflösbar genau wenn  $B = D_y f(x_0, y_0) \in \mathbb{R}^{k \times k}$  invertierbar ist, und zwar lautet die Lösung

$$y = y_0 - B^{-1}A(x - x_0).$$

**Satz 7.2 (implizite Funktionen)** Sei  $g : D \rightarrow \mathbb{R}^k$  stetig differenzierbar auf der offenen Menge  $D \subset \mathbb{R}^m \times \mathbb{R}^k$ , und sei  $(x_0, y_0) \in D$  mit  $g(x_0, y_0) = 0$ . Ist  $D_y g(x_0, y_0)$  invertierbar, so gibt es Umgebungen  $U \subset \mathbb{R}^m$  von  $x_0$  und  $V \subset \mathbb{R}^k$  von  $y_0$  mit folgender Eigenschaft:

- (1) Zu jedem  $x \in U$  gibt es genau ein  $y \in V$  mit  $g(x, y) = 0$ , damit ist die implizite Funktion  $y = u(x)$  definiert.

(2)  $u : U \rightarrow V$  ist stetig differenzierbar mit Ableitung

$$Du(x) = -D_y g(x, u(x))^{-1} D_x g(x, u(x)).$$

Bisher war die Aufteilung in Unbekannte  $y$  und Parameter  $x$  a priori vorgegeben. Das ist nicht immer so, zum Beispiel sind Höhenlinien einfach Nullstellenmengen

$$\Gamma = \{(x, y) \in D : g(x, y) = 0\},$$

die beiden Koordinaten  $(x, y)$  sind völlig gleichberechtigt. Wir können uns dann aussuchen, ob wir nach  $x$  oder nach  $y$  auflösen und die Höhenlinie entweder als Graph  $y = u(x)$  über der  $x$ -Achse oder als Graph  $x = u(y)$  über der  $y$ -Achse schreiben, natürlich im allgemeinen mit einer anderen Funktion  $u$ . Entsprechendes gilt für eine Höhenfläche

$$\Sigma = \{(x, y, z) \in D : g(x, y, z) = 0\}.$$

In der Nähe eines Punkts  $(x_0, y_0, z_0) \in \Sigma$  kann eine Darstellung als Graph  $z = u(x, y)$  existieren, aber auch eine Darstellung über der  $(x, z)$ -Ebene oder der  $(y, z)$ -Ebene. Der Satz über implizite Funktionen liefert lokal diese Darstellungen, wenn die jeweilige Auflösungsbedingung erfüllt ist.

**Beispiel 7.3** Betrachte die zwei-dimensionale Sphäre

$$\mathbb{S}^2 = \{(x, y, z) \in \mathbb{R}^3 : g(x, y, z) = 0\} \quad \text{mit } g(x, y, z) = x^2 + y^2 + z^2 - 1.$$

Es gilt  $\frac{\partial g}{\partial z}(x_0, y_0, z_0) = 2z_0 \neq 0$  für  $z_0 \neq 0$ . Nach dem Satz über implizite Funktionen kann  $\mathbb{S}^2$  in der Nähe jedes Punkts  $(x_0, y_0, z_0) \in \mathbb{S}^2$  als Graph  $z = u(x, y)$  geschrieben werden, es sei denn  $(x_0, y_0, z_0)$  liegt auf dem Äquator ( $z_0 = 0$ ). Tatsächlich haben wir die Darstellungen

$$z = \sqrt{1 - x^2 - y^2} \quad \text{für } z > 0 \quad \text{bzw.} \quad z = -\sqrt{1 - x^2 - y^2} \quad \text{für } z < 0.$$

In der Umgebung eines Punkts  $(x_0, y_0, 0) \in \mathbb{S}^2$  kann die Sphäre tatsächlich nicht als Graph  $z = u(x, y)$  geschrieben werden. Aber man hat eine lokale Graphendarstellung  $x = u(y, z)$ , außer wenn  $x_0 = 0$ , und  $y = u(x, z)$  wenn nicht  $y_0 = 0$ . Da mindestens eine der drei Koordinaten  $x_0, y_0, z_0$  ungleich Null sein muss wegen  $x_0^2 + y_0^2 + z_0^2 = 1$ , hat man eine lokale Darstellung als Graph für jeden Punkt auf  $\mathbb{S}^2$ .

Wir kommen nun zu Extremwertaufgaben mit einer Nebenbedingung. Folgendes Beispiel ist uns schon begegnet.

**Beispiel 7.4** Sei  $A \in \mathbb{R}^{n \times n}$  symmetrische Matrix mit der quadratischen Form

$$f : \mathbb{R}^n \rightarrow \mathbb{R}, f(x) = \langle Ax, x \rangle = \sum_{i,j=1}^n a_{ij} x_i x_j.$$

Wir haben die Funktion  $f(x)$  minimiert unter der Nebenbedingung  $g(x) = |x|^2 = 1$ , also unter Punkten auf der Sphäre  $\mathbb{S}^{n-1}$ .

Im folgenden Satz definiert die Funktion  $g(x)$  bzw. die Menge  $M$  die Nebenbedingung, und  $f(x)$  ist die Zielfunktion, deren Extremwert gesucht wird.

**Satz 7.3 (Lagrange-Multiplikatoren-Regel)** Seien  $f, g \in C^1(\Omega)$  mit  $\Omega \subset \mathbb{R}^n$  offen. Sei  $M = \{x \in \Omega : g(x) = 0\}$  und  $a \in M$  mit

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

Ist dann  $\text{grad } g(a) \neq 0$ , so gibt es ein  $\lambda \in \mathbb{R}$  mit  $\text{grad } f(a) = \lambda \text{grad } g(a)$ .

BEWEIS: Sei  $\gamma(t)$  eine Kurve in  $M$  mit  $\gamma(0) = a$ . Dann hat die Funktion  $f(\gamma(t))$  ein Minimum bei  $t = 0$ , also folgt mit  $V = \gamma'(0)$

$$0 = \frac{d}{dt} f(\gamma(t))|_{t=0} = \langle \text{grad } f(\gamma(0)), \gamma'(0) \rangle = \langle \text{grad } f(a), V \rangle.$$

Die Menge aller Ableitungsvektoren  $V \in \mathbb{R}^n$  solcher Kurven  $\gamma(t)$  nennen wir den Tangentialraum  $T_a M$  von  $M$  im Punkt  $a$ . Da  $g(x)$  sogar identisch Null auf  $M$  ist, folgt

$$\text{grad } f(a), \text{grad } g(a) \perp T_a M.$$

Wir zeigen, dass es  $n - 1$  linear unabhängige Vektoren in  $T_a M$  gibt. Dann sind die beiden Gradienten parallel, und wegen  $\text{grad } g(a) \neq 0$  folgt die Behauptung. Durch Umm Nummerieren der Koordinaten können wir dazu annehmen, dass

$$\frac{\partial g}{\partial x_n}(a) \neq 0.$$

Nach dem Satz über implizite Funktionen ist  $M$  in einer Umgebung von  $a$  als Graph einer  $C^1$ -Funktion  $x_n = u(x')$  darstellbar, wobei  $x' = (x_1, \dots, x_{n-1})$ ; speziell gilt  $a = (a', u(a'))$  wegen  $a \in M$ . Wir wählen nun  $\gamma_i(t) = (a' + te_i, u(a' + te_i)) \in M$  für  $i = 1, \dots, n - 1$ . Es folgt  $\gamma_i(0) = a$  und somit

$$T_a M \ni \frac{d}{dt} \gamma_i(t)|_{t=0} = (e_i, \partial_i u(a')).$$

Die Vektoren sind tatsächlich linear unabhängig, denn es gilt

$$\sum_{i=1}^{n-1} \lambda_i (e_i, \partial_i u(a')) = 0 \quad \Rightarrow \quad \sum_{i=1}^{n-1} \lambda_i e_i = 0 \quad \Rightarrow \quad \lambda_i = 0.$$

□

Für eine beliebige Menge  $M \subset \mathbb{R}^n$  ist  $T_a M$  kein Unterraum. Hier ist aber  $M$  die Nullstellenmenge von  $g(x)$  und  $\text{grad } g(a) \neq 0$ . Der implizite Funktionensatz liefert dann, dass  $T_a M$  ein  $(n - 1)$ -dimensionaler Unterraum ist.  $(n - 1)$ -dimensionaler Unterraum ist. Denn betrachten wir allgemeiner die Kurven  $\gamma(t) = (a' + tv, u(a' + tv))$  mit beliebigem  $v \in \mathbb{R}^{n-1}$ , so folgt

$$\{(v, \partial_v u(a')) : v \in \mathbb{R}^{n-1}\} \subset T_a M.$$

Die Menge links ist aber ein  $(n - 1)$ -dimensionaler Unterraum. Andererseits kann es wegen  $T_a M \perp \text{grad } g(a)$  keine weiteren Tangentialvektoren geben.

**Beispiel 7.5** Für eine symmetrische Matrix  $A \in \mathbb{R}^{n \times n}$  betrachten wir nochmals das Problem, die quadratische Form  $f(x) = \langle Ax, x \rangle$  zu minimieren unter der Nebenbedingung  $g(x) = |x|^2 - 1 = 0$ . Wie in Beispiel 1.6 bewiesen, hat das Minimumproblem eine Lösung, das heißt es gibt ein  $x_0 \in \mathbb{S}^{n-1}$  mit  $f(x_0)$  kleinstmöglich. Da  $\text{grad } g(x_0) = 2x_0 \neq 0$  wegen  $|x_0| = 1$ , gibt es nach Satz 7.3 ein  $\lambda \in \mathbb{R}$  mit  $\text{grad } f(x_0) = \lambda \text{grad } g(x_0)$ . Aber  $\text{grad } f(x_0) = 2Ax_0$  nach Beispiel 3.4, also erhalten wir  $Ax_0 = \lambda x_0$ . Dies hatten wir schon in Kapitel 5, Satz 7.3, gesehen.

In konkreten Anwendungen kann es vorkommen, dass die Bedingung  $\text{grad } g(x) \neq 0$  nicht allen Punkten von  $M = \{x \in \Omega : g(x) = 0\}$  erfüllt ist; diese Punkte müssen dann gesondert betrachtet werden.

# Kapitel 7

## Mehrdimensionale Integration

### 1 Prinzip von Cavalieri und Fubini

In diesem Abschnitt erklären wir kurz die Definition des mehrdimensionalen Riemannsches Integrals. Als erstes definieren wir das Integral stetiger Funktionen auf  $n$ -dimensionalen Quadern  $Q$ , und zwar per Induktion nach der Dimension  $n$ . Wir bezeichnen das Volumen von  $Q$ , also das Produkt der Kantenlängen, mit  $|Q|$ . Für  $n = 1$  haben wir das übliche Riemannsche Integral auf einem Intervall  $Q = [a, b]$ .

Für  $n = 2$  ist  $Q = [a_1, b_1] \times [a_2, b_2]$  ein Rechteck, und wir setzen für eine stetige Funktion  $f = f(x, y)$  auf diesem Rechteck

$$\int_Q f(x, y) dx dy = \int_{a_2}^{b_2} \left( \int_{a_1}^{b_1} f(x, y) dx \right) dy.$$

Das Parameterintegral  $F(y) = \int_{a_1}^{b_1} f(x, y) dx$  ist nach Satz 5.2 in Kapitel 5 eine stetige Funktion, und damit ist das iterierte Integral auf der rechten Seite tatsächlich definiert. Für  $n = 3$ , also  $Q = [a_1, b_1] \times [a_2, b_2] \times [a_3, b_3]$  und eine stetige Funktion  $f = f(x, y, z)$ , setzen wir

$$\int_Q f(x, y, z) dx dy dz = \int_{a_3}^{b_3} \left( \int_{Q'} f(x, y, z) dx dy \right) dz$$

Dabei ist  $Q' = [a_1, b_1] \times [a_2, b_2]$ . Das innere Integral  $F_1(z) = \int_{Q'} f(x, y, z) dx dy$  ist bereits definiert, allerdings müssen wir noch zeigen, dass es stetig von  $z$  abhängt. Es gilt

$$F_1(z) = \int_{a_2}^{b_2} F_2(y, z) dy \quad \text{mit} \quad F_2(y, z) = \int_{a_1}^{b_1} f(x, y, z) dx.$$

Nach Satz 5.2 in Kapitel 5 ist  $F_2(y, z)$  stetig in beiden Variablen, und dann  $F_1(z)$  ebenfalls stetig. Damit ist die Definition auch für  $n = 3$  gerechtfertigt. Allgemein definieren wir für  $n$  beliebig, also  $Q = Q' \times [a_n, b_n]$  und  $f = f(x', x_n)$  stetig mit  $x' \in Q' \subset \mathbb{R}^{n-1}$ ,  $x_n \in [a_n, b_n]$ ,

$$\int_Q f(x) dx = \int_{a_n}^{b_n} \left( \int_{Q'} f(x', x_n) dx' \right) dx_n.$$

Per Induktion sieht man, dass das innere Integral stetig von  $x_n$  abhängt und somit das Integral wohldefiniert ist. Wir schreiben ein  $n$ -dimensionales Integral meistens in der Form

$\int_Q f(x) dx$ , wobei  $x = (x_1, \dots, x_n) \in \mathbb{R}^n$  für alle Koordinaten steht. In einem spezifischen Integral in  $\mathbb{R}^2$  oder  $\mathbb{R}^3$  schreiben wir auch  $\int_Q f(x, y) dx dy$  oder  $\int_Q f(x, y, z) dx dy dz$ , wobei nun  $x, y, z$  die einzelnen Koordinaten bezeichnet.

Es stellt sich die offensichtliche Frage, ob diese Definition von der Reihenfolge abhängt, in der die einzelnen Variablen integriert werden. Wir zeigen deshalb jetzt eine alternative Darstellung des Integrals als Grenzwert Riemannscher Summen. In diesen Riemannschen Summen können die Koordinaten vertauscht werden, woraus dann die Behauptung folgt. Der Einfachheit halber beschränken wir uns auf Zerlegungen, bei denen jedes Intervall  $[a_i, b_i]$  äquidistant durch  $N$  Unterteilungspunkte zerlegt wird. Außerdem nehmen wir als Stützstellen die Unterteilungspunkte.

**Satz 1.1 (Approximation durch Riemannsche Summen)** Für jeden  $n$ -dimensionalen Quader  $Q = [a_1, b_1] \times \dots \times [a_n, b_n]$  und  $f : Q \rightarrow \mathbb{R}$  stetig gilt

$$\int_Q f(x) dx = \lim_{N \rightarrow \infty} \sum_{k_1, \dots, k_n=1}^N f(x_{1,k_1}, \dots, x_{n,k_n}) \Delta x_1 \dots \Delta x_n.$$

Dabei ist  $\Delta x_i = \frac{b_i - a_i}{N}$  und  $x_{i,k} = a_i + k \Delta x_i$ .

BEWEIS: Wir geben den Beweis nur für  $n = 2$ . Für  $n \geq 3$  ist das Argument ganz analog, nur ist der Notationsaufwand wegen der vielen Indizes größer. Wir haben also die Zerlegung

$$\begin{aligned} a_1 = x_0 \leq \dots \leq x_N = b_1 & \quad \text{mit} \quad x_k = a_1 + k \Delta x \\ a_2 = y_0 \leq \dots \leq y_N = b_2 & \quad \text{mit} \quad y_\ell = a_2 + \ell \Delta y, \end{aligned}$$

wobei  $\Delta x = \frac{b_1 - a_1}{N}$ ,  $\Delta y = \frac{b_2 - a_2}{N}$ . Es folgt, wenn  $S_Z(f)$  die obige Riemannsche Summe ist,

$$\begin{aligned} \int_Q f(x, y) dx dy &= \sum_{k, \ell=1}^N \int_{y_{\ell-1}}^{y_\ell} \left( \int_{x_{k-1}}^{x_k} f(x, y) dx \right) dy \\ S_Z(f) &= \sum_{k, \ell=1}^N f(x_k, y_\ell) \Delta x \Delta y = \int_{y_{\ell-1}}^{y_\ell} \left( \int_{x_{k-1}}^{x_k} f(x_k, y_\ell) dx \right) dy. \end{aligned}$$

Wir erinnern nun an den Begriff der Oszillation einer Funktion:

$$\text{osc}(f, \delta) = \sup_{|(x,y)-(x',y')| < \delta} |f(x, y) - f(x', y')|.$$

Für  $N$  groß ist der Durchmesser von  $[x_{k-1}, x_k] \times [y_{\ell-1}, y_\ell]$  kleiner als  $\delta$ , also gilt

$$|f(x, y) - f(x_k, y_\ell)| \leq \text{osc}(f, \delta) \quad \text{für } (x, y) \in [x_{k-1}, x_k] \times [y_{\ell-1}, y_\ell].$$

Mit der Dreiecksungleichung ergibt sich

$$\begin{aligned} \left| \int_Q f(x, y) dx dy - S_Z(f) \right| &\leq \sum_{k, \ell=1}^N \int_{x_{k-1}}^{x_k} \left( \int_{y_{\ell-1}}^{y_\ell} |f(x, y) - f(x_k, y_\ell)| dy \right) dx \\ &\leq \text{osc}(f, \delta) \sum_{k, \ell=1}^N \Delta x \Delta y \\ &\leq \text{osc}(f, \delta) |Q|. \end{aligned}$$

Aber  $\text{osc}(f, \delta) \rightarrow 0$  mit  $\delta \searrow 0$ , siehe Satz 5.1 in Kapitel 5. Damit ist der Satz bewiesen.  $\square$   
 Die Vertauschbarkeit der Integrationsreihenfolge ist direkte Konsequenz. Insbesondere gilt

**Folgerung 1.1 (Fubini)** Sei  $Q = [a_1, b_1] \times \dots \times [a_n, b_n]$  und  $f : Q \rightarrow \mathbb{R}$  stetig. Dann kann das Integral für jedes  $1 \leq i \leq n$  wie folgt berechnet werden:

$$\int_Q f(x) dx = \int_{a_i}^{b_i} \left( \int_{Q'} f(x) dx' \right) dx_i.$$

Dabei ist  $Q' = [a_1, b_1] \times \dots \times \widehat{[a_i, b_i]} \times \dots \times [a_n, b_n]$  und  $dx' = dx_1 \dots \widehat{dx_i} \dots dx_n$ . Das Dach bedeutet, dass dieser Faktor auszulassen ist.

Wir notieren einige Eigenschaften, die direkt aus den Regeln für das eindimensionale Riemannintegral - oder alternativ aus der Darstellung als Grenzwert Riemannscher Summen - folgen.

**Satz 1.2 (Eigenschaften des Integrals)** Auf einem  $n$ -dimensionalen Quader  $Q$  gilt:

- (1) Linearität:  $\int_Q (\lambda f(x) + \mu g(x)) dx = \lambda \int_Q f(x) dx + \mu \int_Q g(x) dx$ .
- (2) Integral einer Konstante: Die konstante Funktion  $f : Q \rightarrow \mathbb{R}$  mit Wert  $f(x) = c$  hat das Integral  $\int_Q f(x) dx = c |Q|$ .
- (3) Monotonie: Aus  $f(x) \leq g(x)$  folgt  $\int_Q f(x) dx \leq \int_Q g(x) dx$ .
- (4) Abschätzung:  $\left| \int_Q f(x) dx \right| \leq \int_Q |f(x)| dx \leq (\sup_Q |f|) |Q|$ .

Nun kommen wir zu allgemeineren Mengen als nur Rechtecken oder Quadern. Eine beliebige Menge im  $\mathbb{R}^n$  approximieren wir von innen und außen durch Vereinigungen von endlich vielen  $n$ -dimensionalen Würfeln. Sei dazu  $\mathcal{Q}$  die Menge aller abgeschlossenen Gitterwürfel  $Q$  mit ganzzahligen Koordinaten und Kantenlänge Eins, und sei  $\mathcal{Q}_k$  die Menge der skalierten Würfel  $2^{-k}Q$  für  $k \in \mathbb{N}_0$ . Setze

$$\begin{aligned} B_k &= \text{Vereinigung der Würfel in } \mathcal{Q}_k, \text{ die im Inneren von } B \text{ enthalten sind,} \\ B^k &= \text{Vereinigung der Würfel in } \mathcal{Q}_k, \text{ die den Abschluss von } B \text{ treffen.} \end{aligned}$$

Beim Übergang von  $k$  zu  $k+1$  kommen bei  $B_k$  Würfel hinzu, bei  $B^k$  fallen welche weg. Somit

$$B_0 \subset B_1 \subset \dots \subset \text{int } B \subset B \subset \overline{B} \subset \dots \subset B^1 \subset B^0.$$

Wegen Monotonie und Beschränktheit existieren somit das innere und äußere Volumen

$$|B|_* := \lim_{k \rightarrow \infty} |B_k| \leq \lim_{k \rightarrow \infty} |B^k| =: |B|^*.$$

**Definition 1.1** Eine beschränkte Menge  $B \subset \mathbb{R}^n$  heißt Jordan-messbar, wenn  $|B|_* = |B|$ . Diese Zahl ist dann das Volumen  $|B|$  von  $B$ .

Die Bedingung ist gleichbedeutend mit  $|\partial B|^* = 0$ . Denn ein Gitterwürfel  $Q \in \mathcal{Q}_k$  trifft genau dann  $\partial B$ , wenn er den Abschluss von  $B$  trifft, aber nicht im Inneren von  $B$  liegt (nachprüfen!). Demnach gilt  $(\partial B)^k = B^k \setminus B_k$  und somit

$$|\partial B|^* = |B|^* - |B|_*.$$

Wir können nun stetige Funktionen auf Jordan-messbaren Mengen integrieren.

**Definition 1.2 (Riemann-Integral II)** Sei  $B \subset \mathbb{R}^n$  Jordan-messbar. Für  $f : B \rightarrow \mathbb{R}$  beschränkt und stetig mit  $f \geq 0$  definieren wir

$$\int_B f(x) dx = \lim_{k \rightarrow \infty} \int_{B_k} f(x) dx.$$

In einem zweiten Schritt setzen wir für eine beliebige beschränkte, stetige Funktion  $f : B \rightarrow \mathbb{R}$

$$\int_B f(x) dx = \int_B f^+(x) dx - \int_B f^-(x) dx.$$

Die Menge  $B_k$  ist die Vereinigung der Gitterwürfel, die im Inneren von  $B$  enthalten sind. Das Integral über  $B_k$  ist damit definiert als Summe der Integrale auf den einzelnen Gitterwürfeln; auf diesen ist die Funktion  $f$  nach Voraussetzung stetig. Beim Übergang von  $B_k$  zu  $B_{k+1}$  kommen Würfel hinzu, damit ist die Folge  $\int_{B_k} f(x) dx$  monoton wachsend für  $f \geq 0$ . Andererseits ist sie nach oben beschränkt durch

$$\int_{B_k} f(x) dx \leq (\sup_B f) |B_k| \leq (\sup_B f) |B|_* < \infty.$$

Wegen Monotonie und Beschränktheit existiert also der Grenzwert der Integrale. Um zu zeigen, dass die Definition sinnvoll ist, müssen wir außerdem nachweisen, dass sie im Fall eines Quaders  $B$  mit unserer ersten Integraldefinition konsistent ist. In diesem Fall sind die  $B_k$  Quader, die in  $B$  enthalten sind. Man kann  $B \setminus B_k$  dann in Quader zerlegen, deren Gesamtvolumen gegen Null geht. Da  $f$  beschränkt ist, gehen die Differenzintegrale gegen Null wie behauptet.

Kurz zum zweiten Teil der Definition: es ist

$$f^+(x) = \begin{cases} f(x) & \text{falls } f(x) \geq 0, \\ 0 & \text{falls } f(x) \leq 0 \end{cases} \quad f^-(x) = \begin{cases} -f(x) & \text{falls } f(x) \leq 0, \\ 0 & \text{falls } f(x) \geq 0 \end{cases}.$$

Insbesondere sind  $f^\pm$  stetig und es gilt  $f = f^+ - f^-$ . Die Definition ist also konsistent mit der Linearität des Integrals.

Wir formulieren nun eine zweite Variante des Satzes von Fubini.

**Satz 1.3 (Cavalieri)** Sei  $B \subset \{(x, z) \in \mathbb{R}^{n-1} \times \mathbb{R} : a \leq z \leq b\}$  Jordan-messbar, und sei  $f : B \rightarrow \mathbb{R}$  stetig. Dann gilt

$$\int_B f(x, z) dx dz = \int_a^b \left( \int_{B_z} f(x, z) dx \right) dz,$$

wobei  $B_z = \{x \in \mathbb{R}^{n-1} : (x, z) \in B\}$  den Querschnitt von  $B$  in Höhe  $z \in [a, b]$  bezeichnet.

**Beispiel 1.1** Der Flächeninhalt  $|D|$  der Kreisscheibe

$$D = \{(x, z) \in \mathbb{R} \times \mathbb{R} : z \in [-1, 1], x^2 + z^2 \leq 1\}$$

ergibt sich wie folgt: es ist  $D_z = [-\sqrt{1-z^2}, \sqrt{1-z^2}]$ , also gilt

$$|D| = \int_{-1}^1 \mathcal{L}(D_z) dz = 2 \int_{-1}^1 \sqrt{1-z^2} dz = 2 \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cos^2 t dt = \pi.$$

Hier haben wir  $z = \sin t$  substituiert. Analog berechnen wir das Volumen  $V(B)$  der Kugel  $B = \{(x, z) \in \mathbb{R}^2 \times \mathbb{R} : |x|^2 + z^2 \leq 1\}$ . Es ergibt sich

$$|B| = \int_{-1}^1 \mathcal{A}(B_z) dz = \pi \int_{-1}^1 (1 - z^2) dz = \frac{4\pi}{3}.$$

BEWEIS: (*des Cavalierischen Prinzips*) Wir können  $f \geq 0$  annehmen, der allgemeine Fall folgt dann durch Zerlegung  $f = f^+ - f^-$ . Da  $B_k$  Vereinigung endlich vieler Gitterwürfel ist und auf jedem einzelnen Gitterwürfel das Integral genau als iteriertes Integral definiert wurde, erhalten wir jedenfalls

$$\int_{B_k} f(x, z) dx dz = \int_a^b \left( \int_{(B_k)_z} f(x, z) dx \right) dz.$$

Man kann nun zeigen, dass die rechte Seite mit  $k \rightarrow \infty$  konvergiert (für die linke Seite ist das klar), so dass folgt

$$\int_B f(x, z) dx dz = \int_a^b \left( \int_{B_z} f(x, z) dx \right) dz.$$

□

Eine Anwendung des Satzes von Fubini bzw. Cavalieri ist die partielle Integration.

**Lemma 1.1** Sei  $B \subset \mathbb{R}^n$  offen und beschränkt, und  $f \in C^1(B)$  mit  $f = 0$  außerhalb einer abgeschlossenen Menge  $K \subset B$ . Dann gilt

$$\int_B \frac{\partial f}{\partial x_i} dx = 0 \quad \text{für } i = 1, \dots, n.$$

BEWEIS: Wir wählen einen Quader  $Q = [a_1, b_1] \times \dots \times [a_n, b_n]$ , so dass  $B$  in  $Q$  enthalten ist, und setzen  $f(x)$  durch Null fort:

$$f_0 : Q \rightarrow \mathbb{R}, f_0(x) = \begin{cases} f(x) & \text{für } x \in B \\ 0 & \text{für } x \in Q \setminus B. \end{cases}$$

Es folgt  $f \in C^1(Q)$  mit  $\frac{\partial f}{\partial x_i} = 0$  auf  $Q \setminus B$ , und  $f = 0$  auf  $\partial Q$ . Mit dem Satz von Fubini und dem Hauptsatz der Differential- und Integralrechnung folgt, wenn wir  $dx' = dx_1 \dots \widehat{dx_i} \dots dx_n$  und  $Q' = [a_1, b_1] \times \dots \times \widehat{[a_i, b_i]} \times \dots \times [a_n, b_n]$  schreiben,

$$\int_B \frac{\partial f}{\partial x_i} dx = \int_Q \frac{\partial f}{\partial x_i} dx = \int_{Q'} \left( \int_{a_i}^{b_i} \frac{\partial f}{\partial x_i}(x) dx_i \right) dx' = \int_{Q'} \underbrace{[f(x)]_{x_i=a_i}^{x_i=b_i}}_{=0} dx' = 0.$$

□

**Satz 1.4 (Partielle Integration)** Sei  $B \subset \mathbb{R}^n$  offen und beschränkt. Für  $f \in C^1(B)$  und  $X \in C^1(B, \mathbb{R}^n)$  sei  $fX = 0$  außerhalb einer abgeschlossenen Menge  $K \subset B$ . Dann gilt

$$\int_B \langle \text{grad } f, X \rangle dx = - \int_B f \text{div } X dx.$$

BEWEIS: Die Funktion  $fX_i$  erfüllt die Voraussetzung von Lemma 1.1, also gilt

$$0 = \int_B \frac{\partial(fX_i)}{\partial x_i} dx = \int_B \frac{\partial f}{\partial x_i} X_i dx + \int_B f \frac{\partial X_i}{\partial x_i} dx.$$

Summation über  $i = 1, \dots, n$  ergibt die Behauptung.

□

## 2 Koordinatentransformationen

Das vielleicht wichtigste Hilfsmittel zur Berechnung mehrdimensionaler Integrale ist der Transformationsatz. Wir erinnern an folgende Aussage, die wir bei der Diskussion der Determinante gezeigt haben.

**Satz 5.9 (Lineare Transformationsformel)** Für  $A \in L(\mathbb{R}^n, \mathbb{R}^n)$  gilt

$$V(A(B)) = |\det(A)| V(B) \quad \text{für alle } B \subset \mathbb{R}^n.$$

Ist  $B$  ein Würfel, so ist  $A(B)$  ein Parallelotop und die Formel folgt aus den Eigenschaften der Determinante durch Scherung. Allgemeine Mengen  $B$  haben wir dann durch Vereinigungen von Gitterwürfeln approximiert, genau wie oben dargestellt. Als Beispiel haben wir das Volumen eines Ellipsoids ausgerechnet, als Bild der Kugel unter einer Diagonalmatrix mit den Halbachsen als Diagonalelementen. Wir suchen nun eine Verallgemeinerung für Abbildungen, die nicht linear sind. Außerdem wollen wir nicht nur Volumina behandeln, sondern beliebige Integrale.

**Definition 2.1** Eine Abbildung  $\phi : U \rightarrow V$  zwischen offenen Mengen  $U, V \subset \mathbb{R}^n$  heißt *Koordinatentransformation (oder Diffeomorphismus)*, wenn  $\phi$  bijektiv ist und  $\phi, \phi^{-1}$  beide  $C^1$ -Abbildungen sind.

Aus  $\phi^{-1}(\phi(x)) = x$  folgt mit der Kettenregel

$$D\phi^{-1}(\phi(x))D\phi(x) = \text{Id} \quad \text{für alle } x \in U.$$

Also ist die Jacobideterminante einer Koordinatentransformation überall ungleich Null.

**Beispiel 2.1** Die Polarkoordinatenabbildung

$$\phi : (0, \infty) \times (0, 2\pi) \rightarrow \mathbb{R}^2 \setminus \{(x, 0) : x \geq 0\}, \quad \phi(r, \theta) = (r \cos \theta, r \sin \theta),$$

ist eine Koordinatentransformation.

**Satz 2.1 (Transformationsformel)** Sei  $\phi : U \rightarrow V$  eine Koordinatentransformation zwischen den offenen Mengen  $U, V \subset \mathbb{R}^n$ . Dann gilt für jede Jordan-messbare Menge  $B \subset U$  und jede stetige Funktion  $f : \phi(B) \rightarrow \mathbb{R}$  die Formel

$$\int_{\phi(B)} f(y) dy = \int_B f(\phi(x)) |\det D\phi(x)| dx.$$

BEWEIS: Wir können  $f \geq 0$  annehmen, sonst zerlegen wir  $f = f^+ - f^-$ . Tatsächlich können wir sogar  $f > 0$  annehmen, indem wir  $f(x)$  durch die Funktionen  $f(x) + \varepsilon$  approximieren. Wir zeigen nun, dass für jeden Würfel  $Q$  im Inneren von  $B$  gilt

$$\int_{\phi(Q)} f(y) dy \leq \int_Q f(\phi(x)) |\det D\phi(x)| dx. \quad (2.1)$$

Wäre das falsch, so gibt es einen Würfel  $Q_0$  und ein  $\theta > 1$  mit

$$\theta \int_{Q_0} f(\phi(x)) |\det D\phi(x)| dx \leq \int_{\phi(Q_0)} f(y) dy.$$

Durch sukzessive Zerlegung in  $2^n$  Teilwürfel finden wir  $Q_0 \supset Q_1 \supset Q_2 \supset \dots$  mit

$$\theta \int_{Q_k} f(\phi(x)) |\det D\phi(x)| dx \leq \int_{\phi(Q_k)} f(y) dy.$$

Da die  $Q_k$  geschachtelt sind und ihr Durchmesser gegen Null geht, gibt es genau ein  $x_0$ , das in allen  $Q_k$  liegt. Wir verwenden nun folgende asymptotischen Aussagen:

- $f(\phi(x)) |\det D\phi(x)| \approx f(\phi(x_0)) |\det D\phi(x_0)|$  (Stetigkeit von  $f \circ \phi$  und  $\det D\phi$ ).
- $f(y) \approx f(\phi(x_0))$  auf  $\phi(Q_k)$  (Stetigkeit von  $f$ ).
- $|\phi(Q_k)| \approx |\det D\phi(x_0)| |Q_k|$  (da  $\text{diam } Q_k \rightarrow 0$ , ist  $\phi(x) \approx \phi(x_0) + D\phi(x_0)(x - x_0)$ , und die Näherung folgt aus der linearen Transformationsformel).

Setzen wir diese asymptotischen Informationen ein, so ergibt sich asymptotisch für  $k \rightarrow \infty$

$$\theta f(\phi(x_0)) |\det D\phi(x_0)| |Q_k| \lesssim f(\phi(x_0)) |\det D\phi(x_0)| |Q_k|.$$

Da  $f(\phi(x_0)) |\det D\phi(x_0)| > 0$ , ergibt sich nach Division  $\theta \lesssim 1$ . Das ist ein Widerspruch zu unserer Annahme  $\theta > 1$ , womit (2.1) gezeigt ist. Wenden wir nun (2.1) auf die Gitterwürfel in  $B_k$  an, so bekommen wir durch Addition

$$\int_{\phi(B_k)} f(y) dy \leq \int_{B_k} f(\phi(x)) |\det D\phi(x)| dx.$$

Die  $B_k$  schöpfen  $B$  mit  $k \rightarrow \infty$  aus, also folgt

$$\int_{\phi(B)} f(y) dy \leq \int_B f(\phi(x)) |\det D\phi(x)| dx.$$

Schließlich sei  $\psi : V \rightarrow U$  die inverse Transformation. Wir wenden das Gezeigte an auf die Funktion  $g(x) = f(\phi(x)) |\det D\phi(x)|$  und bekommen

$$\begin{aligned} \int_B f(\phi(x)) |\det D\phi(x)| dx &= \int_{\psi(\phi(B))} g(x) dx \\ &\leq \int_{\phi(B)} g(\psi(y)) |\det D\psi(y)| dy \\ &= \int_{\phi(B)} f(y) \underbrace{|\det D\phi(\psi(y))| |\det D\psi(y)|}_{=1} dy \\ &= \int_{\phi(B)} f(y) dy. \end{aligned}$$

Damit ist die Transformationsformel gezeigt. □

**Beispiel 2.2 (Polarkoordinaten im  $\mathbb{R}^2$ )** Die Polarkoordinatentransformation ist

$$\phi : U = (0, \infty) \times (0, 2\pi) \rightarrow \mathbb{R}^2 \setminus \{(x, 0) : x \geq 0\} = V, \quad \phi(r, \theta) = (r \cos \theta, r \sin \theta).$$

Die Jacobimatrix und - Determinante sind

$$D\phi(r, \theta) = \begin{pmatrix} \cos \theta & -r \sin \theta \\ \sin \theta & r \cos \theta \end{pmatrix}, \det D\phi(r, \theta) = r > 0.$$

Also besagt der Transformationssatz für  $B \subset U$  und  $f = f(x, y)$

$$\int_{\phi(B)} f(x, y) dx dy = \int_B f(r \cos \theta, r \sin \theta) r dr d\theta.$$

Ein interessanter Fall ist  $f(x, y) = e^{-(x^2+y^2)/2}$  und  $\phi(B) = V$ . Mit Fubini erhalten wir

$$\begin{aligned} \int_{\mathbb{R}^2} e^{-(x^2+y^2)/2} dx dy &= \int_0^\infty \int_0^{2\pi} e^{-r^2/2} r dr d\theta \\ &= -2\pi \int_0^\infty \frac{d}{dr} e^{-r^2} dr \\ &= (-2\pi) \left[ e^{-r^2} \right]_{r=0}^{r=\infty} = 2\pi. \end{aligned}$$

Andererseits gilt, wieder mit Fubini,

$$\int_{\mathbb{R}^2} e^{-(x^2+y^2)/2} dx dy = \int_{-\infty}^\infty e^{-x^2/2} \left( \int_{-\infty}^\infty e^{-y^2/2} dy \right) dx = \left( \int_{-\infty}^\infty e^{-x^2/2} dx \right)^2.$$

Also sehen wir

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

Hieraus ergibt sich die Normierung der Gaußschen Normalverteilung: wählt man die Dichtefunktion  $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ , so hat die Verteilungsfunktion  $F(x) = \int_{-\infty}^x f(y) dy$  Integral Eins. In diesem Beispiel haben wir die Transformationsformel für ein unbeschränktes Gebiet benutzt. Um das zu rechtfertigen, könnte man durch beschränkte Gebiete approximieren, zum Beispiel durch Kreisscheiben  $D_R(0)$ .

**Beispiel 2.3** Sei  $U = (0, \infty) \times (0, \pi) \times (0, 2\pi)$  und  $V = \mathbb{R}^3 \setminus \{(x, 0, z) : x \geq 0\}$ . Dreidimensionale Polarkoordinaten sind gegeben durch

$$\phi : U \rightarrow V, \phi(r, \theta, \varphi) = (r \sin \theta \cos \varphi, r \sin \theta \sin \varphi, r \cos \theta).$$

Die Jacobimatrix von  $\phi$  lautet

$$D\phi(r, \theta, \varphi) = \begin{pmatrix} \sin \theta \cos \varphi & r \cos \theta \cos \varphi & -r \sin \theta \sin \varphi \\ \sin \theta \sin \varphi & r \cos \theta \sin \varphi & r \sin \theta \cos \varphi \\ \cos \theta & -r \sin \theta & 0 \end{pmatrix}.$$

Für die Jacobideterminante ergibt sich  $\det D\phi(r, \theta, \varphi) = r^2 \sin \theta$ . Für ein Gebiet  $\phi(B)$  mit  $B = [r_1, r_2] \times [\theta_1, \theta_2] \times [\varphi_1, \varphi_2]$  liefert der Transformationssatz und der Satz von Fubini

$$|\phi(B)| = \int_{r_1}^{r_2} \int_{\theta_1}^{\theta_2} \int_{\varphi_1}^{\varphi_2} r^2 \sin \theta d\varphi d\theta dr = \frac{r_2^3 - r_1^3}{3} (\cos \theta_1 - \cos \theta_2) (\varphi_2 - \varphi_1).$$

Wir kommen jetzt – endlich – zur Umrechnung von Differentialoperatoren auf krummlinige Koordinaten. Bevor wir das in Angriff nehmen, müssen wir allerdings erst überlegen, was eine solche Transformation bedeutet, insbesondere müssen wir uns über das Transformationsverhalten von skalaren und vektorwertigen Funktionen klar werden. Wir nehmen also an, dass wir eine  $C^1$ -Koordinatentransformation gegeben haben:

$$\phi : U \rightarrow V, y = \phi(x), \quad \text{wobei } U, V \text{ offene Teilmengen des } \mathbb{R}^n.$$

Für Funktionen ist die Sache relativ klar: ist zum Beispiel  $v = v(y)$  eine Temperaturverteilung oder Ladungsdichte, so muss man diese eben am Punkt  $\phi(x)$  auswerten, das heißt

$$\text{Transformation eines Skalars } v = v(y) \text{ ist } u(x) = v(\phi(x)).$$

Hier ist eine Hinweis angesagt: in den Anwendungen werden Funktion und transformierte Funktion oft mit demselben Buchstaben bezeichnet, zum Beispiel  $v(y)$  und  $v(r, \theta, \varphi)$ . Der Grund ist, dass es sich um dieselbe Beobachtungsgröße handelt, zum Beispiel die Temperatur, nur in verschiedenen Koordinaten. Leider kann diese Notation manchmal zur Verwirrung führen, wenn die Kettenregel angewandt wird. Für unsere Rechnung behalten wir die Unterscheidung zwischen  $u$  und  $v$  deshalb bei. Es ist weiter wichtig festzustellen, dass für Vektorgrößen ein anderes Transformationsgesetz gilt als für skalare Funktionen, nämlich

$$\text{Transformation eines Vektorfelds } Y = Y(y) \text{ ist } X(x) \text{ mit } D\phi(x)X(x) = Y(\phi(x)).$$

Betrachten wir als Beispiel die Abbildung, die einem Punkt  $x$  auf einem Stadtplan den realen Punkt  $\phi(x)$  zuordnet. Sei  $x_0$  der Punkt auf der Karte, der Ihrem aktuellen Standort entspricht. Angenommen Sie halten die Karte so, dass der Norden auf der Karte real nach Osten ausgerichtet ist. Dann ist  $\phi(x)$  eine Linksdrehung um  $90^\circ$ , gefolgt von einer Streckung mit dem Maßstab. Wollen wir die Windrichtung  $Y(\phi(x))$  in die Karte eintragen, so müssen wir den Vektor um  $90^\circ$  zurück drehen. Allgemein entspricht bei einer Koordinatentransformation dem Vektor  $X(x)$  der mit der Jacobimatrix „gedrehte“ Vektor  $D\phi(x)X(x)$  an der Stelle  $\phi(x)$ .

Für die Umrechnung von Gradient und Divergenz brauchen wir die Gramsche Matrix

$$g(x) = (g_{ij}(x)) \in \mathbb{R}^{n \times n}, \quad g_{ij}(x) = \left\langle \frac{\partial \phi}{\partial x_i}(x), \frac{\partial \phi}{\partial x_j}(x) \right\rangle.$$

Die  $g_{ij}$  sind die Koeffizienten des Euklidischen Skalarprodukts bezüglich der Basis  $\frac{\partial \phi}{\partial x_i} = D\phi \cdot e_i$ . Für Vektoren  $X, Y \in \mathbb{R}^n$  gilt

$$\langle D\phi \cdot X, D\phi \cdot Y \rangle = \sum_{i,j=1}^n \langle D\phi \cdot e_i, D\phi \cdot e_j \rangle X_i Y_j = \sum_{i,j=1}^n g_{ij} X_i Y_j. \quad (2.2)$$

Weiter haben wir

$$g_{ij}(x) = \langle D\phi(x)e_i, D\phi(x)e_j \rangle = \langle e_i, D\phi(x)^T D\phi(x)e_j \rangle = (D\phi(x)^T D\phi(x))_{ij}.$$

Insbesondere gilt

$$\det g(x) = \det (D\phi(x)^T D\phi(x)) = |\det D\phi(x)|^2 > 0.$$

Die Matrix  $g(x)$  ist somit invertierbar, und wir setzen  $g^{-1}(x) = (g^{ij}(x))$ .

**Lemma 2.1 (Transformation von Vektorfeldern)** Sei  $\phi \in C^1(U, V)$  eine Koordinatentransformation mit Gramscher Matrix  $g(x) = (g_{ij}(x))$ , und sei  $Y \circ \phi = D\phi \cdot X$ . Dann gilt

$$X = \sum_{i,j=1}^n g^{ij} \left\langle Y \circ \phi, \frac{\partial \phi}{\partial x_i} \right\rangle e_j \quad (e_j \text{ Standardbasis}).$$

BEWEIS: Wir berechnen mit (2.2)

$$\left\langle Y \circ \phi, \frac{\partial \phi}{\partial x_i} \right\rangle = \langle D\phi \cdot X, D\phi \cdot e_i \rangle = \sum_{k=1}^n g_{ki} X_k.$$

Multiplizieren wir mit  $g^{ij}$  und summieren über  $i$ , so folgt

$$\sum_{i=1}^n g^{ij} \left\langle Y \circ \phi, \frac{\partial \phi}{\partial x_i} \right\rangle = \sum_{k=1}^n X_k \underbrace{\sum_{i=1}^n g_{ki} g^{ij}}_{=\delta_{jk}} = X_j.$$

□

**Satz 2.2 (Transformation von Differentialoperatoren)** Sei  $\phi \in C^1(U, V)$  eine Koordinatentransformation mit Gramscher Matrix  $(g_{ij}(x))$ . Dann gelten folgende Formeln:

(1) Ist  $v = u \circ \phi$ , so folgt  $(\text{grad } v) \circ \phi = D\phi \cdot \text{grad}_g u$  mit

$$\text{grad}_g u = \sum_{i,j=1}^n g^{ij} \frac{\partial u}{\partial x_i} e_j.$$

(2) Ist  $Y \circ \phi = D\phi \cdot X$ , so folgt  $(\text{div } Y) \circ \phi = \text{div}_g X$  mit

$$\text{div}_g X = \frac{1}{\sqrt{\det g}} \sum_{i=1}^n \frac{\partial}{\partial x_i} (\sqrt{\det g} X_i).$$

(3) Ist  $v = u \circ \phi$ , so folgt  $(\Delta v) \circ \phi = \Delta_g u$  mit

$$\Delta_g u = \frac{1}{\sqrt{\det g}} \sum_{i,j=1}^n \frac{\partial}{\partial x_i} (\sqrt{\det g} g^{ij} \frac{\partial u}{\partial x_j}).$$

BEWEIS: Für die Formel (1) wähle  $Y = \text{grad } v$  in Lemma 2.1. Für das zugehörige Vektorfeld  $X$ , also  $(\text{grad } v) \circ \phi = D\phi \cdot X$ , folgt

$$X_j = \sum_{i=1}^n g^{ij} \left\langle (\text{grad } v) \circ \phi, D\phi \cdot e_i \right\rangle = \sum_{i=1}^n g^{ij} Dv \cdot (D\phi \cdot e_i) = \sum_{i=1}^n g^{ij} \frac{\partial u}{\partial x_i}.$$

Das zeigt (1). Zum Beweis von (2) verwenden wir partielle Integration. Die Divergenz geht dabei über in den Gradienten, dessen Umrechnung wir bereits kennen. Sei  $\varphi \in C^1(U)$  eine

beliebige Funktion mit  $\varphi = 0$  außerhalb einer abgeschlossenen und beschränkten Teilmenge von  $U$ . Mit  $\psi = \varphi \circ \phi^{-1}$  berechnen wir

$$\begin{aligned}
\int_U \varphi(\operatorname{div} Y) \circ \phi \sqrt{\det g} \, dx &= \int_V \psi \operatorname{div} Y \, dy \quad (\text{Transformationsatz}) \\
&= - \int_V \langle \operatorname{grad} \psi, Y \rangle \, dy \quad (\text{partielle Integration}) \\
&= - \int_U \langle (\operatorname{grad} \psi) \circ \phi, Y \circ \phi \rangle \sqrt{\det g} \, dx \quad (\text{Transformationsatz}) \\
&= - \int_U \langle D\phi \cdot \operatorname{grad}_g \varphi, D\phi \cdot X \rangle \sqrt{\det g} \, dx \quad (\text{Umrechnung Gradient, } Y) \\
&= - \int_U \sum_{j,k=1}^n g_{jk} (\operatorname{grad}_g \varphi)_j X_k \sqrt{\det g} \, dx \quad \text{Gleichung (2.2)} \\
&= - \int_U \sum_{i,k=1}^n \sum_{j=1}^n g_{jk} g^{ij} \frac{\partial \varphi}{\partial x_i} X_k \sqrt{\det g} \, dx \quad (\text{Formel (1)}) \\
&= - \int_U \left( \sum_{i=1}^n \frac{\partial \varphi}{\partial x_i} X_i \right) \sqrt{\det g} \, dx \quad (\sum_{j=1}^n g_{jk} g^{ij} = \delta_{ik}) \\
&= \int_U \varphi \sum_{i=1}^n \frac{\partial}{\partial x_i} (\sqrt{\det g} X_i) \, dx \quad (\text{partielle Integration}).
\end{aligned}$$

Da diese Gleichung für alle  $\varphi$  gilt, folgt mit dem üblichen Schluss

$$(\operatorname{div} Y) \circ \phi \sqrt{\det g} = \sum_{i=1}^n \frac{\partial}{\partial x_i} (\sqrt{\det g} X_i).$$

Damit ist (2) bewiesen. Formel (3) ergibt sich direkt aus (1) und (2), es gilt

$$(\Delta v) \circ \phi = (\operatorname{div} \operatorname{grad} v) \circ \phi = \operatorname{div}_g \operatorname{grad}_g u.$$

□

**Beispiel 2.4** Für Polarkoordinaten im  $\mathbb{R}^3$  berechnet man

$$g(r, \theta, \varphi) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & r^2 & 0 \\ 0 & 0 & r^2 \sin^2 \theta \end{pmatrix}.$$

Damit ergeben sich die Operatoren wie folgt:

$$\begin{aligned}
\operatorname{grad}_g u &= \frac{\partial u}{\partial r} e_1 + \frac{1}{r^2} \frac{\partial u}{\partial \theta} e_2 + \frac{1}{r^2 \sin^2 \theta} \frac{\partial u}{\partial \varphi} e_3 \\
\operatorname{div}_g X &= \frac{1}{r^2 \sin \theta} \left( \frac{\partial}{\partial r} (r^2 \sin \theta X_1) + \frac{\partial}{\partial \theta} (r^2 \sin \theta X_2) + \frac{\partial}{\partial \varphi} (r^2 \sin \theta X_3) \right) \\
&= \frac{1}{r^2} \frac{\partial}{\partial r} (r^2 X_1) + \frac{1}{\sin \theta} \frac{\partial}{\partial \theta} (\sin \theta X_2) + \frac{\partial}{\partial \varphi} X_3 \\
\Delta_g u &= \frac{1}{r^2} \frac{\partial}{\partial r} \left( r^2 \frac{\partial u}{\partial r} \right) + \frac{1}{r^2 \sin \theta} \frac{\partial}{\partial \theta} \left( \sin \theta \frac{\partial u}{\partial \theta} \right) + \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2 u}{\partial \varphi^2}.
\end{aligned}$$

Wir können  $\Delta_g$  in der Form  $\Delta_g u = \frac{1}{r^2} \frac{\partial}{\partial r} (r^2 \frac{\partial u}{\partial r}) + \frac{1}{r^2} \Delta_{\mathbb{S}^2} u$  schreiben, wobei

$$\Delta_{\mathbb{S}^2} u = \frac{1}{\sin \theta} \frac{\partial}{\partial \theta} (\sin \theta \frac{\partial u}{\partial \theta}) + \frac{1}{\sin^2 \theta} \frac{\partial^2 u}{\partial \varphi^2}.$$

Die Formel (1) liefert die Entwicklung des Gradienten  $(\text{grad } v) \circ \phi$  bezüglich der Basis  $\frac{\partial \phi}{\partial x_i}$ :

$$(\text{grad } v) \circ \phi = \sum_{i,j=1}^n g^{ij} \frac{\partial u}{\partial x_i} \frac{\partial \phi}{\partial x_j} \quad \text{mit } u = v \circ \phi.$$

Die  $\frac{\partial \phi}{\partial x_i}$  bilden im allgemeinen keine Orthonormalbasis, ihre Skalarprodukte sind ja genau die Gramsche Matrix  $g_{ij}$ . Im Fall der Polarkoordinaten sind die  $\frac{\partial \phi}{\partial x_i}$  immerhin orthogonal. Es gilt

$$(\text{grad } v) \circ \phi = \frac{\partial u}{\partial r} \frac{\partial \phi}{\partial r} + \frac{1}{r^2} \frac{\partial u}{\partial \theta} \frac{\partial \phi}{\partial \theta} + \frac{1}{r^2 \sin^2 \theta} \frac{\partial u}{\partial \varphi} \frac{\partial \phi}{\partial \varphi}.$$

Für orthogonale Koordinaten kann man eine Orthonormalbasis im Bild herstellen, indem man die Vektoren  $\frac{\partial \phi}{\partial x_i}$  normiert. Im Fall der Polarkoordinaten wird diese Basis oft mit  $\vec{e}_r, \vec{e}_\theta, \vec{e}_\varphi$  bezeichnet, die Darstellung des Gradienten lautet dann

$$(\text{grad } v) \circ \phi = \frac{\partial u}{\partial r} \vec{e}_r + \frac{1}{r} \frac{\partial u}{\partial \theta} \vec{e}_\theta + \frac{1}{r \sin \theta} \frac{\partial u}{\partial \varphi} \vec{e}_\varphi.$$

Der Ansatz in Satz 2.2 ist aber aus meiner Sicht systematischer, und er ist auch nicht auf orthogonale Koordinaten beschränkt.

### 3 Der Satz von Gauß

In diesem Abschnitt diskutieren wir den Integralsatz von Gauß (Englisch: *divergence theorem*), der eine mehrdimensionale Verallgemeinerung des Hauptsatzes der Differential- und Integralrechnung ist. Dem Charakter nach handelt es sich um ein Erhaltungsgesetz, deshalb ist der Satz ein wichtiges mathematisches Hilfsmittel zum Beispiel in der Strömungslehre und der Elektrodynamik.

**Satz 3.1 (Integralsatz von Gauß)** Sei  $\Omega \subset \mathbb{R}^n$  ein beschränktes Gebiet mit glattem Rand. Bezeichne mit  $\nu(x)$ ,  $x \in \partial\Omega$ , die nach außen weisende Einheitsnormale am Rand. Ist  $X : \bar{\Omega} \rightarrow \mathbb{R}^n$  Vektorfeld mit stetigen Ableitungen bis zum Rand, so gilt

$$\int_{\Omega} \text{div } X \, dx = \int_{\partial\Omega} \langle X, \nu \rangle \, dA.$$

BEWEIS: Wir geben einen Beweis im Spezialfall dass das zugrundeliegende Gebiet ein  $n$ -dimensionaler Quader  $Q$  ist, das heißt  $Q = (a_1, b_1) \times \dots \times (a_n, b_n)$ . Die äußere Einheitsnormale ist dann wie folgt:

$$\nu(x) = \begin{cases} -e_i & \text{für } x \in \partial Q \text{ mit } x_i = a_i \\ e_i & \text{für } x \in \partial Q \text{ mit } x_i = b_i. \end{cases}$$

Wir setzen  $Q_i = (a_1, b_1) \times \dots \times \widehat{(a_i, b_i)} \times \dots \times (a_n, b_n) \subset \mathbb{R}^{n-1}$ , wobei das Dach bedeutet, dass der Faktor wegzulassen ist. Mit dem Satz von Fubini und dem Hauptsatz der Differential- und Integralrechnung folgt

$$\begin{aligned}
\int_Q \operatorname{div} X \, dx &= \sum_{i=1}^n \int_Q \frac{\partial X_i}{\partial x_i} \, dx \\
&= \sum_{i=1}^n \int_{Q_i} \int_{a_i}^{b_i} \frac{\partial X_i}{\partial x_i}(x_1, \dots, x_i, \dots, x_n) \, dx_i \, dx_1 \dots \widehat{dx_i} \dots dx_n \\
&= \sum_{i=1}^n \int_{Q_i} X_i(x_1, \dots, b_i, \dots, x_n) \, dx_1 \dots \widehat{dx_i} \dots dx_n \\
&\quad - \sum_{i=1}^n \int_{Q_i} X_i(x_1, \dots, a_i, \dots, x_n) \, dx_1 \dots \widehat{dx_i} \dots dx_n \\
&= \sum_{i=1}^n \int_{\{x_i=b_i\}} \langle X(x), e_i \rangle \, dA + \sum_{i=1}^n \int_{\{x_i=a_i\}} \langle X(x), -e_i \rangle \, dA \\
&= \int_{\partial Q} \langle X, \nu \rangle \, dA.
\end{aligned}$$

Wir wollen einen zweiten Beweis andeuten, der für allgemeine Gebiete funktioniert und ebenfalls aufschlussreich ist. Für  $x \in \bar{\Omega}$  sei  $d(x) \geq 0$  der Abstand zum Rand von  $\Omega$ . Man kann zeigen: auf der Parallelfäche  $\{x \in \Omega : d(x) = r\}$  ist die nach außen weisende Einheitsnormale  $\nu(x)$  gleich  $-\operatorname{grad} d(x)$ . Definiere nun die reelle Funktion

$$\eta_\varepsilon : [0, \infty) \rightarrow \mathbb{R}, \quad \eta_\varepsilon(r) = \begin{cases} \frac{r}{\varepsilon} & \text{falls } r \leq \varepsilon, \\ 1 & \text{sonst.} \end{cases}$$

Die Funktion  $\varphi_\varepsilon(x) = \eta_\varepsilon(d(x))$  hat Null-Randwerte, also folgt (siehe Lemma 1.1)

$$0 = \int_\Omega \operatorname{div}(\varphi_\varepsilon X) \, dx = \int_\Omega \varphi_\varepsilon \operatorname{div} X \, dx + \int_\Omega \langle \operatorname{grad} \varphi_\varepsilon, X \rangle \, dx.$$

Es gilt  $\lim_{\varepsilon \searrow 0} \varphi_\varepsilon(x) = 1$  für alle  $x \in \Omega$ , deshalb folgt

$$\lim_{\varepsilon \searrow 0} \int_\Omega \varphi_\varepsilon \operatorname{div} X \, dx = \int_\Omega \operatorname{div} X \, dx.$$

Im zweiten Integral ist  $\operatorname{grad} \varphi_\varepsilon(x) = 0$  für  $d(x) > \varepsilon$ . Für  $0 < d(x) < \varepsilon$  ist

$$\operatorname{grad} \varphi_\varepsilon(x) = \eta'_\varepsilon(d(x)) \operatorname{grad} d(x) = -\frac{1}{\varepsilon} \nu(x).$$

Dabei ist  $\nu(x)$  die Normale auf den Parallelfächen. Für das rechte Integral sollte also gelten

$$\int_\Omega \langle \operatorname{grad} \varphi_\varepsilon, X \rangle \, dx = -\frac{1}{\varepsilon} \int_{\{0 < d(x) < \varepsilon\}} \langle X(x), \nu(x) \rangle \, dx \rightarrow - \int_{\partial \Omega} \langle X, \nu \rangle \, dA.$$

□

Der Satz von Gauß gilt auch für Gebiete mit gewissen Singularitäten, zum Beispiel Kanten. Das können wir hier nicht behandeln. Stattdess machen wir noch ein paar Folgerungen und Anwendungen.

**Beispiel 3.1** Wählen wir im Satz von Gauß als Vektorfeld  $X(x) = x$ , so folgt

$$|\Omega| = \frac{1}{n} \int_{\Omega} \operatorname{div} X \, dx = \frac{1}{n} \int_{\partial\Omega} \langle x, \nu \rangle \, dA.$$

Insbesondere folgt  $|\mathbb{B}^n| = \frac{1}{n} |\mathbb{S}^{n-1}|$ .

**Folgerung 3.1 (Greensche Formel)** Sei  $\Omega$  beschränktes Gebiet im  $\mathbb{R}^n$  mit glattem Rand. Dann gilt für  $u, v \in C^2(\overline{\Omega})$

$$\int_{\Omega} (u\Delta v - v\Delta u) \, dx = \int_{\partial\Omega} \left( u \frac{\partial v}{\partial \nu} - v \frac{\partial u}{\partial \nu} \right) \, dA.$$

BEWEIS: Wegen  $\operatorname{div}(u \operatorname{grad} v) = \langle \operatorname{grad} u, \operatorname{grad} v \rangle + u\Delta v$  liefert der Satz von Gauß

$$\int_{\Omega} (u\Delta v + \langle \operatorname{grad} u, \operatorname{grad} v \rangle) \, dx = \int_{\partial\Omega} u \frac{\partial v}{\partial \nu} \, dA.$$

Die Behauptung folgt nun durch Vertauschen von  $u, v$  und Subtraktion.  $\square$

**Beispiel 3.2 (Strömungen)** Sei  $v : \mathbb{R}^n \times \mathbb{R} \rightarrow \mathbb{R}^n$ ,  $v = v(x, t)$ , ein evtl. zeitabhängiges Geschwindigkeitsfeld auf  $\mathbb{R}^n$ . Wir bezeichnen mit  $\phi^x(t)$  die Bahnkurve eines Partikels, dass zur Zeit  $t = 0$  bei  $x$  startet und im Punkt  $\phi^x(t)$  die Geschwindigkeit  $v(\phi^x(t), t)$  hat. Zusammenfassen aller Bahnkurven ergibt eine Abbildung, den Fluss von  $v$ ,

$$\phi : \mathbb{R}^n \times \mathbb{R} \rightarrow \mathbb{R}^n, \phi(x, t) = \phi^x(t).$$

Nach Definition gilt

$$\frac{\partial \phi}{\partial t}(x, t) = v(\phi(x, t), t) \quad \text{und} \quad \phi(x, 0) = x.$$

Fließen wir erst von  $x$  nach  $\phi(x, t)$ , und dann weiter von  $\phi(x, t)$  nach  $\phi(\phi(x, t), s)$ , so ist das im Ergebnis dasselbe wie ein Fluss direkt nach  $\phi(x, s + t)$ . Also gilt für die Abbildungen  $\phi_s : \mathbb{R}^n \rightarrow \mathbb{R}^n$ ,  $\phi_s(x) = \phi(x, s)$ ,

$$\phi_s \circ \phi_t = \phi_{s+t}.$$

Insbesondere ist  $\phi_s \circ \phi_{-s} = \phi_0 = \operatorname{id}_{\mathbb{R}^n}$ , und die Abbildungen  $\phi_s$  sind bijektiv. Zur Zeit  $t = 0$  berechnen wir

$$\frac{\partial}{\partial t}(D_x \phi) = D_x \left( \frac{\partial \phi}{\partial t} \right) = D_x v,$$

insbesondere wegen  $D_x \phi(x, 0) = \operatorname{Id}_{\mathbb{R}^n}$

$$\frac{\partial}{\partial t}(\det D_x \phi)|_{t=0} = \operatorname{tr} D_x v = \operatorname{div} v.$$

Sei nun  $\varrho : \mathbb{R}^n \times \mathbb{R} \rightarrow \mathbb{R}$  die evtl. zeitabhängige Dichteverteilung der Strömung. Für  $\Omega \subset \mathbb{R}^n$  beschränkt sei  $m_{\Omega}(t)$  die Masse im Gebiet  $\phi_t(\Omega)$ . Mit dem Transformationssatz und Differentiation unter dem Integral folgt

$$\begin{aligned} m'_{\Omega}(0) &= \frac{d}{dt} \int_{\phi_t(\Omega)} \varrho(y, t) \, dy|_{t=0} \\ &= \frac{d}{dt} \int_{\Omega} \varrho(\phi_t(x), t) \det D_x \phi(x, t) \, dx|_{t=0} \\ &= \int_{\Omega} \left( D_x \varrho \cdot v + \frac{\partial \varrho}{\partial t} + \varrho \operatorname{div} v \right) \, dx \\ &= \int_{\Omega} \left( \frac{\partial \varrho}{\partial t} + \operatorname{div}(\varrho v) \right) \, dx. \end{aligned}$$

Für beliebige  $t \in \mathbb{R}$  ergibt sich aus dem Fall  $t = 0$  dieselbe Formel, und zwar

$$m'_{\Omega}(t) = \frac{d}{ds} \int_{\phi_s(\phi_t(\Omega))} \varrho(y, s+t) dy|_{s=0} = \int_{\phi_t(\Omega)} \left( \frac{\partial \varrho}{\partial t} + \operatorname{div}(\varrho v) \right) dx.$$

Gilt daher auf  $\mathbb{R}^n \times \mathbb{R}$  die Kontinuitätsgleichung

$$\frac{\partial \varrho}{\partial t} + \operatorname{div}(\varrho v) = 0,$$

so ist die Funktion  $m_{\Omega}(t)$  für alle  $\Omega$  zeitlich konstant. Wählen wir  $\varrho \equiv 1$ , so ist  $m_{\Omega}(t)$  das Volumen von  $\phi_t(\Omega)$ , und der Satz von Gauß impliziert

$$m'_{\Omega}(0) = \int_{\partial\Omega} \langle v, \nu \rangle dA.$$

Das Flächenelement  $dA$  trägt mit der Normalkomponente  $\langle v, \nu \rangle$  zur Änderung des eingeschlossenen Volumens bei, was eine sehr anschauliche Deutung des Gaußschen Satzes ist.