0 Übersicht

1. Euklidische und unitäre Vektorräume. (Nachtrag "Lineare Algebra")
Skalarprodukt, Norm, Abstand, Bilinearform, Sesquilinearform, orthogonale und unitäre Endomorphismen
Literatur: G. Fischer, "Lineare Algebra," Vieweg 2005.

2. Differentialrechnung im $\mathbb{R}^n$
Grenzwerte und Stetigkeit, Kompaktheit, partielle Ableitungen, totales Differential, Taylor-Formel
Literatur: O. Forster, "Analysis 2," Vieweg 2005.

3. Integralrechnung im $\mathbb{R}^n$
Funktionen mit kompaktem Träger, Variablentransformation, partielle Integration, Lebesgue-Integral, Fourier-Integrale, Gaußscher Integralsatz
Literatur: K. Königsberger, "Analysis 2," Springer 2004.

1 Euklidische und unitäre Vektorräume

Im Vektorraum $\mathbb{R}^n$ und $\mathbb{C}^n$ läßt sich als Zusatzstruktur ein *Skalarprodukt* einführen, welches dann insbesondere einen *Abstand* definiert. Damit lassen sich geometrische Größen wie Längen und Winkel berechnen, aber auch die Analysis vom $\mathbb{R}^1$ auf den $\mathbb{R}^n$ verallgemeinern.

1.1 Kanonische Skalarprodukte im $\mathbb{R}^n$ und $\mathbb{C}^n$

Definition 1.1 Die durch

$$\langle x,y\rangle := x_1y_1 + x_2y_2 + \cdots + x_ny_n\,, \qquad \text{für} \quad \begin{array}{l} x=(x_1,\dots,x_n)\in\mathbb{R}^n\,,\\ y=(y_1,\dots,y_n)\in\mathbb{R}^n \end{array}$$

definierte Abbildung $\langle\ ,\ \rangle : \mathbb{R}^n\times\mathbb{R}^n\to\mathbb{R}$ heißt das *kanonische Skalarprodukt* im $\mathbb{R}^n$.

Fassen wir $x,y\in\mathbb{R}^n$ als Spaltenvektoren auf, dann gilt $\langle x,y\rangle = x^t\cdot y = y^t\cdot x$ als Matrixmultiplikation. Für $x,x',y,y'\in\mathbb{R}^n$ und $\lambda,\lambda'\in\mathbb{R}$ gilt:

$$\begin{aligned} &\langle\lambda x+\lambda' x',y\rangle = \lambda\langle x,y\rangle + \lambda'\langle x',y\rangle\,, \qquad \langle x,\lambda y+\lambda' y'\rangle = \lambda\langle x,y\rangle + \lambda'\langle x,y'\rangle\,,\\ &\langle x,y\rangle = \langle y,x\rangle\,,\\ &\langle x,x\rangle \geq 0\,, \qquad\qquad \langle x,x\rangle = 0 \quad\Leftrightarrow\quad x=0\,. \end{aligned}$$

Die letzte Eigenschaft erlaubt die Definition einer Abbildung

$$\|\ \| : \mathbb{R}^n\to\mathbb{R}_+\,, \qquad x\mapsto \|x\| := \sqrt{\langle x,x\rangle}\,,$$

der *Norm.* Für $x \in \mathbb{R}^1$ gilt $\|x\| = |x|$, was die Bezeichnung motiviert. Die Norm führt dann auf die Definition des *Abstands*

$$d : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}_+ \,, \qquad d(x,y) := \|x-y\| \,.$$

Man erhält $d(x,y) = \sqrt{(x_1-y_1)^2 + \cdots + (x_n-y_n)^2}$, also den *Satz des Pythagoras* in n Dimensionen.

Norm und Abstand haben folgende Eigenschaften (für $x,y,z \in \mathbb{R}^n$ und $\lambda \in \mathbb{R}$):

(N1) $\|x\| = 0 \quad \Leftrightarrow \quad x = 0$,

(N2) $\|\lambda x\| \leq |\lambda| \, \|x\|$,

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

und

(D1) $d(x,y) = 0 \quad \Leftrightarrow \quad x = y$,

(D2) $d(x,y) = d(y,x)$ (Symmetrie),

(D3) $d(x,z) \leq d(x,y) + d(y,z)$ (Dreiecksungleichung).

Einzig die Dreiecksungleichung folgt nicht sofort aus der Definition des Skalarprodukts. Wir beweisen deshalb

Satz 1.1 (Ungleichung von Cauchy-Schwarz) *Für $x,y \in \mathbb{R}^n$ gilt $|\langle x,y\rangle| \leq \|x\| \, \|y\|$ und $|\langle x,y\rangle| = \|x\| \, \|y\|$ genau dann, wenn x,y linear abhängig sind.*

Beweis. Sei $x \neq 0$ (für $x = 0$ ist die Ungleichung offensichtlich erfüllt). Dann gilt für beliebiges $\lambda \in \mathbb{R}$

$$\begin{aligned} 0 \leq \langle \lambda x - y, \lambda x - y\rangle &= \lambda^2 \|x\|^2 - 2\lambda \langle x,y\rangle + \|y\|^2 \\ &= \|x\|^2 \Big(\lambda - \frac{\langle x,y\rangle}{\|x\|^2}\Big)^2 + \frac{1}{\|x\|^2}\Big(\|x\|^2\|y\|^2 - (\langle x,y\rangle)^2\Big) \,. \end{aligned}$$

Für $\lambda = \frac{\langle x,y\rangle}{\|x\|^2}$ wird direkt die Ungleichung von Cauchy-Schwarz erhalten.

Gilt das Gleichheitszeichen, ist also $\langle x,y\rangle = \pm\|x\| \, \|y\|$, dann gilt für $\lambda = \pm\frac{\|y\|}{\|x\|}$ die Identität $\langle \lambda x - y, \lambda x - y\rangle = 0$, also $y = \lambda x$. $\square$

Die Dreiecksungleichung folgt nun aus Cauchy-Schwarz durch

$$\begin{aligned} \|x+y\|^2 = \langle x+y, x+y\rangle &= \|x\|^2 + 2\langle x,y\rangle + \|y\|^2 \leq \|x\|^2 + 2|\langle x,y\rangle| + \|y\|^2 \\ &\leq (\|x\| + \|y\|)^2 \,. \end{aligned}$$

Das Skalarprodukt definiert neben dem Abstand $d(x,y)$ zweier Punkte $x,y \in \mathbb{R}^n$ auch den *Winkel* zwischen den Vektoren x,y durch

$$\measuredangle(x,y) = \arccos \frac{\langle x,y\rangle}{\|x\| \, \|y\|} \,.$$

Es sei erwähnt, daß man (N1)–(N3) und (D1)–(D3) als abstrakte Definition einer Norm bzw. eines Abstands ansehen kann. Auf diese Weise lassen sich *normierte Räume* $(V, \| \ \|)$ und *metrische Räume* (X, d) definieren, welche nicht notwendigerweise über ein Skalarprodukt bzw. eine Norm erzeugt werden. Ein metrischer Raum muß nicht einmal ein Vektorraum sein.

Die analoge Konstruktion im komplexen Vektorraum $\mathbb{C}^n$ ist:

Definition 1.2 Die durch

$$\langle w, z\rangle := \overline{w_1}z_1 + \overline{w_2}z_2 + \cdots + \overline{w_n}z_n\ , \qquad \text{für} \quad \begin{array}{l} w = (w_1, \dots, w_n) \in \mathbb{C}^n\ , \\ z = (z_1, \dots, z_n) \in \mathbb{C}^n \end{array}$$

definierte Abbildung $\langle\ ,\ \rangle : \mathbb{C}^n \times \mathbb{C}^n \to \mathbb{C}$ heißt das *kanonische Skalarprodukt* im $\mathbb{C}^n$.

Wir verzichten auf eine symbolische Unterscheidung zum reellen Skalarprodukt. Sollte die Kennzeichnung des Vektorraums notwendig sein, so schreiben wir $\langle\ ,\ \rangle_{\mathbb{R}^n}$ bzw. $\langle\ ,\ \rangle_{\mathbb{C}^n}$. Für $w, w', z, z' \in \mathbb{C}^n$ und $\lambda, \lambda' \in \mathbb{C}$ gilt:

$$\begin{aligned} \langle \lambda w + \lambda' w', z\rangle &= \overline{\lambda}\langle w, z\rangle + \overline{\lambda'}\langle w', z\rangle\ , \quad & \langle w, \lambda z + \lambda' z'\rangle &= \lambda\langle w, z\rangle + \lambda'\langle w, z'\rangle\ , \\ \langle w, z\rangle &= \overline{\langle z, w\rangle}\ , & & \\ \langle z, z\rangle &\in \mathbb{R}_+\ , & \langle z, z\rangle = 0 \quad &\Leftrightarrow \quad z = 0\ . \end{aligned}$$

Wir haben in Definition 1.2 die Konvention der Physiker verwendet, daß die komplexe Konjugation im *linken* Eintrag des Skalarprodukts auftritt. Mathematiker bevorzugen die komplexe Konjugation im rechten Eintrag. Die physikalische Konvention erweist sich als sehr hilfreich in einem auf Dirac zurückgehenden symbolischen bra–ket Kalkül für Hilbert-Räume ($\mathbb{C}^n$ mit dem kanonischen Skalarprodukt ist ein Hilbert-Raum) und Operatoren auf Hilbert-Räumen.

Das komplexe Skalarprodukt definiert wie im Reellen eine Norm

$$\| \ \| : \mathbb{C}^n \to \mathbb{R}_+\ , \qquad z \mapsto \|z\| := \sqrt{\langle z, z\rangle}\ .$$

Die Normaxiome (N1)–(N3) sind erfüllt, und insbesondere gilt:

Satz 1.2 (Ungleichung von Cauchy-Schwarz) *Für $w, z \in \mathbb{C}^n$ gilt $|\langle w, z\rangle| \leq \|w\|\,\|z\|$ und $|\langle w, z\rangle| = \|w\|\,\|z\|$ genau dann, wenn w, z linear abhängig sind.*

Beweis. Sei $w \neq 0$ (für $w = 0$ ist die Ungleichung offensichtlich erfüllt). Dann gilt für beliebiges $\lambda \in \mathbb{C}$

$$\begin{aligned} 0 \leq \langle \lambda w - z, \lambda w - z\rangle &= \lambda\overline{\lambda}\|w\|^2 - \overline{\lambda}\langle w, z\rangle - \lambda\langle z, w\rangle + \|z\|^2 \\ &= \|w\|^2\left|\lambda - \frac{\langle w, z\rangle}{\|w\|^2}\right|^2 + \frac{1}{\|w\|^2}\Big(\|w\|^2\|z\|^2 - \big|\langle w, z\rangle\big|^2\Big)\ . \end{aligned}$$

Für $\lambda = \frac{\langle w,z\rangle}{\|w\|^2}$ wird direkt die Ungleichung von Cauchy-Schwarz erhalten.

Gilt das Gleichheitszeichen, ist also $\langle w, z\rangle = \mathrm{e}^{\mathrm{i}\alpha}\|w\|\,\|z\|$ für $\alpha \in [0, 2\pi]$, dann gilt für $\lambda = \mathrm{e}^{\mathrm{i}\alpha}\frac{\|z\|}{\|w\|}$ die Identität $\langle \lambda w - z, \lambda w - z\rangle = 0$, also $z = \lambda w$. □

Schließlich wird der $\mathbb{C}^n$ mit dem Abstand

$$d : \mathbb{C}^n \times \mathbb{C}^n \to \mathbb{R}_+ \;, \qquad d(z, w) := \|z - w\|$$

zu einem metrischen Raum (die Axiome (D1)–(D3) sind erfüllt).

Interessant ist die Verbindung zwischen $\mathbb{C}^n$ und dem als reeller Vektorraum identischen $\mathbb{R}^{2n}$ bezüglich der jeweiligen kanonischen Skalarprodukte. Identifizieren wir

$$\begin{aligned} & z = (x_1 + \mathrm{i}y_1,\; x_2 + \mathrm{i}y_2, \ldots, x_n + \mathrm{i}y_n)^t \in \mathbb{C}^n \\ \Leftrightarrow \quad & \tilde{z} = (x_1, y_1, x_2, y_2, \ldots, x_n, y_n)^t \in \mathbb{R}^{2n} \;, \end{aligned}$$

(jeweils als Spaltenvektoren aufgefaßt) dann gilt

$$\langle w, z\rangle_{\mathbb{C}^n} = \langle \tilde{w}, \tilde{z}\rangle_{\mathbb{R}^{2n}} + \mathrm{i}\tilde{w}^t \cdot J \cdot \tilde{z} \;, \quad J = \left(\begin{array}{c|c|c|c} \begin{matrix} 0 & 1 \\ -1 & 0 \end{matrix} & 0 & \cdots & 0 \\ \hline 0 & \begin{matrix} 0 & 1 \\ -1 & 0 \end{matrix} & \cdots & 0 \\ \hline \vdots & \vdots & \ddots & \vdots \\ \hline 0 & 0 & \cdots & \begin{matrix} 0 & 1 \\ -1 & 0 \end{matrix} \end{array}\right) .$$

Der zweite Term ist antisymmetrisch in $\tilde{w}, \tilde{z} \in \mathbb{R}^{2n}$, d.h. $\tilde{w}^t \cdot J \cdot \tilde{z} = -\tilde{z}^t \cdot J \cdot \tilde{w}$. Deshalb gilt $\tilde{z}^t \cdot J \cdot \tilde{z} = 0$, so daß die Normen und die Abstände im $\mathbb{C}^n$ und im $\mathbb{R}^{2n}$ übereinstimmen:

$$\|z\| = \|\tilde{z}\| \;, \qquad d(z, w) = d(\tilde{z}, \tilde{w}) \;.$$

1.2 Bilinearformen und Dualräume

Bilinearformen stellen wichtige Verallgemeinerungen der Skalarprodukte dar.

Definition 1.3 Sei V ein Vektorraum über dem Körper K. Eine Abbildung $s : V \times V \to K$ heißt *Bilinearform*, wenn

(B1) $s(\lambda v + \lambda' v', w) = \lambda s(v, w) + \lambda' s(v', w) \quad \forall v, v', w \in V, \quad \forall \lambda, \lambda' \in K$

(B2) $s(v, \lambda w + \lambda' w') = \lambda s(v, w) + \lambda' s(v, w') \quad \forall v, w, w' \in V, \quad \forall \lambda, \lambda' \in K$

Eine Bilinearform s heißt *symmetrisch*, wenn

(BS) $s(v, w) = s(w, v) \quad \forall v, w \in V,$

und *alternierend* (oder *antisymmetrisch*, *schiefsymmetrisch*), wenn

(BA) $s(v,w) = -s(w,v) \quad \forall v,w \in V$.

Beispiele für symmetrische Bilinearformen sind das kanonische Skalarprodukt im $\mathbb{R}^n$ oder für den Vektorraum $V = \mathcal{C}([a,b],\mathbb{R})$ der stetigen reellwertigen Funktionen über dem Intervall $[a,b]$ die Abbildung

$$s : V \times V \to \mathbb{R} , \qquad s(f,g) = \int_a^b dt\ f(t)\, g(t) .$$

In endlich-dimensionalen Vektorräumen können Bilinearformen durch Matrizen beschrieben werden. Ist $\mathcal{B} = (v_1, \ldots, v_n)$ eine Basis von V, dann legt die *darstellende Matrix*

$$M_{\mathcal{B}}(s) = (s_{ij}) \in M(n \times n, K) , \qquad s_{ij} := s(v_i, v_j)$$

die Bilinearform s eindeutig fest. Denn haben $v, w \in V$ bezüglich der Basis $\mathcal{B}$ die Koordinaten $x = \Phi_{\mathcal{B}}^{-1}(v) = (x_1, \ldots, x_n) \in K^n$ bzw. $y = \Phi_{\mathcal{B}}^{-1}(w) = (y_1, \ldots, y_n) \in K^n$, dann folgt aus der Bilinearität von s die Identität

$$s(v,w) = \sum_{i,j=1}^{n} s_{ij} x_i y_j .$$

Offenbar ist s genau dann (anti-)symmetrisch, wenn (s_{ij}) eine (anti-)symmetrische Matrix ist.

Umgekehrt definiert zu gegebener Basis $\mathcal{B} = (v_1, \ldots, v_n)$ von V jede Matrix $A = (a_{ij}) \in M(n \times n, K)$ eine Bilinearform auf V durch $s(v_i, v_j) = a_{ij}$.

Bilinearformen stehen in Verbindung zum *Dualraum* eines Vektorraumes:

Definition 1.4 Ist V ein Vektorraum über K, dann heißt

$$V^* := \mathrm{Hom}_K(V,K) = \{\varphi : V \to K \text{ linear}\}$$

der *Dualraum* von V. Die Elemente $\varphi \in V^*$ heißen *Linearformen* oder *lineare Funktionale* auf V.

Zum Beispiel ist für den Vektorraum $V = \mathcal{C}([a,b],\mathbb{R})$ der stetigen reellwertigen Funktionen über dem Intervall $[a,b]$ das Integral

$$\phi = \int_a^b dt \in V^* , \qquad \phi(f) = \int_a^b dt\ f(t)$$

ein lineares Funktional.

Ist V ein endlich-dimensionaler Vektorraum und $\mathcal{B} = (v_1, \ldots, v_n)$ eine Basis von V, dann gibt es nach Satz 2.24 aus dem letzten Semester zu jedem Index $1 \le i \le n$ genau eine lineare Abbildung $v_i^* : V \to K$ mit $v_i^*(v_j) = \delta_{ij}$.

Satz 1.3 *Ist $\mathcal{B} = (v_1, \dots, v_n)$ Basis eines endlich-dimensionalen Vektorraums V über K, dann ist $\mathcal{B}^* = (v_1^*, \dots, v_n^*)$ eine Basis (die duale Basis) des Dualraums V^*.*

Damit definiert jede Basis $\mathcal{B} = (v_1, \dots, v_n)$ von V einen Isomorphismus $\Psi_{\mathcal{B}} : V \to V^*$ durch $\Psi_{\mathcal{B}}(v_i) = v_i^*$.

Beweis. Wir nutzen folgende äquivalente Definition (nach Satz 2.4 aus dem letzten Semester) einer Basis $\mathcal{C} = (w_1, \dots, w_r)$ von V^*: Zu jedem $\varphi \in V^*$ gibt es eindeutig bestimmte $\lambda_1, \dots, \lambda_r \in K$ mit $\varphi = \lambda_1 w_1 + \cdots + \lambda_r w_r$. Tatsächlich findet man für $w_i \mapsto v_i^*$, $1 \leq i \leq r = n$, die Beziehung

$$\varphi(v_i) = \lambda_i \ ,$$

so daß die Koeffizienten λ_i eindeutig bestimmt sind. $\square$

Unter Basiswechsel von $\mathcal{B} = (v_1, \dots, v_n)$ nach $\mathcal{A} = (w_1, \dots, w_n)$, beschrieben durch die Transformationsmatrix $M^{\mathcal{B}}_{\mathcal{A}}(\mathrm{id}_V) = (a_{ij})$ mit

$$(\mathrm{id}_V)(v_j) = v_j = \sum_{i=1}^n a_{ij} \cdot w_i \ ,$$

gilt für die Transformationsmatrix $M^{\mathcal{B}^*}_{\mathcal{A}^*}(\mathrm{id}_{V^*}) = (a^*_{kl})$ der dualen Basen (von $\mathcal{B}^* = (v_1^*, \dots, v_n^*)$ nach $\mathcal{A}^* = (w_1^*, \dots, w_n^*)$):

$$\delta_{lj} = v_l^*(v_j) = \Big(\sum_{k=1}^n a^*_{kl} w_k^*\Big)\Big(\sum_{i=1}^n a_{ij} w_i\Big) = \sum_{i,k=1}^n a^*_{kl} a_{ij} \delta_{ik} = \sum_{i=1}^n a^*_{il} a_{ij} \ ,$$

also

$$M^{\mathcal{B}^*}_{\mathcal{A}^*}(\mathrm{id}_{V^*}) = \Big(\big(M^{\mathcal{B}}_{\mathcal{A}}(\mathrm{id}_V)\big)^{-1}\Big)^t \ .$$

Jede Bilinearform $s : V \times V \to K$ definiert eine lineare Abbildung $\tilde{s} : V \to V^*$ durch

$$\big(\tilde{s}(w)\big)(v) := s(v, w) \ , \qquad \forall v, w \in V \ .$$

Die Bilinearität von s stellt sicher, daß $\tilde{s}$ eine lineare Abbildung ist und daß jedes $\tilde{s}(w)$ ein lineares Funktional auf V ist. Ist $\mathcal{B} = (v_1, \dots, v_n)$ eine Basis von V, dann haben wir

$$\big(\tilde{s}(v_j)\big)(v_k) = s(v_k, v_j) = s_{kj} = \Big(\sum_{i=1}^n s_{ij} v_i^*\Big)(v_k) \ , \qquad \text{also} \quad M_{\mathcal{B}}(s) = M^{\mathcal{B}}_{\mathcal{B}^*}(\tilde{s}) \ .$$

Die darstellende Matrix der Bilinearform s bezüglich der Basis $\mathcal{B}$ stimmt also mit der darstellenden Matrix von $\tilde{s} : V \to V^*$ bezüglich der Basis $\mathcal{B}$ und ihrer dualen Basis $\mathcal{B}^*$ überein. Bei Basiswechsel von $\mathcal{B}$ nach $\mathcal{A}$ gilt deshalb:

$$M_{\mathcal{B}}(s) = M^{\mathcal{B}}_{\mathcal{B}^*}(\tilde{s}) = M^{\mathcal{A}^*}_{\mathcal{B}^*}(\mathrm{id}_{V^*}) \cdot M^{\mathcal{A}}_{\mathcal{A}^*}(\tilde{s}) \cdot M^{\mathcal{B}}_{\mathcal{A}}(\mathrm{id}_V) = \big(T^{\mathcal{B}}_{\mathcal{A}}\big)^t \cdot M_{\mathcal{A}}(s) \cdot T^{\mathcal{B}}_{\mathcal{A}} \ ,$$

mit $T^{\mathcal{B}}_{\mathcal{A}} := M^{\mathcal{B}}_{\mathcal{A}}(\mathrm{id}_V)$. Zum Vergleich: Die Transformationsformel für die darstellende Matrix von Endomorphismen $F \in \mathrm{End}(V)$ ist

$$M^{\mathcal{B}}_{\mathcal{B}}(F) = T^{\mathcal{A}}_{\mathcal{B}}(\mathrm{id}_V) \cdot M^{\mathcal{A}}_{\mathcal{A}}(F) \cdot M^{\mathcal{B}}_{\mathcal{A}}(\mathrm{id}_V) = \big(T^{\mathcal{B}}_{\mathcal{A}}\big)^{-1} \cdot M^{\mathcal{A}}_{\mathcal{A}}(F) \cdot T^{\mathcal{B}}_{\mathcal{A}} .$$

Definition 1.5 Sei V ein endlich-dimensionaler Vektorraum über K. Eine Bilinearform auf V heißt *nicht ausgeartet*, wenn $\mathrm{rang}(\tilde{s}) = \dim(V)$.

Übersetzt in die darstellenden Matrizen: die Bilinearform s ist genau dann nicht ausgeartet, wenn $\mathrm{rang}(M_{\mathcal{B}}(s)) = n$ für eine beliebige Basis $\mathcal{B} = (v_1, \dots, v_n)$ von V. Aus den allgemeinen Dimensionssätzen folgt, daß jede nichtausgeartete Bilinearform auf V einen Isomorphismus $\tilde{s} : V \to V^*$ definiert. Dessen Inverses $\tilde{s}^{-1} : V^* \to V$ definiert dann eine nichtausgeartete Bilinearform s^{-1} auf V^* durch

$$s^{-1}(v^*, w^*) = v^*\big(\tilde{s}^{-1}(w^*)\big) .$$

Dann gilt

$$M_{\mathcal{B}^*}(s^{-1}) = M^{\mathcal{B}^*}_{\mathcal{B}}(\tilde{s}^{-1}) = \big(M^{\mathcal{B}}_{\mathcal{B}^*}(\tilde{s})\big)^{-1} = \big(M_{\mathcal{B}}(s)\big)^{-1} .$$

1.3 Sesquilinearformen

Selbstverständlich kann man Bilinearformen auch über komplexen Vektorräumen betrachten. Es stellt sich aber heraus, daß die nun zu definierenden Sesquilinearformen wichtiger sind, da sie (unter gewissen Bedingungen) eine Norm induzieren.

Definition 1.6 Seien V, W komplexe Vektorräume. Eine Abbildung $F : V \to W$ heißt *semilinear*, wenn

$$F(\lambda v + \lambda' v') = \overline{\lambda} F(v) + \overline{\lambda'} F(v') \qquad \forall v, v' \in V , \quad \forall \lambda, \lambda' \in \mathbb{C} .$$

Eine Abbildung $s : V \times V \to \mathbb{C}$ heißt *Sesquilinearform*, wenn

($\overline{\mathrm{B}}$1) $s(\lambda v + \lambda' v', w) = \overline{\lambda} s(v, w) + \overline{\lambda'} s(v', w) \qquad \forall v, v', w \in V , \quad \forall \lambda, \lambda' \in \mathbb{C}$

(B2) $s(v, \lambda w + \lambda' w') = \lambda s(v, w) + \lambda' s(v, w') \qquad \forall v, w, w' \in V \quad \forall \lambda, \lambda' \in \mathbb{C}$

Eine Sesquilinearform s heißt *hermitesch*, wenn

(BH) $s(v, w) = \overline{s(w, v)} \quad \forall v, w \in V$.

Wir haben wieder die physikalische Konvention zur komplexen Konjugation gewählt. Das Standardbeispiel für eine Sesquilinearform ist das kanonische Skalarprodukt im $\mathbb{C}^n$.

Sesquilinearformen auf endlich-dimensionalen (komplexen) Vektorräumen werden durch Matrizen beschrieben: Ist $\mathcal{B} = (v_1, \dots, v_n)$ eine (komplexe) Basis von V, dann ist

$$M_{\mathcal{B}}(s) = (s_{ij}) \in M(n \times n, \mathbb{C}) , \qquad s_{ij} := s(v_i, v_j) ,$$

die *darstellende Matrix* der Sesquilinearform s. Haben $v, w \in V$ bezüglich $\mathcal{B}$ die Koordinaten $x = \Phi_{\mathcal{B}}^{-1}(v) = (x_1, \ldots, x_n) \in \mathbb{C}^n$ bzw. $y = \Phi_{\mathcal{B}}^{-1}(w) = (y_1, \ldots, y_n) \in \mathbb{C}^n$, dann ist

$$s(v, w) = \sum_{i,j=1}^{n} s_{ij} \overline{x_i} y_j \ .$$

Außerdem ist s genau dann hermitesch, wenn $s_{ij} = \overline{s_{ji}}$. (Die Abbildung

$$^* : M(n, \mathbb{C}) \to M(n, \mathbb{C}) \ , \qquad A \mapsto A^* := \overline{A^t}$$

heißt *hermitesche Konjugation* und Matrizen $A \in M(n, \mathbb{C})$ mit $A = A^*$ heißen hermitesch.) Beim Basiswechsel von $\mathcal{B}$ nach $\mathcal{A}$ gilt

$$M_{\mathcal{B}}(s) = \left(T_{\mathcal{A}}^{\mathcal{B}}\right)^* \cdot M_{\mathcal{A}}(s) \cdot T_{\mathcal{A}}^{\mathcal{B}} \ .$$

1.4 Skalarprodukte

Definition 1.7 Sei $\mathbb{K} = \mathbb{R}$ oder $\mathbb{K} = \mathbb{C}$ und sei V ein Vektorraum über $\mathbb{K}$. Eine symmetrische Bilinearform (bzw. hermitesche Form) $s : V \times V \to \mathbb{K}$ heißt *positiv definit*, wenn

$$s(v, v) > 0 \qquad \forall v \in V \text{ mit } v \neq 0 \ .$$

Man beachte, daß für hermitesche Formen stets $s(v, v) \in \mathbb{R}$ ist. Die Bedingung für positive Definitheit kann in dieser Form natürlich nicht überprüft werden. Insbesondere genügt es nicht, sie für alle Vektoren v_i einer Basis $(v_1, \ldots, v_n)$ zu überprüfen! Wir werden später ein brauchbares Kriterium für positive Definitheit finden.

Vereinfachend nennen wir eine positiv definite symmetrische Bilinearform bzw. eine positiv definite hermitesche Form ein *Skalarprodukt*, für das wir wie zuvor $s(v, w) \mapsto \langle v, w \rangle$ schreiben.

Definition 1.8 Ein reeller bzw. komplexer Vektorraum mit Skalarprodukt heißt *euklidischer* bzw. *unitärer* Vektorraum.

Jedes Skalarprodukt auf V definiert eine Norm auf V durch $\|v\| = \sqrt{\langle v, v \rangle}$. Für allgemeine Skalarprodukte gilt unverändert die Ungleichung von Cauchy-Schwarz $|\langle v, w \rangle| \leq \|v\| \, \|w\|$ (Gleichheit bei linearer Abhängigkeit). Des weiteren erhalten wir einen Abstand $d(v, w) = \|v - w\|$ zwischen Vektoren $v, w \in V$ und die Winkelformel $\cos \measuredangle(v, w) = \frac{\langle v, w \rangle}{\|v\| \, \|w\|}$. Besonders interessant sind orthogonale Vektoren:

Definition 1.9 Sei V ein euklidischer bzw. unitärer Vektorraum.

i) Zwei Vektoren $v, w \in V$ heißen *orthogonal (bezüglich des Skalarprodukts)*, geschrieben $v \perp w$, wenn $\langle v, w \rangle = 0$.

ii) Zwei Untervektorräume $U, W \subset V$ heißen orthogonal, geschrieben $U \perp W$, wenn $u \perp w$ für alle $u \in U$ und $w \in W$.

iii) Ist $U \subset V$ ein Untervektorraum, dann heißt

$$U^\perp := \{v \in V \ : \ v \perp u \ \forall u \in U\}$$

das *orthogonale Komplement* von U in V.

iv) Eine Familie $(v_1, \dots, v_n)$ von Vektoren $v_i \in V$ heißt orthogonal, wenn $v_i \perp v_j$ für alle $i \neq j$. Die Familie heißt *orthonormal*, falls zusätzlich $\|v_i\| = 1$ für alle $1 \leq i \leq n$, und *Orthonormalbasis*, falls $(v_1, \dots, v_n)$ außerdem eine Basis von V ist.

Dazu einige Bemerkungen (die sehr leicht zu beweisen sind):

- Ist U Untervektorraum von V, dann ist $U^\perp$ wieder ein Untervektorraum von V.
- Ist in einer orthogonalen Familie $(v_1, \dots, v_n)$ der Nullvektor nicht enthalten, $v_i \neq 0$, so ist $(\frac{1}{\|v_1\|}v_1, \dots, \frac{1}{\|v_n\|}v_n)$ eine orthonormale Familie.
- Ist in einer orthogonalen Familie $(v_1, \dots, v_n)$ der Nullvektor nicht enthalten, $v_i \neq 0$, dann ist $(v_1, \dots, v_n)$ linear unabhängig: Sei $0 = \lambda_1 v_1 + \dots + \lambda_n v_n$, dann ergibt das Skalarprodukt mit sich selbst

$$0 = |\lambda_1|^2\|v_1\|^2 + \dots |\lambda_n|^2\|v_n\|^2 \qquad \Rightarrow \quad \lambda_1 = \dots = \lambda_n = 0 \; .$$

- Ist $(v_1, \dots, v_n)$ eine Orthonormalbasis, dann gilt

$$v = \lambda_1 v_1 + \dots + \lambda_n v_n \qquad \Rightarrow \quad \lambda_i = \langle v_i, v\rangle \; .$$

Außerdem ist $\|v\|^2| = \lambda_1|^2 + \dots |\lambda_n|^2$.

Satz 1.4 (Orthonormalisierungssatz) *Sei V ein endlich-dimensionaler euklidischer oder unitärer Vektorraum und sei $W \subset V$ ein Untervektorraum mit Orthonormalbasis $(w_1, \dots, w_m)$. Dann gibt es eine Ergänzung zu einer Orthonormalbasis $(w_1, \dots, w_m, w_{m+1}, \dots, w_n)$ von V.*

Insbesondere besitzt jeder endlich-dimensionale euklidische oder unitäre Vektorraum eine Orthonormalbasis.

Beweis (Schmidtsches Orthonormalisierungsverfahren). Für $W = V$ ist alles klar. Andernfalls gibt es einen Vektor $v \in V$ mit $v \notin W$. Seine Projektion auf W ist

$$\tilde{v} := \langle w_1, v\rangle w_1 + \dots + \langle w_m, v\rangle w_m \; .$$

Dann gilt: $(v - \tilde{v}) \perp W$ (leicht nachzurechnen) und $v - \tilde{v} \neq 0$ (sonst wäre $v \in W$). Also ist mit $w_{m+1} := \frac{1}{\|v-\tilde{v}\|}(v - \tilde{v})$ die Familie $(w_1, \dots, w_m, w_{m+1})$ orthonormal, und $W' := \mathrm{span}(w_1, \dots, w_m, w_{m+1})$ ist $(m+1)$-dimensionaler Untervektorraum von V mit Orthonormalbasis $(w_1, \dots, w_{m+1})$. Durch Wiederholung des Verfahrens erhält man eine Orthonormalbasis von V. $\square$

1.5 Orthogonale und unitäre Endomorphismen

Definition 1.10 Sei V ein endlich-dimensionaler euklidischer bzw. unitärer Vektorraum. Ein Endomorphismus $F \in \mathrm{End}(V)$ heißt *orthogonal* bzw. *unitär*, wenn

$$\langle F(v), F(w)\rangle = \langle v, w\rangle \qquad \forall v, w \in V \ .$$

Satz 1.5 *Ist $F \in \mathrm{End}(V)$ orthogonal bzw. unitär, so gilt*

i) $\|F(v)\| = \|v\| \qquad \forall v \in V,$

ii) $v \perp w \quad \Rightarrow \quad F(v) \perp F(w),$

iii) *F ist Isomorphismus und F^{-1} ist ebenfalls orthogonal bzw. unitär,*

iv) *Ist $\lambda \in \mathbb{K}$ Eigenwert von F, so ist $|\lambda| = 1$.*

Beweis. i) und ii) sind klar.

iii) Aus i) folgt, daß F injektiv ist. Als lineare Abbildung zwischen endlich-dimensionalen Vektorräumen ist F damit auch bijektiv. Die Ersetzung $v \mapsto F^{-1}(v)$ liefert Orthogonalität/Unitarität von F^{-1}.

iv) Ist v Eigenvektor von F zum Eigenwert λ, so ist

$$\|v\| = \|F(v)\| = \|\lambda v\| = |\lambda|\|v\| \qquad \Rightarrow \quad |\lambda| = 1 \text{ wegen } v \neq 0 \ . \qquad \square$$

Damit erhalten orthogonale/unitäre Endomorphismen die Normen, Abstände und Winkel und haben somit eine wichtige geometrische Interpretation als *Isometrien*. Es gilt sogar die Umkehrung:

Satz 1.6 *Sei V ein endlich-dimensionaler euklidischer bzw. unitärer Vektorraum. Ist $\|F(v)\| = \|v\|$ für alle $v \in V$, so ist F orthogonal bzw. unitär.*

Beweis.

$$\begin{aligned}
\langle F(v), F(w)\rangle &= \frac{1}{2}\Big(\langle F(v)+F(w), F(v)+F(w)\rangle - \langle F(v), F(v)\rangle - \langle F(w), F(w)\rangle\Big) \\
&= \frac{1}{2}\Big(\|F(v+w)\|^2 - \|F(v)\|^2 - \|F(w)\|^2\Big) \\
&= \frac{1}{2}\Big(\|v+w\|^2 - \|v\|^2 - \|w\|^2\Big) \\
&= \frac{1}{2}\Big(\langle v+w, v+w\rangle - \langle v, v\rangle - \langle w, w\rangle\Big) = \langle v, w\rangle \ . \qquad \square
\end{aligned}$$

Definition 1.11 Eine Matrix $A \in GL(n, \mathbb{R})$ heißt *orthogonal*, falls $A^{-1} = A^t$. Eine Matrix $A \in GL(n, \mathbb{C})$ heißt *unitär*, falls $A^{-1} = A^*$ $(= \overline{A^t})$.

Für $A \in M(n, \mathbb{K})$ gilt

$$|\det A|^2 = (\det A)\overline{(\det A)} = (\det A)(\det \overline{A}) = (\det A)(\det \overline{A^t}) = \det(A \cdot A^*) \ .$$

Damit folgt $|\det A| = 1$ für orthogonales bzw. unitäres A. Die Bedingungen $A^t \cdot A = E_n$ bzw. $A^* \cdot A = E_n$ bedeuten, daß die Spalten von A eine Orthonormalbasis von $\mathbb{K}^n$ bilden. Analog bedeutet $A \cdot A^t = E_n$ bzw. $A \cdot A^* = E_n$, daß die Zeilen von A eine Orthonormalbasis von $\mathbb{K}^n$ bilden.

Satz 1.7

- $O(n) := \{A \in GL(n;\mathbb{R}) \;:\; A^{-1} = A^t\}$ *ist Untergruppe der* $GL(n;\mathbb{R})$ *(orthogonale Gruppe).*
- $SO(n) := \{A \in O(n) \;:\; \det A = 1\}$ *ist Untergruppe der* $GL(n;\mathbb{R})$ *(spezielle orthogonale Gruppe).*
- $U(n) := \{A \in GL(n;\mathbb{C}) \;:\; A^{-1} = A^*\}$ *ist Untergruppe der* $GL(n;\mathbb{C})$ *(unitäre Gruppe).*
- $SU(n) := \{A \in U(n) \;:\; \det A = 1\}$ *ist Untergruppe der* $GL(n;\mathbb{R})$ *(spezielle unitäre Gruppe).*

Beweis. Für $A, B \in O(n)$ ist $(AB)^{-1} = B^{-1}A^{-1} = B^tA^t = (AB)^t$ sowie $(A^{-1})^{-1} = A = (A^{-1})^t$, damit ist $O(n)$ Untergruppe. Für $A, B \in SO(n)$ ist $\det(AB) = (\det A)(\det B) = 1$, damit ist $SO(n)$ Untergruppe. Analog für $U(n)$ und $SU(n)$. □

Satz 1.8 *Sei V ein endlich-dimensionaler euklidischer bzw. unitärer Vektorraum. Ein Endomorphismus $F \in \mathrm{End}(V)$ ist genau dann orthogonal bzw. unitär, wenn seine darstellende Matrix $M_{\mathcal{B}}^{\mathcal{B}}$ bezüglich einer beliebigen Orthonormalbasis $\mathcal{B}$ von V orthogonal bzw. unitär ist.*

Beweis. Die Orthonormalbasis $\mathcal{B}$ definiert einen Isomorphismus $\Phi_{\mathcal{B}} : \mathbb{K}^n \to V$. Für $\Phi_{\mathcal{B}}(x) = v$ und $\Phi_{\mathcal{B}}(y) = w$ ist

$$\langle v, w\rangle = x^* \cdot y \;,$$

wenn die Vektoren $x, y \in \mathbb{K}$ als Spaltenvektoren aufgefaßt werden (für reelle Vektoren ist $x^* = x^t$). Die darstellende Matrix des Endomorphismus F bezüglich $\mathcal{B}$ ist

$$A := M_{\mathcal{B}}^{\mathcal{B}}(F) = \Phi_{\mathcal{B}}^{-1} \circ F \circ \Phi_{\mathcal{B}} \;.$$

Damit gilt $\Phi_{\mathcal{B}}(A \cdot x) = F(v)$ und $\Phi_{\mathcal{B}}(A \cdot y) = F(w)$ sowie

$$\langle F(v), F(w)\rangle = (A \cdot x)^* \cdot (A \cdot y) \;.$$

Daraus folgt die Behauptung. □

Beispiel 1.1 Im $\mathbb{R}^2$ mit dem kanonischen Skalarprodukt ist die Standardbasis eine Orthonormalbasis. Die Bedingung für orthogonale Matrizen lautet

$$O(2) := \left\{ A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \;:\; \begin{array}{l} a, b, c, d \in \mathbb{R} \;,\; A^t \cdot A = E_2 \\ \Leftrightarrow \quad a^2 + b^2 = c^2 + d^2 = 1 \text{ und } ac + bd = 0 \end{array} \right\} .$$

Setzen wir $a = \cos\phi$, $b = \sin\phi$, $c = \sin\theta$, $d = \cos\theta$, so verbleibt $\sin(\phi + \theta) = 0$ mit den beiden verschiedenen Lösungen $\phi = -\theta$ und $\phi = \pi - \theta$. Somit gilt:

$$\begin{aligned} O(2) = &\left\{ \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \;:\; \theta \in [0, 2\pi[\right\} \\ &\cup \left\{ \begin{pmatrix} \cos\phi & \sin\phi \\ \sin\phi & -\cos\phi \end{pmatrix} \;:\; \phi \in [0, 2\pi[\right\} . \end{aligned}$$

Die erste Zeile beschreibt Drehungen im $\mathbb{R}^2$ und die zweite Zeile eine Kombination aus Drehungen und einer Spiegelung. Das sind gerade die Transformationen im $\mathbb{R}^2$, welche Winkel und Abstände erhalten. Die Gruppe $SO(2)$ besteht nur aus den in der ersten Zeile von $O(2)$ gegebenen Matrizen.

1.6 Selbstadjungierte Endomorphismen

Definition 1.12 Sei V ein euklidischer bzw. unitärer Vektorraum und $F \in \mathrm{End}(V)$. Dann ist der zu F *adjungierte Endomorphismus* $F^{\mathrm{ad}} \in \mathrm{End}(V)$ definiert durch

$$\langle v, F(w)\rangle = \langle F^{\mathrm{ad}}(v), w\rangle \qquad \forall v, w \in V \; .$$

Der adjungierte Endomorphismus ist wohldefiniert: Gäbe es zwei Lösungen F_1^{ad} und F_2^{ad} mit $\langle v, F(w)\rangle = \langle F_1^{\mathrm{ad}}(v), w\rangle = \langle F_2^{\mathrm{ad}}(v), w\rangle$ für alle $v, w \in V$, so folgt aus der Bi/Sesquilinearität

$$0 = \langle F_1^{\mathrm{ad}}(v), w\rangle - \langle F_2^{\mathrm{ad}}(v), w\rangle = \langle F_1^{\mathrm{ad}}(v) - F_2^{\mathrm{ad}}(v), w\rangle$$

für alle v, w, insbesondere für $w = F_1^{\mathrm{ad}}(v) - F_2^{\mathrm{ad}}(v)$. Damit ist $F_1^{\mathrm{ad}} = F_2^{\mathrm{ad}}$.

Satz 1.9 *Sei F Endomorphismus eines endlich-dimensionalen euklidischen bzw. unitären Vektorraums V. Dann gilt:*

i) $\ker(F^{\mathrm{ad}}) = (\mathrm{im}(F))^\perp$

ii) $\mathrm{im}(F^{\mathrm{ad}}) = (\ker(F))^\perp$

iii) *Ist $\mathcal{B}$ eine Orthonormalbasis von V, so gilt $M_{\mathcal{B}}^{\mathcal{B}}(F^{\mathrm{ad}}) = \big(M_{\mathcal{B}}^{\mathcal{B}}(F)\big)^*$.*

Beweis. i) Sei $v \in \big(\mathrm{im}(F)\big)^\perp$, so ist $0 = \langle v, F(w)\rangle = \langle F^{\mathrm{ad}}(v), w\rangle$ für alle $w \in V$, und damit $\ker(F^{\mathrm{ad}}) = \big(\mathrm{im}(F)\big)^\perp$.

ii) Sei $w \in \ker(F)$, dann ist $0 = \langle v, F(w)\rangle = \langle F^{\mathrm{ad}}(v), w\rangle$ für alle $v \in V$. Das bedeutet $\mathrm{im}(F^{\mathrm{ad}}) = (\ker(F))^\perp$.

iii) Für $\mathcal{B} = (v_1, \ldots, v_n)$ sei $F(v_j) = \sum_{k=1}^n a_{kj} v_k$ und $F^{\mathrm{ad}}(v_i) = \sum_{l=1}^n b_{li} v_l$. Dann ist

$$a_{ij} = \langle v_i, F(v_j)\rangle = \langle F^{\mathrm{ad}}(v_i), v_j\rangle = \overline{b_{ji}} \; .$$

Andererseits ist $\big(M_{\mathcal{B}}^{\mathcal{B}}(F^{\mathrm{ad}})\big)_{ij} = b_{ij} = \overline{a_{ji}}$. $\square$

Definition 1.13 Ein Endomorphismus F eines euklidischen bzw. unitären Vektorraums V heißt *selbstadjungiert*, wenn $F = F^{\mathrm{ad}}$ gilt, d.h.

$$\langle v, F(w)\rangle = \langle F(v), w\rangle \qquad \forall v, w \in V \;.$$

Aus Satz 1.9.iii) folgt, daß $F \in \mathrm{End}(V)$ genau dann selbstadjungiert ist, wenn die darstellende Matrix $M^{\mathcal{B}}_{\mathcal{B}}(F)$ bezüglich einer beliebigen Orthonormalbasis $\mathcal{B}$ von V symmetrisch bzw. hermitesch ist.

Satz 1.10 *Ist $F \in \mathrm{End}(V)$ selbstadjungiert, dann sind alle Eigenwerte von F reell. Insbesondere hat eine hermitesche Matrix nur reelle Eigenwerte.*

Beweis. Ist $F(v) = \lambda v$ mit $v \neq 0$, so gilt

$$\lambda\langle v, v\rangle = \langle v, F(v)\rangle = \langle F(v), v\rangle = \overline{\lambda}\langle v, v\rangle$$

und damit $\lambda = \overline{\lambda}$. □

Theorem 1.1 *Ist F ein selbstadjungierter Endomorphismus eines endlichdimensionalen euklidischen bzw. unitären Vektorraums V, so gibt es eine Orthonormalbasis von V, die aus Eigenvektoren von F besteht. Damit sind selbstadjungierte Endomorphismen stets diagonalisierbar.*

Beweis. Zunächst sei $\mathbb{K} = \mathbb{C}$. Dann zerfällt das charakteristische Polynom von F in Linearfaktoren,

$$P_F(t) = (-1)^n (t - \lambda_1)(t - \lambda_2)\cdots(t - \lambda_n) \;,$$

wobei sämtliche Nullstellen $\lambda_1, \dots, \lambda_n$ reell sind. Der weitere Beweis wird durch Induktion nach $n = \dim(V) \geq 1$ geführt. Für $n = 1$ ist nichts zu zeigen. Sei $n \geq 2$ und v_1 ein Eigenvektor zum Eigenwert λ_1 mit $\|v_1\| = 1$. Wir betrachten

$$W := \{w \in V \;:\; \langle w, v_1\rangle = 0\} \;.$$

Dann ist W mit dem aus V eingeschränkten Skalarprodukt wieder ein unitärer Vektorraum. Dieser ist invariant unter F, also $F(W) \subset W$, denn

$$\langle F(w), v_1\rangle = \langle w, F(v_1)\rangle = \lambda_1\langle w, v_1\rangle = 0 \qquad \forall w \in W \;.$$

Damit ist die Einschränkung $F\big|_W$ von F auf W ein Endomorphismus des $(n-1)$-dimensionalen unitären Vektorraums W. Nach Induktionsannahme gibt es eine Orthonormalbasis von W aus Eigenvektoren von $F\big|_W$, welche um v_1 ergänzt eine Orthonormalbasis von V ergibt.

Für $\mathbb{K} = \mathbb{R}$ können wir dennoch die darstellende Matrix als komplexe Matrix auffassen, deren charakteristisches Polynom in n *reelle* Linearfaktoren zerfällt. Der weitere Induktionsbeweis verläuft wie in komplexen Fall. □

Durch Zusammenfassung der Eigenvektoren von (selbstadjungiertem) F zu Eigenräumen zum gleichen Eigenwert erhält man

$$V = \mathrm{Eig}(F;\lambda_1) \oplus \cdots \oplus \mathrm{Eig}(F;\lambda_k)\,,$$

wobei die Eigenräume paarweise orthogonal zueinander sind: Ist $v \in \mathrm{Eig}(F;\lambda)$ und $w \in \mathrm{Eig}(F;\mu)$ mit $\lambda \neq \mu$, dann ist

$$\lambda\langle w, v\rangle = \langle w, F(v)\rangle = \langle F(w), v\rangle = \overline{\mu}\langle w, v\rangle = \mu\langle w, v\rangle\ ,$$

also $\langle w, v\rangle = 0$ für $\lambda \neq \mu$. Dementsprechend kann eine Orthonormalbasis aus Eigenvektoren eines selbstadjungierten Endomorphismus F wie folgt bestimmt werden:

1. Bestimme die Linearfaktorzerlegung des charakteristischen Polynoms $P_F = \pm(t-\lambda_1)^{r_1}\cdots(t-\lambda_k)^{r_k}$ mit $\lambda_i \in \mathbb{R}$ paarweise verschieden.
2. Zu jedem Eigenwert λ_i der Vielfachheit r_i bestimme man eine Basis von $\mathrm{Eig}(F;\lambda_i)$ durch Lösen des linearen Gleichungssystems $F(v) = \lambda_i v$. Durch das Schmidtsche Orthonormalisierungsverfahren wird daraus eine Orthonormalbasis $\mathcal{B}_i$. Dann ist $\mathcal{B}_i = (\mathcal{B}_1, \dots, \mathcal{B}_k)$ die gesuchte Orthonormalbasis von V aus Eigenvektoren.

Satz 1.11 *Ist $A \in M(n,\mathbb{K})$ eine reelle symmetrische bzw. hermitesche Matrix, dann gibt es eine orthogonale bzw. unitäre Matrix $T \in O(n)$ bzw. $T \in U(n)$ mit*

$$A = T\Lambda T^*\ , \qquad \Lambda = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}\ ,$$

wobei $\lambda_i \in \mathbb{R}$ die Eigenwerte von A sind.

Beweis. Die Matrix $A = A^t = (a_{ij})$ definiert einen selbstadjungierten Endomorphismus $F = F^{\mathrm{ad}} \in \mathrm{End}(\mathbb{K}^n)$ durch $F(e_j) = \sum_{i=1}^n a_{ij}e_i$, wobei $(e_1,\dots,e_n)$ die Standardbasis ist. Sei $(v_1,\dots,v_n)$ eine orthonormale Basis von $\mathbb{K}^n$ aus Eigenvektoren von F, d.h. $F(v_i) = \lambda_i v_i$. Wir zerlegen v_i nach der Standardbasis: $v_i = \sum_{k=1}^n t_{ki}e_k$. Aus der Orthonormalität folgt $\langle v_i, v_j\rangle = \sum_{k=1}^n \overline{t_{ki}}t_{kj} = \delta_{ij}$, und damit $T^*T = E_n$ für $T = (t_{ij})$. Somit ist $T \in O(n)$ bzw. $T \in U(n)$, und insbesondere ist $\sum_{k=1}^n t_{jk}\overline{t_{ik}} = \delta_{ij}$ wegen $TT^* = E_n$. Nun gilt

$$F(v_k) = \sum_{j=1}^n t_{jk}F(e_j) = \sum_{i,j=1}^n t_{jk}a_{ij}e_i = \lambda_k v_k = \sum_{l=1}^n \lambda_k t_{lk}e_l\ .$$

Da die e_i linear unabhängig sind, folgt $\sum_{j=1}^n a_{ij}t_{jk} = t_{ik}\lambda_k$ für alle k. Wir multiplizieren mit $\overline{t_{lk}}$ und summieren über k:

$$\sum_{k,j=1}^n a_{ij}t_{jk}\overline{t_{lk}} = a_{il} = \sum_{k=1}^n t_{ik}\lambda_k\overline{t_{lk}} = (T\Lambda T^*)_{il}\ . \qquad \square$$

1.7 Hauptachsentransformation im $\mathbb{R}^n$

Wir interessieren uns nun für die Frage, wie man entscheiden kann, ob eine symmetrische Bilinearform positiv definit ist (und somit ein Skalarprodukt definiert). Dabei beschränken wir uns auf den Fall $V = \mathbb{R}^n$.

Gegeben ist also eine symmetrische Bilinearform $s : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$, welche durch eine symmetrische Matrix $A = A^t = (a_{ij})$ mit $a_{ij} = s(e_i, e_j)$ bestimmt ist. Damit gilt $s(x,y) = x^t \cdot A \cdot y = \langle x, Ay \rangle$ für alle Spaltenvektoren $x, y \in \mathbb{R}^n$. Fassen wir A als selbstadjungierten Endomorphismus $A : \mathbb{R}^n \to \mathbb{R}^n$ auf, dann ist A diagonalisierbar, und es gibt eine Orthonormalbasis $\mathcal{B} = (v_1, \dots, v_n)$ (bezüglich des kanonischen Skalarprodukts) aus Eigenvektoren von A:

$$Av_i = \lambda_i v_i \;, \quad \lambda_i \in \mathbb{R} \qquad \Rightarrow \quad s(v_i, v_j) = \langle Av_i, v_j \rangle = \lambda_i \delta_{ij}$$

Geometrisch beschreibt die Basis $\mathcal{B}$ aus Eigenvektoren eine Rotation des durch die Standardbasis gegebenen Koordinatensystems in ein der Matrix A angepaßtes Koordinatensystem, welches gerade durch die Basis $\mathcal{B}$ gegeben ist.

Durch Umnumerierung und Reskalierung der Eigenvektoren zu $\lambda_i \neq 0$ läßt sich eine orthogonale Basis $\mathcal{A} = (w_1, \dots, w_k, w_{k+1}, \dots, w_{k+l}, w_{k+l+1}, \dots w_n)$ finden, so daß

$$M_{\mathcal{A}}(s) = \begin{pmatrix} E_k & & \\ & -E_l & \\ & & 0 \end{pmatrix} .$$

Dabei sind die Dimensionen k, l der Unterräume zu positiven/negativen λ_i unabhängig von der Basiswahl (hier ohne Beweis).

Satz 1.12 *Eine symmetrische Matrix $A = A^t \in M(n \times n, \mathbb{R})$ ist genau dann positiv definit, wenn ihre Eigenwerte positiv sind.*

Beweis. Sei $\mathcal{B} = (v_1, \dots, v_n)$ eine Orthonormalbasis von $\mathbb{R}^n$ aus Eigenvektoren von A zu den Eigenwerten $\lambda_1, \dots, \lambda_n$, dann ist für $x = \mu_1 v_1 + \dots + \mu_n v_n$

$$s(x,x) = \langle x, Ax \rangle = \lambda_1 \mu_1^2 + \dots + \lambda_n \mu_n^2 \;.$$

Daraus folgt die Behauptung. □

Allerdings ist die Bestimmung der Eigenwerte meist nur numerisch möglich. Ein einfacheres Kriterium kann wie folgt erhalten werden.

Satz 1.13 *Sei $A \in M(n \times n, \mathbb{R})$ eine symmetrische Matrix und $A_k \in M(k \times k, \mathbb{R})$ die linke obere Teilmatrix von A aus k Zeilen und Spalten. Dann ist A genau dann positiv definit, wenn* $\det A_k > 0$ *für alle* $1 \leq k \leq n$.

Beweis. ($\Rightarrow$) Ist $A = A^t \in M(n, \mathbb{R})$ positiv definit, dann ist $\det A = \det(T \Lambda T^t) = \det \Lambda (\det T)^2 = \lambda_1 \cdot \dots \cdot \lambda_n > 0$. Die Matrix A_k beschreibt die Einschränkung der durch A definierten Bilinearform auf den Unterraum

$$V_k := \{x = (x_1, \dots, x_n) \in \mathbb{R}^2 \;:\; x_{k+1} = \dots = x_n = 0\} \;.$$

Ist A positiv definit, dann ist auch $\langle x, Ax\rangle > 0$ für alle $x \in V_k$ mit $x \neq 0$. Folglich ist $\det A_k > 0$.

($\Leftarrow$) wird durch Induktion nach n bewiesen. Sei in Blockdarstellung

$$A = A^t = \left(\begin{array}{c|c} B & b \\ \hline b^t & c \end{array}\right) \in M(n,\mathbb{R}) \qquad B = B^t \in M(n-1,\mathbb{R})\,, \quad c \in \mathbb{R}$$
$$b \in \mathbb{R}^{n-1} \text{ (Spaltenvektor)}\,.$$

Nach Induktionsvoraussetzung folgt aus $\det A_k = \det B_k > 0$ für $1 \leq k \leq n-1$, daß B positiv definit ist. Damit gilt $B = T\Lambda T^t$, wobei $T \in O(n-1)$ und $\Lambda \in M(n-1,\mathbb{R})$ eine Diagonalmatrix mit *positiven Eigenwerten* $\lambda_1,\dots,\lambda_{n-1}$ von B auf der Diagonale ist. Dann ist

$$\begin{aligned} A &= \left(\begin{array}{c|c} T & 0 \\ \hline 0 & 1 \end{array}\right) \left(\begin{array}{c|c} \Lambda & T^t b \\ \hline b^t T & c \end{array}\right) \left(\begin{array}{c|c} T & 0 \\ \hline 0 & 1 \end{array}\right)^t \\ &= \left(\begin{array}{c|c} T & 0 \\ \hline 0 & 1 \end{array}\right) \left(\begin{array}{c|c} E_{n-1} & 0 \\ \hline b^t T\Lambda^{-1} & 1 \end{array}\right) \left(\begin{array}{c|c} \Lambda & 0 \\ \hline 0 & c - b^t T\Lambda^{-1}T^t b \end{array}\right) \left(\begin{array}{c|c} E_{n-1} & \Lambda^{-1}T^t b \\ \hline 0 & 1 \end{array}\right) \left(\begin{array}{c|c} T & 0 \\ \hline 0 & 1 \end{array}\right)^t . \end{aligned}$$

Nun ist $\det A = (\det T)^2(\det\Lambda)(c - b^t T\Lambda^{-1}T^t b)$ genau dann positiv, wenn $c - b^t T\Lambda^{-1}T^t b > 0$. Andererseits ist $x^t A x > 0$ für alle $x \in \mathbb{R}^n$ genau dann, wenn $y^t \left(\begin{array}{c|c} \Lambda & 0 \\ \hline 0 & c - b^t T\Lambda^{-1}T^t b \end{array}\right) y > 0$ für alle $y \in \mathbb{R}^n$. $\square$

Achtung: In obiger Darstellung für A ist $\left(\begin{array}{c|c} E_{n-1} & \Lambda^{-1}T^t b \\ \hline 0 & 1 \end{array}\right)$ keine orthogonale Matrix für $b \neq 0$, so daß $\lambda_1,\dots,\lambda_{n+1}, c - b^t T\Lambda^{-1}T^t b$ *nicht* die Eigenwerte von A sind!

2 Differentialrechnung im $\mathbb{R}^n$

2.1 Metrische Räume

Die grundlegende Struktur, mit der wir die Differentialrechnung vom $\mathbb{R}^1$ in den $\mathbb{R}^n$ übertragen, ist die *Metrik* bzw. der *Abstand*. Zur Erinnerung:

Definition 2.1 Eine Menge X mit einer Abbildung $d : X \times X \to \mathbb{R}$, dem *Abstand*, heißt *metrischer Raum*, wenn für alle $x, y, z \in X$ gilt

(D1) $d(x, y) = 0 \quad \Leftrightarrow \quad x = y$,

(D2) $d(x, y) = d(y, x)$ (Symmetrie),

(D3) $d(x, z) \leq d(x, y) + d(y, z)$ (Dreiecksungleichung).

Die Axiome liefern insbesondere $d(x, y) \geq 0$. Jede Norm auf einem Vektorraum V definiert einen Abstand auf V. Umgekehrt gibt es Abstände, die nicht aus einer Norm zu erhalten sind, und insbesondere muß die Grundmenge X kein Vektorraum sein.

Definition 2.2

i) Sei (X, d) ein metrischer Raum, $a \in X$ und $r > 0$. Dann heißt die Teilmenge

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

die *offene Kugel* in (X, d) mit Mittelpunkt a und Radius r.

ii) Eine Teilmenge $U \subset X$ eines metrischen Raumes (X, d) heißt *Umgebung* eines Punktes $x \in X$, wenn ein $\epsilon > 0$ existiert, so daß $B_\epsilon(x) \subset U$. Die offene Kugel $B_\epsilon(x)$ heißt auch die *ϵ-Umgebung* von x.

iii) Eine Teilmenge U eines metrischen Raumes heißt *offen*, wenn jeder Punkt $x \in U$ eine in U enthaltene ϵ-Umgebung besitzt, d.h.

$$\forall x \in U \ \exists \epsilon > 0 \ : \ B_\epsilon(x) \subset U \ .$$

Satz 2.1 *Für die offenen Teilmengen eines metrischen Raumes (X, d) gilt:*

i) *Die gesamte Menge X und die leere Menge $\varnothing$ sind offen.*

ii) *Ist $(U_1, \dots, U_n)$ eine endliche Familie offener Teilmengen, dann ist auch der Durchschnitt $\bigcap_{i=1}^n U_i := U_1 \cap \dots \cap U_n$ eine offene Teilmenge von X.*

iii) *Ist $(U_i)_{i \in I}$ eine (möglicherweise unendliche) Familie offener Teilmengen, dann ist auch die Vereinigung $\bigcup_{i \in I} U_i$ eine offene Teilmenge von X.*

Außerdem gilt (Hausdorffsches Trennungsaxiom):

iv) *Zu je zwei Punkten $x, y \in X$ mit $x \neq y$ gibt es disjunkte offene Umgebungen U von x und V von y, d.h. $U \cap V = \varnothing$.*

Wir verzichten auf den Beweis. Es sei bemerkt, daß man durch i)–iii) die offenen Teilmengen einer Menge X axiomatisch definieren kann. Eine Menge X mit einer Familie von Teilmengen, die i)-iii) erfüllen, heißt dann *topologischer Raum*. Gilt zusätzlich iv), dann heißt X ein *Hausdorff-Raum*. Zur Charakterisierung der *Stetigkeit* genügen topologische Räume, für Differenzierbarkeit benötigen wir aber die Metrik.

Definition 2.3 Sei (X, d) ein metrischer Raum.

i) Ein Teilmenge $Y \subset X$ heißt *abgeschlossen*, wenn das Komplement $X \setminus Y$ offen ist.

ii) Sei $Y \subset X$ eine Teilmenge. Ein Punkt $y \in X$ heißt *Randpunkt* von Y, wenn es in jeder Umgebung von y sowohl Punkte aus Y als auch aus $X \setminus Y$ gibt. Die Menge aller Randpunkte von Y heißt der *Rand* von Y und wird mit ∂Y bezeichnet.

Beispiel 2.1 Die n-dimensionale Einheitskugel $B^n := \{x \in \mathbb{R}^n \ : \ \|x\| \leq 1\}$ ist abgeschlossen im $\mathbb{R}^n$. Ihr Rand ist die Einheitssphäre $S^{n-1} := \partial B^n = \{x \in \mathbb{R}^n \ : \ \|x\| = 1\}$, sie ist ebenfalls abgeschlossen im $\mathbb{R}^n$. Es ist dann sinnvoll, offene Teilmengen von S^{n-1} zu betrachten. Jeder Durchschnitt der S^{n-1} mit einer offenen Teilmenge des $\mathbb{R}^n$ definiert eine offene Teilmenge von S^{n-1}. Im Sinne dieser Teilmengen ist S^{n-1} dann offen und abgeschlossen zugleich.

Satz 2.2 *Sei (X, d) ein metrischer Raum und $Y \subset X$. Dann gilt:*

i) *$Y \setminus \partial Y$ ist offen in X.*

ii) *$Y \cup \partial Y$ ist abgeschlossen in X.*

iii) *∂Y ist abgeschlossen in X.*

2.2 Grenzwerte und Konvergenz in metrischen Räumen

Definition 2.4 Sei (X, d) ein metrischer Raum und $(x_k)_{k \in \mathbb{N}}$ eine Folge von Punkten aus X. Die Folge (x_k) heißt *konvergent* gegen den *Grenzwert* $a \in X$, geschrieben

$$\lim_{k \to \infty} x_k = a ,$$

wenn

- zu jeder Umgebung U von a ein $n \in \mathbb{N}$ existiert mit $x_k \in U$ für alle $k \geq n$, bzw.
- zu jedem $\epsilon > 0$ ein $n \in \mathbb{N}$ existiert mit $d(a, x_k) < \epsilon$ für alle $k \geq n$.

Im $\mathbb{R}^n$ konvergiert eine Punktfolge $(x_k)_{k \in \mathbb{N}}$ mit $x_k = (x_{k1}, \ldots, x_{kn})$ genau dann gegen $a = (a_1, \ldots, a_n) \in \mathbb{R}^n$ wenn für jede Koordinate $1 \leq i \leq n$ die Folge x_{ki} gegen a_i konvergiert.

Satz 2.3 *Sei (X, d) ein metrischer Raum. Eine Teilmenge $Y \subset X$ ist genau dann abgeschlossen in X, wenn für jede in X konvergente Folge $(y_k)_{k\in\mathbb{N}}$ mit $y_k \in Y$ gilt* $\lim_{k\to\infty} y_k = x \in Y$.

Beweis. ($\Rightarrow$) Sei $Y \subset X$ abgeschlossen und $x = \lim_{k\to\infty} y_k$ mit $y_k \in Y$. Angenommen, $x \notin Y$, also $x \in X \setminus Y$. Da $X \setminus Y$ offen ist, gibt es in $X \setminus Y$ eine ϵ-Umgebung U_ϵ um x. Da x der Grenzwert der Folge (y_k) ist, gibt es ein $n \in \mathbb{N}$, so daß $y_k \in U_\epsilon \subset X \setminus Y$ für alle $k \geq n$. Das ist ein Widerspruch.

($\Leftarrow$) Die Grenzwerte aller konvergenten Folgen von Punkten aus Y liegen in Y. Wir zeigen: $X \setminus Y$ ist offen. Angenommen, es gibt einen Punkt $\tilde{x} \in X \setminus Y$, der keine ϵ-Umgebung in $X \setminus Y$ besitzt, d.h. jede ϵ-Umgebung enthält einen Punkt aus Y. Wählen wir z.B. die Folge der Umgebungen mit $\epsilon_k = \frac{1}{k+1}$, dann gibt es zu jedem $k \in \mathbb{N}$ ein $y_k \in Y$ mit $d(\tilde{x}, y_k) < \epsilon_k$. Auf diese Weise wird eine Folge $(y_k)_{k\in\mathbb{N}}$ konstruiert, die gegen $\tilde{x}$ konvergiert. Damit wäre $\tilde{x} \notin X \setminus Y$, so daß im Widerspruch zur Annahme jeder Punkt aus $X \setminus Y$ eine ϵ-Umgebung in $X \setminus Y$ besitzt. $\square$

Definition 2.5 Sei (X, d) ein metrischer Raum. Eine Folge $(x_k)_{k\in\mathbb{N}}$ von Punkten aus X heißt *Cauchy-Folge*, wenn zu jedem $\epsilon > 0$ ein $k \in \mathbb{N}$ existiert, so daß $d(x_n, x_m) < \epsilon$ für alle $m, n \geq k$.

Satz 2.4 *Jede konvergente Folge in einem metrischen Raum ist eine Cauchy-Folge.*

Beweis. Sei $(x_k)_{k\in\mathbb{N}}$ konvergent in (X, d) mit $\lim_{k\to\infty} x_k = x$. Dann gibt es zu jedem $\epsilon > 0$ ein $k \in \mathbb{N}$, so daß $d(x_n, x) < \frac{\epsilon}{2}$ und $d(x_m, x) < \frac{\epsilon}{2}$ für alle $m, n \geq k$. Dann folgt aus der Dreiecksungleichung $d(x_n, x_m) \leq d(x_n, x) + d(x, x_m) < \epsilon$. $\square$

Der Sinn der Cauchy-Folge ist, daß man den Grenzwert gar nicht kennen muß. Ob eine Folge eine Cauchy-Folge ist, kann man allein aus der Gesamtheit der Folgenelemente entscheiden, ohne Abstände zu einem fiktiven Grenzwert zu untersuchen. Das wirft die Frage auf, wann eine Cauchy-Folge auch konvergent ist.

Definition 2.6 Ein metrischer Raum (X, d) heißt *vollständig*, wenn jede Cauchy-Folge in ihm konvergiert.

Ein vollständiger normierter Vektorraum heißt *Banach-Raum*.

Satz 2.5 $\mathbb{R}^n$ *ist ein Banach-Raum. Insbesondere konvergiert im* $\mathbb{R}^n$ *jede Cauchy-Folge.*

Der $\mathbb{R}^1$ wurde gerade durch Vervollständigung der rationalen Zahlen bezüglich des natürlichen Abstands definiert. Komponentenweise überträgt sich dann die Vollständigkeit auf $\mathbb{R}^n$.

Definition 2.7 Für eine Teilmenge $Y \subset X$ eines metrischen Raumes (X, d) wird der *Durchmesser* definiert als $\mathrm{diam}(Y) := \sup_{x,y \in Y} d(x, y)$. Eine Teilmenge $Y \subset X$ heißt *beschränkt*, wenn $\mathrm{diam}(Y) < \infty$.

2.3 Stetigkeit in metrischen Räumen

Definition 2.8 Eine Abbildung $f : X \to Y$ zwischen metrischen Räumen X, Y heißt *stetig im Punkt* $a \in X$, wenn für jede Folge $(x_k)_{k \in \mathbb{N}}$ von Punkten aus X mit $\lim_{k\to\infty} x_k = a$ gilt

$$\lim_{k\to\infty} f(x_k) = f(a) .$$

In diesem Fall schreiben wir auch $\lim_{x\to a} f(x) = f(a)$.

Die Abbildung $f : X \to Y$ heißt *stetig auf* X, wenn f in jedem Punkt $a \in X$ stetig ist.

Satz 2.6 (ϵ-δ-Kriterium) *Seien (X, d_X) und (Y, d_Y) metrische Räume und $a \in X$. Eine Abbildung $f : X \to Y$ ist genau dann stetig in a, wenn für alle $\epsilon > 0$ ein $\delta > 0$ existiert, so daß*

$$d_Y\big(f(x), f(a)\big) < \epsilon \qquad \text{für alle } x \in X \text{ mit } d_X(x, a) < \delta .$$

Beweis. ($\Rightarrow$) Sei $f : X \to Y$ stetig in $a \in X$ und das ϵ-δ-Kriterium verletzt. Wir können dann ein $\epsilon > 0$ finden, so daß für jedes $\delta > 0$ ein $x \in X$ existiert mit $d_X(x, a) < \delta$ aber $d_Y\big(f(x), f(a)\big) > \epsilon$. Wählen wir $\delta = \frac{1}{k+1}$, dann finden wir eine Folge $(x_k)_{k\in\mathbb{N}}$ von Punkten aus X mit $d_X(x_k, a) < \frac{1}{k+1}$ (welche also gegen a konvergiert), so daß $d_Y\big(f(x_k), f(a)\big) > \epsilon$ ist für alle $k \in \mathbb{N}$. Dann wäre f nicht stetig in a.

($\Leftarrow$) Sei $(x_k)_{k\in\mathbb{N}}$ eine gegen a konvergierende Folge und $\epsilon > 0$ gegeben. Das ϵ-δ-Kriterium bestimmt dann ein $n \in \mathbb{N}$, so daß $d_X(x_k, a) < \delta$ für alle $k \geq n$. Dann ist $d_Y\big(f(x_k), f(a)\big) < \epsilon$ für alle $k \geq n$. Damit gilt $\lim_{k\to\infty} f(x_k) = f(a)$. □

Die Bedeutung des ϵ-δ-Kriteriums besteht unter anderem darin, daß sich damit die Stetigkeit von Abbildungen allein unter Verwendung des Begriffs der offenen Teilmengen definieren läßt: Eine Abbildung $f : X \to Y$ zwischen topologischen Räumen heißt *stetig* im Punkt $a \in X$, wenn zu jeder offenen Teilmenge $V \subset Y$ mit $f(a) \in V$ eine offene Teilmenge $U \subset X$ mit $a \in U$ existiert, so daß $f^{-1}(V) \subset U$ gilt.

Satz 2.7 *Seien X, Y, Z metrische Räume und seien die Abbildungen $f : X \to Y$ stetig in $a \in X$ und $g : Y \to Z$ stetig in $f(a) \in Y$. Dann ist $g \circ f : X \to Z$ stetig in a.*

Definition 2.9 Eine Abbildung $f : X \to Y$ zwischen metrischen Räumen X, Y heißt *Homöomorphismus*, wenn gilt:

i) f ist bijektiv,

ii) $f : X \to Y$ und $f^{-1} : Y \to X$ sind stetig.

Zwei metrische Räume X, Y heißen *homöomorph*, wenn es einen Homöomorphismus $f : X \to Y$ gibt.

In dieser Definition können die metrischen Räume durch topologische Räume ersetzt werden.

Beispiel 2.2

- $X = \mathbb{R}^n, \quad Y = B^n \setminus S^{n-1} := \{y \in \mathbb{R}^n \ : \ \|y\| < 1\}$
$f : X \to Y$ mit $f(x) = \frac{x}{1+\|x\|}$ ist Homöomorphismus mit $f^{-1}(y) = \frac{y}{1-\|y\|}$
- $X = [0, 2\pi[\subset \mathbb{R}, \quad Y = S^1 = \{(x, y) \in \mathbb{R}^2 \ : \ x^2 + y^2 = 1\}$
$f : X \to Y$ mit $f(t) = (\cos t, \sin t)$ ist bijektiv und stetig, aber $f^{-1} : Y \to X$ ist nicht stetig in $(1, 0)$. Die Folge von Punkten
$$p_k := \big(\cos(2\pi - \tfrac{1}{k+1}),\ \sin(2\pi - \tfrac{1}{k+1})\big)\ , \qquad k \in \mathbb{N}\ ,$$
konvergiert gegen $(1, 0) = f(0)$, aber $f^{-1}(p_k) = 2\pi - \frac{1}{k+1}$ konvergiert nicht gegen 0.

Definition 2.10 Seien $f : X \to Y$ sowie $f_k : X \to Y$ für $k \in \mathbb{N}$ Abbildungen zwischen metrischen Räumen X, Y. Die Folge der Abbildungen f_k *konvergiert gleichmäßig* gegen die Abbildung f, wenn zu jedem $\epsilon > 0$ ein $n \in \mathbb{N}$ existiert mit $d\big(f_k(x), f(x)\big) < \epsilon$ für alle $x \in X$ und alle $k \geq n$.

Satz 2.8 *Seien X, Y metrische Räume und $f_k : X \to Y$, für $k \in \mathbb{N}$, eine Folge stetiger Abbildungen, die gleichmäßig gegen eine Abbildung $f : X \to Y$ konvergiert. Dann ist f stetig.*

Beweis. Wir zeigen, daß f in jedem Punkt $a \in X$ stetig ist. Sei $\epsilon > 0$ gegeben. Wegen der gleichmäßigen Konvergenz gibt es ein $n \in \mathbb{N}$, so daß $d_Y\big(f_k(x), f(x)\big) < \frac{\epsilon}{3}$ für alle $k \geq n$ und alle $x \in X$. Da f_k in a stetig ist, gibt es ein $\delta > 0$ mit $d_Y\big(f_k(x), f_k(a)\big) < \frac{\epsilon}{3}$ für alle $x \in X$ mit $d_X(x, a) < \delta$. Dann folgt aus der Dreiecksungleichung
$$d_Y\big(f(x), f(a)\big) \leq d_Y\big(f(x), f_k(x)\big) + d_Y\big(f_k(x), f_k(a)\big) + d_Y\big(f_k(a), f(a)\big) < \epsilon$$
für alle $x \in X$ mit $d_X(x, a) < \delta$. □

Zum Abschluß noch einige Bemerkungen zur Stetigkeit von linearen Abbildungen.

Satz 2.9 *Seien $(V, \|\ \|_V)$ und $(W, \|\ \|_W)$ normierte Vektorräume über $\mathbb{K} = \mathbb{R}$ oder $\mathbb{K} = \mathbb{C}$. Eine lineare Abbildung $A : V \to W$ ist genau dann stetig, wenn es eine reelle Zahl $C \geq 0$ gibt mit $\|A(x)\|_W \leq C\|x\|_V$ für alle $x \in V$.*

Beweis. ($\Rightarrow$) Sei A stetig in $0 \in V$. Wegen $A(0) = 0$ gibt es dann zu $\epsilon = 1$ ein $\delta > 0$, so daß

$$d_W\big(A(z), A(0)\big) = \|A(z)\|_W < 1 \qquad \text{für alle } z \in V \text{ mit } d_V(z,0) = \|z\|_V < \delta\ .$$

Für $x \in V \setminus \{0\}$ setzen wir $z := \frac{\delta}{2\|x\|_V} \cdot x$, mit $\|z\|_V \leq \frac{\delta}{2} < \delta$, und erhalten

$$\|A(x)\|_W = \big\|\tfrac{2\|x\|_V}{\delta} A(z)\big\|_W \leq \tfrac{2\|x\|_V}{\delta} \|A(z)\|_W < \tfrac{2}{\delta}\|x\|_V \qquad \text{für alle } x \in V \setminus \{0\}\ .$$

Damit ist $C = \frac{2}{\delta}$. Die Ungleichung $\|A(x)\|_W \leq C\|x\|_V$ gilt auch für $x = 0$.

($\Leftarrow$) Für $x, a \in V$ gilt

$$d_W\big(A(x), A(a)\big) = \big\|A(x) - A(a)\big\|_W = \big\|A(x-a)\big\|_W \leq C\|x - a\|_V\ .$$

Ist ϵ gegeben, dann wählen wir $\delta = \frac{\epsilon}{C}$ für $C \neq 0$ (und δ beliebig für $C = 0$). Dann gilt für alle $x \in V$ mit $d_V(x,a) < \delta$ die Ungleichung $d_W\big(A(x), A(a)\big) < \epsilon$, und A ist stetig. □

Offenbar bildet die Menge aller stetigen linearen Abbildungen zwischen zwei normierten Vektorräumen V, W selbst wieder einen Vektorraum, der mit $L(V,W)$ bezeichnet wird.

Definition 2.11 Seien $(V, \|\ \|_V)$ und $(W, \|\ \|_W)$ normierte Vektorräume über $\mathbb{K} = \mathbb{R}$ oder $\mathbb{K} = \mathbb{C}$ und $A : V \to W$ eine lineare stetige Abbildung. Dann wird die Norm von A definiert als

$$\|A\| := \sup\big\{\|A(x)\|_W \ :\ x \in V \text{ mit } \|x\|_V \leq 1\big\}\ .$$

Man kann zeigen, daß die so erklärte Norm von stetigen linearen Abbildungen wirklich die Axiome einer Norm erfüllt, so daß $L(V,W)$ ein normierter Vektorraum wird, und damit insbesondere ein metrischer Raum. Es können dann wieder Cauchy-Folgen von Punkten aus $L(V,W)$ betrachtet werden, und durch Vervollständigung (Hinzunahme der Grenzwerte aller Cauchy-Folgen) wird $L(V,W)$ zu einem Banach-Raum.

2.4 Kompaktheit

Eine Teilmenge des $\mathbb{R}^n$ ist genau dann kompakt, wenn sie beschränkt und abgeschlossen ist. Stetige Funktionen über solchen Teilmengen haben viele nützliche Eigenschaften, die man gerne auf Teilmengen beliebiger metrischer Räume verallgemeinern möchte. Dazu wird der Begriff der Kompaktheit recht abstrakt definiert.

Definition 2.12 Sei A Teilmenge eines metrischen Raumes X. Unter einer *offenen Überdeckung* von A versteht man eine Familie $(U_i)_{i\in I}$ von offenen Teilmengen $U_i \subset X$ mit $A \subset \bigcup_{i\in I} U_i$.

Eine Teilmenge eines metrischen Raumes X heißt *kompakt*, wenn es zu jeder offenen Überdeckung $(U_i)_{i\in I}$ von A endlich viele Indizes $i_1, \ldots, i_n \in I$ gibt, so daß $A \subset U_{i_1} \cup \cdots \cup U_{i_n}$.

Es wird also nicht gefordert, daß man A durch endlich viele offene Teilmengen von X überdecken kann. Das geht immer, denn A läßt sich durch X selbst überdecken, und X ist offen. Die Forderung ist, daß man jede exotische unendliche Überdeckung von A auf eine endliche Überdeckung reduzieren kann.

Satz 2.10 *Sei X ein metrischer Raum und $(x_k)_{k\in\mathbb{N}}$ eine Folge von Punkten aus X, die gegen einen Grenzwert $a \in X$ konvergiert. Dann ist die Teilmenge*

$$A := \{x_k \ : \ k \in \mathbb{N}\} \cup \{a\}$$

kompakt in X.

Beweis. Sei $(U_i)_{i\in I}$ eine beliebige offene Überdeckung von A. Dann gibt es einen Index $m \in I$ mit $a \in U_m$. Die Teilmenge $U_m \subset X$ ist offen, enthält also die offene Kugel $B_\epsilon(a)$ für ein $\epsilon > 0$. Da $(x_k)_{k\in\mathbb{N}}$ gegen a konvergiert, gibt es einen Index $n \in \mathbb{N}$, so daß $x_k \in B_\epsilon(a)$ für alle $k \geq n$. Jeder Punkt x_k mit $0 \leq k < n$ liegt in irgendeiner Umgebung U_{i_k} mit $i_k \in I$. Damit gilt

$$A \subset U_{i_0} \cup U_{i_1} \cup \cdots \cup U_{i_{n-1}} \cup U_m \ ,$$

so daß A kompakt ist. □

Ganz entscheidend im Beweis ist die Tatsache, daß der Grenzwert a zu A gehört:

Satz 2.11 *Es sei $A := \{\frac{1}{n+1} \ : \ n \in \mathbb{N}\} \subset \mathbb{R}$. Dann ist A nicht kompakt.*

Beweis. Wir geben eine Überdeckung von A an, die nicht auf eine endliche Überdeckung reduziert werden kann. Dazu sei

$$U_0 :=]\tfrac{3}{4}, 2[\ , \qquad U_n :=]\tfrac{1}{n+2}, \tfrac{1}{n}[\quad \text{für } n \geq 1 \ .$$

Alle $U_k \subset \mathbb{R}$ sind offen und enthalten genau einen Punkt von A, nämlich $\frac{1}{k+1}$. (Z.B. ist $\frac{1}{2} \in U_1 =]\frac{1}{3}, 1[$, $\frac{1}{3} \in U_2 =]\frac{1}{4}, \frac{1}{2}[$, $\frac{1}{4} \in U_3 =]\frac{1}{5}, \frac{1}{3}[$, $\ldots$) Damit kann keine Teilmenge U_k weggelassen werden, ohne die Überdeckungseigenschaft zu verlieren. □

Satz 2.12 *Jede kompakte Teilmenge A eines metrischen Raumes X ist beschränkt und abgeschlossen.*

Beweis. Zu beliebigem $a \in A$ ist $\bigcup_{i=0}^{\infty} B_{i+1}(a) = X$, so daß $\{B_{i+1}(a)\}_{i\in\mathbb{N}}$ eine Überdeckung von A ist. Da A kompakt ist, reichen bereits endlich viele $B_{i+1}(a)$ zur Überdeckung. Also gibt es einen Index $n \in \mathbb{N}$ mit $A \subset B_{n+1}(a)$. Damit ist A beschränkt.

Sei $x \in X \setminus A$ beliebig. Für $i \in \mathbb{N}$ betrachten wir $U_i := \{y \in X \ : \ d(y,x) > \frac{1}{i+1}\}$. Die U_i sind offen als Komplement der abgeschlossenen Kugel vom Radius $\frac{1}{i+1}$ um x. Dann gilt $\bigcup_{i=0}^{\infty} U_i = X \setminus \{x\} \supset A$. Da A kompakt, gibt es ein $n \in \mathbb{N}$ mit $A \subset \bigcup_{i=0}^{n} U_i$. Dann ist $B_{\frac{1}{n}}(x) \subset X \setminus A$. Also ist A abgschlossen. □

Satz 2.13 *Jede konvergente Folge in einem metrischen Raum ist beschränkt.*

Sei $(x_k)_{k\in\mathbb{N}}$ konvergent mit $\lim_{k\to\infty} x_k = a$. Nach Satz 2.10 ist $\{x_k \ : \ k \in \mathbb{N}\} \cup \{a\}$ kompakt, damit nach Satz 2.12 beschränkt. Die durch Weglassen von $\{a\}$ entstehende Menge bleibt beschränkt. □

Satz 2.14 *Sei X ein metrischer Raum, $Y \subset X$ eine kompakte Teilmenge und $A \subset Y$ eine abgeschlossene Teilmenge. Dann ist A kompakt.*

Beweis. $X \setminus A$ ist offen. Ist $(U_i)_{i\in I}$ eine offene Überdeckung von A, dann ist $Y \subset X = (X \setminus A) \cup \bigcup_{i\in I} U_i$. Da Y kompakt, genügen endlich viele U_i zur Überdeckung von Y und damit auch von A. □

Wir zeigen nun, daß im $\mathbb{R}^n$ auch die Umkehrung von Satz 2.12 gilt.

Satz 2.15 (Heine-Borel) *Eine Teilmenge $A \subset \mathbb{R}^n$ ist genau dann kompakt, wenn sie beschränkt und abgeschlossen ist.*

Wegen Satz 2.12 ist nur die Richtung ($\Leftarrow$) zu zeigen. Wegen Satz 2.14 genügt es zu zeigen, daß der abgeschlossene Quader

$$Q = [a_1, b_1] \times \cdots \times [a_n, b_n] = \{(x_1, \ldots, x_n) \in \mathbb{R}^n \ : \ a_\nu \leq x_\nu \leq b_\nu\} \subset \mathbb{R}^n$$

kompakt ist, denn jede beschränkte abgeschlossene Teilmenge von $\mathbb{R}^n$ liegt in einem abgeschlossenen Quader.

Sei $(U_i)_{i\in I}$ eine unendliche offene Überdeckung von $Q_0 = Q$, die nicht auf eine endliche reduziert werden kann. Durch Halbierung aller Kanten zerlegen wir Q in 2^n gleich große Teilquader der halben Größe. Es gibt dann mindestens einen abgeschlossenen Teilquader, den wir mit Q_1 bezeichnen, der nicht durch endlich viele U_i überdeckt werden kann. Durch Wiederholung des Verfahrens finden wir eine Folge

$$Q = Q_0 \supset Q_1 \supset Q_2 \supset \ldots$$

von abgeschlossenen Quadern mit $\operatorname{diam}(Q_k) = \frac{1}{2^k} \operatorname{diam}(Q)$, so daß jeder von ihnen nicht durch endlich viele U_i überdeckt werden kann.

Wir zeigen: Es gibt einen Punkt $x \in Q_k$ für alle k. Jeder Quader Q_k enthält irgendeinen Punkt x_k. Sind $Q_k, Q_l \subset Q_m$, dann ist $d(x_k, x_l) \leq \operatorname{diam}(Q_m) <$

$\frac{2\,\mathrm{diam}(Q)}{2^m} = \epsilon$. Damit bilden die $(x_k)_{k\in I}$ eine Cauchy-Folge. Da im $\mathbb{R}^n$ jede Cauchy-Folge konvergiert, ist $x = \lim_{k\to\infty} x_k \in \mathbb{R}^n$. Da Q_k abgeschlossen ist und $x_l \in Q_k$ für alle $l \geq k$, ist $x \in Q_k$ für alle k.

Der gemeinsame Punkt x liegt in irgendeiner Umgebung U_j mit $j \in I$. Da U_j offen, gibt es ein $\epsilon > 0$, so daß $B_\epsilon(x) \subset U_i$. Dann finden wir aber auch ein $p \in \mathbb{N}$ mit $\mathrm{diam}(Q_p) = \frac{1}{2^p}\,\mathrm{diam}(Q) < \epsilon$. Somit gilt $Q_p \subset U_j$, d.h. die Quader lassen sich im Widerspruch zur Annahme durch endlich viele U_i überdecken. Also ist Q kompakt. □

Vorsicht: In allgemeinen metrischen Räumen gilt die Umkehrung von Satz 2.12 nicht!

Satz 2.16 *Sei $f : X \to Y$ eine stetige Abbildung zwischen metrischen Räumen X, Y. Ist $A \subset X$ kompakt, dann ist auch $f(A) \subset Y$ kompakt.*

Beweis. Sei $(U_i)_{i\in\mathbb{N}}$ eine offene Überdeckung von $f(A)$. Aus der Stetigkeit von f folgt, daß $V_i := f^{-1}(U_i)$ offen ist. Dann ist $A \subset \bigcup_{i\in I} V_i$, aber tatsächlich genügen endlich viele V_i zur Überdeckung: $A \subset V_{i_1} \cup \cdots \cup V_{i_n}$. Dann ist $f(A) \subset U_{i_1} \cup \cdots \cup U_{i_n}$. □

Wir können nun einige sehr wichtige Eigenschaften kompakter Teilmengen beweisen:

Satz 2.17 *Sei $Y \subset X$ ein kompakte Teilmenge eines metrischen Raumes X und $f : X \to \mathbb{R}$ eine stetige Funktion. Dann ist die Einschränkung $f|_Y$ auf Y beschränkt (d.h. $|f(y)| < \infty$ für alle $y \in Y$) und nimmt ihr Supremum und Infimum auf Y an, d.h. es gibt $p, q \in Y$ mit*

$$f(p) = \sup\{f(y) \;:\; y \in Y\} \qquad \textit{und} \qquad f(p) = \inf\{f(y) \;:\; y \in Y\}\ .$$

Beweis. Nach Satz 2.16 ist $f(Y) \subset \mathbb{R}$ kompakt und nach Satz 2.12 beschränkt (und abgeschlossen). Damit gibt es eine obere und untere Schranke für alle $f(y)$. Nun gibt es eine Folge $(y_k)_{k\in\mathbb{N}}$, die gegen das Supremum $\sup\{f(y) \;:\; y \in Y\} \in X$ konvergiert. Da $f(Y)$ abgeschlossen, liegt der Grenzwert in $f(Y)$. Aus der Stetigkeit folgt die Existenz von $p \in Y$, wo das Supremum angenommen wird. Analog für das Infimum. □

Satz 2.18 (Bolzano-Weierstraß) *Sei $Y \subset X$ ein kompakte Teilmenge eines metrischen Raumes X und $\{y_k\}_{k\in\mathbb{N}}$ eine Folge von Punkten aus Y. Dann gibt es einen Punkt $a \in Y$ und eine Teilfolge $\{y_{k_l}\}_{l\in\mathbb{N}}$ von $\{y_k\}_{k\in\mathbb{N}}$, die gegen a konvergiert.*

Beweis. Wäre das nicht der Fall, dann gibt es zu jedem Punkt $a \in Y$ eine offene Umgebung U_a, in der nur endlich viele Punkte y_k der Folge liegen. Nun ist $Y \subset \bigcup_{a\in Y} U_a$, da aber Y kompakt ist, wird Y bereits durch endlich viele Umgebungen überdeckt: $Y \subset U_{a_1} \cup \cdots \cup U_{a_m}$. Damit kann es insgesamt nur endlich viele Punkte y_k geben, im Widerspruch zur Annahme. $\square$

Einige Folgerungen aus dem Satz von Bolzano-Weierstraß:

- Jede beschränkte Folge im $\mathbb{R}^n$ besitzt eine konvergente Teilfolge: Wir können die Folge in einen abgeschlossenen Quader einbetten, welcher kompakt ist.
- Die unendlich-dimensionale Einheitssphäre ist beschränkt und abgeschlossen, aber nicht kompakt: Es sei

  $$\text{“}\mathbb{R}^\infty\text{”} = \ell^\infty := \big\{(x_1, x_2, x_3, \dots) \ : x_i \in \mathbb{R} \text{ beschränkt } \big\}\ .$$

  Mit dem kanonischen Skalarprodukt auf ℓ^∞ ist $S^\infty := \{x \in \ell^\infty \ : \ \|x\| = 1\}$. Wir betrachten die Folge

  $$y_0 = (1, 0, 0, 0, \dots)\,, \quad y_1 = (0, 1, 0, 0, \dots)\,, \quad y_2 = (0, 0, 1, 0, \dots)\,, \quad \dots .$$

  Es gilt $y_k \in S^\infty$ für alle k mit $d(y_k, y_l) = \sqrt{2}$ für alle $k \neq l$, so daß die Folge $(y_k)_{k\in\mathbb{N}}$ auf S^∞ nicht konvergiert. Nach dem Satz von Bolzano-Weierstraß kann S^∞ nicht kompakt sein. Ebenso ist die unendlich-dimensionale abgeschlossene Einheitskugel nicht kompakt.

Definition 2.13 Seien (X, d_X) und (Y, d_Y) metrische Räume. Eine Abbildung $f : X \to Y$ heißt ***gleichmäßig stetig***, wenn zu jedem $\epsilon > 0$ ein $\delta > 0$ existiert, so daß

$$d_Y\big(f(x_1), f(x_2)\big) < \epsilon \quad \text{für alle } x_1, x_2 \in X \text{ mit } d_X(x_1, x_2) < \delta\ .$$

Satz 2.19 *Seien X, Y metrische Räume und sei X kompakt. Dann ist jede stetige Abbildung $f : X \to Y$ auch gleichmäßig stetig.*

Beweis. Sei $\epsilon > 0$ global vorgegeben. Wegen der Stetigkeit von f gibt es zu jedem Punkt $a \in X$ ein $\delta(a) > 0$, so daß $d_Y\big(f(x), f(a)\big) < \frac{\epsilon}{2}$ für alle $x \in B_{2\delta(a)}(a)$. Zunächst ist $X = \bigcup_{a\in X} B_{\delta(a)}(a)$. Da X aber kompakt ist, gibt es endlich viele Punkte $a_1, \dots, a_k$ mit $X = B_{\delta(a_1)}(a_1) \cup \cdots \cup B_{\delta(a_k)}(a_k)$. Wir setzen $\delta := \min(\delta(a_1), \dots, \delta(a_k))$.

Seien nun zwei Punkte $x_1, x_2 \in X$ mit $d_X(x_1, x_2) < \delta$ gegeben. Der Punkt x_1 liege in der j-ten Umgebung, d.h. $x_1 \in B_{\delta(a_j)}(a_j)$. Dann ist $d_X(x_1, a_j) < \delta(a_j)$ und $d_X(x_2, a_j) \leq d_X(x_2, x_1) + d_X(x_1, a_j) < \delta(a_j) + \delta < 2\delta(a_j)$. Aus der Stetigkeit im Punkt a_j folgt:

$$d_Y\big(f(x_1), f(x_2)\big) \leq d_Y\big(f(x_1), f(a_j)\big) + d_Y\big(f(x_2), f(a_j)\big) < \epsilon$$

für alle x_1, x_2 mit $d_X(x_1, x_2) < \delta$. $\square$

2.5 Tangentialvektoren an Kurven im $\mathbb{R}^n$

Definition 2.14 Sei $I \subset \mathbb{R}$ ein Intervall, das nicht nur aus einem Punkt besteht. Eine Funktion $f : I \to \mathbb{R}$ heißt *differenzierbar* im Punkt $t \in I$, falls für jede gegen a konvergierende Folge $(x_k)_{k\in\mathbb{N}}$ von Punkten aus I eine reelle Zahl b und eine Folge $(g_k)_{k\in\mathbb{N}}$ existiert, so daß

$$f(x_k) - f(t) = b \cdot (x_k - t) + g_k \quad \text{mit} \quad \lim_{\substack{k\to\infty \\ x_k\neq t}} \frac{g_k}{x_k - t} = 0 \; .$$

Dann heißt

$$f'(t) := b = \lim_{\substack{k\to\infty \\ x_k\neq t}} \frac{f(x_k) - f(t)}{x_k - t}$$

die *Ableitung* der Funktion f im Punkt t.

Die Funktion f heißt differenzierbar auf I, wenn sie in jedem Punkt von I differenzierbar ist. Die Funktion f heißt *stetig differenzierbar* auf I, wenn ihre Ableitung $f' : I \to \mathbb{R}$ stetig ist.

Eine Funktion $f : I \to \mathbb{R}$, die im Punkt $t \in I$ differenzierbar ist, ist in t auch stetig.

Beispiel 2.3 i) $f(t) = t^n$ für $n \in \mathbb{N}$. Sei $(x_k)_{k\in\mathbb{N}}$ eine gegen t konvergierende Folge, mit $x_k \neq t$ für alle k. Nach binomischer Formel ist

$$\begin{aligned} f(x_k) = \big(t + (x_k - t)\big)^n &= \sum_{j=0}^{n} \binom{n}{j} t^j (x_k - t)^{n-j} \\ &= \underline{t^n} + \underline{\underline{n t^{n-1}(x_k - t)}} + \sum_{j=0}^{n-2} \binom{n}{j} t^j (x_k - t)^{n-j} \; . \end{aligned}$$

Damit ist $b = f'(t) = nt^{n-1}$ und $g_k = \sum_{j=0}^{n-2} \binom{n}{j} t^j (x_k - t)^{n-j}$ mit

$$\lim_{x_k\to t} \frac{g_k}{x_k - t} = \lim_{x_k\to t} \sum_{j=0}^{n-2} \binom{n}{j} t^j (x_k - t)^{n-j-1} = 0 \; .$$

ii) $f(t) = \cos t$. Nach dem Additionstheorem ist

$$\begin{aligned} \cos x_k &= \cos\big(t + (x_k - t)\big) \\ &= \underline{\cos t} - \cos t(1 - \cos(x_k{-}t)) \underline{\underline{- \sin t\,(x_k{-}t)}} - \sin t \big(\sin(x_k{-}t) - (x_k{-}t)\big) \; . \end{aligned}$$

Damit ist $b = f'(t) = \sin t$ und $g_k = -\cos t(1 - \cos(x_k{-}t)) - \sin t \big(\sin(x_k{-}t) - (x_k{-}t)\big)$ mit

$$\lim_{x_k\to t} \frac{g_k}{(x_k{-}t)} = \lim_{x_k\to t} \Big(- \frac{\cos t \sin^2 \frac{(x_k-t)}{2}}{2(x_k{-}t)} - \sin t \Big(\frac{\sin(x_k{-}t)}{(x_k{-}t)} - 1\Big)\Big) = 0 \; .$$

Oft schreibt man auch $f'(x) = \frac{df(x)}{dx}$. Wichtige Eigenschaften der Ableitung sind (Differenzierbarkeit aller Funktionen vorausgesetzt):

- Linearität: $h(t) = f(t) + g(t) \quad \Rightarrow \quad h'(t) = f'(t) + g'(t)$
- Leibniz-Regel: $h(t) = f(t) \cdot g(t) \quad \Rightarrow \quad h'(t) = f'(t) \cdot g(t) + f(t) \cdot g'(t)$
- Kettenregel: $h(t) = f\big(g(t)\big) \quad \Rightarrow \quad h'(t) = f'(g(t)) \cdot g'(t)$
  symbolisch: $\dfrac{df(g(t))}{dt} = \dfrac{df(g)}{dg}\dfrac{dg(t)}{dt}$

Definition 2.15 Unter einer *Kurve* im $\mathbb{R}^n$ versteht man eine stetige Abbildung $f : I \to \mathbb{R}^n$ (wobei $I \subset \mathbb{R}^n$ aus mehr als einem Punkt besteht). Die Kurve heißt *(stetig) differenzierbar* im Punkt $t \in I$, wenn jede der durch $f = (f_1, \dots, f_n)$ definierten Funktionen $f_k : I \to \mathbb{R}$ (stetig) differenzierbar ist.

Ist die Kurve $f = (f_1, \dots, f_n) : I \to \mathbb{R}^n$ differenzierbar in t, dann heißt

$$f'(t) := \big(f_1'(t), \dots, f_n'(t)\big) \in \mathbb{R}^n$$

der *Tangentialvektor* an f im Punkt $f(t)$.

Wir gehen auf zwei geometrische Konstruktionen mit Tangentialvektoren näher ein.

Gegeben seien zwei Kurven $f_1 : I_1 \to \mathbb{R}^n$ und $f_2 : I_2 \to \mathbb{R}^n$, die sich schneiden im Punkt $f_1(t_1) = f_2(t_2)$ mit $t_i \in I_i$. Sind die Kurven f_i differenzierbar in t_i, dann ist ihr Schnittwinkel θ gegeben durch

$$\cos\theta = \frac{\langle f_1'(t_1), f_2'(t_2)\rangle}{\|f_1'(t_1)\| \, \|f_2'(t_2)\|} \;.$$

Beispiel 2.4 $f : \mathbb{R} \to \mathbb{R}^2$ mit $f(t) = (t^2-1, t^3-t)$. Es gilt $f(1) = f(-1) = 0$ und $f'(t) = (2t, 3t^2 - 1)$, insbesondere $f'(-1) = (-2, 2)$ und $f'(1) = (2, 2)$. Dann ist der Winkel zwischen beiden Tangentialvekoren $f'(-1)$ und $f'(1)$ gegeben durch $\cos\theta = 0$.

Ist die Kurve $f : [a, b] \to \mathbb{R}^n$ stetig differenzierbar auf dem Intervall $[a, b]$, dann ist die *Bogenlänge* L der Kurve zwischen den Punkten $f(a)$ und $f(b)$ gegeben durch

$$L = \int_a^b \|f'(t)\| \, dt \;.$$

Tatsächlich läßt sich die Bogenlänge auch unter schwächeren Bedingungen als stetige Differenzierbarkeit definieren, z.B. stückweise stetige Differenzierbarkeit.

Beispiel 2.5 Wir betrachten die Schraubenlinie $f : \mathbb{R} \to \mathbb{R}^3$ gegeben durch $f(t) := (r\cos t, r\sin t, at)$. Dann ist $f'(t) = (-r\sin t, r\cos t, a)$ und

$$\|f'(t)\| = \sqrt{r^2\sin^2 t + r^2\cos^2 t + a^2} = \sqrt{r^2 + a^2}$$

ist unabhängig von t. Die Bogenlänge einer vollen Umdrehung $t \in [0, 2\pi]$, also von $(r, 0, 0)$ nach $(r, 0, a)$ ist also $L = \int_0^{2\pi} \sqrt{r^2 + a^2}\, dt = 2\pi\sqrt{r^2 + a^2}$.

2.6 Partielle Ableitungen

Es geht nun um Funktionen mehrerer Veränderlicher, also Abbildungen $f : U \to \mathbb{R}$ mit $U \subset \mathbb{R}^n$. Sei $x = (x_1, \ldots, x_n) \in U$. Wir halten alle Koordinaten bis auf die j-te fest und betrachten folgende in U eingebettete Intervalle:

$$I_j = \{(x_1, \ldots, x_{j-1}, t, x_{j+1}, x_n) \in U \ : \ t \in \tilde{I}_j \subset \mathbb{R} \text{ mit } x_j \in \tilde{I}_j\} \ .$$

Die Einschränkung der Funkion $f : U \to \mathbb{R}$ auf I_j wird dann zu einer gewöhnlichen Funktion $f\big|_{I_j} : I_j \to \mathbb{R}$ einer Veränderlicher mit $f\big|_{I_j}(t) = f(x_1, \ldots, x_{j-1}, t, x_{j+1}, x_n)$. Für diese Funktion können wir die Differenzierbarkeit im Punkt x_j betrachten. Das Ergebnis ist die partielle Ableitung von f in der j-ten Koordinatenrichtung:

Definition 2.16 Sei $U \subset \mathbb{R}^n$ eine offene Teilmenge. Eine Funktion $f : U \to \mathbb{R}$ heißt im Punkt $x \in U$ *partiell differenzierbar* in der j-ten Koordinatenrichtung, falls der Grenzwert

$$(\partial_j f)(x) := \lim_{\substack{h \to 0 \\ h \neq 0}} \frac{1}{h}\big(f(x + he_j) - f(x)\big)$$

existiert. Dabei ist $e_j \in \mathbb{R}^n$ der j-te Einheitsvektor und h ist so zu wählen, daß $x + he_j \in U$. Der Grenzwert $(\partial_j f)(x)$ heißt die *j-te partielle Ableitung* von f in x.

Eine Funktion $f : U \to \mathbb{R}$ heißt *partiell differenzierbar*, falls $(\partial_j f)(x)$ für alle $x \in U$ und alle $1 \leq j \leq n$ existiert, und *stetig partiell differenzierbar*, falls alle Funktionen $\partial_j f : U \to \mathbb{R}$ stetig sind.

Oft schreibt man auch $\frac{\partial f}{\partial x_j}$ an Stelle von $\partial_j f$. Die partielle Ableitung erfüllt die Leibniz-Regel: Seien $f, g : U \to \mathbb{R}$ partiell differenzierbare Funktionen auf U, dann gilt

$$(\partial_j(f \cdot g))(x) = (\partial_j f)(x) \cdot g(x) + f(x) \cdot (\partial_j g)(x) \ , \qquad x \in U \ .$$

Beispiel 2.6 Die partielle Ableitung des Radius $r(x) := \|x\| = \sqrt{x_1^2 + \cdots + x_n^2}$ in der j-ten Koordinatenrichtung ist nach der Kettenregel

$$\frac{\partial r}{\partial x_j} = \frac{d\sqrt{r^2}}{d(r^2)} \frac{\partial r^2}{\partial x_j} = \frac{1}{2\sqrt{r^2}} \cdot 2x_j = \frac{x_j}{r} \ .$$

Damit ist $r : \mathbb{R}^n \setminus \{0\} \to \mathbb{R}$ partiell differenzierbar mit $(\partial_j r)(x) = \frac{x_j}{r}$.

Entsprechend ist nach der Kettenregel jede differenzierbare Funktion $f(r)$ des Radius, aufgefaßt als Funktion $f : \mathbb{R}^n \setminus \{0\} \to \mathbb{R}$, partiell differenzierbar mit $(\partial_i f)(x) = x_j \frac{f'(r)}{r}$. Zum Beispiel sind für $f(r) = \mathrm{e}^{-ar^2}$ die partiellen Ableitungen gegeben durch $(\partial_j f)(x) = -2ax_j f(r)$.

Das folgende Beispiel zeigt, daß aus der partiellen Differenzierbarkeit einer Funktion nicht die Stetigkeit folgt.

Beispiel 2.7 Es sei $f : \mathbb{R}^2 \to \mathbb{R}$ gegeben durch

$$f(x_1, x_2) = \begin{cases} \frac{x_1 x_2}{x_1^2 + x_2^2} & \text{für } (x_1, x_2) \neq (0,0) \\ 0 & \text{für } (x_1, x_2) = (0,0) \end{cases}$$

Dann ist f auf $\mathbb{R}^2 \setminus \{(0,0)\}$ partiell differenzierbar mit

$$(\partial_1 f)(x_1, x_2) = \frac{x_2}{x_1^2 + x_2^2} - \frac{2x_1^2 x_2}{(x_1^2 + x_2^2)^2} = \frac{x_2(x_2^2 - x_1^2)}{(x_1^2 + x_2^2)^2}$$

und analog

$$(\partial_2 f)(x_1, x_2) = \frac{x_1(x_1^2 - x_2^2)}{(x_1^2 + x_2^2)^2} \ .$$

Im Nullpunkt haben wir

$$(\partial_1 f)(0,0) = \lim_{h \to 0} \frac{f(h,0) - f(0,0)}{h} = 0 \ , \quad (\partial_1 f)(0,0) = \lim_{h \to 0} \frac{f(0,h) - f(0,0)}{h} = 0 \ ,$$

so daß f auf dem gesamten $\mathbb{R}^2$ partiell differenzierbar ist. Jedoch ist f nicht stetig in $(0,0)$. Die Folge $(y_k)_{k \in \mathbb{N}}$ mit $y_k = (\frac{1}{k+1}, \frac{1}{k+1})$ konvergiert gegen $(0,0)$, aber

$$f(y_k) = \frac{\frac{1}{k+1} \cdot \frac{1}{k+1}}{(\frac{1}{k+1})^2 + (\frac{1}{k+1})^2} = \frac{1}{2}$$

konvergiert nicht gegen $f(0) = 0$.

Die partielle Ableitung einer partiell differenzierbaren Funktion kann nochmals partiell differenziert werden, usw. Induktiv definieren wir:

Definition 2.17 Sei $U \subset \mathbb{R}^n$ offen und $k \in \mathbb{N} \setminus \{0\}$. Eine Funktion $f : U \to \mathbb{R}$ heißt $(k{+}1)$-*mal partiell differenzierbar*, wenn sie k-mal partiell differenzierbar ist und alle partiellen Ableitungen $\partial_{i_k} \dots \partial_{i_1} f$ partiell differenzierbar sind. Sind die partiellen Ableitungen $\partial_{i_{k+1}} \partial_{i_k} \dots \partial_{i_1} f$ stetig, so heißt f eine $(k{+}1)$-*mal stetig partiell differenzierbare Funktion.*

Satz 2.20 (Schwarz) *Sei $U \subset \mathbb{R}^n$ offen und $f : U \to \mathbb{R}$ eine zweimal stetig partiell differenzierbare Funktion. Dann vertauschen die zweiten partiellen Ableitungen, d.h. für alle $a \in U$ und alle $i, j = 1, \dots, n$ gilt*

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

Zunächst erinnern wir an:

Lemma 1 (Mittelwertsatz der Differentialrechnung) *Sei $g : [a,b] \to \mathbb{R}$ stetig und zusätzlich $g :]a,b[\to \mathbb{R}$ differenzierbar. Dann gibt es einen Punkt $c \in]a,b[$ mit $g'(c) = \frac{g(b)-g(a)}{b-a}$.*

Beweis. Wir beweisen das Lemma zunächst für $g(a) = g(b)$ (Satz von Rolle). Nach Satz 2.17 hat g auf der kompakten Menge $[a, b]$ ein Maximum und ein Minimum. Ist g nicht konstant (für $g = \text{const}$ ist alles klar), so nimmt g ein Maximum oder Minimum in einem inneren Punkt $c \in]a, b[$ an. Nun ist g im Punkt c differenzierbar, und damit gilt $g'(c) = 0$.

Im allgemeinen Fall setzen wir $h(x) = g(x) - \frac{g(b)-g(a)}{b-a}(x - a)$. Dann ist $h(a) = h(b)$, so daß es einen Punkt $c \in]a, b[$ gibt mit $h'(c) = g'(c) - \frac{g(b)-g(a)}{b-a} = 0$. □

Beweis des Satzes von Schwarz. Der Übersichtlichkeit wegen sei $i = 1, j = 2$ (kann durch Umnumerieren der Koordinaten immer erreicht werden) und dann $n = 2$ (die weiteren Komponenten sind festgehalten und spielen keine Rolle).

Für gegebenes $\delta > 0$ sei $Q_\delta(a) \subset U \subset \mathbb{R}^2$ der offene Quader mit Kantenlänge δ und Mittelpunkt $a = (x_0, y_0)$, und sei (x, y) ein beliebiger Punkt von Q, d.h. $|x - x_0| < \delta$ und $|y - y_0| < \delta$. Für festgehaltenes y sei $F_y(x) = f(x, y) - f(x, y_0)$. Nach dem Mittelwersatz gibt es einen Punkt $\xi \in]x_0 - \delta, x_0 + \delta[$, so daß

$$\begin{aligned} & F_y(x) - F_y(x_0) = F'_y(\xi) \cdot (x - x_0) \\ \Rightarrow \quad & f(x,y) - f(x,y_0) - f(x_0,y) + f(x_0,y_0) = \big((\partial_1 f)(\xi, y) - (\partial_1 f)(\xi, y_0)\big) \cdot (x - x_0) \,. \end{aligned}$$

Wir nutzen den Mittelwertsatz nochmals für die Funktion $G_\xi(y) := (\partial_1 f)(\xi, y)$. Es gibt also ein $\eta \in]y_0 - \delta, y_0 + \delta[$, so daß

$$\begin{aligned} & G_\xi(y) - G_\xi(y_0) = G'_\xi(\eta) \cdot (y - y_0) \\ \Rightarrow \quad & (\partial_1 f)(\xi, y) - (\partial_1 f)(\xi, y_0) = (\partial_2 \partial_1 f)(\xi, \eta) \cdot (y - y_0) \,. \end{aligned}$$

Insgesamt gibt es somit ein $(\xi, \eta) \in Q$ mit

$$f(x,y) - f(x,y_0) - f(x_0,y) + f(x_0,y_0) = (\partial_2 \partial_1 f)(\xi, \eta) \cdot (x - x_0)(y - y_0) \,.$$

Wir können aber auch erst x festhalten und den Mittelwertsatz in y anwenden, und als letztes den Mittelwertwsatz in x. Im Ergebnis gibt es einen neuen Punkt $(\tilde{\xi}, \tilde{\eta}) \in Q$ mit

$$f(x,y) - f(x,y_0) - f(x_0,y) + f(x_0,y_0) = (\partial_1 \partial_2 f)(\tilde{\xi}, \tilde{\eta}) \cdot (x - x_0)(y - y_0) \,.$$

Somit gilt $(\partial_2 \partial_1 f)(\xi, \eta) = (\partial_1 \partial_2 f)(\tilde{\xi}, \tilde{\eta})$. Lassen wir δ gegen 0 streben, so konvergieren (ξ, η) und $(\tilde{\xi}, \tilde{\eta})$ gegen (x, y), und aus der vorausgesetzten Stetigkeit der zweiten partiellen Ableitungen folgt $(\partial_2 \partial_1 f)(x, y) = (\partial_1 \partial_2 f)(x, y)$. □

Entsprechend können bei k-mal stetig partiell differenzierbaren Funktionen die partiellen Ableitungen in beliebiger Reihenfolge geschrieben werden:

$$\partial_{i_1} \dots \partial_{i_k} f = \partial_{\pi(i_1)} \dots \partial_{\pi(i_k)} f$$

für eine beliebige Permutation π der Indizes $i_1, \ldots, i_k$, denn jede Permutation läßt sich durch Vertauschen benachbarter Elemente darstellen. Es ist deshalb auch üblich, die mehrfachen partiellen Ableitungen zu schreiben als

$$\partial_{i_1} \ldots \partial_{i_k} f = \frac{\partial^k f}{\partial x_{i_1} \ldots \partial x_{i_k}} .$$

2.7 Gradient, Divergenz, Rotation, Laplace

Definition 2.18 Sei $U \subset \mathbb{R}^n$ offen und $f : U \to \mathbb{R}$ eine partiell differenzierbare Funktion. Dann heißt der Vektor

$$(\operatorname{grad} f)(x) = \Big(\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n}\Big) \in \mathbb{R}^n$$

der *Gradient* von f im Punkt $x \in U$.

Geometrisch ist der Gradient in Richtung des steilsten Anstiegs der Funktion gerichtet, und die Norm $\|(\operatorname{grad} f)(x)\|$ ist ein Maß für die Stärke des Anstiegs. Zum Beispiel gilt $\operatorname{grad} r = \frac{x}{r}$, der steilste Anstieg ist also radial nach außen gerichtet und vom Betrag her überall (außer im Nullpunkt) konstant. Im Beispiel $\operatorname{grad} \frac{1}{r} = -\frac{x}{r^3}$ ist der steilste Anstieg radial nach innen gerichtet.

Der Gradient ist linear und erfüllt die Leibniz-Regel: Sind $f, g : U \to \mathbb{R}$ partiell differenzierbar, so gilt

$$\begin{aligned}(\operatorname{grad}(f+g))(x) &= (\operatorname{grad} f)(x) + (\operatorname{grad} g)(x) , \\ (\operatorname{grad}(f \cdot g))(x) &= (\operatorname{grad} f)(x) \cdot g(x) + f(x) \cdot (\operatorname{grad} g)(x) .\end{aligned}$$

Der Gradient $(\operatorname{grad} f)(x)$ ordnet jedem Punkt $x \in U$ einen Vektor zu. So etwas nennt man ein Vektorfeld:

Definition 2.19 Unter einem *Vektorfeld* auf einer offenen Teilmenge $U \subset \mathbb{R}$ versteht man eine Abbildung $v : U \to \mathbb{R}^n$. Das Vektorfeld heißt stetig/partiell differenzierbar/..., wenn alle Komponenten von v stetig/partiell differenzierbar/... sind.

Die Vorstellung ist, daß an jedem Punkt $x = (x_1, \ldots, x_n) \in U$ ein Vektor $v(x) = (v_1(x), \ldots, v_n(x))$ angeheftet ist. Eine nützliche Konstruktion besteht darin, diese Vektoren als Tangentialvektoren an Kurven $f : I \to \mathbb{R}^n$ durch $x = f(t_0)$ zu betrachten. Ist $f(t) = (f_1(t), \ldots, f_n(t))$, dann ist also

$$\frac{df_i(t)}{dt} = v_i(f(t)) , \quad f_i(t_0) = x_i \qquad \text{für } i = 1, \ldots, n \text{ und } t, t_0 \in I$$

Die Kurve selbst ist dann durch Lösen (Integration) dieser $2n$ Gleichungen zu erhalten und heißt *Integralkurve* des Vektorfeldes durch x. In der Physik werden die Bilder der Integralkurven auch *Feldlinien* bzw. *Stromlinien* des Vektorfelds genannt. Unter recht schwachen Voraussetzungen an das Vektorfeld (Lipschitz-stetig) kann man in einer genügend kleinen Umgebung von x die Integralkurven immer finden, aber nicht unbedingt auf ganz U, weil das an Singularitäten des Vektorfeldes scheitern kann. Eine Art von Singularität ist die Divergenz:

Definition 2.20 Sei $v = (v_1, \dots, v_n) : U \to \mathbb{R}^n$ ein partiell differenzierbares Vektorfeld auf einer offenen Teilmenge $U \subset \mathbb{R}^n$. Dann heißt die Funktion

$$\mathrm{div}(v) := \frac{\partial v_1}{\partial x_1} + \dots \frac{\partial v_n}{\partial x_n}$$

die *Divergenz* des Vektorfeldes v.

Die Divergenz ist linear, $\mathrm{div}(v + w) = \mathrm{div}(v) + \mathrm{div}(w)$ und erfüllt folgendes Analogon zur Leibniz-Regel: Sei v ein partiell differenzierbares Vektorfeld und f eine partiell differenzierbare Funktion auf $U \subset \mathbb{R}^n$, dann gilt

$$\mathrm{div}(f \cdot v) = \langle \mathrm{grad}(f), v \rangle + f \cdot \mathrm{div}(v) \; .$$

Die Divergenz eines Vektorfeldes ist ein Maß für die Gesamtbeschleunigung der Integralkurven. Das kann einerseits dadurch erreicht werden, daß die Feldlinien von der Parallelität abweichen (was wahrscheinlich den Namen motiviert hat), oder durch Geschwindigkeitszunahme entlang der Feldlinien. Für $x \in \mathbb{R}^n$ haben wir z.B. $\mathrm{div}\, x = n$. Hier sind die Feldlinien radial vom Nullpunkt nach außen gerichtet und gerade durch den Radius parametrisiert. Ein anderes Beispiel ist

$$\mathrm{div} \frac{x}{r} = \left\langle \mathrm{grad} \frac{1}{r}, x \right\rangle + \frac{1}{r} \cdot \mathrm{div}\, x = \frac{n-1}{r} \qquad \text{für } x \neq 0 \; .$$

Die Vektoren $v(x) = \frac{x}{\|x\|}$ haben in jedem Punkt $x \in \mathbb{R}^n \setminus \{0\}$ die Länge 1, aber die Feldlinien sind wieder radial nach außen gerichtet und damit nicht parallel.

Gradient und Divergenz sind wichtige Hilfsmittel der Theoretischen Physik. In der Mechanik läßt sich ein konservatives Kraftfeld als (negativer) Gradient eines Potentials schreiben, $F = -\mathrm{grad}\, V$. In der Elektrodynamik wird die Divergenz von Vektorfeldern zur Formulierung der Maxwellschen Gleichungen gebraucht. Dort gibt es noch eine weitere Konstruktion mit partiellen Ableitungen, die *Rotation*. Die Rotation und das *Vektorprodukt* beruhen auf einer nur im $\mathbb{R}^3$ möglichen Identifikation von sogenannten *Differentialformen*, deren Ideen wir im folgenden vorstellen.

Definition 2.21 Sei V ein reeller Vektorraum und $\{e_1, \dots, e_n\}$ eine Basis von V. Dann heißt

$$\Lambda^k(V) := \begin{cases} V & \text{für } k = 0 \\ \mathrm{span}\big(e_{i_1} \wedge \dots \wedge e_{i_k} \; : \; 1 \le i_1 < i_2 < \dots < i_k \le n\big) & \text{für } 1 \le k \le n \\ 0 & \text{für } k > n \end{cases}$$

der *Vektorraum der k-fach antisymmetrischen Tensoren über V*. Dabei ist die Verknüpfung $\wedge$ als formelles Symbol anzusehen.

Sei $U \subset \mathbb{R}^n$ offen. Eine stetige/partiell differenzierbare/... Abbildung $\omega : U \to \Lambda^k(\mathbb{R}^n)$ heißt *stetige/partiell differenzierbare/... k-Form* auf U. Der Raum der k-Formen auf U wird mit $\Omega^k(U)$ bezeichnet.

Eine k-Form ist also die Zuordnung eines k-fach antisymmetrischen Tensors zu jedem Punkt von U. Insbesondere können 1-Formen auf U mit Vektorfeldern auf U identifiziert werden, und 0-Formen auf U sind nichts anderes als Funktionen auf U. Im allgemeinen nimmt man an, daß die Formen beliebig oft differenzierbar sind; sie heißen dann auch Differentialformen. Ist $\dim(V) = n$, so gilt $\dim(\Lambda^k(V)) = \binom{n}{k}$.

Definition 2.22 Sei $\{e_1, \ldots, e_n\}$ eine Basis von V. Die durch die Vereinbarung $e_i \wedge e_j = -e_j \wedge e_i$ für alle $i, j = 1, \ldots, n$ und lineare Fortsetzung entstehende assoziative Multiplikation

$$\wedge : \Lambda^k(V) \times \Lambda^l(V) \to \Lambda^{k+l}(V)$$

heißt das *äußere Tensorprodukt* (auch *wedge-Produkt* wegen der englischen Bezeichnung "wedge" (Keil) für $\wedge$). Entsprechend heißt

$$\wedge : \Omega^k(U) \times \Omega^l(U) \to \Omega^{k+l}(U)$$

das *äußere Formenprodukt*.

Sei z.B. $n \geq 3$ und $\omega_1 = f \cdot e_2$ und $\omega_2 = g \cdot e_1 \wedge e_3$ für Funktionen $f, g : U \to \mathbb{R}$, d.h. $\omega_1(x) = f(x)\, e_2$ und $\omega_2(x) = g(x)\, e_1 \wedge e_3$ für $x \in U$, dann ist

$$\omega_1 \wedge \omega_2 = (f \cdot g) \cdot (e_2 \wedge e_1 \wedge e_3) = -(f \cdot g) \cdot (e_1 \wedge e_2 \wedge e_3) \; .$$

Allgemein gilt für $\omega_k \in \Omega_k(U)$ und $\tilde{\omega}_l \in \Omega_l(U)$ die graduierte Kommutativität

$$\omega_k \wedge \tilde{\omega}_l = (-1)^{kl} \tilde{\omega}_l \wedge \omega_k \; .$$

Definition 2.23 Sei $(e_1, \ldots, e_n)$ die Standardbasis von $\mathbb{R}^n$. Die durch

$$d\omega := \sum_{i=1}^{n} e_i \wedge \partial_i(\omega)$$

für eine partiell differenzierbare k-Form $\omega \in \Omega^k(U)$ definierte Abbildung $d : \Omega^k(U) \to \Omega^{k+1}(U)$ heißt *äußere Ableitung* oder *äußeres Differential*.

Sei z.B. $n = 3$ und $\omega = f_1 \cdot (e_2 \wedge e_3) + f_2 \cdot (e_1 \wedge e_3)$, dann ist $d\omega = (\partial_1 f_1) \cdot (e_1 \wedge e_2 \wedge e_3) - (\partial_2 f_2) \cdot (e_1 \wedge e_2 \wedge e_3)$.

Satz 2.21
- *Ist $f : U \to \mathbb{R}$ eine partiell differenzierbare Funktion, dann gilt $df = \text{grad}(f)$.*
- *Sind $\omega_k \in \Omega^k(U)$ und $\tilde{\omega}_l \in \Omega^l(U)$ partiell differenzierbare Formen, dann gilt die* graduierte Leibniz-Regel

$$d(\omega_k \wedge \tilde{\omega}_l) = (d\omega_k) \wedge \tilde{\omega}_l + (-1)^k \omega_k \wedge d\tilde{\omega}_l \ .$$

- *Ist $\omega \in \Omega^k(U)$ zweimal stetig partiell differenzierbar, dann gilt $d(d\omega) = 0$.*

Bemerkungen zum Beweis. i) ist klar. ii) folgt aus der Leibniz-Regel für die partielle Ableitung und der graduierte Kommutativität des $\wedge$-Produkts. iii) folgt aus dem Satz von Schwarz über die Vertauschbarkeit der zweiten partiellen Ableitungen zusammen mit der Antisymmetrie $e_i \wedge e_j = -e_j \wedge e_i$. □

Definition 2.24 Sei $(e_1, \ldots, e_n)$ die Standardbasis von $\mathbb{R}^n$. Die durch

$$*(e_{i_1} \wedge \cdots \wedge e_{i_k}) := \omega \quad \text{falls } \omega \wedge (e_{i_1} \wedge \cdots \wedge e_{i_k}) = e_1 \wedge e_2 \wedge \cdots \wedge e_n$$

und lineare Fortsetzung definierte Abbildung $* : \Omega^k(U) \to \Omega^{n-k}(U)$ heißt die *Hodge-Abbildung*.

Ist z.B. $n = 3$, so gilt $*(e_1 \wedge e_3) = -e_2$, denn $(-e_2) \wedge (e_1 \wedge e_3) = e_1 \wedge e_2 \wedge e_3$. Die Hodge-Abbildung ist ein Isomorphismus von Vektorräumen, denn $\dim(\Lambda^k(\mathbb{R}^n)) = \binom{n}{k} = \binom{n}{n-k} = \dim(\Lambda^{n-k}(\mathbb{R}^n))$. Das Inverse der Hodge-Abbildung ist, bis auf ein mögliches Vorzeichen, die Hodge-Abbildung selbst: $*(*\omega_k) = (-1)^{k(n-k)}\omega_k$ für $\omega_k \in \Omega^k(U)$ und $U \subset \mathbb{R}^n$.

Definition 2.25
- Sei $U \subset \mathbb{R}^n$. Die durch $\delta\omega_k := -(-1)^{nk} * (d(*\omega_k))$ für eine partiell differenzierbare k-Form $\omega_k \in \Omega^k(U)$ definierte Abbildung $\delta : \Omega^k(U) \to \Omega^{k-1}(U)$ heißt *Kodifferential*.
- Sei $U \subset \mathbb{R}^n$ und $v \in \Omega^1(U)$ ein partiell differenzierbares Vektorfeld. Dann heißt die Funktion

$$\text{div}(v) := *(d(*v)) = (-1)^{n+1}\delta v \in \Omega^0(U)$$

die *Divergenz von* v.
- Sei $U \subset \mathbb{R}^3$ und $v, w \in \Omega^1(U)$ Vektorfelder auf U. Dann heißt das Vektorfeld

$$v \times w := *(v \wedge w) \in \Omega^1(U)$$

das *Vektorprodukt* von v, w.

- Sei $U \subset \mathbb{R}^3$ und $v \in \Omega^1(U)$ ein partiell differenzierbares Vektorfeld auf U. Dann heißt das Vektorfeld

$$\mathrm{rot}(v) := *(dv) \in \Omega^1(U)$$

die *Rotation* von v.

Wir rechnen nach, das unsere Definition des Vektorprodukts mit der üblichen Vorschrift zusammenfällt. Sei $v = v_1e_1 + v_2e_2 + v_3e_3$ und $w = w_1e_1 + w_2e_2 + w_3e_3$, dann ist

$$\begin{array}{llllll} v \wedge w = v_1w_2\, e_1 \wedge e_2 & + v_1w_3\, e_1 \wedge e_3 & - v_2w_1\, e_1 \wedge e_2 & + v_2w_3\, e_2 \wedge e_3 & - v_3w_1\, e_1 \wedge e_3 & - v_3w_2\, e_2 \wedge e_3 \\ *(v \wedge w) = v_1w_2\, e_3 & - v_1w_3\, e_2 & - v_2w_1\, e_3 & + v_2w_3\, e_1 & + v_3w_1\, e_2 & - v_3w_2\, e_1 \end{array}$$

also tatsächlich

$$(v \times w)_1 = v_2w_3 - v_3w_2\,, \quad (v \times w)_2 = v_3w_1 - v_1w_3\,, \quad (v \times w)_3 = v_1w_2 - v_2w_1\,.$$

Entsprechend rechnen wir die Rotation aus: Sei $v = v_1e_1 + v_2e_2 + v_3e_3$, wobei v_i nun Funktionen sind, dann gilt

$$\begin{array}{llllll} dv = \frac{\partial v_2}{\partial x_1}\, e_1 \wedge e_2 & + \frac{\partial v_3}{\partial x_1}\, e_1 \wedge e_3 & - \frac{\partial v_1}{\partial x_2}\, e_1 \wedge e_2 & + \frac{\partial v_3}{\partial x_2}\, e_2 \wedge e_3 & - \frac{\partial v_1}{\partial x_3}\, e_1 \wedge e_3 & - \frac{\partial v_2}{\partial x_3}\, e_2 \wedge e_3 \\ *(dv) = \frac{\partial v_2}{\partial x_1}\, e_3 & - \frac{\partial v_3}{\partial x_1}\, e_2 & - \frac{\partial v_1}{\partial x_2}\, e_3 & + \frac{\partial v_3}{\partial x_2}\, e_1 & + \frac{\partial v_1}{\partial x_3}\, e_2 & - \frac{\partial v_2}{\partial x_3}\, e_1 \end{array}$$

also tatsächlich

$$(\mathrm{rot}\, v)_1 = \frac{\partial v_3}{\partial x_2} - \frac{\partial v_2}{\partial x_3}\,, \quad (\mathrm{rot}\, v)_2 = \frac{\partial v_1}{\partial x_3} - \frac{\partial v_3}{\partial x_1}\,, \quad (\mathrm{rot}\, v)_3 = \frac{\partial v_2}{\partial x_1} - \frac{\partial v_1}{\partial x_2}\,.$$

Satz 2.22 *Sei $U \subset \mathbb{R}^3$ offen, $f : U \to \mathbb{R}$ eine zweifach stetig partiell differenzierbare Funktion und $v : U \to \mathbb{R}^3$ ein zweifach stetig partiell differenzierbares Vektorfeld. Dann gilt*

$$\mathrm{rot}(\mathrm{grad}\, f) = *(d(df)) = 0\,, \qquad \mathrm{div}(\mathrm{rot}\, v) = *(d(*(*(dv)))) = 0\,.$$

Ein wichtige Rolle innerhalb der Mathematik spielen Fragestellungen der folgenden Art (übersetzt in physikalische Begriffe):

i) Gegeben sei ein rotationsfreies Vektorfeld $v : U \to \mathbb{R}^3$ mit $U \subset \mathbb{R}^3$ und $\mathrm{rot}(v) = 0$. Unter welchen Voraussetzungen an U existiert eine Funktion $f : U \to \mathbb{R}$, so daß $v = \mathrm{grad}(f)$ gilt?

ii) Gegeben sei ein divergenzfreies Vektorfeld $v : U \to \mathbb{R}^3$ mit $U \subset \mathbb{R}^3$ und $\mathrm{div}(v) = 0$. Unter welchen Voraussetzungen an U existiert ein Vektorfeld $w : U \to \mathbb{R}^n$, so daß $v = \mathrm{rot}(w)$ gilt?

In der Formensprache scheiben sich die Maxwellschen Gleichungen wie folgt:

$$\begin{aligned} dB &= 0 & \delta B + \dot{E} &= -j \\ dE + \dot{B} &= 0 & \delta E &= \rho \end{aligned}$$

Dabei bezeichnet der Punkt über einer Form die Zeitableitung, und die folgenden Formen treten auf:

- $E(t) \in \Omega^1(U)$ ist das elektromagnetische Feld
- $B(t) \in \Omega^2(U)$ ist das Magnetfeld (die Physik arbeitet mit $B_{\text{Physik}} = *B \in \Omega^1(U)$)
- $j(t) \in \Omega^1(U)$ ist die Stromdichte
- $\rho(t) \in \Omega^0(U)$ ist die Ladungsdichte

Wegen $\delta\delta = \pm * d * *d* = 0$ gilt die Kontinuitätsgleichung $\delta j + \dot{\rho} = 0$.

Die links stehenden Gleichungen kann man unter geeigneten Voraussetzungen an U durch $B = dA$ und $E = -d\phi - \dot{A}$ lösen mit $A \in \Omega^1(U)$ und $\phi \in \Omega^0(U)$. Unter der Voraussetzung $\delta A = -\dot{\phi}$ ergibt sich für die rechten Gleichungen

$$\ddot{A} - (\delta d + d\delta)A = j \qquad \ddot{\phi} - \delta d\phi = \rho \ .$$

Definition 2.26 Der durch $\Delta\omega := \delta(d\omega) + d(\delta\omega)$ für $\omega \in \Omega^k(U)$ definierte Operator $\Delta = \delta d + d\delta : \Omega^k(U) \to \Omega^k(U)$ heißt *Laplace-Beltrami-Operator*. Im Spezialfall $k = 0$ heißt $\Delta f = \delta(df) = \operatorname{div}(\operatorname{grad} f)$ der *Laplace-Operator* für eine zweifach partiell differenzierbare Funktion $f : U \to \mathbb{R}$.

Für den Laplace-(Beltrami-)Operator gibt es die Darstellung

$$\Delta\omega = \frac{\partial^2\omega}{\partial x_1^2} + \cdots + \frac{\partial^2\omega}{\partial x_n^2} , \qquad \omega \in \Omega^k(U) \ , \ U \subset \mathbb{R}^n \ ,$$

was zumindest für Funktionen $\omega = f \in \Omega^0(U)$ leicht zu überprüfen ist.

Folgende Differentialgleichungen enthalten den Lapace-Operator:

$\ddot{f} - \Delta f = \rho$	inhomogene Wellengleichung
$\ddot{f} - \Delta f = 0$	homogene Wellengleichung
$\Delta f = -\rho$	Poisson-Gleichung
$\Delta f = 0$	Potentialgleichung
$c\dot{f} - \Delta f = 0$	Wärmeleitungsgleichung

Die Maxwellschen Gleichungen im Vakuum ($j = 0$ und $\rho = 0$) führen also auf insgesamt vier skalare Wellengleichungen (elektromagnetische Wellen). Zeitunabhängige elektrische Felder werden durch die Poisson-Gleichung beschrieben. Gibt es zusätzlich keine Ladungensdichte, dann ist die Potentialgleichung zu lösen. Die allgemeine Lösungstheorie für solche partiellen Differentialgleichungen wird im 4. Semester behandelt.

Wichtig ist noch zu zeigen, daß man die Bedingung $\delta A + \dot{\phi} = 0$, die uns die Differentialgleichungen für A und ϕ entkoppelt hat, wirklich lösen kann. Dabei ist entscheidend, daß A und ϕ durch E und B nicht eindeutig festgelegt sind; es gibt eine sogenannte Eichfreiheit: Setzt man $A \mapsto A + df$ und $\phi \mapsto -\dot{f}$ für $f \in \Omega^0(U)$, so bleiben E, B unverändert. Die Separierungsgleichung geht aber über in

$$\delta A + \dot{\phi} \mapsto \delta A + \dot{\phi} - (\ddot{f} - \Delta f) \ .$$

Man bestimmt also zunächst A, ϕ als Lösung der inhomogenen Wellengleichungen und berechnet daraus E und B. Die Eichbedingung $\delta A + \dot{\phi} = 0$ wird zunächst verletzt sein, wir müssen aber nur wissen, daß die Eichung f so gewählt werden kann, daß $\delta A + \dot{\phi} = 0$ zu erfüllen ist. Dann ist die Herleitung der Wellengleichungen für die eichtransformierten Potentiale A, ϕ korrekt. Da aber E und B durch f nicht geändert werden, bestimmen sich die Feldstärken E, B aus Lösungen A, ϕ der Wellengleichungen auch ohne Berücksichtigung der Eichbedingung.

2.8 Totales Differential

Definition 2.27 Sei $U \subset \mathbb{R}$ eine offene Teilmenge. Eine Abbildung $f : U \to \mathbb{R}^m$ heißt *total differenzierbar* (oder einfach nur *differenzierbar*) im Punkt $x \in U$, falls es eine lineare Abbildung $A(x) : \mathbb{R}^n \to \mathbb{R}^m$ und eine auf einer offenen Umgebung V von $0 \in \mathbb{R}^n$ definierte Abbildung $\phi : V \to \mathbb{R}^m$ gibt, so daß

$$f(x+\xi) = f(x) + A(x) \circ \xi + \phi(\xi) \quad \text{mit} \quad \lim_{\substack{\xi \to 0 \in \mathbb{R}^n \\ \xi \neq 0}} \frac{\phi(\xi)}{\|\xi\|} = 0 \qquad \text{für alle } \xi \in V \;.$$

Dann heißt die lineare Abbildung $(Df)(x) := A(x)$ das *totale Differential* (oder einfach nur das *Differential*) von f im Punkt x.

Einige Bemerkungen:

- In Definition 2.14 haben wir die Differenzierbarkeit im $\mathbb{R}^1$ so eingeführt, daß das totale Differential die direkte Verallgemeinerung ist.
- Um den Restterm nicht ganz so mühsam zu charakterisieren, schreibt man einfach $o(|\xi\|)$ und meint eine Abbildung $\phi : V \to \mathbb{R}^m$ mit obigen Eigenschaften.
- Die lineare Abbildung $A(x) : \mathbb{R}^n \to \mathbb{R}^n$ ist bezüglich der Standardbasis durch eine $(m \times n)$-Matrix gegeben, die wir mit dem gleichen Buchstaben bezeichnen, $A(x) = (a_{ij}(x)) \in M(m \times n, \mathbb{R})$. Variiert man den Punkt $x \in U$, so ist Df also durch $m \cdot n$ Funktionen $a_{ij} : U \to \mathbb{R}$ bestimmt.
- Of schreibt man auch df für das totale Differential. Wir reservieren d für das äußere Differential. Für $f : U \to \mathbb{R}$ stimmen df und Df überein (siehe nächster Satz), aber bei mehrfachen Differentiationen ist $DDf \neq ddf = 0$.

Satz 2.23 *Sei $U \subset \mathbb{R}^n$ offen und sei $f : U \to \mathbb{R}^m$ im Punkt $x \in U$ total differenzierbar mit $f(x+\xi) = f(x) + A(x) \cdot \xi + o(\|\xi\|)$ und $A(x) = (a_{ij}(x))$. Dann gilt:*

i) *f ist im Punkt $x \in U$ stetig.*

ii) *Alle Komponenten $f_i : U \to \mathbb{R}$ von $f = (f_1, \dots, f_m)$ sind im Punkt x partiell differenzierbar mit $(\partial_j f_i)(x) = a_{ij}(x)$.*

Beweis. i) Wegen $\lim_{\xi\to 0} A(x)\cdot\xi = 0$ und $\lim_{\xi\to 0} o(\|\xi\|) = 0$ gilt $\lim_{\xi\to 0} f(x+\xi) = f(x)$. Damit ist f stetig.

ii) Ist e_k der k-te Basisvektor der Standardbasis, dann ist $A(x)\circ e_j = \sum_{i=1}^m a_{ij}(x)e_i$, so daß für $\xi = h\cdot e_j$ gilt

$$f_i(x+h\cdot e_j) = f_i(x) + a_{ij}(x)\cdot h + o(h) \qquad \Rightarrow \qquad (\partial_j f_i)(x) = a_{ij}(x)\ . \qquad \square$$

Satz 2.24 *Sei $U\subset\mathbb{R}^n$ offen und $f : U\to\mathbb{R}$ partiell differenzierbar auf U. Sind alle partiellen Ableitungen $\partial_j f$ stetig im Punkt $x\in U$, dann ist f im Punkt x total differenzierbar.*

Beweis. Da U offen, gibt es ein $\delta > 0$ mit $B_\delta(x)\subset U$. Wir wählen ein $\xi = (\xi_1,\dots,\xi_n)\in\mathbb{R}^n$ mit $\|\xi\| < \delta$ und betrachten die Punkte $z^{(k)} := x + \sum_{j=1}^k \xi_j e_j$. Es gilt $z^{(0)} = x$ und $z^{(n)} = x+\xi$. Da sich benachbarte $z^{(k-1)}$ und $z^{(k)}$ nur in der k-ten Koordinate unterscheiden, können wir den Mittelwertsatz der Differentialrechnung anwenden: Es gibt also ein $\eta^{(k)}\in\mathbb{R}$ mit $|\eta^{(k)}| < \xi_k$, so daß

$$f(z^{(k)}) - f(z^{(k-1)}) = \xi_k\cdot(\partial_k f)(y^{(k)})\ , \qquad y^{(k)} := z^{(k-1)} + \eta^{(k)}e_k\ .$$

Das bedeutet

$$\begin{aligned} f(x+\xi) &= f(x) + \sum_{k=1}^n \xi_k\cdot(\partial_k f)(y^{(k)}) \\ &= f(x) + \sum_{k=1}^n (\partial_k f)(x)\cdot\xi_k + \underbrace{\sum_{k=1}^n \big((\partial_k f)(y^{(k)}) - (\partial_k f)(x)\big)\cdot\xi_k}_{\phi(\xi)}\ . \end{aligned}$$

Für $\xi\to 0$ strebt y_k gegen x. Aus der Stetigkeit der partiellen Ableitungen folgt $\lim_{\xi\to 0}(\partial_k f)(y^{(k)}) = (\partial_k f)(x)$ und damit $\lim_{\substack{\xi\to 0\in\mathbb{R}^n\\ \xi\neq 0}} \dfrac{\phi(\xi)}{\|\xi\|} = 0.$ $\square$

Durch Kombination der Sätze 2.23 und 2.24 folgt, daß jede stetig partiell differenzierbare Funktion $f : U\to\mathbb{R}$ auf U auch stetig ist. Außerdem gelten folgende Implikationen:

$$\begin{aligned} & f : U\to\mathbb{R} \text{ stetig partiell differenzierbar} \\ \Rightarrow\quad & f : U\to\mathbb{R} \text{ total differenzierbar} \\ \Rightarrow\quad & f : U\to\mathbb{R} \text{ partiell differenzierbar} \end{aligned}$$

Die Umkehrungen gelten im allgemeinen nicht.

Satz 2.25 (Kettenregel) *Seien $U\subset\mathbb{R}^n$ und $V\subset\mathbb{R}^m$ offene Mengen sowie $f : U\to\mathbb{R}^m$ und $g : V\to\mathbb{R}^k$ Abbildungen mit $f(U)\subset V$. Die Abbildung f sei im Punkt $x\in U$ differenzierbar und g im Punkt $f(x)\in V$. Dann ist die Abbildung $g\circ f : U\to\mathbb{R}^k$ im Punkt $x\in U$ differenzierbar, und es gilt*

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

Beweis. Nach Voraussetzung gilt

$$f(x+\xi) = f(x) + (Df)(x)\cdot\xi + o(\|\xi\|)\ , \qquad g(y+\eta) = g(y) + (Dg)(y)\cdot\eta + o(\|\eta\|)\ .$$

Wir berechnen

$$\begin{aligned}
(g\circ f)(x+\xi) = g(f(x+\xi)) &= g\Big(\underbrace{f(x)}_{y} + \underbrace{(Df)(x)\cdot\xi + o(\|\xi\|)}_{\eta}\Big) \\
&= g(f(x)) + (Dg)(f(x))\cdot\big((Df)(x)\cdot\xi + o(\|\xi\|)\big) \\
&\quad + o\big(\big\|(Df)(x)\cdot\xi + o(\|\xi\|)\big\|\big) \\
\text{Linearität} \Rightarrow \quad &= (g\circ f)(x) + \big((Dg)(f(x))\cdot(Df)(x)\big)\cdot\xi \\
&\quad + (Dg)(f(x))\cdot o(\|\xi\|) + o\big(\big\|(Df)(x)\cdot\xi + o(\|\xi\|)\big\|\big)\ .
\end{aligned}$$

Da die letzte Zeile wieder $o(\|\xi\|)$ ist, folgt die Behauptung. $\square$

Definition 2.28 Sei $U\subset\mathbb{R}^n$ offen, $x\in U$ und $v\in\mathbb{R}^n$ ein Vektor mit $\|v\| = 1$. Dann heißt der Differentialquotient

$$(D_v f)(x) := \lim_{\substack{t\to 0\\ t\neq 0}} \frac{f(x+tv) - f(x)}{t}$$

die *Richtungsableitung* der Funktion $f : U\to\mathbb{R}$ im Punkt x in Richtung v.

Insbesondere sind die partiellen Ableitungen die Richtungsableitungen in Richtung der Standardbasisvektoren, $(D_{e_i}f)(x) = (\partial_i f)(x)$.

Satz 2.26 *Sei $U\subset\mathbb{R}^n$ offen und $f : U\to\mathbb{R}$ stetig (partiell) differenzierbar. Dann gilt für jeden Punkt $x\in U$ und jeden Vektor $v\in\mathbb{R}^n$ mit $\|v\| = 1$*

$$(D_v f)(x) = \langle v, (\operatorname{grad} f)(x)\rangle\ .$$

Beweis. Eine stetig differenzierbare Funktion ist nach Satz 2.24 auch total differenzierbar, so daß $(D_v f)(x) = (df)(x)\cdot v$. Nach Satz 2.23 ist $(df)(x)$ eine $(1\times n)$-Matrix $(df)(x) = (a_{1j}(x))$ mit $a_{1j}(x) = (\partial_j f)(x)$. Ist $v = (v_1,\dots,v_j)$, dann gilt

$$(df)(x)\cdot v = \sum_{j=1}^n (\partial_j f)(x)\cdot v_j = \langle v, (\operatorname{grad} f)(x)\rangle\ . \qquad \square$$

Ist $\operatorname{grad} f \neq 0$ und ist θ der Winkel zwischen v und $\operatorname{grad} f$, dann gilt für das Skalarprodukt

$$(D_v f)(x) = \|(\operatorname{grad} f)(x)\|\cos\theta\ .$$

Damit ist die Richtungsableitung maximal, wenn v in Richtung $\operatorname{grad} f$ zeigt. Das bedeutet, daß $\operatorname{grad} f$ die Richtung des stärksten Anstiegs von f angibt.

2.9 Die Taylor-Formel

Um die im folgenden auftretenden vielen Indizes übersichtlicher zu gestalten, hat sich eine abkürzende Schreibweise eingebürgert. Sei $\alpha = (\alpha_1, \dots, \alpha_n) \in \mathbb{N}^n$ ein Multiindex, d.h. ein n-Tupel von Indizes $\alpha_i \in \mathbb{N}$. Dann setzt man

$$|\alpha| := \alpha_1 + \alpha_2 + \cdots + \alpha_n \;, \qquad \alpha! := \alpha_1!\,\alpha_2!\cdots\alpha_n! \;.$$

Für eine $|\alpha|$-mal stetig partiell differenzierare Funktion $f : U \to \mathbb{R}$ schreibt sich eine mehrfache partielle Ableitung wie folgt:

$$(\partial^\alpha f)(x) := (\partial_1^{\alpha_1} \dots \partial_n^{\alpha_n} f)(x) = \frac{\partial^{|\alpha|} f}{\partial x_1^{\alpha_1} \dots \partial x_n^{\alpha_n}}(x) \;.$$

Dabei ist $\partial_1^{\alpha_i} f = \underbrace{\partial_i \dots \partial_i}_{\alpha_i \text{mal}} f$. Ebenso setzt man $x^\alpha := x_1^{\alpha_1} \cdots x_n^{\alpha_n}$.

Satz 2.27 *Sei $U \subset \mathbb{R}$ offen und sei $f : U \to \mathbb{R}$ eine k-mal stetig differenzierbare Funktion. Zu $x \in U$ sei ein Vektor $\xi \in \mathbb{R}^n$ so gewählt, daß $x + t\xi \in U$ für alle $t \in [0,1]$. Dann ist die Funktion einer Veränderlichen*

$$g : [0,1] \mapsto \mathbb{R} \;, \qquad g(t) := f(x + t\xi)$$

k-mal stetig differenzierbar auf $[0,1]$, und es gilt

$$\frac{d^k g}{dt^k}(t) = \sum_{|\alpha|=k} \frac{k!}{\alpha!} \xi^\alpha \,(\partial^\alpha f)(x + t\xi) \;.$$

Dabei läuft die Summe über alle Multiindizes α mit $|\alpha| = k$.

Beweis. Zunächst verwenden wir nicht den Satz von Schwarz, d.h. wir betrachten $\partial_i\partial_j$ und $\partial_j\partial_i$ als verschieden. Es gilt $x + t\xi = (x_1 + t\xi_1, \dots, x_n + t\xi_n)$. Durch Subtraktion und sofortige Addition geeigneter Zwischenterme und Ausnutzen der Kettenregel erhalten wir

$$\begin{aligned}
\frac{dg}{dt}(t) &= \lim_{h\to 0} \frac{1}{h}\Big(f(x_1{+}(t{+}h)\xi_1, \dots, x_n{+}(t{+}h)\xi_n) - f(x_1{+}t\xi_1, \dots, x_n{+}t\xi_n)\Big) \\
&= \lim_{h\to 0} \frac{1}{h}\Big(f(x_1{+}(t{+}h)\xi_1, x_2{+}(t{+}h)\xi_2, \dots, x_n{+}(t{+}h)\xi_n) \\
&\qquad\qquad - f(x_1{+}t\xi_1, x_2{+}(t{+}h)\xi_2, \dots, x_n{+}(t{+}h)\xi_n) \\
&\qquad + . - . + \dots \\
&\qquad + f(x_1{+}t\xi_1, \dots, x_{n-1}{+}t\xi_{n-1}, x_n{+}(t{+}h)\xi_n) \\
&\qquad\qquad - f(x_1{+}t\xi_1, \dots, x_{n-1}{+}t\xi_{n-1}, x_n{+}t\xi_n)\Big) \\
&= \xi_1(\partial_1 f)(x + t\xi) + \cdots + \xi_n(\partial_n f)(x + t\xi) \;.
\end{aligned}$$

Führt man diese Schritte k-mal hintereinander aus, so entsteht

$$\frac{d^k g}{dt^k}(t) = \sum_{i_1,i_2,\ldots,i_k=1}^{n} \xi_{i_k} \cdots \xi_{i_2}\xi_{i_1}(\partial_{i_k} \ldots \partial_{i_2}\partial_{i_1} f)(x+t\xi) \ .$$

Nach dem Satz von Schwarz können wir die partiellen Ableitungen ordnen und zu ∂^α zusammenfassen. Ebenso fassen sich die Produkte der ξ_i zu ξ^α zusammen. Es gibt $\frac{k!}{\alpha_1!\cdots\alpha_n!}$ verschiedene $\partial_{i_k} \ldots \partial_{i_2}\partial_{i_1}$, die nach Umordnung das gleiche ∂^α ergeben. □

Satz 2.28 (Taylor) *Sei $U \subset \mathbb{R}$ offen und sei $f : U \to \mathbb{R}$ eine $(k+1)$-mal stetig differenzierbare Funktion. Zu $x \in U$ sei ein Vektor $\xi \in \mathbb{R}^n$ so gewählt, daß $x + t\xi \in U$ für alle $t \in [0,1]$. Dann gibt es ein $\theta \in [0,1]$, so daß*

$$f(x+\xi) = \sum_{|\alpha|\le k} \frac{\xi^\alpha}{\alpha!}(\partial^\alpha f)(x) + \sum_{|\alpha|=k+1} \frac{\xi^\alpha}{\alpha!}(\partial^\alpha f)(x+\theta\xi) \ .$$

Beweis. Wir betrachten die Funktion einer Veränderlichen $g : [0,1] \to \mathbb{R}$, die durch $g[t] := f(x+t\xi)$ gegeben ist. Nach dem vorigen Satz ist g eine $(k+1)$ mal stetig differenzierbare Funktion, auf die wir den eindimensionalen Satz von Taylor anwenden können: Es gibt also ein $\theta \in [0,1]$ mit

$$g(1) = \sum_{j=0}^{k} \frac{1}{j!}\frac{d^j g}{dt^j}(0) + \frac{1}{(k+1)!}\frac{d^{k+1} g}{dt^{k+1}}(\theta) \ .$$

Einsetzen von $g(t) = f(x+t\xi)$ und Verwenden von Satz 2.27 liefert die Behauptung. □

Der Satz von Taylor ist wichtig bei Abschätzungen der folgenden Art:

Satz 2.29 *Sei $U \subset \mathbb{R}^n$ offen und $f : U \to \mathbb{R}$ eine k-mal stetig differenzierbare Funktion. Dann gilt für jedes $x \in U$*

$$f(x+\xi) = \sum_{|\alpha|\le k} \frac{\xi^\alpha}{\alpha!}(\partial^\alpha f)(x) + o(\|\xi\|^k) \qquad \textit{für } \xi \to 0 \ .$$

Beweis. Da U offen, gibt es ein $\delta > 0$, so daß $B_\delta(x) \subset U$. Dann gibt es zu jedem $\xi \in \mathbb{R}^n$ mit $\|\xi\| < \delta$ ein $\theta \in [0,1]$ mit

$$\begin{aligned} f(x+\xi) &= \sum_{|\alpha|\le k-1} \frac{\xi^\alpha}{\alpha!}(\partial^\alpha f)(x) + \sum_{|\alpha|=k} \frac{\xi^\alpha}{\alpha!}(\partial^\alpha f)(x+\theta\xi) \\ &= \sum_{|\alpha|\le k} \frac{\xi^\alpha}{\alpha!}(\partial^\alpha f)(x) + \sum_{|\alpha|=k} \frac{\xi^\alpha}{\alpha!}\big((\partial^\alpha f)(x+\theta\xi) - (\partial^\alpha f)(x)\big) \ . \end{aligned}$$

Es gilt $|\xi^\alpha| = |\xi_1|^{\alpha_1} \cdots |\xi_n|^{\alpha_n} \leq \|\xi\|^{\alpha_1} \cdots \|\xi\|^{\alpha_n} = \|\xi\|^{|\alpha|}$. Da $\partial^\alpha f$ stetig ist, gilt

$$\lim_{\xi \to 0} \frac{1}{\|\xi\|^k} \sum_{|\alpha|=k} \frac{\xi^\alpha}{\alpha!} \big((\partial^\alpha f)(x + \theta\xi) - (\partial^\alpha f)(x) \big) = 0 \ .$$

Das ist genau die Behauptung. □

Für $k = 1$ ist die Menge aller Multiindizes α mit $|\alpha| = 1$ gerade die Menge der Standardbasisvektoren (e_i). Somit erhalten wir die Formel für das totale Differential

$$f(x + \xi) = f(x) + \langle \xi, (\mathrm{grad} f)(x) \rangle + o(\|\xi\|) \ .$$

Für $k = 2$ ergibt sich

$$f(x + \xi) = f(x) + \langle \xi, (\mathrm{grad} f)(x) \rangle + \frac{1}{2} \sum_{i,j=1}^{n} \xi_i \xi_j (\partial_i \partial_j f)(x) + o(\|\xi\|^2) \ .$$

Dabei gibt es die Möglichkeiten $\alpha = (0, \dots, 0, 2, 0, \dots, 0)$ mit $\alpha! = 2$ und $\alpha = (0, \dots, 0, 1, 0, \dots, 0, 1, 0, \dots, 0)$ mit $\alpha! = 1$, wobei die Einsen an der i-ten und j-ten Stelle stehen mit $i < j$. Da eine in i, j symmetrische Funktion summiert wird, kann die Summe mit $i < j$ durch die halbe Summe mit $i \neq j$ ersetzt werden. Zusammen mit der Summe über $i = j$ von $\alpha = (0, \dots, 0, 2, 0, \dots, 0)$ ergibt sich die Beziehung.

Definition 2.29 Sei $U \subset \mathbb{R}^n$ offen und $f : U \to \mathbb{R}$ eine zweifach stetig differenzierbare Funktion. Dann heißt die symmetrische $n \times n$-Matrix

$$(\mathrm{Hess}\, f)(x) := \big((\partial_i \partial_j f)(x) \big)_{1 \leq i,j \leq n}$$

die *Hessesche Matrix* von f im Punkt x.

Somit gilt

$$f(x + \xi) = f(x) + \langle a, \xi \rangle + \frac{1}{2} \langle \xi, A\xi \rangle + o(\|\xi\|^2)$$
$$\text{mit } a = (\mathrm{grad}\, f)(x) \ , \ A = (\mathrm{Hess}\, f)(x) \ .$$

Die Hessesche Matrix ist wichtig bei der Untersuchung von lokalen Extrema einer Funktion.

Definition 2.30 Sei $U \subset \mathbb{R}^n$ offen. Ein Punkt $x \in U$ heißt *lokales Maximum* bzw. *lokales Minimum*, wenn es eine Umgebung $V \subset U$ von x gibt, so daß

$$f(x) \geq f(y) \quad \text{bzw.} \quad f(x) \leq f(y) \qquad \text{für alle } y \in V \ .$$

Gilt in diesem Fall $f(x) = f(y) \quad \Leftrightarrow \quad x = y$, so heißt das lokale Maximum (bzw. Minimum) *strikt*. Ein *lokales Extremum* ist ein lokales Minimum oder lokales Maximum.

Der folgende Satz liefert eine notwendige Bedingung für ein lokales Extremum:

Satz 2.30 *Sei $U \subset \mathbb{R}^n$ offen und $f : U \to \mathbb{R}$ partiell differenzierbar. Besitzt f in $x \in U$ ein lokales Extremum, so gilt* $(\operatorname{grad} f)(x) = 0$.

Beweis. Es genügt, die n Funktionen einer Veränderlichen $g_i := f(x + te_i)$ zu betrachten, wobei e_i der i-te Standardbasisvektor ist und $t \in [-\epsilon, \epsilon]$. Hat f ein lokales Extremum in x, so hat g_i ein lokales Extremum in 0. Dann gilt $0 = g'(0) = (\partial_i f)(x)$. Da das für jedes $1 \leq i \leq n$ gilt, folgt die Behauptung. □

Wir beweisen eine hinreichende Bedingung für lokale Extrema unter Verwendung der Hesseschen Matrix. Dazu erinnern wir an die Definitheit einer symmetrischen Matrix:

Definition 2.31 Eine symmetrische Matrix $A = A^t \in M(n \times n, \mathbb{R})$ heißt

- *positiv definit*, falls $\langle \xi, A\xi \rangle > 0 \quad \forall \xi \in \mathbb{R}^n \setminus \{0\}$,
- *positiv semidefinit*, falls $\langle \xi, A\xi \rangle \geq 0 \quad \forall \xi \in \mathbb{R}^n \setminus \{0\}$,
- *negativ (semi)definit*, falls $-A$ positiv (semi)definit ist,
- *indefinit*, falls es $\xi, \eta \in \mathbb{R}^n$ gibt mit $\langle \xi, A\xi \rangle > 0$ und $\langle \eta, A\eta \rangle < 0$.

Für praktische Bestimmungen der Definitheit ist das Determinantenkriterium (Satz 1.13) hilfreich.

Satz 2.31 *Sei $U \subset \mathbb{R}^n$ offen und $f : U \to \mathbb{R}$ zweimal stetig differenzierbar. In einem Punkt $x \in U$ gelte* $(\operatorname{grad} f)(x) = 0$.

i) *Ist* $(\operatorname{Hess} f)(x)$ *positiv definit, so besitzt f in x ein striktes lokales Minimum.*

ii) *Ist* $(\operatorname{Hess} f)(x)$ *negativ definit, so besitzt f in x ein striktes lokales Maximum.*

iii) *Ist* $(\operatorname{Hess} f)(x)$ *indefinit, so besitzt f in x kein lokales Extremum.*

Beweis. Zur Vereinfachung der Schreibweise sei $A := (\operatorname{Hess} f)(x)$. In einer Umgebung V von x gilt

$$f(x + \xi) = f(x) + \frac{1}{2}\langle \xi, A\xi \rangle + \phi(\xi) \qquad \text{mit } \phi(\xi) = o(\|\xi\|^2) \, .$$

Es gibt also zu jedem $\epsilon > 0$ ein $\delta > 0$, so daß $|\phi(\xi)| < \epsilon \|\xi\|^2$ für alle $\xi \in \mathbb{R}^n$ mit $\|\xi\| < \delta$.

i) Die $(n-1)$-Sphäre $S^{n-1} := \{\eta \in \mathbb{R}^n \ : \ \|\eta\| = 1\}$ ist kompakt. Nach Satz 2.17 nimmt die stetige Funktion $g : S^{n-1} \to \mathbb{R}$ mit $g(\eta) := \langle \eta, A\eta \rangle$ auf S^{n-1} ihr Supremum und ihr Infimum an. Es gibt also ein $\mu > 0$ mit

$$\mu := \min_{\eta \in S^{n-1}} \{\langle \eta, A\eta \rangle\} \, .$$

Da für beliebiges $\xi \in \mathbb{R}^n \setminus \{0\}$ gilt $\frac{1}{\|\xi\|}\xi \in S^{n-1}$, folgt $\langle \xi, A\xi\rangle \geq \mu\|\xi\|^2$ für alle $\xi \in \mathbb{R}^n$ (für $\xi = 0$ trivialerweise). Wählen wir δ so klein, daß $|\phi(\xi)| \leq \frac{\mu}{4}\|\xi\|^2$, so gilt $\frac{1}{2}\langle \xi, A\xi\rangle + \phi(\xi) \geq \frac{\mu}{4}\|\xi\|^2$. Folglich haben wir

$$f(x+\xi) \geq f(x) + \frac{\mu}{4}\|\xi\|^2 \qquad \text{für alle } \xi \in \mathbb{R}^n \text{ mit } \|\xi\| < \delta\ .$$

Also hat f in x ein striktes lokales Minimum. Analog beweist man ii).

iii) Wir zeigen, daß es $y', y'' \in U$ gibt mit $f(y') < f(x) < f(y'')$. Nach Voraussetzung gibt es ein $\xi \in \mathbb{R}^n$ mit $\langle \xi, A\xi\rangle := \mu > 0$. Dann gilt für hinreichend kleine $t > 0$

$$f(x+t\xi) = f(x) + \frac{\mu}{2}t^2 + \phi(t\xi)\ .$$

Wie zuvor finden wir ein $\delta > 0$, so daß $|\phi(t\xi)| \leq \frac{\mu}{4}t^2$ gilt für alle $0 < t < \delta$. Dann ist $f(x+t\xi) > f(x)$ für alle $0 < t < \delta$. Ebenso folgt aus der Existenz eines $\eta \in \mathbb{R}^n$ mit $\langle \eta, A\eta\rangle := -\mu' < 0$, daß $f(x+t'\eta) < f(x)$ für alle $0 < t' < \delta'$. $\square$

Ist die Hessesche Matrix im Punkt x positiv oder negativ semidefinit, so muß man höhere Ordnungen in der Taylorsche Formel betrachten, um Aussagen über Extrema von f mit $(\operatorname{grad} f)(x) = 0$ zu gewinnen.

Beispiel 2.8 $f : \mathbb{R}^2 \to \mathbb{R}$ gegeben durch $f(x,y) = 1 + x^2 + y^2$
Es gilt $(\operatorname{grad} f)(x,y) = (2x, 2y)$, folglich kann f nur in $(0,0)$ ein lokales Extremum haben. Wir testen $(\operatorname{Hess} f)(x,y) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$. Da $(\operatorname{Hess} f)(0,0)$ positiv definit, hat f im Punkt $(0,0)$ ein lokales Minimum.

$f : \mathbb{R}^2 \to \mathbb{R}$ gegeben durch $f(x,y) = 1 + x^2 - y^2$
Es gilt $(\operatorname{grad} f)(x,y) = (2x, -2y)$, folglich kann f nur in $(0,0)$ ein lokales Extremum haben. Wir testen $(\operatorname{Hess} f)(x,y) = \begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix}$. Damit ist $(\operatorname{Hess} f)(0,0)$ indefinit, so daß f im Punkt $(0,0)$ kein Extremum hat.

2.10 Der Satz über implizite Funktionen

Es geht nun um Funktionen, die implizit definiert sind, z.B. durch Gleichungen der Form $0 = F(x, f(x)) = (f(x))^2 + x^2 - 1$. Wir werden untersuchen, unter welchen Bedingungen sich derartige Gleichungen zumindest im Prinzip nach $f(x)$ auflösen lassen und welche Differenzierbarkeitseigenschaften die Lösungen haben. Im obigen Beispiel ist offenbar $f(x) = \pm\sqrt{1-x^2}$. Differentiation von $F(x, f(x))$ nach der Kettenregel liefert

$$\begin{aligned} 0 &= F'(x,f(x)) = (\partial_1 F)(x,f(x)) + (\partial_2 F)(x,f(x))f'(x) \\ \Rightarrow \quad f'(x) &= \frac{(\partial_1 F)(x,f(x))}{(\partial_2 F)(x,f(x))} = \frac{x}{f(x)}\ . \end{aligned}$$

Satz 2.32 (Banachscher Fixpunktsatz) *Sei $A \subset X$ eine abgeschlossene Teilmenge eines Banachraums X (d.h. eines vollständigen normierten Vektorraums $(X, \| \ \|)$). Die Abbildung $\Phi : A \to A$ sei eine* Kontraktion*, d.h. es gibt eine Konstante $\theta \in]0,1[$, so daß*

$$\|\Phi(f) - \Phi(g)\| \leq \theta |f - g\| \qquad \textit{für alle } f, g \in A \ .$$

Dann besitzt Φ genau einen Fixpunkt f_*, d.h. es gibt ein eindeutig bestimmtes $f_* \in A$ mit $\Phi(f_*) = f_*$. Für einen belieigen Anfangspunkt $g \in A$ konvergiert die Folge $(f_k)_{k\in\mathbb{N}}$ definiert durch

$$f_0 := g \ , \quad f_{k+1} = \Phi(f_k) \textit{ für } k \in \mathbb{N}$$

gegen f_*, d.h. $\lim_{k\to\infty} f_k = f_*$.

Beweis. i) Wir beweisen zunächst die Eindeutigkeit des Fixpunktes. Gäbe es zwei Fixpunkte $f_*, g*$, dann ist

$$\|f_* - g_*\| = \|\Phi(f_*) - \Phi(g_*)\| \leq \theta \|f_* - g_*\| \ ,$$

und damit $\|f_* - g_*\| = 0$ wegen $0 < \theta < 1$. Das bedeutet $f_* = g_*$.

ii) Sei $(f_k)_{k\in\mathbb{N}}$ wie oben definiert. Dann ist

$$\|f_{k+1} - f_k\| = \|\Phi(f_k - \Phi(f_{k-1})\| \leq \theta \|f_k - f_{k-1}\| \leq \cdots \leq \theta^k \|f_1 - f_0\| \ .$$

Für $m > l$ betrachten wir

$$\begin{aligned} \|f_m - f_l\| &= \Big\| \sum_{k=l}^{m-1} (f_{k+1} - f_k) \Big\| \leq \sum_{k=l}^{m-1} \|f_{k+1} - f_k\| \leq \sum_{k=l}^{m-1} \theta^k \|f_1 - f_0\| \\ &= \frac{(1-\theta^m) - (1-\theta^l)}{1-\theta} \|f_1 - f_0\| \leq \theta^l \frac{\|f_1 - f_0\|}{1-\theta} \ . \end{aligned}$$

Damit ist $(f_k)_{k\in\mathbb{N}}$ eine Cauchy-Folge, die wegen der Vollständigkeit des Banach-Raums gegen einen Punkt f_* konvergiert. Da A abgeschlossen, liegt der Grenzwert f_* sogar in A. Aus $f_* = \lim_{k\to\infty} f_k$ und $f_{k+1} = \Phi(f_k)$ folgt $\Phi(f_*) = f_*$. □

Satz 2.33 *Sei $U \subset \mathbb{R}^n$ offen. Der Vektorraum $\mathcal{C}_b(U, \mathbb{R}^m)$ der beschränkten stetigen Abbildungen $f : U \to \mathbb{R}^m$, zusammen mit der Norm $\|f\| := \sup_{x\in U} \|f(x)\|$, ist ein Banachraum.*

Beweis. Zu zeigen ist die Vollständigkeit. Sei $(f_k)_{k\in N}$ bezüglich dieser Norm eine Cauchy-Folge von Abbildungen $f_k \in \mathcal{C}_b(U, \mathbb{R}^m)$, d.h. für jedes $\epsilon > 0$ gibt es ein $k \in \mathbb{N}$ mit $\sup_{x\in U} \|f_i(x) - f_j(x)\| < \epsilon$ für alle $i, j \geq k$. Dann ist auch $\|f_i(x) - f_j(x)\| < \epsilon$ für alle $i, j \geq k$ und alle $x \in U$. Da zu gegebenem $x \in U$ gilt $f_i(x) \in \mathbb{R}^m$ und jede

Cauchy-Folge im $\mathbb{R}^n$ konvergiert, wird durch $f(x) := \lim_{k\to\infty} f_k(x)$ punktweise eine Abbildung $f : U \to \mathbb{R}^m$ definiert.

Zu zeigen ist, daß f stetig ist und daß $(f_k)_{k\in\mathbb{N}}$ in $C_b(U, \mathbb{R}^m)$ gegen f konvergiert. Da $(f_k)_{k\in\mathbb{N}}$ eine Cauchy-Folge in $C_b(U, \mathbb{R}^m)$ ist, gibt es zu jedem $\epsilon > 0$ ein $k \in \mathbb{N}$, so daß für alle $i, j \geq k$ und einen beliebigen Punkt $x \in U$ gilt $\|f_i(x) - f_j(x)\| < \epsilon$. Das gilt aber auch für den Grenzwert $j \to \infty$:

$$\lim_{j\to\infty} \|f_i(x) - f_j(x)\| = \|f_i(x) - f(x)\| < \epsilon \quad \text{für alle } x \in U \;.$$

Das bedeutet, daß $(f_k)_{k\in\mathbb{N}}$ gleichmäßig gegen f konvergiert. Nach Satz 2.8 ist f dann auch stetig. Da $\|f_i(x) - f(x)\| < \epsilon$ für $i \geq k$ und f_i beschränkt ist, ist f ebenfalls beschränkt, und

$$\|f_i - f\| = \sup_{x\in U} \|f_i(x) - f(x)\| < \epsilon \qquad \text{für alle } i \geq k \;.$$

Also gilt $\lim_{k\to\infty} f_k = f \in \mathcal{C}_b(U, \mathbb{R}^m)$. □

Satz 2.34 (über implizite Funktionen) *Seien $U_1 \subset \mathbb{R}^n$ und $U_2 \subset \mathbb{R}^m$ offene Teilmengen und $F = (F_1, \dots, F_m) : U_1 \times U_2 \to \mathbb{R}^m$ eine stetig differenzierbare Abbildung. Es sei $(a, b) \in U_1 \times U_2$ ein Punkt mit $F(a, b) = 0$, und im Punkt (a, b) sei die Jacobi-Matrix*

$$\Big(\frac{\partial F}{\partial y}\Big)(a,b) := \begin{pmatrix} \frac{\partial F_1}{\partial y_1} & \cdots & \frac{\partial F_1}{\partial y_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial F_m}{\partial y_1} & \cdots & \frac{\partial F_m}{\partial y_m} \end{pmatrix}(a,b) \in GL(m, \mathbb{R})$$

invertierbar. Dann gibt es eine offene Umgebung $V_1 \subset U_1$ von a und eine Umgebung $V_2 \subset U_2$ von b sowie eine stetig differenzierbare Abbildung $g : V_1 \to V_2 \subset \mathbb{R}^m$, so daß

$$F(x, g(x)) = 0 \qquad \text{für alle } x \in V_1 \;.$$

Mit anderen Worten: Ist eine implizit gegebene Gleichung $F(x, g(x)) = 0$ in einem Punkt a lösbar mit $g(x) = b$ und $F(a, b) = 0$, dann ist sie (unter den gegebenen Voraussetzungen) sogar in einer Umgebung V_1 von a lösbar.

Beweis. Wir setzen $B := \big(\frac{\partial F}{\partial y}\big)(a, b) \in GL(m, \mathbb{R})$. Damit werde eine Abbildung $G : U_1 \times U_2 \to \mathbb{R}^m$ definiert durch

$$G(x, y) := y - B^{-1} \cdot F(x, y) \;, \qquad x \in U_1 \;,\; y \in U_2 \subset \mathbb{R}^m \;.$$

Die Jacobi-Matrix von G im Punkt $(x, y) \in U_1 \times U_2$ ist

$$\Big(\frac{\partial G}{\partial y}\Big)(x,y) = E_m - B^{-1} \cdot \Big(\frac{\partial G}{\partial y}\Big)(x,y) \qquad \Rightarrow \qquad \Big(\frac{\partial G}{\partial y}\Big)(a,b) = 0 \in M(m\times m, \mathbb{R}) \;.$$

Da die Matrixmultiplikation als lineare stetige Abbildung $M(m\times m,\mathbb{R})\times\mathbb{R}^m\to\mathbb{R}^m$ aufgefaßt werden kann, wird nach Definition 2.11 eine Norm auf $M(m\times m,\mathbb{R})$ erklärt. Insbesondere gilt $\|Ax\|\le\|A\|\|x\|$ für alle $x\in\mathbb{R}^m$ und $A\in M(m\times m,\mathbb{R})$. Da F auf $U_1\times U_2$ stetig differenzierbar ist, gibt es Umgebungen $W_1\subset U_1$ von a und $W_2\subset U_2$ von b, so daß

$$\left\|\Big(\frac{\partial G}{\partial y}\Big)(x,y)\right\|\le\frac{1}{2}\qquad\text{für alle } (x,y)\in W_1\times W_2\ . \tag{1}$$

Entscheidend ist die Beobachtung

$$F(x,y)=0\qquad\Leftrightarrow\qquad y=G(x,y)$$

und insbesondere $b=G(a,b)$. Damit führen wir die Lösung von $F(x,y)=0$ nach y auf ein Fixpunktproblem zurück.

(1. Konstruktion der Abbildung g) Wir geben zunächst einen vereinfachten Beweis, der die Abbildung g nur punktweise konstruiert und keine Informationen über Stetigkeit liefert. Die Gebiete können dann etwas größer gewählt werden.

Nach dem Satz von Taylor gilt

$$G(x,y)-G(x,\eta)=\Big(\frac{\partial G}{\partial y}\Big)(x,\eta)\cdot(y-\eta)+\phi(y-\eta)\ ,\quad \phi(y-\eta)=o(\|y-\eta\|)\ .$$

für alle $x,y\in W_1\times W_2$. Es gibt also ein $r>0$, so daß

$$\|\phi(y-\eta)\|\le\frac{1}{4}\|y-\eta\|\qquad\text{für alle } y\in W_2 \text{ mit } \|y-\eta\|<\frac{5}{2}r\ . \tag{2}$$

Es sei $V_2:=\{y\in W_2\ :\ \|y-b\|\le r\}\subset W_2$ die abgeschlossene Kugel mit Mittelpunkt b und Radius r. Damit gilt

$$\|G(x,y)-G(x,\eta)\|\le\left\|\Big(\frac{\partial G}{\partial y}\Big)(x,\eta)\right\|\|y-\eta\|+\|\phi(y-\eta)\|\le\frac{3}{4}\|y-\eta\| \tag{3}$$

für alle $x\in W_1$ und alle $y,\eta\in V_2$. Folglich ist für beliebiges, aber festes $x\in W_1$ die Abbildung $\Phi:V_2\to\mathbb{R}^m$ definiert durch $\Phi(y):=G(x,y)$ eine Kontraktion auf der abgeschlossenen Teilmenge $V_2\subset\mathbb{R}^m$, so daß es nach dem Banachschen Fixpunktsatz zu jedem $x\in W_1$ genau einen Fixpunkt $y_*\in V_2$ gibt mit $y_*=\Phi(x,y_*)=G(x,y_*)$. Dann ist die so definierte Funktion $g:V_1\to V_2$ mit $g(x):=y_*$ aber Lösung des Problems $F(x,g(x))$.

(2. Stetigkeit von g) Wie zuvor sei $V_2:=\{y\in W_2\ :\ \|y-b\|\le r\}\subset W_2$ mit r aus (2). Da $G(a,b)=b$, gibt es wegen der Stetigkeit von F eine offene Umgebung $V_1\subset W_1$ von a, so daß

$$\sup_{x\in V_1}\|G(x,b)-b\|\le\frac{r}{4}\ .$$

Zusammen mit (3) folgt

$$\sup_{(x,y)\in V_1\times V_2} \|G(x,y)-b\| \leq \sup_{(x,y)\in V_1\times V_2} \Big(\|G(x,y)-G(x,b)\| + \|G(x,b)-b\|\Big) \leq r\ . \tag{4}$$

Wir betrachten jetzt stetige und beschränkte Abbildungen $\gamma \in \mathcal{C}_b(V_1,\mathbb{R}^m)$ mit $\|\gamma\| = \sup_{x\in V_1} \|\phi(x)\|$. Zu gegebenem $\phi \in \mathcal{C}_b(V_1,\mathbb{R}^m)$ werde eine Abbildung $\psi : V_1 \to \mathbb{R}^m$ definiert durch

$$\psi(x) := G(x,\gamma(x)) = \gamma(x) - B^{-1}\cdot F(x,\gamma(x))\ .$$

Wegen der Stetigkeit von F ist ψ ebenfalls stetig. Außerdem ist ψ beschränkt: Falls $\|\gamma - b\| \leq r$, dann gilt nach (4) auch $\|\psi - b\| \leq r$. Dabei wird b als die konstante Funktion aufgefaßt. Damit wird durch $\Phi(\gamma) := \psi$ eine Abbildung $\Phi : A \to A$ der abgeschlossenen Teilmenge

$$A := \{\gamma \in \mathcal{C}_b(V_1,\mathbb{R}^m)\ :\ \|\gamma - b\| \leq r\} \subset \mathcal{C}_b(V_1,\mathbb{R}^m)$$

eines Banach-Raumes auf sich selbst definiert. Nach (3) gilt für alle $\gamma \in A$

$$\begin{aligned}
\|\Phi(\gamma_1)-\Phi(\gamma_2)\| &= \sup_{x\in V_1} \|G(x,\gamma_1(x)) - G(x,\gamma_2(x))\| \\
&\leq \frac{3}{4}\sup_{x\in V_1}\|\gamma_1(x)-\gamma_2(x)\| = \frac{3}{4}\|\gamma_1-\gamma_2\|\ .
\end{aligned}$$

Folglich ist $\Phi : A \to A$ eine Kontraktion und hat nach dem Banachschen Fixpunktsatz genau einen Fixpunkt $\gamma_* \in \mathcal{C}_b(V_1,\mathbb{R}^m)$. Diese eindeutig bestimmte stetige und beschränkte Funktion $g = \gamma_* : V_1 \to \mathbb{R}^n$ löst die Gleichung $F(x,g(x)) = 0$ für alle $x \in V_1$.

(3. partielle Differenzierbarkeit der Lösung) Die Jacobi-Matrix $(\frac{\partial F}{\partial y})(x,y) \in M(m\times m,\mathbb{R})$ ist genau dann invertierbar, wenn ihre Determinante ungleich Null ist. Da die Determinante als Polynom der Matrixelemente eine stetige Funktion der Matrixelemente ist und $\det B = \det((\frac{\partial F}{\partial y})(a,b)) \neq 0$ ist, gibt es eine Umgebung $V_1' \subset V_1$ von a, so daß $(\frac{\partial F}{\partial y})(x,g(x))$ invertierbar ist für alle $x \in V_1'$.

Über die Lösung $g = (g_1,\dots,g_m) : V_1' \to \mathbb{R}^m$ definieren wir eine Abbildung $\tilde{F} : V_1' \to \mathbb{R}^m$ durch $\tilde{F}(x) := F(x,g(x))$. Da $\tilde{F}(x) = 0$ für alle $x = (x_1,\dots,x_n) \in V_1'$, verschwinden auch alle partiellen Ableitungen von $\tilde{F}$:

$$0 = \Big(\frac{\partial \tilde{F}}{\partial x_j}\Big)(x) = \Big(\frac{\partial F}{\partial x_j}\Big)(x,y)\Big|_{y=g(x)} + \sum_{i=1}^m \Big(\frac{\partial F}{\partial y_i}\Big)(x,y)\Big|_{y=g(x)} \cdot \Big(\frac{\partial g_i}{\partial x_j}\Big)(x)\ .$$

Sind $K_{il}(x,y)$ die Matrixelemente der inversen Matrix $J^{-1} = (K_{il}(x,y))$ der Jacobi-Matrix

$$J = (J_{li}(x,y)) \in M(m\times m,\mathbb{R})\ , \qquad J_{li}(x,y) = \Big(\frac{\partial F_l}{\partial y_i}\Big)(x,y)\ ,$$

von F im Punkt (x,y), dann erhalten wir

$$\Big(\frac{\partial g_i}{\partial x_j}\Big)(x) = -\sum_{l=1}^{m} K_{il}(x,y)\Big(\frac{\partial F_l}{\partial x_j}\Big)(x,y)\Big|_{y=g(x)} .$$

Insbesondere ist $g : V_1' \to \mathbb{R}^m$ partiell differenzierbar.

(4. totale Differenzierbarkeit von g) Es genügt, die totale Differenzierbarkeit im Punkt $x = a$ zu zeigen. Dazu sei

$$B := (B_{ij}) \in GL(m,\mathbb{R}) , \qquad B_{ij} = \Big(\frac{\partial F_i}{\partial y_j}\Big)(a,b) ,$$
$$C := (C_{kl}) \in M(m \times n,\mathbb{R}) , \qquad C_{kl} = \Big(\frac{\partial F_k}{\partial x_l}\Big)(a,b) .$$

Da F nach Voraussetzung differenzierbar ist, haben wir nach der Taylor-Formel

$$F(x,y) = \underbrace{F(a,b)}_{=0} + C\cdot(x-a) + B\cdot(y-b) + \phi(x,y) , \qquad \phi(x,y) = o(\|(x-a,y-b)\|) .$$

Ist $g(x)$ die Lösung von $F(x,g(x)) = 0$ für $x \in V_1'$, so folgt

$$g(x) = g(a) - B^{-1}\cdot C\cdot(x-a) - B^{-1}\cdot\phi(x,g(x)) . \tag{5}$$

Also bleibt zu beweisen, daß $\phi(x,g(x)) = o(\|x-a\|)$. Wegen $\phi(x,y) = o(\|(x-a,y-b)\|)$ gibt es zu jedem $\epsilon > 0$ eine offene Umgebung V_1^ϵ von a und eine offene Umgebung V_2^ϵ von b, so daß

$$\|\phi(x,y)\| \leq \epsilon(\|x-a\| + \|y-b\|) \qquad \text{für alle } (x,y) \in V_1^\epsilon \times V_2^\epsilon .$$

Wegen der Stetigkeit von g gibt es eine offene Umgebung $V_1^{\epsilon\prime} \subset V_1^\epsilon$, so daß $(x,g(x)) \subset V_1^{\epsilon\prime} \times V_2^\epsilon$. Also ist

$$\|\phi(x,g(x))\| \leq \epsilon(\|x-a\| + \|g(x)-b\|) \qquad \text{für alle } x \in V_1^{\epsilon\prime} .$$

Für $\epsilon = \frac{1}{2\|B^{-1}\|}$ folgen aus (5) die Abschätzungen

$$\begin{aligned} \|g(x)-b\| &\leq \|B^{-1}\|\|C\|\|x-a\| + \|B^{-1}\|\|\phi(x,g(x))\| \\ &\leq \|B^{-1}\|\|C\|\|x-a\| + \frac{1}{2}(\|x-a\| + \|g(x)-b\|) \\ &\qquad \text{für alle } x \in V_1^{1/(2\|B\|)\prime} \\ \Rightarrow \quad \|g(x)-b\| &\leq \big(2\|B^{-1}\|\|C\| + 1\big)\|x-a\| \qquad \text{für alle } x \in V_1^{1/(2\|B\|)\prime} . \end{aligned}$$

Somit gibt es zu jedem $\epsilon > 0$ eine Umgebung $V_1^{\epsilon\prime\prime} = V_1^\epsilon \cap V_1^{1/(2\|B\|)\prime}$, so daß

$$\|\phi(x,g(x))\| \leq \epsilon\big(2\|B^{-1}\|\|C\| + 2\big)\|x-a\| \qquad \text{für alle } x \in V_1^{\epsilon\prime\prime} .$$

Also ist g in a differenzierbar. $\square$

Eine erste nützliche Anwendung des Satzes über implizite Funktionen sind die Höhenlinien von Funktionen zweier Veränderlicher $f : U \to \mathbb{R}$ mit $U \subset \mathbb{R}^2$ offen. Eine *Niveaumenge* von f ist die Menge

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

für $c \in \mathbb{R}$. Dabei kann $N_f(c)$ leer sein, aus einem Punkt gestehen oder eine eigentliche Höhenlinie darstellen. Im letzten Fall kann der Satz über implizite Funktionen genutzt werden, um die Höhenline lokal zu konstruieren.

Es werde vorausgesetzt, daß eine Lösung $(a, b) \in U$ mit $f(a, b) = c$ bekannt ist. Wir können $(\operatorname{grad} f)(a, b) \neq 0$ voraussetzen (sonst gibt es ein lokales Extremum oder einen Sattelpunkt, was wir schon vorher untersucht hatten). Ist $(\frac{\partial f}{\partial x})(a, b) \neq 0$, dann konstruiert für $F(x, y) := f(x, y) - c$ der Satz über implizite Funktionen ein offenes Intervall I mit $a \in I$ und eine eindeutig bestimmte (partiell) differenzierbare Lösung $g : I \to \mathbb{R}$, so daß $f(x, g(x)) = c$ für alle $x \in I$ gilt. Ist $(\frac{\partial f}{\partial y})(a, b) \neq 0$, so kann/muß man die Rolle von x und y vertauschen: Für $F(y, x) := f(x, y) - c$ konstruiert der Satz über implizite Funktionen für ein offenes Intervall J mit $b \in J$ eine eindeutig bestimmte (partiell) differenzierbare Lösung $h : J \to \mathbb{R}$, so daß $f(h(y), y) = c$ für alle $y \in J$.

Dadurch wird eine offene Teilmenge V der Höhenlinie konstruiert. Man kann einen so gewonnenen Punkt der Teilmenge als neuen Startpunkt für den Satz über implizite Funktionen nehmen und so die Höhenlinie über die Teilmenge V hinaus ausdehnen. Das Konstruktionsverfahren läßt sich auf dem Computer implementieren und kann zur dreidimensionalen Visualisierung von Funktionen $f : U \to \mathbb{R}$ verwandt werden.

Eine weitere wichtige Anwendung des Satzes über implizite Funktionen ist das lokale Invertieren einer Abbildung $f : V \to U$ mit $U, V \subset \mathbb{R}^n$.

Satz 2.35 *Sei $V \subset \mathbb{R}^n$ offen und die Abbildung $f : V \to \mathbb{R}^n$ stetig differenzierbar auf V. In einem Punkt $b \in V$ sei das totale Differential von f invertierbar, d.h. $(Df)(b) \in GL(n, \mathbb{R})$. Dann gibt es eine offene Umgebung $V_0 \subset V$ von b und eine offene Umgebung $U_0 \subset \mathbb{R}^n$ von $a := f(b)$ so daß $f : V_0 \to U_0$ bijektiv ist und die Umkehrabbildung $g = f^{-1} : U_0 \to V_0$ stetig differenzierbar ist. Außerdem gilt $(Dg)(a) = ((Df)(b))^{-1}$.*

Beweis. Wir verwenden den Satz über implizite Funktionen mit $F : \mathbb{R}^n \times V \to \mathbb{R}^n$ gegeben durch $F(x, y) := x - f(y)$. Ziel ist lokale Auflösung nach $y = g(x) = f^{-1}(x)$.

Es gilt $F(a, b) = 0$ und $(\frac{\partial F}{\partial y})(a, b) = -(Df)(b) \in GL(n, \mathbb{R})$. Nach dem Satz über implizite Funktionen existiert eine offene Umgebung U' von a, eine Umgebung $V' \subset V$ von b und eine eindeutig bestimmte stetig differenzierbare Abbildung $g : U' \to V'$, so daß

$$F(x, g(x)) = x - f(g(x)) = 0 \qquad \text{für alle } x \in U' .$$

Wegen der Stetigkeit von f gibt es eine offene Umgebung $V_0 \subset V$ von b mit $f(V_0) := U_0 \subset U'$. Aus der Stetigkeit von g folgt, daß U_0 offen ist. Also ist $f : V_0 \to U_0$ bijektiv mit $f^{-1} = g : U_0 \to V_0$.

Aus $y = g(f(y))$ und der Kettenregel folgt

$$(D(g \circ f))(y) = E_n = (Dg)(f(y)) \cdot (Df)(y) \qquad \text{für alle } y \in V_0$$

und insbesondere $(Dg)(a) = ((Df)(b))^{-1}$. □

Als Anwendung der lokalen Invertierbarkeit betrachten wir die Umrechnung zwischen kartesischen Koordinaten und Polarkoordinaten. Dazu sei $f : \mathbb{R}^*_+ \times \mathbb{R} \to \mathbb{R}^2$ gegeben durch $f(r, \phi) := (r \cos \phi, r \sin \phi)$. Das totale Differential (Jacobi-Matrix) ist

$$(Df)(r, \phi) = \begin{pmatrix} \frac{\partial f_1}{\partial r} & \frac{\partial f_1}{\partial \phi} \\ \frac{\partial f_2}{\partial r} & \frac{\partial f_2}{\partial \phi} \end{pmatrix} (r, \phi) = \begin{pmatrix} \cos \phi & -r \sin \phi \\ \sin \phi & r \cos \phi \end{pmatrix}$$

Damit gilt $\det((Df)(r, \phi)) = r > 0$, die Jacobi-Matrix ist also in jedem Punkt $(r, \phi) \in \mathbb{R}^*_+ \times \mathbb{R}$ invertierbar. Folglich ist f in jedem Punkt von $\mathbb{R}^2 \setminus \{0\}$ lokal invertierbar. Das ist in diesem Fall auch direkt zu erhalten: Ist $f(r, \phi) = (x, y)$, dann ist $r = \sqrt{x^2 + y^2}$ und $\cos \phi = \frac{x}{r}$, $\sin \phi = \frac{y}{r}$. Damit erhalten wir die Jacobi-Matrix für *eine* lokale Umkehrung $g : U \to V$ mit $f(g(x, y)) = (x, y)$ zu

$$(Dg)(x, y) = \big((Df)(r, \phi)\big)^{-1} = \begin{pmatrix} \cos \phi & \sin \phi \\ -\frac{\sin \phi}{r} & \frac{\cos \phi}{r} \end{pmatrix} = \begin{pmatrix} \frac{x}{\sqrt{x^2+y^2}} & \frac{y}{\sqrt{x^2+y^2}} \\ -\frac{y}{x^2+y^2} & \frac{x}{x^2+y^2} \end{pmatrix} .$$

Eine Bijektion läßt sich z.B. finden zwischen

$$f : \mathbb{R}^*_+ \times \left]-\tfrac{\pi}{2}, \tfrac{\pi}{2}\right[\to \mathbb{R}^*_+ \times \mathbb{R} .$$

Für $(x, y) \in \mathbb{R}^*_+ \times \mathbb{R}$ setzen wir dann $g(x, y) = \Big(\sqrt{x^2 + y^2}, \arctan \frac{y}{x}\Big)$. Es existiert aber keine globale Bijektion $f : \mathbb{R}^*_+ \times \mathbb{R} \to \mathbb{R}^2$, da f periodisch im Winkel ϕ ist.

2.11 Untermannigfaltigkeiten

Definition 2.32 Eine Teilmenge $M \subset \mathbb{R}^{n+k}$ heißt *n-dimensionale Untermannigfaltigkeit*, wenn zu jedem Punkt $a \in M$ eine offene Umgebung $U \subset \mathbb{R}^{n+k}$ von a und eine stetig differenzierbare Abbildung $f : U \to \mathbb{R}^k$ existieren, so daß

i) $U \cap M = f^{-1}(0)$

ii) für alle $x \in U$ mit $f(x) = 0 \in \mathbb{R}^k$ hat das Differential $(Df)(x) \in M(k \times (n+k), \mathbb{R})$ den maximalen Rang k.

Mit diesen Bezeichnungen heißt k die *Kodimension* von M.

Beispiel 2.9 (Sphäre S^n) Dazu sei $U := \mathbb{R}^{n+1} \setminus \{0\}$ und $f : U \to \mathbb{R}^1$ gegeben durch $f(x) = x_1^2 + \cdots + x_{n+1}^2 - 1$. Dann ist $\mathrm{rang}((Df)(x)) = \mathrm{rang}(2(x_1, \ldots, x_{n+1})) = 1$ für alle $x \in U$. Somit ist

$$U \cap S^n = S^n = f^{-1}(0) = \{(x_1, \ldots, x_{n+1}) \in \mathbb{R}^{n+1} \ : \ x_1^2 + \cdots + x_{n+1}^2 = 1\}$$

eine n-dimensionale Untermannigfaltigkeit. Die Kodimension ist 1.

Definition 2.33 Sei $M \subset \mathbb{R}^{n+k}$ Untermannigfaltigkeit, $a \in M$ und $f : U \to \mathbb{R}^k$ die die Untermannigfaltigkeit M definierende differenzierbare Abbildung mit $a \in U \subset \mathbb{R}^{n+k}$. Dann heißt der Untervektorraum

$$T_a(M) := \ker((Df)(a)) = \{v \in \mathbb{R}^{n+k} \ : \ (Df)(a) \cdot v = 0\} \subset \mathbb{R}^{n+k}$$

der *Tangentialraum* von M im Punkt $a \in M$. Sei orthogonales Komplement bezüglich des kanonischen Skalarprodukts $\langle \ , \ \rangle$ im $\mathbb{R}^{n+k}$,

$$N_a(M) := T_a(M)^\perp := \{w \in \mathbb{R}^{n+k} \ : \ \langle v, w\rangle = 0 \text{ für alle } v \in T_a(M)\}$$

heißt der *Normalenvektorraum* von M im Punkt a. Elemente $v \in T_a(M)$ bzw. $w \in N_a(M)$ heißen *Tangentialvektoren* bzw. *Normalenvektoren* an M im Punkt a.

Zu bemerken ist, daß der Kern einer linearen Abbildung $F : V \to W$ ein Untervektorraum ist von V ist. Wegen $\dim(V) = \dim(\ker(F)) + \mathrm{rang}(F)$ folgt mit $F = (Df)(a)$, daß $T_a(M)$ ein n-dimensionaler Vektorraum ist. Damit ist $N_a(M)$ ein k-dimensionaler Vektorraum.

Definition 2.34 Sei $T \subset \mathbb{R}^n$ offen. Eine stetig differenzierbare Abbildung $\phi : T \to \mathbb{R}^{n+k}$ heißt *Immersion*, wenn der Rang des totalen Differentials

$$(D\phi)(t) = (a_{ij}(t)) \in M((n+k) \times n, \mathbb{R}) \ , \qquad a_{ij} = \frac{\partial \phi_i}{\partial t_j}(t)$$

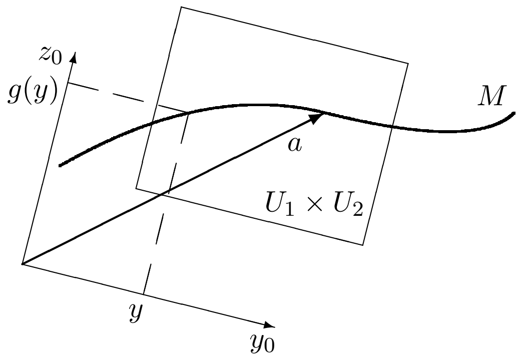
in jedem Punkt $t \in T$ gleich n ist.

Satz 2.36 (lokales Koordinatensystem) *Sei $M \subset \mathbb{R}^{n+k}$ eine n-dimensionale Untermannigfaltigkeit. Dann gibt es zu jedem Punkt $a \in M$ eine offene Umgebung $V \subset M$, eine offene Umgebung $T \subset \mathbb{R}^n$ und eine Immersion $\phi : T \to \mathbb{R}^{n+k}$, die T homöomorph auf V abbildet.*

Ein Homöomorphismus war eine bijektive stetige Abbildung ϕ mit stetigem Inversen ϕ^{-1}.

Lemma 2 (angepaßte Koordinaten) *Sei $M \subset \mathbb{R}^{n+k}$ eine n-dimensionale Untermannigfaltigkeit. Dann gibt es zu jedem $a \in M$ eine Zerlegung $\mathbb{R}^{n+k} = T_a(M) \oplus N_a(M) \simeq \mathbb{R}^n \times \mathbb{R}^k$ mit offenen Mengen $U_1 \subset T_a(M) \simeq \mathbb{R}^n$ und $U_2 \subset N_a(M) \simeq \mathbb{R}^k$ sowie eine differenzierbare Abbildung $g : U_1 \to U_2$, so daß $M \cap (U_1 \times U_2) = \{(y, g(y)) \ : \ y \in U_1\}$.*

Beweis: Es werde ein Koordinatensystem im $\mathbb{R}^{n+k}$ gewählt, daß die ersten n Koordinatenrichtungen den Tangentialraum $T_a(M)$ aufspannen und die letzten k Koordinatenrichtungen den Normalenvektorraum $N_a(M)$. In diesen Koordinaten habe $x \in M$ die Darstellung $x = (y, z)$ mit $y \in \mathbb{R}^n$ und $z \in \mathbb{R}^k$. Speziell ist $a = (y_0, z_0)$. Seien $V_1 \subset \mathbb{R}^n$ eine offene Umgebung von y_0 und $V_2 \subset \mathbb{R}^k$ eine offene Umgebung von z_0, so daß $(Df)((y, z)) \in M((n + k) \times k)$ für alle $(y, z) \in M \cap (V_1 \times V_2)$ maximalen Rang k hat.



Es gilt $(Df)(x, y) = (a_{ij}(y, z))$ mit $a_{ij}(y, z) = \frac{\partial f_i}{\partial y_j}(y, z)$ für $1 \leq j \leq n$ und $a_{i,n+j}(y, z) = \frac{\partial f_i}{\partial z_j}(y, z)$ für $1 \leq j \leq k$. Dann gilt für die Jacobi-Matrix $\frac{\partial f}{\partial z}(y_0, z_0) = (b_{ij}) \in GL(k, \mathbb{R})$, mit $b_{ij} = \frac{\partial f_i}{\partial z_j}(y_0, z_0)$, denn $\frac{\partial f}{\partial z}(y_0, z_0) \cdot w = (Df(y_0, z_0)) \cdot (0, w)$ und $(0, w)$ liegt im Komplement von $\ker((Df)(y_0, z_0))$. Nach dem Satz über implizite Funktionen existiert eine offene Umgebung $U_1 \subset V_1$ von y_0 und eine offene Umgebung $U_2 \subset V_2$ von z_0 sowie eine eindeutig bestimmte stetig differenzierbare Abbildung $g : U_1 \to U_2$ mit $f(y, g(y)) = 0$ für alle $y \in U_1$. $\square$

Beweis von Satz 2.36. Nach Lemma 2 existieren offene Teilmengen $U_1 \subset \mathbb{R}^n$ und $U_2 \subset \mathbb{R}^k$ sowie eine stetig differenzierbare Abbildung $g : U_1 \to U_2$, so daß $M \cap (U_1 \times U_2) = \{(y, g(y)) \ , \ y \in U_1\}$. Wir setzen $V = M \cap (U_1 \times U_2)$ und $T = U_1$ sowie $\phi(y) = (y, g(y))$. Surjektivität von $\phi : T \to V$ folgt aus der Konstruktion und Injektivität aus der Eindeutigkeit von g. Also ist ϕ bijektiv und stetig. Stetigkeit von $\phi^{-1} : V \to U$ folgt aus der Möglichkeit beliebiger Verkleinerungen von U, V (das Urbild offener Mengen muß offen sein).

Schließlich gilt $(D\phi)(y) = \begin{pmatrix} E_{nn} \\ (Dg)(y) \end{pmatrix}$. Da g differenzierbar auf T ist, ist auch ϕ differenzierbar in y, und es gilt $\operatorname{rang}((D\phi)(y)) = n$. Damit ist $\phi : T \to V$ eine Immersion. $\square$

Es sei bemerkt, daß auch die Umkehrung von Satz 2.36 gilt: Ist eine Immersion ϕ mit diesen Eigenschaften gegeben, dann kann man zeigen, daß M Untermannigfaltigkeit ist.

Satz 2.37 *Sei $M \subset \mathbb{R}^{n+k}$ eine n-dimensionale Untermannigfaltigkeit, lokal definiert durch die Abbildung $f = (f_1, \ldots, f_k) : U \to \mathbb{R}^k$, wobei $U \subset \mathbb{R}^{n+k}$ offen ist. Sei $a \in M \cap U$ ein Punkt und $\phi = (\phi_1, \ldots, \phi_{n+k}) : T \to \mathbb{R}^{n+k}$ die aus f konstruierte Immersion, die $T \subset \mathbb{R}^n$ homöomorph auf eine Umgebung $V \subset M$ von $\phi(t_a) = a \in M$ abbildet. Dann gilt:*

i) *Zu jedem Tangentialvektor $v \in T_a(M)$ im Sinne von Definition 2.34 gibt es eine Kurve $\gamma :]-\epsilon, \epsilon[\to M$ durch $a = \gamma(0)$, so daß v Tangentialvektor an γ im Punkt a ist.*

ii) *Ist umgekehrt v ein Tangentialvektor an eine Kurve $\gamma : I \to M$ durch $a \in M$, dann gilt $v \in \ker(Df)(a)$.*

iii) *Die Familie $\big((\partial_j\phi)(t_a)\big)_{j=1,\ldots,n}$ der Vektoren $(\partial_j\phi)(t_a) \in \mathbb{R}^{n+k}$ ist eine Basis von $T_a(M)$.*

iv) *Die Familie $\big((\operatorname{grad} f_i)(a)\big)_{i=1,\ldots,k}$ der Vektoren $(\operatorname{grad} f_i)(a) \in \mathbb{R}^{n+k}$ ist eine Basis von $N_a(M)$.*

Beweis. Sei $\gamma : I \to M$ eine Kurve auf M durch $\gamma(0) = a$, wobei das Intervall $I =]-\epsilon, \epsilon[$ so gewählt sei, daß die Kurve $s = \phi^{-1} \circ \gamma : I \to T$ durch $t_a = s(0) = \phi^{-1}(a)$ vollständig in T liegt. Wegen $f(\phi(t)) = 0$ für alle $t \in T$ gilt

$$0 = (D(f \circ \phi))(t_a) = (Df)(a) \cdot (D\phi)(t_a) \in M(k \times n, \mathbb{R})\ . \tag{*}$$

Schreiben wir $(D\phi)(t_a) = (v_1, \ldots, v_n) \in M((n+k) \times n, \mathbb{R})$, $v_j = (\partial_j\phi)(t_a) \in \mathbb{R}^{n+k}$, so folgt $v_j \in T_a(M)$ für alle $1 \leq j \leq n$.

iii) Wegen $\operatorname{rang}((D\phi)(t)) = \operatorname{rang}((\partial_j\phi_i)(t)) = n$ ist die Familie $(v_j)_{j=1,\ldots,n}$ linear unabhängig und spannt somit einen n-dimensionalen Untervektorraum von $\mathbb{R}^{n+k}$ auf, der wegen der Gleichheit der Dimensionen identisch mit $T_a(M)$ ist. Also ist $\big((\partial_j\phi)(t_a)\big)_{j=1,\ldots,n}$ eine Basis von $T_a(M)$.

i) Damit läßt sich jeder Vektor $v \in T_a(M)$ schreiben als $v = \sum_{i=1}^n c_i v_i$. Zum entsprechenden Vektor $c = (c_1, \ldots, c_n) \in \mathbb{R}^n$ wählen wir die Kurve $\gamma(\tau) = \phi(t_a + c\tau)$. Dann ist

$$\gamma'(0) = (\frac{d}{d\tau}(\phi \circ s))(0) = \sum_{i=1}^n (\partial_i\phi)(t_a) s_i'(0) = \sum_{i=1}^n (\partial_i\phi)(t_a) c_i = v\ .$$

ii) Nach der Kettenregel gilt für den Tangentialvektor v and die Kurve γ

$$v = \gamma'(0) = \Big(\frac{d}{s\tau}(\phi \circ s)\Big)(0) = \sum_{i=1}^n (\partial_i\phi)(t_a)\, s_i'(0)\ .$$

Damit ist $v \in T_a(M)$.

iv) Wegen $(Df)(a) = (a_{ij}) \in M(k \times (n+k), \mathbb{R})$, $a_{ij} = (\partial_j f_i)(a)$ sind die Zeilen von $(Df)(a)$ gegeben durch die Vektoren $(\operatorname{grad} f_i)(a)$, $i = 1, \dots, n$. Nach (*) gilt $(Df)(a) \cdot v = 0$ für alle $v \in T_a(M)$ und somit $(\operatorname{grad} f_i)(a) \in N_a(M)$. Wegen $\operatorname{rang}((Df)(a)) = k$ ist die Familie $\big((\operatorname{grad} f_i)(a)\big)_{i=1,\dots,n}$ linear unabhängig und damit wegen der Gleichheit der Dimensionen eine Basis von $N_a(M)$. $\square$

Satz 2.38 (Extrema mit Nebenbedingungen) *Sei $M \subset \mathbb{R}^{n+k}$ eine n-dimensionale Untermannigfaltigkeit, lokal definiert durch die Abbildung $f = (f_1, \dots, f_k) : U \to \mathbb{R}^k$, für $U \subset \mathbb{R}^{n+k}$ offen, mit $M \cap U = f^{-1}(0)$ und $(\operatorname{rang}(Df)(x)) = k$ für alle $x \in M \cap U$.*

Gegeben sei eine Funktion $F : U \to \mathbb{R}$, so daß die Einschränkung $F\big|_M : M \cap U \to \mathbb{R}$ im Punkt $a \in M$ ein lokales Maximum (bzw. Minimum) besitzt, d.h. es gibt eine Umgebung $V \subset M \cap U$ von a, so daß

$$F(b) \leq F(a) \qquad \textit{bzw. } F(b) \geq F(a) \qquad \textit{für alle } b \in V\ .$$

Dann gilt $(\operatorname{grad} F)(a) \in N_a(M)$, es gibt also Konstanten $\lambda_1, \dots, \lambda_k$, so daß

$$(\operatorname{grad} F)(a) = \sum_{i=1}^{n} \lambda_i (\operatorname{grad} f_i)(a)\ .$$

Diese Konstanten $\lambda_1, \dots, \lambda_k$ heißen Lagrange-Multiplikatoren.

Beweis. Da $F\big|_M$ in $a \in M$ ein lokales Extremum hat, hat die nochmalige Einschränkung auf eine Kurve $\gamma :]{-\epsilon}, \epsilon[\to M$ durch $\gamma(0) = a$ ein lokales Extremum in $t = 0$:

$$0 = \Big(\frac{d}{d\tau}(F \circ \gamma)\Big)(0) = (DF)(a) \cdot \gamma'(0)\ .$$

Wegen $\gamma'(0) \in T_a(M)$ gilt $(DF)(a) = (\operatorname{grad} F)(a) \in N_a(M)$. $\square$

Mit der Methode der Lagrange-Multiplikatoren läßt sich z.B. beweisen, daß für eine symmetrische Matrix $A = A^t \in M(n \times n, \mathbb{R})$ die Einschränkung der Funktion $F : \mathbb{R}^n \to \mathbb{R}$, $F(x) = \langle x, Ax \rangle$, auf die Einheitssphäre ihr Maximum und Minimum in einem Eigenvektor annimmt. Der Lagrange-Multiplikator ist dann der Eigenwert (Übungsaufgabe).

Untermannigfaltigkeiten spielen eine wichtige Rolle in der Mechanik. Gegeben sei ein mechanisches System aus N Teilchen (gleicher Masse m). Ein Zustand des Systems wird beschrieben durch einen Punkt $x = (x_1, \dots, x_{3N}) \in \mathbb{R}^{3N}$ (Angabe aller Koordinaten zu gegebenem Zeitpunkt). Dem System werden k holonome Zwangsbedingungen auferlegt, beschrieben durch k Gleichungen $f_1(x) = 0$, $\dots$, $f_k(x) = 0$. Wir setzen voraus, daß für die aus den partiellen Ableitungen $a_{ij}(x) := (\frac{\partial f_i}{\partial x_j})(x)$ gebildete Matrix der partiellen Ableitungen maximalen Rang hat, $\operatorname{rang}(a_{ij}(x)) = k$ für alle $x \in M$. Dann definieren die Gleichungen eine $n = (3N-k)$-dimensionale Untermannigfaltigkeit $M \subset \mathbb{R}^{n+k}$. Zustände des Systems mit Zwangsbedingungen sind dann durch Punkte aus M zu beschreiben. Die Dimension von M entspricht der Zahl der Freiheitsgrade.

Nach Satz 2.36 gibt es zu jedem Punkt $a \in M$ eine offene Umgebung $V \subset M$ sowie eine Umgebung $T \subset \mathbb{R}^n$ und eine Immersion $\phi : T \to M$, die T homöomorph auf V abbildet. Die Koordinaten $(q_1, \dots, q_n)$ eines Punktes $q \in T$ heißen verallgemeinerte Koordinaten.

Sei $(K_j(a))_{j=1,\dots,3N}$ eine Familie von Kräften, die auf die Teilchen wirken. Dann wird die Beschleunigung der Teilchen beschrieben durch das d'Alembertsche Prinzip

$$\langle m\ddot{x}_j(a) - K_j(a), v\rangle = 0 \;,$$

wobei $v \in T_aM$ ein beliebiger Tangentialvektor ist. Die Forderung besagt, daß die durch $Z_j(a) := m\ddot{x}_j(a) - K_j$ definierten Zwangskräfte keine virtuelle Arbeit verrichten bzw. Normalenvektoren sind. Somit gibt es zu jeder Komponente $Z_j(a)$ der Zwangskraft Lagrange-Multiplikatoren λ_{jl}, $l = 1, \dots, k$ mit $Z_j(a) = \sum_{l=1}^k \lambda_{jl}(a)\,(\operatorname{grad} f_l)(a)$. Es liegt dann nahe, die Zwangskraft mit dem Gradienten einer Funktion $W : U \to \mathbb{R}$ in Verbindung zu setzen, deren Einschränkung auf M im Punkt a ein lokales Extremum hat. Diese Funktion ist die Wirkung und die Extremalitätsforderung heißt *Hamiltonsches Prinzip*.

Beispiel 2.10 Eine eindimensionale Untermannigfaltigkeit $M \subset \mathbb{R}^3$ werde durch $f(x_1, x_2, x_3) = \begin{pmatrix} x_2 \\ x_1^2 + x_3^2 - l^2 \end{pmatrix}$ definiert. Das ist relevant für die Dynamik eines Pendels, bestehend aus einem Massenpunkt aufgehängt an einem masselosen Seil der Länge l, und zusätzlicher Beschränkung der Bewegung auf die Ebene $x_2 = 0$. Es gilt

$$(Df)(x) = \begin{pmatrix} 0 & 1 & 0 \\ 2x_1 & 0 & 2x_3 \end{pmatrix} \;, \qquad \operatorname{rang}((Df)(x)) = 2 \qquad \text{für } x_1^2 + x_3^2 \neq 0 \;.$$

Damit bilden $n_1(x) := (\operatorname{grad} f_1)(x) = (0,1,0)$ und $n_2(x) := (\operatorname{grad} f_2)(x) = (2x_1, 0, 2x_3)$ eine Basis von $N_{(x_1,x_2,x_3)}(M)$ und $v(x) = (-x_3, 0, x_1)$ eine Basis von $T_{(x_1,x_2,x_3)}(M)$. Auf den Massepunkt wirke die Kraft $(K_1, K_2, K_3) = (0, 0, -mg)$. Das d'Alembertsche Prinzip liefert

$$m\ddot{x}_1 = -2\lambda_2(x)\,x_1 \;, \qquad m\ddot{x}_2 = \lambda_1(x) \;, \qquad m\ddot{x}_3 = -mg - 2\lambda_2(x)\,x_3 \;.$$

Die Zwangskraft $Z = (-2\lambda_2 x_1, 0, -2\lambda_2 x_3)$ ist dann die Seilspannung.

2.12 Die Euler-Lagrange-Gleichungen

Wir zeigen zunächst, daß man bei geeigneten Voraussetzungen Integration und partielle Differentiation vertauschen kann.

Satz 2.39 *Sei $I = [a, b] \subset \mathbb{R}$ ein kompaktes Intervall, $U \subset \mathbb{R}^n$ eine Umgebung und $f : I \times U \to \mathbb{R}$ eine stetige Funktion. Dann ist die durch $\phi(x) := \int_a^b dt\ f(t, x)$ definierte Funktion $\phi : U \to \mathbb{R}$ stetig.*

Beweis. Es sei $(x_k)_{k\in N}$ eine Folge von Punkten $x_k \in U$, die gegen $x \in U$ konvergiert. Wir betrachten die Folge der Funktionen $F_k : I \to \mathbb{R}$, $F_k(t) := f(t, x_k)$. Da die Menge $X := I \times (\{x_k \ :\ k \in \mathbb{N}\} \cup \{x\})$ kompakt ist, ist nach Satz 2.19 die Abbildung $f : X \to \mathbb{R}$ gleichmäßig stetig. Zu gegebenem $\epsilon > 0$ gibt es also ein $\delta > 0$, so daß für alle $(t, x), (t', x') \in X$ mit $\|(t, x) - (t', x')\| < \delta$ gilt $\|f(t, x) - f(t', x')\| < \epsilon$. Insbesondere gibt es ein $l \in \mathbb{N}$, so daß $\|f(t, x) - f(t, x_k)\| < \epsilon$ für alle $t \in I$ und alle $k \geq l$. Folglich ist $(F_k)_{k\in\mathbb{N}}$ gleichmäßig stetig, so daß Integration und Grenzwertbildung vertauschen:

$$\lim_{k\to\infty} \int_a^b dt\ f(t, x_k) = \int_a^b dt\big(\lim_{k\to\infty} f(t, x_k)\big) = \int_a^b dx\ f(t, x)\ .$$

Damit ist $\phi : U \to \mathbb{R}$ stetig. □

Satz 2.40 *Seien $I, J \subset \mathbb{R}$ kompakte Intervalle und $f : I \times J \to \mathbb{R}$ eine stetige Funktion, so daß $f(t, x)$ nach x stetig partiell differenzierbar ist. Dann ist die stetige Funktion $\phi : J \to \mathbb{R}$ mit $\phi(x) := \int_I dt\ f(t, x)$ stetig differenzierbar, und es gilt*

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

Beweis. Es sei $(x_k)_{k\in\mathbb{N}}$ eine Folge von Punkten $x_k \in J$, die gegen $x \in J$ konvergiert, wobei $x_k \neq x$ für alle $k \in \mathbb{N}$. Nach dem Mittelwertsatz gibt es zu jedem $t \in I$ und $k \in \mathbb{N}$ ein $y_k \in J$ zwischen x und x_k, so daß

$$\frac{f(t, x_k) - f(t, x)}{x_k - x} = (\partial_2 f)(t, y_k)\ .$$

Da $I \times J$ kompakt ist, ist $(\partial_2 f)(t, x)$ auf $I \times J$ gleichmäßig stetig. Also gibt es zu jedem $\epsilon > 0$ ein $\delta > 0$, so daß

$$\|(\partial_2 f)(t, x) - (\partial_2 f)(t', x')\| < \epsilon \text{ für alle } (t, x), (t', x') \in I \times J \text{ mit } \|(t, x) - (t', x')\| < \delta\ .$$

Insbesondere gibt es ein $l \in \mathbb{N}$, so daß

$$\|(\partial_2 f)(t, x) - (\partial_2 f)(t, y_k)\| < \epsilon \qquad \text{für alle } t \in I \text{ und alle } k \geq l\ .$$

Also ist die Folge $((\partial_2 f)(t, y_k))_{k\in\mathbb{N}}$ gleichmäßig konvergent auf I, so daß Integration und Grenzwertbildung vertauschen:

$$\begin{aligned}\phi'(x) &= \lim_{k\to\infty} \frac{\phi(x_k) - \phi(x)}{x_k - x} = \lim_{k\to\infty} \frac{1}{x_k - x}\Big(\int_I dt\ f(t, x_k) - \int_I dt\ f(t, x)\Big)\\ &= \int_I dt\Big(\lim_{k\to\infty} \frac{f(t, x_k) - f(t, x)}{x_k - x}\Big) = \int_I dt\ (\partial_2 f)(t, x)\ .\end{aligned}$$

□

Das Ergebnis überträgt sich sofort auf mehrere Variablen, da diese bei partiellen Ableitungen festgehalten werden: Ist $I \subset \mathbb{R}$ ein kompaktes Intervall, $U \subset \mathbb{R}^n$ offen und $f : I \times [a,b] \to \mathbb{R}$ eine stetige Funktion, die stetig partiell differenzierbar nach $x_1, \dots, x_k$ ist, dann gilt

$$\frac{\partial}{\partial x_i} \int_a^b dt\; f(t, x_1, \dots, x_n) = \int_a^b dt\; \Big(\frac{\partial f}{\partial x_i}\Big)(t, x_1, \dots, x_n)\;.$$

Mit $\mathcal{C}^2[a,b]$ werde der Vektorraum aller zweimal stetig differenzierbaren Funktionen über dem kompakten Intervall $[a,b] \subset \mathbb{R}$ bezeichnet. Zu $\alpha, \beta \in \mathbb{R}$ betrachten wir die Teilmenge

$$\mathcal{C}^2_{\alpha,\beta}[a,b] := \{f \in \mathcal{C}^2[a,b] \;:\; f(a) = \alpha\;,\; f(b) = \beta\} \subset \mathcal{C}^2[a,b]$$

der zweimal stetig differenzierbaren Funktionen mit vorgegebenen Werten am Rand des Intervalls. Für eine zweimal stetig differenzierbare Funktion $L : [a,b] \times \mathbb{R}^n \times \mathbb{R}^n$ betrachten wir die Abbildung

$$\begin{aligned} &S : (\mathcal{C}^2_{\alpha,\beta}[a,b])^n \to \mathbb{R}\;, \\ &S(\phi_1, \dots, \phi_n) := \int_a^b dt\; L\big(t, \phi_1(t), \dots, \phi_n(t), \phi_1'(t), \dots, \phi_n'(t)\big)\;, \end{aligned}$$

mit $(\mathcal{C}^2_{\alpha,\beta}[a,b])^n := \mathcal{C}^2_{\alpha_1,\beta_1}[a,b] \times \cdots \times \mathcal{C}^2_{\alpha_n,\beta_n}[a,b]$. In physikalischen Anwendungen heißt L die *Lagrange-Funktion* und S die Wirkung. Zur Fixierung der partiellen Ableitungen setzen wir $L(t, q_1, \dots, q_n, v_1, \dots, v_n)$ und entsprechend $(\partial_{1+i}L)(t,q,v) = \frac{\partial L}{\partial q_i}(t,q,v)$ und $(\partial_{1+n+i}L)(t,q,v) = \frac{\partial L}{\partial v_i}(t,q,v)$. Eine typische Problemstellung der Variationsrechnung ist es, das Infimum von S auf $(\mathcal{C}^2_{\alpha,\beta}[a,b])^n$ zu suchen:

Satz 2.41 *Mit diesen Bezeichnungen gilt: Eine notwendige Bedingung für die Realisierung* $S(\phi) = \inf_{\psi \in (\mathcal{C}^2_{\alpha,\beta}[a,b])^n} S(\psi)$ *des Infimums durch die Funktion* $\phi \in (\mathcal{C}^2_{\alpha,\beta}[a,b])^n$ *ist die Erfüllung der Euler-Lagrange-Gleichungen*

$$\Big(\frac{d}{dt}\frac{\partial L}{\partial v_i}\Big)(t, \phi(t), \phi'(t)) - \Big(\frac{\partial L}{\partial q_i}\Big)(t, \phi(t), \phi'(t)) = 0\;, \qquad 1 \leq i \leq n\;.$$

Beweis. Es sei $\phi \in (\mathcal{C}^2_{\alpha,\beta}[a,b])^n$ eine Funktion mit

$$S(\phi) \leq S(\psi) \qquad \text{für alle } \psi \in (\mathcal{C}^2_{\alpha,\beta}[a,b])^n\;.$$

Zu $g \in \mathcal{C}^2_{0,0}[a,b]$ sei $\delta_i = (\underbrace{0, \dots, 0}_{i-1}, g_i, \underbrace{0, \dots, 0}_{n-i}) \in (\mathcal{C}^2_{0,0}[a,b])^n$. Dann ist $\psi + \epsilon\delta_i \in (\mathcal{C}^2_{\alpha,\beta}[a,b])^n$ mit $S(\phi) \leq S(\phi + \epsilon\delta_i)$. Dann besitzt $F(\epsilon) := S(\phi + \epsilon\delta_i)$ ein Minimum bei $\epsilon = 0$. Also gilt notwendigerweise $(\frac{d}{d\epsilon}F)(0) = 0$. Die Voraussetzungen von

Satz 2.40 für Differentiation unter dem Integral sind erfüllt, so daß wir mit der Kettenregel erhalten:

$$
\begin{aligned}
(\frac{d}{d\epsilon}F)(\epsilon) &= \int_a^b dt\, \frac{\partial}{\partial\epsilon} L\big(t, \phi(t) + \epsilon\delta_i(t), \phi'(t) + \epsilon\delta_i'(t)\big) \\
&= \int_a^b dt\, \Big(\frac{\partial L}{\partial q_i}(\dots) \cdot g(t) + \frac{\partial L}{\partial v_i}(\dots) \cdot g'(t)\Big) .
\end{aligned}
$$

Dabei steht $(\dots)$ für $\big(t, \phi(t) + \epsilon\delta_i(t), \phi'(t) + \epsilon\delta_i'(t)\big)$. Im letzten Term integrieren wir partiell unter Verwendung von $g(a) = g(b) = 0$:

$$
\begin{aligned}
&\int_a^b dt\, \frac{\partial L}{\partial v_i}(\dots) \cdot g'(t) \\
&= \frac{\partial L}{\partial v_i}(a, \phi(a) + \epsilon\delta_i(a), \phi'(a) + \epsilon\delta_i'(a)) \cdot \underbrace{g(a)}_{=0} - \frac{\partial L}{\partial v_i}(b, \phi(b) + \epsilon\delta_i(b), \phi'(b) + \epsilon\delta_i'(b)) \cdot \underbrace{g(b)}_{=0} \\
&- \int_a^b dt\, \Big(\frac{d}{dt}\frac{\partial L}{\partial v_i}\Big)(\dots) \cdot g(t) .
\end{aligned}
$$

Somit folgt

$$
0 = (\frac{d}{d\epsilon}F)(0) = \int_a^b dt\, \Big(\frac{\partial L}{\partial q_i}(t, \phi(t), \phi'(t)) - \Big(\frac{d}{dt}\frac{\partial L}{\partial v_i}\Big)(t, \phi(t), \phi'(t))\Big) \cdot g(t) .
$$

Das ergibt die Euler-Lagrange-Gleichungen, falls wir zeigen können, daß aus $\int_a^b f(t)g(t)$ für alle $g \in \mathcal{C}^2_{0,0}[a,b]$ für die stetige Funktion $f : [a,b] \to \mathbb{R}$ folgt $f(t) = 0$ für alle $t \in [a,b]$. Wegen der Stetigkeit von f genügt es, $f(t) = 0$ für alle $t \in\,]a,b[$ zu zeigen.

Angenommen, es gibt ein $t_0 \in\,]a,b[$ mit $f(t_0) = \epsilon > 0$. (Der Beweis für $f(t_0) = -\epsilon < 0$ ist analog.) Wegen der Stetigkeit von f gibt es dann ein $\delta > 0$, so daß $B_\delta(t_0) \in\,]a,b[$ und $f(t) > \frac{\epsilon}{2}$ für alle $t \in B_\epsilon(t_0)$. Dann gibt es (sogar beliebig oft) differenzierbare nichtnegative Funktionen, die außerhalb $B_\delta(t_0)$ identisch verschwinden. Eine Wahl ist z.B.

$$
g(t) = \begin{cases} 0 & \text{für } t \leq t_0 - \delta \text{ oder } t \geq t_0 + \delta \\ e^{-\frac{1}{t-(t_0-\delta)} - \frac{1}{(t_0+\delta)-t}} & \text{für } t_0 - \delta \leq t \leq t_0 + \delta \end{cases}
$$

Damit ist

$$
\int_a^b dt\, f(t)g(t) = \int_{t_0-\delta}^{t_0+\delta} dt\, f(t)g(t) > \frac{\epsilon}{2}\int_{t_0-\delta}^{t_0+\delta} dt\, g(t) > 0 ,
$$

im Widerspruch zur Annahme. Daraus folgt $f(t) = 0$ für alle t und schließlich die Gültigkeit der Euler-Lagrange-Gleichungen. □

Das Supremum $S(\phi) = \sup_{\psi \in (\mathcal{C}^2_{\alpha,\beta}[a,b])^n} S(\psi)$ führt offenbar auf dieselben Euler-Lagrange-Gleichungen. Die Erfahrung zeigt, daß Bewegungsgleichungen für ein physikalisches System sich oft als Euler-Lagrange-Gleichungen zu einer Wirkung erhalten lassen. Entsprechend fordert man das *Hamiltonsche Prinzip*: Von allen möglichen Bahnkurven eines physikalischen Systems zwischen festgehaltenen Anfangs- und Endkonfigurationen ist jene Bahn realisiert, für die die Wirkung extremal ist.

In der einfachsten Situation ist die Lagrange-Funktion gegeben durch die Differenz aus kinetischer Energie und potentieller Energie. Wir diskutieren den Fall ohne potentielle Energie, aber mit allgemeinem Skalarprodukt in der kinetischen Energie.

Beispiel 2.11 (Geodätengleichung) Durch eine in jedem Punkt $x \in \mathbb{R}^n$ positiv definite symmetrische Matrix $G(x) = G^t(x) = (g_{ij}(x)) \in GL(n,\mathbb{R})$ werde punktweise ein allgemeines Skalarprodukt $\langle\ ,\ \rangle_{G(x)} : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ mit $\langle v(x), w(x)\rangle_{G(x)} := \langle v(x), G(x)\cdot w(x)\rangle$ definiert. Für Bahnkurven $\gamma : \mathbb{R} \to \mathbb{R}^n$ mit $t \mapsto x(t) = (x_1(t), \dots, x_n(t))$ betrachten die Lagrange-Dichte

$$L(t,q,v) := \frac{m}{2}\langle \dot{x}(t), \dot{x}(t)\rangle_{G(x)} = \frac{m}{2}\sum_{i,j=1}^n g_{ij}(x(t))\, \dot{x}_i(t)\dot{x}_j(t) \;.$$

Die Euler-Lagrange-Gleichungen sind dann

$$\begin{aligned}
0 &= \frac{d}{dt}\Big(m\sum_{j=1}^n g_{kj}(x(t))\dot{x}_j(t)\Big) - \frac{m}{2}\sum_{i,j=1}^n \frac{\partial g_{ij}}{\partial x_k}(x(t))\dot{x}_i(t)\dot{x}_j(t) \\
&= m\sum_{j=1}^n g_{kj}(x(t))\,\ddot{x}_j(t) + m\sum_{i,j=1}^n \frac{\partial g_{kj}}{\partial x_i}(x(t))\dot{x}_i(t)\dot{x}_j(t) - \frac{m}{2}\sum_{i,j=1}^n \frac{\partial g_{ij}}{\partial x_k}(x(t))\dot{x}_i(t)\dot{x}_j(t) \\
&= m\sum_{j=1}^n g_{kj}(x(t))\ddot{x}_j(t) + \frac{m}{2}\sum_{i,j=1}^n \Big(\frac{\partial g_{kj}}{\partial x_i}(x(t)) + \frac{\partial g_{ki}}{\partial x_j}(x(t)) - \frac{\partial g_{ij}}{\partial x_k}(x(t))\Big)\dot{x}_i(t)\dot{x}_j(t) \;.
\end{aligned}$$

Durch Multiplikation mit der inversen Matrix $G^{-1}(x) = (h_{lk}(x)) \in GL(n,\mathbb{R})$ entsteht die *Geodätengleichung*

$$\begin{aligned}
0 &= \ddot{x}_l(t) + \sum_{i,j=1}^n \Gamma^l_{ij}(x(t))\dot{x}_i(t)\dot{x}_j(t) \;, \\
\Gamma^l_{ij}(x(t)) &:= \frac{1}{2}\sum_{k=1}^n h_{lk}(x(t))\Big(\frac{\partial g_{kj}}{\partial x_i}(x(t)) + \frac{\partial g_{ki}}{\partial x_j}(x(t)) - \frac{\partial g_{ij}}{\partial x_k}(x(t))\Big) \;.
\end{aligned}$$

Im Fall des kanonischen Skalarprodukts $G = E_n$ ergibt sich das Trägheitsgesetz $\ddot{x}_l = 0$, da auf das Teilchen dann keine Kraft wirkt. In der allgemeinen Relativitätstheorie beschreibt die *Metrik* $G \neq E_n$ das Gravitationsfeld. Dann beschreiben die Geodätengleichungen die Bahnen von Testteilchen im Gravitationsfeld. Für geeignete Wahl von G, welche dem Gravitationsfeld einer kugelsymmetrischen Masseverteilung entspricht, ergeben sich aus den Bahngleichungen berühmte Effekte wie die Lichtablenkung an der Sonne, beobachtbar bei einer totalen Sonnenfinsternis, und die Periheldrehung des Merkur.

3 Integralrechnung

3.1 Treppenfunktionen

Ein n-dimensionaler Quader $Q \subset \mathbb{R}^n$ ist das direkte Produkt $Q = I_1 \times \cdots \times I_n$ zusammenhängender beschränkter nichtleerer Intervalle $I_1, \ldots, I_n \subset \mathbb{R}^n$. Dabei dürfen die Intervalle, offen, abgeschlossen, halboffen und zu einem Punkt entartet sein. Die Länge eines Intervalls I ist definiert als

$$|I| := \operatorname{diam}(I) = \sup_{t_1, t_2 \in I} |t_1 - t_2| \,.$$

Dann definieren wir das Volumen eines Quaders zu

$$v(I_1 \times \cdots \times I_n) := |I_1| \cdots |I_n| \,.$$

Diese Definition ist bereits nichttrivial, denn z.B. haben das offene Intervall $]a, b[$ und das abgeschlossene Intervall $[a, b]$ das gleiche Volumen $|a - b|$. Wichtig ist, daß die Volumina sich additiv unter Zerlegungen verhalten. Sind I, I_1, I_2 zusammenhängende Intervalle mit $I = I_1 \cup I_2$, dann ist $|I| = |I_1| + |I_2|$. Damit gilt für die Quader $v(Q_1 \cup Q_2) = v(Q_1) + v(Q_2)$. Ausgeartete Quader, für die mindestens eine Kante aus einem zu einem Punkt entarteten Intervall $I = \{t\}$, $t \in \mathbb{R}$, besteht, haben das Volumen Null.

Definition 3.1 Eine Funktion $f : \mathbb{R}^n \to \mathbb{R}$ heißt *Treppenfunktion*, falls es endlich viele paarweise disjunkte Quader $Q_1, \ldots, Q_k$ gibt, so daß

- i) Für jedes $1 \leq i \leq k$ ist die Einschränkung $f|_{Q_i}$ von f auf Q_i eine konstante Funktion.
- ii) $f(x) = 0$ für alle $x \in \mathbb{R}^n \setminus (\bigcup_{i=1}^n Q_i)$.

Durch Teilungen der Quader kann man stets erreichen, daß zwei beliebige Treppenfunktion f_1, f_2 bezüglich des gleichen Satzes paarweise disjunkter Quader $\{Q_1, \ldots, Q_k\}$ definiert sind. Daraus folgt, daß endliche Linearkombinationen von Treppenfunktionen wieder eine Treppenfunktion ist. Somit bildet die Menge aller Treppenfunktionen einen Vektorraum. Sind f_1, f_2 Treppenfunktionen, so auch $|f_1|$ sowie $\max(f_1, f_2)$ und $\min(f_1, f_2)$.

Es ist bequem, die *charakteristische Funktion* einer Teilmenge $A \subset \mathbb{R}^n$ einzuführen als

$$\delta_A : \mathbb{R}^n \to \mathbb{R} \,, \qquad \delta_A(a) := \begin{cases} 1 & \text{für } x \in A \,, \\ 0 & \text{für } x \notin A \,. \end{cases}$$

Dann kann jede Treppenfunktion f als Linearkombination $f = \sum_{i=1}^k c_i \delta_{Q_i}$ mit $c_i \in \mathbb{R}$ geschrieben werden. Wegen der Möglichkeit der Zerteilung von Quadern ist eine solche Darstellung aber nicht eindeutig.

Definition 3.2 Das *Integral einer Treppenfunktion* $f = \sum_{i=1}^{k} c_i \delta_{Q_i} : \mathbb{R}^n \to \mathbb{R}$ ist definiert als

$$\int_{\mathbb{R}^n} dx\; f(x) := \sum_{i=1}^{k} c_i\, v(Q_i)\ .$$

Satz 3.1 *Das so definierte Integral einer Treppenfunktion ist unabhängig von der Darstellung der Treppenfunktion. Außerdem gilt für Treppenfunktionen f, g*

i) $\displaystyle\int_{\mathbb{R}^n} dx\; (\alpha f + \beta g)(x) = \alpha \int_{\mathbb{R}^n} dx\; f(x) + \beta \int_{\mathbb{R}^n} dx\; g(x)\ , \qquad \alpha, \beta \in \mathbb{R}\ ,$

ii) $\displaystyle\Big|\int_{\mathbb{R}^n} dx\; f(x)\Big| \leq \int_{\mathbb{R}^n} dx\; |f(x)|\ ,$

iii) *ist $f \leq g$, so folgt* $\displaystyle\int_{\mathbb{R}^n} dx\; f(x) \leq \int_{\mathbb{R}^n} dx\; g(x)\ .$

Die Aussagen sind anschaulich klar. Entscheidend ist die Additivität der Volumina bei Zerteilung von Quadern. Allerdings ist es sehr umständlich, daraus einen mathematisch strengen Beweis zu machen, so daß wir auf einen solchen Beweis verzichten.

Wir verallgemeinern Treppenfunktionen nun zu Hüllreihen durch folgende Schritte:

- Die Quader dürfen sich überlappen. Die Integraldefinition einer Treppenfunktion $f = \sum_{i=1}^{k} c_i \delta_{Q_i}$ für nichtdisjunkte Q_i bleibt unverändert, und die Eigenschaften aus Satz 3.1 bleiben erhalten.
- Wir lassen abzählbare (unendliche) Linearkombinationen von δ_{Q_i} zu, $F = \sum_{i\in\mathbb{N}} c_i \delta_{Q_i}$ (das sind dann i.A. keine Treppenfunktionen mehr).
- “Unendlich” wird als Funktionswert zugelassen. Dabei setzt man $c+\infty = \infty$ für alle $c \in \mathbb{R}\cup\{\infty\}$ und $0\cdot\infty = 0$ sowie $c\cdot\infty = \infty$ für $c \in \mathbb{R}^* \cup \{\infty\}$.

Definition 3.3 Eine Reihe $F = \sum_{k\in\mathbb{N}} c_k \delta_{Q_k}$ heißt *Hüllreihe* einer Funktion $f : \mathbb{R}^n \to \mathbb{R}\cup\{\infty\}$, wenn gilt:

i) Die Quader $Q_k \in \mathbb{R}^n$ sind offen und $c_k \in \mathbb{R}$ mit $c_k \geq 0$.

ii) Für jedes $x \in \mathbb{R}$ gilt $|f(x)| \leq F(x) = \sum_{k\in\mathbb{N}} c_k \delta_{Q_k}(x)$.

Der *Inhalt* einer Hüllreihe $F = \sum_{k\in\mathbb{N}} c_k \delta_{Q_k}$ ist definiert als $I(F) := \sum_{k\in\mathbb{N}} c_k\, v(Q_k) \in \mathbb{R}^+ \cup \{\infty\}$.

Jede Funktion $f : \mathbb{R}^n \to \mathbb{R}\cup\{\infty\}$ besitzt eine Hüllreihe, z.B. $F = \sum_{k=1}^{\infty} \delta_{\hat{Q}_k}$, wobei $\hat{Q}_k \subset \mathbb{R}^n$ der offene Quader mit Mittelpunkt 0 und Kantenlänge k ist.

Definition 3.4 Sei $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ eine beliebige Funktion. Dann ist die L^1-*Halbnorm* von f erklärt als das Infimum der Inhalte von Hüllreihen zu f,

$$\|f\|_1 := \inf\{I(F) \ : \ F \text{ ist Hüllreihe zu } f\} \ .$$

Die L^1-Halbnorm ist keine Norm, denn $\|f\| = 0 \ \not\Rightarrow \ f = 0$. Ein Beispiel ist die charakteristische Funktion $f = \delta_Q$ eines ausgearteten Quaders Q (eine Kante $I_j = \{a\}$ hat die Länge Null). Diese (unstetige) Funktion ist nicht die Nullfunktion. Sei $Q^{(k)}$ ein offener Quader, der Q enthält, wobei die j-te Kante von $Q^{(k)}$ gegeben ist durch $I_j^{(k)} =]a - \frac{1}{k+1}, a + \frac{1}{k+1}[$. Dann ist $F_k = \delta_{Q^{(k)}}$ eine Hüllreihe für jedes $k \in \mathbb{N}$. Da das Volumen solcher Quader für $k \to \infty$ gegen Null strebt, ist $\|\delta_Q\|_1 = 0$.

Satz 3.2 *Für $f, g : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ und $C \in \mathbb{R}$ gilt:*

i) $\|cf\|_1 = |c|\|f\|_1$

ii) $\|f + g\|_1 \leq \|f\|_1 + \|g\|_1$

iii) *aus $|f(x)| \leq |g(x)|$ für alle $x \in \mathbb{R}^n$ folgt $\|f\|_1 \leq \|g\|_1$*

Satz 3.3 *Für einen abgeschlossenen Quader $A \subset \mathbb{R}^n$ gilt*
$\|\delta_A\|_1 = v(A) = \displaystyle\int_{\mathbb{R}^n} dx\ \delta_A(x)$.

Beweis. Die Aussage erscheint auf den ersten Blick offensichtlich, da eine Hüllreihe aus einem einzigen offenen Quader $Q \supset A$ verwendet werden kann mit $v(Q) \leq v(A) + \epsilon$ und $\epsilon \to 0$.

Allerdings ist zu zeigen, daß auch durch eine unendliche Zerteilung von A die entsprechende Hüllreihe $F = \sum_{k=0}^{\infty} c_k \delta_{Q_k}$ zu δ_A, mit Q_k offen, keinen kleineren Inhalt haben kann. Wegen $F(x) \geq 1$ für alle $x \in A$ gibt es zu gegebenem $\epsilon > 0$ einen Index $N(x)$ mit $\sum_{k=0}^{N(x)} c_k \delta_{Q_k}(x) > 1 - \epsilon$. Da die Q_k offen sind und der Durchschnitt endlich vieler offener Mengen offen ist, gibt es eine Umgebung $U(x) \subset A$ von x, so daß wegen der Konstantheit von δ_{Q_k} auf Q_k gilt

$$\sum_{k=0}^{N(x)} c_k \delta_{Q_k}(x) > 1 - \epsilon \qquad \text{für alle } x \in U(x) \ .$$

Da A kompakt ist, wird A durch endlich viele $U(x_1), \dots, U(x_p)$ überdeckt. Also gibt es ein $N = \max(N(x_1), \dots, N(x_p))$, so daß $\sum_{k=0}^{N} c_k \delta_{Q_k}(x) > 1 - \epsilon = (1 - \epsilon)\delta_A(x)$ für alle $x \in A$. Dann folgt mit Satz 3.1.iii)

$$I(F) = \sum_{k=0}^{\infty} c_k v(Q_k) \geq \sum_{k=0}^{N} c_k v(Q_k) \geq (1 - \epsilon) v(A) \ .$$

Also gilt $\|\delta_A\|_1 = v(A) = \displaystyle\int_{\mathbb{R}^n} dx\ \delta_A(x)$. □

Satz 3.4 *Für jede Treppenfunktion $f = \sum_{i=1}^{k} c_i \delta_{Q_i}$ gilt*

$$\|f\|_1 = \int_{\mathbb{R}^n} dx\ |f(x)|\ .$$

Beweis. Da $\|f\|_1 = \||f|\|_1$, können wir $f(x) \geq 0$ annehmen. Wir zerlegen die beliebigen Quader Q_i, deren Kanten offen, abgeschlossen, halboffen und entartet sein dürfen, in disjunkte offene Quader und disjunkte Ränder, so daß

$$f = \sum_{i=1}^{k'} c'_i \delta_{Q_i^o} + \sum_{j=1}^{l} d_j \delta_{R_j}\ .$$

Zu jedem der entarteten Quader R_j mit Volumen 0 wählen wir einen nichtentarteten offenen Quader $R_j^o \supset R_j$ mit $v(R_j^0) < \epsilon$. Alle c'_i, d_j sind positiv. Dann ist ${}_\epsilon F := \sum_{i=0}^{k'} c'_i \delta_{Q_i^o} + \sum_{j=0}^{l} d_j \delta_{R_j^0}$ eine Hüllreihe zu f mit

$$\|f\|_1 \leq \inf_\epsilon I(F_\epsilon) = \inf_\epsilon \Big(\sum_{i=1}^{k'} c'_i v(Q_i^o) + \epsilon \sum_{j=1}^{l} d_j \Big) = \int_{\mathbb{R}^n} f(x)\ .$$

Sei A ein abgeschlossener Quader derart, daß $f(x) = 0$ für alle $x \notin A$, und sei $m = \max_{x\in A} f(x)$. Dann ist $g := m\delta_A - f$ wieder eine nichtnegative Treppenfunktion, so daß $-\displaystyle\int_{\mathbb{R}^n} g(x) \leq -\|g\|_1$ und daraus

$$\begin{aligned}\int_{\mathbb{R}^n} f(x) &= \int_{\mathbb{R}^n} (m\delta_A(x) - g(x)) = m \int_{\mathbb{R}^n} \delta_A(x) - \int_{\mathbb{R}^n} g(x) \\ &\leq \|m1_A\|_1 - \|g\|_1 = \|f+g\|_1 - \|g\|_1 \leq \|f\|_1\ .\end{aligned}$$

□

3.2 Das Lebesgue-Integral

Definition 3.5 Eine Funktion $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ heißt *(Lebesgue-)integrierbar*, wenn es eine Folge von Treppenfunktionen $(f_k)_{k\in\mathbb{N}}$ gibt mit $\lim_{k\to\infty} \|f - f_k\|_1 = 0$. In diesem Fall heißt

$$\int_{\mathbb{R}^n} dx\ f(x) := \lim_{k\to\infty} \int_{\mathbb{R}^n} dx\ f_k(x)$$

das *Lebesgue-Integral* von f.

Zur Definition ist zu bemerken, daß die Folge der Integrale $\Big(\displaystyle\int_{\mathbb{R}^n} dx\ f_k(x)\Big)_{k\in\mathbb{N}}$ eine Cauchy-Folge ist. Wegen der L^1-Konvergenz gegen f gibt es zu jedem $\epsilon > 0$ ein $k \in \mathbb{N}$, so daß $\|f - f_l\|_1 < \frac{\epsilon}{2}$ für alle $l \geq k$. Dann gilt für alle $m, l \geq k$

$$\begin{aligned}\Big| \int_{\mathbb{R}^n} dx\ f_l(x) - \int_{\mathbb{R}^n} dx\ f_m(x) \Big| &= \Big| \int_{\mathbb{R}^n} dx\ (f_l - f_m)(x) \Big| \leq \int_{\mathbb{R}^n} dx\ |(f_l - f_m)(x)| \\ &= \|f_l - f_m\|_1 \leq \|f_l - f\|_1 + \|f - f_m\|_1 < \epsilon\ .\end{aligned}$$

Da im $\mathbb{R}^n$ jede Cauchy-Folge konvergiert, existiert der Grenzwert und damit das Lebesgue-Integral von f. Offenbar ist jede Treppenfunktion integrierbar.

Satz 3.5 *Ist $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ integrierbar, dann ist auch $|f|$ integrierbar, und es gilt*

$$\Big|\int_{\mathbb{R}^n} dx\; f(x)\Big| \leq \int_{\mathbb{R}^n} dx\; |f(x)| = \|f\|_1\;.$$

Beweis. Sei $(f_k)_{k\in\mathbb{N}}$ eine Folge von Treppenfunktionen mit $\lim_{k\to\infty}\|f-f_k\|_1 = 0$. Wegen $|(|f|-|f_k|)| \leq |f-f_k|$ gilt $\|(|f|-|f_k|)\|_1 \leq \|f-f_k\|_1$ und insbesondere $\lim_{k\to\infty}\|(|f|-|f_k|)\|_1 = 0$. Also ist $|f|$ integrierbar mit

$$\begin{aligned}\Big|\int_{\mathbb{R}^n} dx\; f(x)\Big| &= \Big|\lim_{k\to\infty}\int_{\mathbb{R}^n} dx\; f_k(x)\Big| = \lim_{k\to\infty}\Big|\int_{\mathbb{R}^n} dx\; f_k(x)\Big| \leq \lim_{k\to\infty}\int_{\mathbb{R}^n} dx\; |f_k(x)| \\ &= \int_{\mathbb{R}^n} dx\; |f|(x)\;.\end{aligned}$$

Wegen $\|f_k\|_1 \leq \|f_k-f\|_1 + \|f\|_1 \leq 2\|f_k-f\|_1 + \|f_k\|_1$ gilt mit Satz 3.4

$$\Big(\int_{\mathbb{R}^n} dx\; |f_k|(x)\Big) - \|f_k-f\|_1 \leq \|f\|_1 \leq \Big(\int_{\mathbb{R}^n} dx\; |f_k|(x)\Big) + \|f_k-f\|_1\;.$$

Für $k\to\infty$ erhalten wir $\|f\|_1 = \displaystyle\int_{\mathbb{R}^n} dx\; |f(x)|$. □

Satz 3.6 *Das Lebesgue-Integral erfüllt die folgenden Rechenregeln für integrierbare Funktionen $f,g : \mathbb{R}^n \to \mathbb{R}\cup\{\infty\}$ und $\alpha,\beta\in\mathbb{R}$:*

i) $\displaystyle\int_{\mathbb{R}^n} dx\; (\alpha f+\beta g)(x) = \alpha\int_{\mathbb{R}^n} dx\; f(x) + \beta\int_{\mathbb{R}^n} dx\; g(x)\;.$

ii) *Ist $f(x)\leq g(x)$ für alle $x\in\mathbb{R}^n$, so folgt* $\displaystyle\int_{\mathbb{R}^n} dx\; f(x) \leq \int_{\mathbb{R}^n} dx\; g(x)$.

iii) *Ist g beschränkt, so ist $f\cdot g$ integrierbar.*

Beweis. Nur iii) ist nicht sofort offensichtlich. Sei $|g(x)| \leq M < \infty$ und f_k, g_k Treppenfunktionen, die gegen f,g konvergieren: Für $\epsilon > 0$ gibt es ein $k\in\mathbb{N}$, so daß $\|f-f_k\|_1 < \frac{\epsilon}{2M}$. Sei dann $\mu := \max_{x\in\mathbb{R}^n}|f_k(x)|$. Dann gibt es ein $l\in\mathbb{N}$, so daß $\|g-g_l\|_1 < \frac{\epsilon}{2\mu}$. Es gilt

$$|(fg-f_kg_l)(x)| = |(f-f_k)(x)g(x)+f_k(x)(g-g_l)(x)| \leq M|(f-f_k)(x)|+\mu|(g-g_l)(x)|$$

und deshalb

$$\|fg-f_kg_l\|_1 \leq \|(M|f-f_k|+\mu|g-g_l|)\| \leq M\|f-f_k\|_1 + \mu\|g-g_l\|_1 < \epsilon\;.$$ □

Insbesondere sind für f,g integrierbar auch $\max(f,g) = \frac{1}{2}(f+g+|f-g|)$ und $\min(f,g) = \frac{1}{2}(f+g-|f-g|)$ integrierbar.

Die Integration einer Funktion $f : A \to \mathbb{R} \cup \{\infty\}$ über eine Teilmenge $A \subset \mathbb{R}^n$ wird zurückgeführt auf die Integration im $\mathbb{R}^n$ durch

$$f_A(x) := \begin{cases} f(x) & \text{für } x \in A\,, \\ 0 & \text{für } x \notin A\,. \end{cases}$$

Dann heißt f integrierbar über $A \subset \mathbb{R}^n$, wenn f_A integrierbar ist, und wir setzen $\displaystyle\int_A dx\; f(x) := \int_{\mathbb{R}^n} dx\; f_A(x)$.

3.3 Beispiele integrierbarer Funktionen

Wir zeigen zunächst, daß Regelfunktionen über einem kompakten Intervall $[a,b] \subset \mathbb{R}$ integrierbar sind.

Definition 3.6 Eine Funktion $f : [a,b] \to \mathbb{R}$ heißt *Regelfunktion*, wenn es eine Folge $(f_k)_{k\in\mathbb{N}}$ von Treppenfunktionen gibt, die gleichmäßig gegen f konvergiert.

Satz 3.7 *Eine Regelfunktion $f : [a,b] \to \mathbb{R}$ ist über $[a,b]$ Lebesgue-integrierbar, und das Regelintegral stimmt mit dem Lebesgue-Integral überein, $\displaystyle\int_{[a,b]} dx\; f(x) = \int_a^b dx\; f(x)$.*

Beweis. Wegen $|g(x)| \leq \sup_{x\in[a,b]} |g(x)| = (\sup_{t\in[a,b]} |g(t)|)1_{[a,b]}(x)$ für eine beliebige Funktion $g : [a,b] \to \mathbb{R}$ gilt

$$\|g\|_1 \leq (\sup_{t\in[a,b]} |g(t)|)\|1_{[a,b]}\|_1 = |b-a| \cdot \sup_{t\in[a,b]} |g(t)|\,.$$

Ist f eine Regelfunktion, dann gibt es eine Folge $(f_k)_{k\in\mathbb{N}}$ von Treppenfunktionen, so daß zu jedem $\epsilon > 0$ ein $k \in \mathbb{N}$ existiert mit $|f(x) - f_l(x)| < \epsilon$ für alle $l \geq k$ und alle $x \in [a,b]$. Dann ist auch $\sup_{x\in[a,b]} |f(x) - f_l(x)| < \epsilon$ und damit

$$\|f - f_l\|_1 \leq |b-a| \cdot \sup_{t\in[a,b]} |g(t)| < \epsilon\,|b-a|\,.$$

Damit ist f integrierbar, und es gilt

$$\begin{aligned} \int_{[a,b]} dx\; f(x) &= \int_{\mathbb{R}} dx\; f_{[a,b]}(x) = \lim_{k\to\infty} \int_{\mathbb{R}} dx\; f_{k,[a,b]}(x) = \lim_{k\to\infty} \int_a^b dx\; f_k(x) \\ &= \int_a^b dx\; f(x)\,. \end{aligned}$$

□

Wir geben nun ein Approximationskriterium durch Ober- und Untersummen an:

Satz 3.8 *Zu $f : \mathbb{R}^n \to \mathbb{R}^n \cup \{\infty\}$ gebe es eine monoton wachsende oder monoton fallende Folge $(f_k)_{k\in\mathbb{N}}$ von Treppenfunktionen, so daß (f_k) punktweise gegen f konvergiert und die Folge der Integrale $\displaystyle\int_{\mathbb{R}^n} dx\ f_k(x)$ beschränkt ist. Dann ist f integrierbar, und es gilt $\displaystyle\int_{\mathbb{R}^n} dx\ f(x) = \lim_{k\to\infty} \int_{\mathbb{R}^n} dx\ f_k(x)$.*

Beweis. Sei $(f_k)_{k\in\mathbb{N}}$ monoton wachsend. Für $l > k$ gilt $(f_l - f_k)(x) = \sum_{i=k}^{l-1}(f_{i+1} - f_i)(x)$. Wegen der punktweisen Konvergenz von $f_l(x)$ gegen $f(x)$ erhalten wir $(f - f_k)(x) = \sum_{i=k}^{\infty}(f_{i+1} - f_i)(x)$. Nach der Dreiecksungleichung gilt dann

$$\begin{aligned}
\|f - f_k\|_1 &= \Big\| \sum_{i=k}^{\infty}(f_{i+1} - f_i)\Big\|_1 \le \lim_{l\to\infty} \sum_{i=k}^{l-1} \big\|(f_{i+1} - f_i)\big\|_1 \\
&= \lim_{l\to\infty} \sum_{i=k}^{l-1} \int_{\mathbb{R}^n} dx\ |f_{i+1} - f_i|(x) = \lim_{l\to\infty} \sum_{i=k}^{l-1} \int_{\mathbb{R}^n} dx\ (f_{i+1} - f_i)(x) \\
&= \Big(\lim_{l\to\infty} \int_{\mathbb{R}^n} dx\ f_l(x)\Big) - \int_{\mathbb{R}^n} dx\ f_k(x)\ .
\end{aligned}$$

Da die Folge der Integrale $\Big(\displaystyle\int_{\mathbb{R}^n} dx\ f_i(x)\Big)_{i\in\mathbb{N}}$ monoton und beschränkt ist, konvergiert sie gegen einen Grenzwert F, d.h. es gilt $\|f - f_k\|_1 \le F - \displaystyle\int_{\mathbb{R}^n} dx\ f_k(x)$. Damit gilt $\displaystyle\lim_{k\to\infty} \|f - f_k\|_1 = 0$, so daß f integrierbar ist mit $\displaystyle\int_{\mathbb{R}^n} dx\ f(x) = F$. □

Wir untersuchen nun die Integrierbarkeit von stetigen Funktionen über offenen und kompakten Teilmengen. Der Beweis erfolgt in mehreren Zwischenschritten:

Lemma 1 *Sei $U \subset \mathbb{R}^n$ offen und $K \subset U$ kompakt. Sei $f : K \to \mathbb{R}$ stetig mit $f \ge 0$. Dann gibt es zu jedem $\epsilon > 0$ Treppenfunktionen $f_\epsilon, g_\epsilon \ge 0$ mit*

$$\begin{aligned}
f(x) \le g_\epsilon(x) \le f(x) + \epsilon \quad &\text{für } x \in K\ , \quad g_\epsilon(x) = 0 \quad \text{für } x \in \mathbb{R} \setminus U\ , \\
f(x) - \epsilon \le f_\epsilon(x) \le f(x) \quad &\text{für } x \in K\ , \quad f_\epsilon(x) = 0 \quad \text{für } x \in \mathbb{R} \setminus U\ .
\end{aligned}$$

Beweis. Nach Satz 2.19 ist $f : K \to \mathbb{R}$ gleichmäßig stetig, d.h. für jedes $\epsilon > 0$ gibt es ein $\delta > 0$, so daß

$$\|f(x) - f(x')\| < \epsilon \qquad \text{für alle } x, x' \in K \text{ mit } \|x - x'\| < \delta\ .$$

Für jeden Punkt $x \in U$ werde ein abgeschlossener Würfel $W_x \subset U$ mit Mittelpunkt x und Kantenlänge $0 < l \le \delta$ gewählt. Sei $W_i^o := W_i \setminus \partial W_i$ das (offene) Innere eines solchen Würfels. Dann ist $K \subset \bigcup_{x\in U} W_i^o$. Da K kompakt ist, gibt es endlich viele Würfel $W_1^0, \dots, W_p^0$, die K überdecken, und insbesondere ist

$K \subset W_1 \cup \cdots \cup W_p \subset U$. Nun ist auch $W_i \cap K$ kompakt, so daß f auf $W_i \cap K$ ein Maximum M_i und Minimum m_i annimmt. Dann sind

$$g_k := \max(M_1 \delta_{W_1}, \ldots M_p \delta_{W_p}) , \qquad f_\epsilon := \min(m_1 \delta_{W_1}, \ldots m_q \delta_{W_p})$$

Treppenfunktionen mit den geforderten Eigenschaften. □

Lemma 2 *Sei $U \subset \mathbb{R}^n$ offen und $K \subset \mathbb{R}^n$ kompakt. Seien $f : U \to \mathbb{R}$ und $g : K \to \mathbb{R}$ stetig mit $f, g \geq 0$. Dann gibt es*

- *eine monoton wachsende Folge $(f_k)_{k\in\mathbb{N}}$ von Treppenfunktionen mit $f_k(x) \geq 0$, die gegen f_U konvergiert,*
- *eine monoton fallende Folge $(g_k)_{k\in\mathbb{N}}$ von Treppenfunktionen mit $g_k(x) \geq 0$, die gegen g_K konvergiert.*

Beweis. i) Wir wählen kompakte Mengen $K_k \subset U$ mit $U = \bigcup_{k=0}^\infty K_k$ und nach Lemma 1 Treppenfunktionen f'_k mit $f(x) - \frac{1}{2^k} \leq f'_k(x) \leq f(x)$ für alle $x \in K_k$ und $f'_k(x) = 0$ für $x \in \mathbb{R}^n \setminus U$. Dann ist die Folge $(f_k)_{k\in\mathbb{N}}$ mit $f_k := \max(f'_0, \ldots, f'_k)$ eine monoton wachsende Folge von Treppenfunktionen, die gegen f_U konvergiert.

ii) Wir wählen offene Mengen $U_k \subset \mathbb{R}^n$ mit $K = \bigcap_{k=0}^\infty U_k$ und nach Lemma 1 Treppenfunktionen g'_k mit $g(x) \leq g'_k(x) \leq g(x) + \frac{1}{2^k}$ für alle $x \in K$ und $g'_k(x) = 0$ für $x \in \mathbb{R}^n \setminus U_k$. Dann ist die Folge $(g_k)_{k\in\mathbb{N}}$ mit $g_k := \min(g'_0, \ldots, g'_k)$ eine monoton fallende Folge von Treppenfunktionen, die gegen g_K konvergiert. □

Satz 3.9 *Sei $U \subset \mathbb{R}^n$ eine beschränkte offene Teilmenge und $K \subset \mathbb{R}^n$ eine kompakte Teilmenge. Dann ist jede beschränkte stetige Funktion $f : U \to \mathbb{R}$ über U integrierbar und jede stetige Funktion $g : K \to \mathbb{R}$ über K integrierbar.*

Beweis. Wegen $f = f^+ - f^-$ mit $f^+(x) := \max(f(x), 0) \geq 0$ und $f^-(x) := \max(-f(x), 0) \geq 0$ genügt es, die Integrierbarkeit von nichtnegativen Funktion f zu zeigen.

i) Nach Lemma 2 gibt es eine monoton wachsende Folge $(f_k)_{k\in\mathbb{N}}$ von nichtnegativen Treppenfunktionen, die gegen f_U konvergiert. Da U beschränkt ist, gibt es einen abgeschlossenen Quader $Q \subset \mathbb{R}^n$ mit $U \subset Q$. Da f beschränkt ist, gibt es ein $M \in \mathbb{R}$ mit $0 \leq f_k(x) \leq f_U(x) \leq M\delta_Q(x)$ für alle $x \in \mathbb{R}^n$. Also gilt

$$\int_{\mathbb{R}^n} dx\ f_k(x) \leq M v_Q < \infty ,$$

so daß f_U nach Satz 3.8 integrierbar ist.

ii) Nach Lemma 2 gibt es eine monoton fallende Folge $(g_k)_{k\in\mathbb{N}}$ von nichtnegativen Treppenfunktionen, die gegen g_K konvergiert. Die Folge der Integrale $\Big(\int_{\mathbb{R}^n} dx\ g_k(x)\Big)_{k\in\mathbb{N}}$ ist nach unten durch 0 beschränkt, so daß g_K nach Satz 3.8 integrierbar ist. □

Wir wissen, daß aus der Integrierbarkeit von $f : \mathbb{R}^n \to \mathbb{R} \cup \infty$ auch die Integrierbarkeit von $f_+ = \max(f, 0)$ und $f_- = \max(-f, 0)$ folgt. Sind umgekehrt für eine Funktion f die Anteile f_+, f_- integrierbar, so ist auch $f = f_+ - f_-$ integrierbar und auch $|f| = f_+ + f_-$. Dagegen folgt aus der Integrierbarkeit von $|f|$ noch nicht die Integrierbarkeit von f_+, f_-, denn z.B. könnte f in einer beschränkten Teilmenge überabzählbar oft zwischen 1 und -1 springen, so daß f nicht durch Treppenfunktionen approximiert werden kann. Dagegen wäre $|f|$ konstant und somit über die beschränkte Teilmenge integrierbar. Wir zeigen nun, daß für Regelfunktionen (in einer Dimension) das nicht passieren kann:

Satz 3.10 *Sei $f :]a, b[\to \mathbb{R}$ Regelfunktion, wobei $a = \infty$ und/oder $b = \infty$ zugelassen ist. Dann ist f genau dann über $]a, b[$ Lebesgue-integrierbar, wenn das uneigentliche Regelintegral von f über $]a, b[$ absolut integrierbar ist, d.h. für Folgen $[a_k, b_k] \subset]a, b[$ mit $\lim_{k\to\infty} a_k = a$ und $\lim_{k\to\infty} b_k = b$ gilt $\lim\limits_{k\to\infty} \int_{a_k}^{b_k} dx\, |f(x)| < \infty$.*

In diesem Fall gilt $\int_{]a,b[} dx\, f(x) = \int_a^b dx\, f(x)$.

Beweis. Wir wählen kompakte Intervalle $I_k = [a_k, b_k] \subset]a, b[$ mit $I_k \subset I_{k+1}$ und $\bigcup_{k=0}^{\infty} I_k =]a, b[$.

i) Sei f über $]a, b[$ Lebesgue-integrierbar, dann ist auch $|f|$ Lebesgue-integrierbar, und wegen $|f|_{[a_k,b_k]} \leq |f|_{]a,b]}$ gilt

$$\int_{a_k}^{b_k} dx\, |f(x)| = \int_{[a_k,b_k]} dx\, |f(x)| \leq \int_{]a,b[} dx\, |f(x)|\, .$$

Damit existiert $\int_a^b dx\, |f(x)| := \lim\limits_{k\to\infty} \int_{a_k}^{b_k} dx\, |f(x)|$ und ist durch $\int_{]a,b[} dx\, |f(x)|$ beschränkt.

ii) Für jedes I_k gibt es eine Treppenfunktion f'_k mit $\|f(x) - f'_k(x)\| \leq \frac{1}{2^{k+1}}$ für alle $x \in I_k$. Dann gilt für $f'_{k+} := \max(f'_k, 0)$ und $f_+ := \max(f, 0)$ auch $\|f_+(x) - f'_{k+}(x)\| \leq \frac{1}{2^{k+1}}$, denn wenn $f(x)$ und $f'_k(x)$ verschiedene Vorzeichen haben, gilt $\|f(x) - 0\| \leq \|f_+(x) - f'_{k+}(x)\|$ und $\|0 - f'_k(x)\| \leq \|f_+(x) - f'_{k+}(x)\|$. Analog für $f'_{k-} := \max(-f'_k, 0)$ und $f_- := \max(-f, 0)$.

Da es also approximierende Treppenfunktionen für f_+, f_- gibt, folgt aus der Existenz des Regelintegrals $\int_a^b dx\, |f(x)| := \lim\limits_{k\to\infty} \int_{a_k}^{b_k} dx\, |f(x)|$, daß auch die Regelintegrale $\int_a^b dx\, f_\pm(x)$ existieren. Wir können also $f \geq 0$ voraussetzen.

Dann ist $f''_k := f_k - \sup_{x\in I_k} \|f(x) - f'_k(x)\|$ eine Treppenfunktion mit $f(x) - \frac{1}{2^k} \leq f''_k(x) \leq f(x)$ für alle $x \in I_k$. Die Funktionen f''_k werden trivial auf $\mathbb{R} \setminus I_k$ fortgesetzt. Dann ist die Folge $(f_k)_{k\in\mathbb{N}}$ mit $f_k := \max(f''_0, \dots, f''_k)$ eine monoton wachsende Folge von Treppenfunktionen, die gegen $f_{]a,b[}$ konvergiert. Da die Folge der Integrale $\int_{\mathbb{R}^n} dx\ f_k(x)$ beschränkt ist durch $\int_a^b dx\ f(x)$, ist f nach Satz 3.8 Lebesgue-integrierbar mit

$$\int_{]a,b[} dx\ f(x) = \lim_{k\to\infty} \int_{\mathbb{R}^n} dx\ f_k(x) \leq \int_a^b dy\ f(x)\ .$$

Zusammen mit i) ist damit der Satz bewiesen. $\square$

Die absolute Konvergenz des Regelintegrals kann nicht auf einfache Konvergenz reduziert werden. Z.B. gilt für das uneigentliche Regelintegral $\displaystyle\int_{-\infty}^{\infty} dx\ \frac{\sin x}{x} = \pi$, aber $\frac{\sin x}{x}$ ist nicht Lebesgue-integrierbar über $\mathbb{R}$. Denn mit $\frac{\sin x}{x}$ wäre auch $\left|\frac{\sin x}{x}\right|$ Lebesgue-integrierbar, was nicht der Fall ist.

3.4 Der Satz von Fubini

Es geht nun darum, die Integration einer Funktion $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ auf getrennte Integrationen über $\mathbb{R}^p$ und $\mathbb{R}^{n-p}$ zurückzuführen mit $0 < p < n$. Letztendlich genügt es also, eindimensionale Integrale ausrechnen zu können, und diese reduzieren sich unter geeigneten Bedingungen (Satz 3.7 und Satz 3.10) auf die bekannten Regelintegrale. Zunächst diskutieren wir Treppenfunktionen.

Satz 3.11 *Es sei* $f = \sum_{i=1}^k c_i \delta_{Q_i} : \mathbb{R}^p \times \mathbb{R}^{n-p} \to \mathbb{R}$ *eine Treppenfunktion, geschrieben* $f(x,y)$ *für* $x \in \mathbb{R}^p$ *und* $y \in \mathbb{R}^{n-p}$. *Dann ist auch* $F : \mathbb{R}^{n-p} \to \mathbb{R}$ *mit* $F(y) := \displaystyle\int_{\mathbb{R}^p} dx\ f(x,y)$ *eine Treppenfunktion, und es gilt*

$$\int_{\mathbb{R}^p \times \mathbb{R}^{n-p}} d(x,y)\ f(x,y) = \int_{\mathbb{R}^{n-p}} dy \Big(\int_{\mathbb{R}^p} dx\ f(x,y) \Big)\ .$$

Beweis. Die entsprechende Zerlegung der Quader sei $Q_i = Q_i' \times Q_i''$ mit $Q_i' \subset \mathbb{R}^p$ und $Q_i'' \subset \mathbb{R}^{n-p}$. Dann gilt $v(Q_i) = v(Q_i') \cdot v(Q_i'')$. Für festes $y \in \mathbb{R}^{n-p}$ definieren wir

$$f_y(x) := \sum_{i=1}^k \big(c_i \delta_{Q_i''}(y)\big) \delta_{Q_i'}(x)$$

$$\Rightarrow \quad F(y) := \int_{\mathbb{R}^p} dx\ f_y(x) = \sum_{i=1}^k \big(c_i \delta_{Q_i''}(y)\big) v(Q_i') = \sum_{i=1}^k \big(c_i v(Q_i')\big) \delta_{Q_i''}(y)\ .$$

Also ist $F = \sum_{i=1}^k \big(c_i v(Q_i')\big) \delta_{Q_i''} : \mathbb{R}^{n-p} \to \mathbb{R}$ eine Treppenfunktion. Ihr Integral ist

$$\int_{\mathbb{R}^{n-p}} dx\ F(y) = \sum_{i=1}^k \big(c_i v(Q_i')\big) v(Q_i'') = \sum_{i=1}^k c_i v(Q_i) = \int_{\mathbb{R}^p \times \mathbb{R}^{n-p}} d(x,y)\ f(x,y)\ .$$

$\square$

Offenbar gilt für diese Situation

$$\int_{\mathbb{R}^p\times\mathbb{R}^{n-p}} d(x,y)\; f(x,y) = \int_{\mathbb{R}^{n-p}} dy\Big(\int_{\mathbb{R}^p} dx\; f(x,y)\Big) = \int_{\mathbb{R}^p} dx\Big(\int_{\mathbb{R}^{n-p}} dy\; f(x,y)\Big)\;.$$

Satz 3.12 (Fubini) *Es sei $A \subset \mathbb{R}^p \times \mathbb{R}^{n-p}$ eine kompakte Teilmenge oder eine beschränkte offene Teilmenge. Für festes $y \in \mathbb{R}^{n-p}$ sei $A_y := \{x \in \mathbb{R}^p \;:\; (x,y) \in A\} \subset \mathbb{R}^p$ und für festes $x \in \mathbb{R}^p$ sei $A_x := \{y \in \mathbb{R}^{n-p} \;:\; (x,y) \in A\} \subset \mathbb{R}^{n-p}$. Eine Funktion $f : A \to \mathbb{R}$ sei stetig und beschränkt. Dann gilt:*

i) *Ist $A_y \neq \varnothing$, dann ist die durch $f_y(x) := f(x,y)$ definierte Funktion $f_y : A_y \to \mathbb{R}$ über A_y integrierbar, und die durch*

$$F(y) := \begin{cases} \displaystyle\int_{A_y} dx\; f(x,y) & \text{für } A_y \neq \varnothing \\ 0 & \text{für } A_y = \varnothing \end{cases}$$

definierte Funktion $F : \mathbb{R}^{n-p} \to \mathbb{R}$ ist über $\mathbb{R}^{n-p}$ integrierbar.

ii) *Ist $A_x \neq \varnothing$, dann ist die durch $f_x(y) := f(x,y)$ definierte Funktion $f_x : A_x \to \mathbb{R}$ über A_x integrierbar, und die durch*

$$G(x) := \begin{cases} \displaystyle\int_{A_x} dy\; f(x,y) & \text{für } A_x \neq \varnothing \\ 0 & \text{für } A_x = \varnothing \end{cases}$$

definierte Funktion $G : \mathbb{R}^p \to \mathbb{R}$ ist über $\mathbb{R}^p$ integrierbar.

iii) *Es gilt*

$$\int_A d(x,y)\; f(x,y) = \int_{\mathbb{R}^{n-p}} dy\; F(y) = \int_{\mathbb{R}^n} dx\; G(x)\;.$$

Beweis. Wir beweisen den Satz für A offen und beschränkt und können wie üblich $f \geq 0$ annehmen. Nach Lemma 2 gibt es eine monoton wachsende Folge $(f_k)_{k\in\mathbb{N}}$ von nichtnegativen Treppenfunktionen $f_k : \mathbb{R}^p \times \mathbb{R}^{n-p} \to \mathbb{R}$, die gegen f_A konvergiert. Dann ist für festgehaltenes y die durch $f_{ky} := f_k(x,y)$ definierte Folge von nichtnegativen Treppenfunktionen $f_{ky} : \mathbb{R}^p \to \mathbb{R}$ monoton wachsend und gegen f_y konvergent. Durch Einbettung in einen genügend großen abgeschlossenen Quader folgt analog zum Beweis von Satz 3.9.i), daß die Folge der Integrale $F_k(y) := \displaystyle\int_{\mathbb{R}^p} dx\; f_{yk}(x)$ auch beschränkt ist. Nach Satz 3.8 gilt somit

$$F(y) = \lim_{k\to\infty} F_k(y) = \lim_{k\to\infty} \int_{\mathbb{R}^p} dx\; f_k(x,y)\;.$$

Die Integrale $F_k(y) : \mathbb{R}^{n-p} \to \mathbb{R}$ sind selbst wieder nichtnegative Treppenfunktionen, die monoton wachsend gegen $F(y)$ konvergieren. Wie zuvor ist die Folge der Integrale $\int_{\mathbb{R}^{n-p}} dy\ F_k(y)$ beschränkt, so daß nach Satz 3.8 gilt

$$\int_{\mathbb{R}^{n-p}} dy\ F(y) = \lim_{k\to\infty} \int_{\mathbb{R}^{n-p}} dy\ F_k(y) = \lim_{k\to\infty} \int_{\mathbb{R}^{n-p}} dy\ \Big(\int_{\mathbb{R}^p} dx\ f_k(x,y) \Big) .$$

Mit Satz 3.11 und $\lim_{k\to\infty} \int_{\mathbb{R}^n} d(x,y)\ f_k(x,y) = \int_{\mathbb{R}^n} d(x,y)\ f(x,y)$ folgt die Behauptung.

Für A kompakt ist f durch eine monoton fallende Folge von nichtnegativen Treppenfunktionen zu approximieren, ansonsten ist der Beweis identisch. □

Auf diese Weise können wir die Integration von stetigen und beschränkten Funktionen $f : A \to \mathbb{R}$ über beschränkte offene oder kompakte Teilmengen $A \subset \mathbb{R}^n$ schrittweise auf eindimensionale Integrationen zurückführen ($p = 1$). Für $p = 1$ sei $B := \{y \in \mathbb{R}^{n-1} : A_y \neq \varnothing\} \subset \mathbb{R}^{n-1}$ und (im einfachsten Fall) $A_y = \bigcup_{k=1}^N [x_{1k}(y), x_{2k}(y)] \subset \mathbb{R}$. Dann gilt

$$\int_A d(x,y)\ f(x,y) = \int_B dy \Big(\sum_{k=1}^N \int_{x_{1k}(y)}^{x_{2k}(y)} dx\ f(x,y) \Big) .$$

Wenn A_y aus offenen Intervallen besteht, ist Satz 3.10 zu beachten.

Insbesondere gilt für die Integration stetiger Funktionen über das Rechteck $[a,b] \times [c,d] \subset \mathbb{R}^2$

$$\int_{[a,b]\times[c,d]} d(x,y)\ f(x,y) = \int_c^d dy \Big(\int_a^b dx\ f(x,y) \Big) = \int_a^b dx \Big(\int_c^d dy\ f(x,y) \Big) .$$

Ist $A = \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 \leq r\}$ der abgeschlossene Vollkreis, so ist $B = [-r, r]$ und $A_y = [-\sqrt{r^2 - y^2}, \sqrt{r^2 - y^2}]$. Somit gilt für eine stetige Funktion $f : A \to \mathbb{R}$

$$\int_A d(x,y)\ f(x,y) = \int_{-r}^r dy\ \Big(\int_{-\sqrt{r^2-y^2}}^{\sqrt{r^2-y^2}} dx\ f(x,y) \Big) .$$

3.5 Meßbarkeit

Definition 3.7 Eine Menge $A \subset \mathbb{R}^n$ heißt *(Lebesgue-) meßbar*, wenn die Funktion 1 über A integrierbar ist. In diesem Fall heißt

$$v_n(A) := \int_A dx\ 1(x) = \int_A dx = \int_{\mathbb{R}^n} dx\ \delta_A(x)$$

das *n-dimensionale Volumen* bzw. das *Lebesgue-Maß* von A. Die leere Menge hat das Lebesgue-Maß Null.

Aus Satz 3.9 folgt, daß jede kompakte Teilmenge $K \subset \mathbb{R}^n$ und jede offene und beschränkte Teilmenge $U \subset \mathbb{R}^n$ integrierbar ist.

Für Quader $Q \subset \mathbb{R}^n$ ist offenbar $v_n(Q) = v(Q)$ das zu Beginn eingeführte Volumen. Für den Vollkreis $A = \{(x,y) \in \mathbb{R}^2 \;:\; x^2 + y^2 \leq r\}$ gilt

$$\begin{aligned} v_2(A) &= \int_{-r}^{r} dy \,\Big(\int_{-\sqrt{r^2-y^2}}^{\sqrt{r^2-y^2}} dx \Big) = \int_{-r}^{r} dy \,\Big(2\sqrt{r^2-y^2} \Big) \\ &\overset{y \mapsto r \sin t}{=} 2r^2 \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} dt \; \cos^2 t = 2r^2 \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} dt \; \cos t \,\frac{d}{dt}(\sin t) \\ &= 2r^2 \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} dt \; \frac{d}{dt}(\cos t \sin t) + 2r^2 \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} dt \; \sin^2 t \\ &= 2r^2 \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} dt \; (1 - \cos^2 t) = r^2 \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} dt = \pi r^2 \;. \end{aligned}$$

Satz 3.13 *Sind $A, B \subset \mathbb{R}^n$ meßbar, so gilt*

i) *$A \cup B$ und $A \cap B$ sind meßbar, und es gilt $v_n(A \cup B) = v_n(A) + v_n(B) - v_n(A \cap B)$.*

ii) *Aus $A \subset B$ folgt $v_n(A) \leq v(B)$.*

Ist eine Funktion f über $A \subset \mathbb{R}^n$ integrierbar und ist $B \subset \mathbb{R}^n$ meßbar, dann ist f auch über $A \cap B$ integrierbar, und es gilt

$$\Big| \int_{A \cap B} dx \; f(x) \Big| \leq \int_A dx \; \big| f(x) \big| \;.$$

Beweis. i) folgt aus $\delta_{A\cap B} = \delta_A \cdot \delta_B$ und $\delta_{A \cup B} = \delta_A + \delta_B - \delta_A \cdot \delta_B$ und ii) aus $\delta_A \leq \delta_B$.

Die letzte Behauptung folgt aus $f_{A\cap B} = f_A \cdot \delta_B$. Da f_A und δ_B integrierbar sind und δ_B beschränkt ist, ist $f_{A \cap B}$ integrierbar nach Satz 3.6.iii). Die Abschätzung folgt aus $|f_{A\cap B}| \leq |f_A|$. □

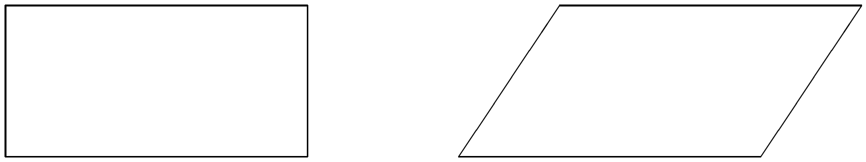
Der Satz von Fubini ermöglicht wieder die schrittweise Berechnung von Volumina. Ist $A \subset \mathbb{R}^p \times \mathbb{R}^{n-p}$ und $A_y := \{x \in \mathbb{R}^p \;:\; (x,y) \in A\} \subset \mathbb{R}^p$ die in Satz 3.12 eingeführte Schnittmenge. Dann gilt

$$v_n(A) = \int_{\mathbb{R}^{n-p}} dy \; v_p(A_y) \;.$$

Insbesondere folgt

Satz 3.14 (Prinzip von Cavalieri) *Seien $A, B \subset \mathbb{R}^p \times \mathbb{R}^{n-p}$ zwei kompakte Mengen, und es gelte $v_p(A_y) = v_p(B_y)$ für alle $y \in \mathbb{R}^{n-p}$. Dann gilt $v_n(A) = v_n(B)$.*

Zum Beispiel haben ein Rechteck und ein Parallelogramm mit gleicher Grundlänge und gleicher Höhe das gleiche zweidimensionale Volumen (Flächeninhalt):



Das Volumen der dreidimensionalen Vollkugel wir in Aufgabe 2 berechnet, und das Volumen des dreidimensionalen Volltorus in Aufgabe 3.

3.6 Nullmengen und Vollständigkeit des Lebesgue-Integrals

Wir haben bisher vor allem stetige Funktionen integriert. Wir zeigen nun, daß Singularitäten, die nur auf eine Teilmenge vom Maß Null auftreten, keine Rolle spielen.

Definition 3.8 Eine Menge $N \subset \mathbb{R}^n$ heißt *(Lebesgue-) Nullmenge*, wenn sie eine der folgenden äquivalenten Eigenschaften hat:

i) N ist meßbar mit $v_n(N) = 0$.

ii) Die charakteristische Funktion von N hat die L^1-Halbnorm Null: $\|\delta_N\|_1 = 0$.

($\Rightarrow$) $0 = v_n(N) = \int_{\mathbb{R}^n} dx\ \delta_N(x) = \int_{\mathbb{R}^n} dx\ |\delta_N(x)| = \|\delta_N\|_1$
($\Leftarrow$) Für die Treppenfunktionen $f_k = 0$ gilt $\|\delta_N - f_k\|_1 = 0$ und damit $v_n(N) = \int_{\mathbb{R}^n} dx\ \delta_N(x) = \lim_{k\to\infty} \int_{\mathbb{R}^n} dx\ f_k(x) = 0$.

Offensichtlich ist jeder entartete Quader eine Nullmenge. Wichtig ist, daß die Vereinigung abzählbar vieler Nullmengen wieder eine Nullmenge ist: Ist $N = \bigcup_{k=1}^\infty N_k$ mit $v_n(N_k) = 0$, dann ist $0 \leq \delta_N(x) \leq \sum_{k=1}^\infty \delta_{N_k}(x)$ und somit $\|\delta_N\|_1 \leq \sum_{k=1}^\infty \|\delta_{N_k}\|_1 = 0$. Zum Beispiel ist die Menge der rationalen Zahlen $\mathbb{Q} \subset \mathbb{R}$ eine Nullmenge.

Satz 3.15 *Es sei $A \subset \mathbb{R}^{n-1}$ eine offene oder abgeschlossene Teilmenge und $\Gamma = \{(x, g(x))\ :\ x \in A\ ,\ g : A \to \mathbb{R}\ \text{stetig}\} \subset \mathbb{R}^n$ der Graph einer stetigen Funktion. Dann ist Γ eine Nullmenge.*

Beweis. Wir beweisen die Aussage zunächst für A kompakt. Das Bild einer kompakten Menge unter einer stetigen Funktion ist ein kompaktes Intervall. Also ist Γ kompakt, und das Volumen ist nach dem Satz von Fubini $v_n(\Gamma) = \displaystyle\int_A dx \int_{g(x)}^{g(x)} dy\ 1(x) = 0$.

Die Situation des Satzes wird auf kompakte Teilmengen zurückgeführt. Sei $\overline{B_r(0)} \subset \mathbb{R}^{n-1}$ die kompakte Kugel mit Mittelpunkt 0 und Radius r. Eine beliebige offene Teilmenge $A \subset \mathbb{R}^{n-1}$ läßt sich schreiben als $A = \bigcup_{k=1}^{\infty} A_k$ mit den kompakten Teilmengen $A_k := \{x \in A \; : \mathrm{dist}(x, \partial A) \geq \frac{1}{k}\} \cap \overline{B_k(0)}$. Eine beliebige abgeschlossene Teilmenge $A \subset \mathbb{R}^{n-1}$ läßt sich schreiben als $A = \bigcup_{k=1}^{\infty} A_k$ mit den kompakten Teilmengen $A_k := A \cap \overline{B_k(0)}$. □

Definition 3.9 Es sei E eine Eigenschaft, die einer Teilmenge $A \subset \mathbb{R}^n$ zukommt. Wir sagen, die Eigenschaft E gilt *fast überall* auf A, wenn die Menge der Punkte, für die E nicht gilt, eine Nullmenge ist.

Satz 3.16 *Eine Funktion $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ mit $\|f\|_1 < \infty$ ist fast überall endlich, d.h. $N := \{x \in \mathbb{R}^n \; : \; f(x) = \infty\}$ ist eine Nullmenge. Insbesondere ist jede integrierbare Funktion fast überall endlich.*

Beweis. Für alle $\epsilon > 0$ gilt $\delta_N \leq \epsilon |f|$, also $\|\delta_N\|_1 \leq \epsilon \|f\|_1$ und damit $\|\delta_N\|_1 = 0$, da $\|f\|_1 < \infty$. □

Satz 3.17 (Modifikationssatz) *Seien $f, g : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ Funktionen, die fast überall gleich sind, und f sei integrierbar. Dann ist auch g integrierbar, und es gilt $\displaystyle\int_{\mathbb{R}^n} dx \;\, f(x) = \int_{\mathbb{R}^n} dx \;\, g(x)$. Insbesondere gibt es zu jeder integrierbaren Funktion f eine integrierbare Funktion g mit gleichem Lebesgue-Integral, die fast überall mit f übereinstimmt und nur Werte $\neq \infty$ annimmt.*

Beweis. Zu f gibt es eine Folge $\{f_k\}$ von Treppenfunktionen mit $\lim\limits_{k\to\infty} \|f - f_k\|_1 = 0$. Es sei $N := \{x \in \mathbb{R}^n \; : \; f(x) \neq g(x)\}$. Dann gilt $|g - f| \leq \sum_{j=1}^{\infty} \delta_N$ und somit

$$0 \leq \|g - f_k\|_1 \leq \|g - f\|_1 + \|f - f_k\|_1 \leq \Big(\sum_{j=1}^{\infty} \|\delta_N\|_1\Big) + \|f - f_k\|_1 = \|f - f_k\|_1 \; .$$

Also gilt $\lim_{k\to\infty} \|g - f_k\|_1 = 0$, d.h. g ist integrierbar, mit $\displaystyle\int_{\mathbb{R}^n} dx \;\, g(x) = \lim_{k\to\infty} \int_{\mathbb{R}^n} dx \; f_k(x) = \int_{\mathbb{R}^n} dx \; f(x)$. □

Ein Standardbeispiel ist die Funktion $\delta_{\mathbb{Q}} : \mathbb{R} \to \mathbb{R}$ mit $\int_{\mathbb{R}} dx \; \delta_{\mathbb{Q}}(x) = 0$, da die rationalen Zahlen eine Nullmenge bilden und $\delta_{\mathbb{Q}}$ fast überall gleich der Funktion 0 ist.

Der Modifikationssatz ist auch hilfreich bei Zusammensetzungen von Gebieten: Sei f über A und über B integrierbar und sei $A \cap B$ eine Nullmenge, dann ist f auch über $A \cup B$ integrierbar, da $f_{A\cup B}$ fast überall mit der Funktion $f_A + f_B$ übereinstimmt. Also gilt $\displaystyle\int_{A\cup B} dx \; f(x) = \int_A dx \; f(x) + \int_B dx \; f(x)$. Insbesondere können wir Funktionen auf einer beschränkten offenen oder kompakten Teilmenge, für die die Menge der Unstetigkeitsstellen eine Nullmenge ist, integrieren. Das betrifft z.B. stückweise stetige und beschränkte Funktionen.

Satz 3.18 *Für eine Funktion $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ gilt $\|f\|_1 = 0$ genau dann, wenn f fast überall gleich Null ist. Insbesondere verschwindet jede nichtnegative integrierbare Funktion f mit $\int_{\mathbb{R}^n} dx\, f(x) = 0$ fast überall.*

Beweis. Die Richtung ($\Leftarrow$) ist klar. Sei also $\|f\|_1 = 0$. Wir betrachten $N := \{x \in \mathbb{R}^n \;:\; f(x) \neq 0\}$. Dann gilt $N = \bigcup_{k=1}^{\infty} N_k$ mit $N_k := \{x \in \mathbb{R}^n \;:\; |f(x)| \geq \frac{1}{k}\}$. Dann ist $\delta_{N_k} \leq k \cdot |f|$ und somit $0 \leq \|\delta_{N_k}\|_1 \leq k\|f\|_1 = 0$. Also ist jedes N_k und damit N eine Nullmenge. $\square$

Mit Hilfe der Eigenschaften von Nullmengen zeigen wir jetzt, daß das Lebesgue-Integral in gewisser Weise vollständig ist.

Definition 3.10 Eine Folge $(f_k)_{k\in\mathbb{N}}$ von Funktionen $f_k : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ heißt *L^1-konvergent* gegen eine Funktion $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$, wenn $\lim_{k\to\infty} \|f - f_k\|_1 = 0$ gilt. Die Funktion f heißt dann der *L^1-Grenzwert* von $(f_k)_{k\in\mathbb{N}}$.

Die Folge $(f_k)_{k\in\mathbb{N}}$ heißt *L^1-Cauchyfolge*, wenn es zu jedem $\epsilon > 0$ ein $N \in \mathbb{N}$ gibt, so daß $\|f_k - f_l\|_1 < \epsilon$ für alle $k, l \geq N$.

Der L^1-Grenzwert kann nicht eindeutig sein, denn aus $\lim_{k\to\infty} \|f - f_k\|_1 = \lim_{k\to\infty} \|g - f_k\|_1 = 0$ folgt $\|f - g\|_1 = 0$, d.h. f und g sind nur fast überall gleich. Wie üblich ist jede L^1-konvergente Folge eine L^1-Cauchyfolge. Für integrierbare Funktionen gilt aber auch die Umkehrung:

Satz 3.19 (Riesz-Fischer) *Es sei $\mathcal{L}^1(\mathbb{R}^n)$ der Vektorraum der über $\mathbb{R}^n$ integrierbaren Funktion. Jede L^1-Cauchyfolge $(f_k)_{k\in\mathbb{N}}$ integrierbarer Funktionen $f_k \in \mathcal{L}^1(\mathbb{R}^n)$ besitzt einen Grenzwert $f \in \mathcal{L}^1(\mathbb{R}^n)$, und es gilt*

i) $\displaystyle\int_{\mathbb{R}^n} dx\, f(x) = \lim_{k\to\infty} \int_{\mathbb{R}^n} dx\, f_k(x)$

ii) *Es gibt eine Teilfolge von $(f_k)_{k\in\mathbb{N}}$, die fast überall punktweise gegen f konvergiert.*

Beweis. Es werden Indizes $k_1 < k_2 < \dots$ so gewählt, daß $\|f_k - f_{k_\nu}\|_1 < \frac{1}{2^\nu}$ für alle $k \geq k_\nu$. Dann gilt $\sum_{\nu=1}^{\infty} \|f_{k_\nu} - f_{k_{\nu+1}}\|_1 \leq 1$. Abkürzend sei $g_\nu := f_{k_\nu} - f_{k_{\nu+1}}$ und $g = \sum_{\nu=1}^{\infty} |g_\nu|$. Nach Satz 3.16 ist wegen $\|g\|_1 \leq 1$ die Menge $N = \{x \in \mathbb{R}^n \;:\; g(x) = \infty\}$ eine Nullmenge. Außerdem gibt es eine Nullmenge N_1 mit $f_{k_1}(x) \neq \infty$ für alle $x \notin N_1$. Dann setzen wir

$$f(x) := \begin{cases} \lim\limits_{\nu\to\infty} f_{k_\nu} = f_{k_1} + \sum\limits_{n=1}^{\infty} g_\nu & \text{für } x \in \mathbb{R}^n \setminus (N \cup N_1) \\ 0 & \text{für } x \in N \cup N_1 \end{cases}$$

Damit ist $f(x) \neq \infty$, und die Teilfolge (f_{k_ν}) konvergiert fast überall gegen f, d.h. ii) ist gezeigt.

Die Teilfolge ist so gewählt, daß es zu jedem $\epsilon > 0$ ein $\rho \in \mathbb{N}$ gibt, so daß $\sum_{\nu=\rho}^{\infty} \|g_\nu\|_1 < \epsilon$ und $\|f_k - f_{k_\rho}\|_1 < \epsilon$ für alle $k \geq k_\rho$. Da f_{k_ρ} integrierbar ist, gibt es eine Treppenfunktion ϕ mit $\|f_{k_\rho} - \phi\| < \epsilon$. Somit gilt für $k \geq k_\rho$

$$\|f - \phi\|_1 \leq \|f - f_{k_\rho}\|_1 + \|f_{k_\rho} - \phi\|_1 < \Big\| \sum_{\nu=\rho}^{\infty} g_\nu \Big\|_1 + \epsilon < 2\epsilon \,,$$

d.h. f ist integrierbar. Für $k \geq k_\rho$ gilt

$$\|f - f_k\|_1 \leq \|f - f_{k_\rho}\|_1 + \|f_{k_\rho} - f_k\|_1 < 2\epsilon \,,$$

d.h. (f_k) ist L^1-konvergent gegen f. Schließlich gilt

$$\Big| \int_{\mathbb{R}^n} dx\; f(x) - \int_{\mathbb{R}^n} dx\; f_k(x) \Big| \leq \int_{\mathbb{R}^n} dx\; \big|f(x) - f_k(x)\big| = \|f - f_k\|_1 < 2\epsilon \,.$$

also die in i) behauptete Konvergenz der Integrale. □

Eine integrierbare Funktion ist L^1-Grenzwert einer Folge von Treppenfunktionen. Nach dem Satz von Riesz-Fischer kann man erwarten, daß fast überall auch punktweise Konvergenz zu erreichen ist:

Satz 3.20 *Jede integrierbare Funktion $f \in \mathcal{L}^1(\mathbb{R})$ ist L^1-Grenzwert einer Folge $(f_k)_{k\in\mathbb{N}}$ von Treppenfunktionen mit*

i) $\displaystyle\sum_{k=0}^{\infty} \|f_{k+1} - f_k\|_1 < \infty$

ii) (f_k) *konvergiert fast überall punktweise gegen f.*

Beweis. Es gibt eine Folge (g_l) von Treppenfunktionen mit $\lim_{l\to\infty} \|f - g_l\|_1 = 0$. Nach dem Satz von Riez-Fischer gibt es eine Teilfolge (f_k) mit Eigenschaft i), die fast überall punktweise gegen eine integrierbare Funktion $\tilde{f}$ konvergiert. Nach Konstruktion sind beide L^1-Grenzwerte $f, \tilde{f}$ sind fast überall gleich. □

Wir wissen, daß die L^1-Halbnorm keine Norm auf dem Vektorraum $\mathcal{L}(\mathbb{R}^n)$ aller über $\mathbb{R}^n$ integrierbaren Funktionen ist: aus $\|f\| = 0$ folgt nach Satz 3.18 nur, daß f fast überall Null ist. Es bietet sich deshalb an, fast überall gleiche Funktionen zu *Äquivalenzklassen* zusammenzufassen. Sei dazu $\mathcal{N}(\mathbb{R}^n) = \{f \in \mathcal{L}^1(\mathbb{R}^n) \;:\; \|f\|_1 = 0\}$. Offenbar ist $\mathcal{N}(\mathbb{R}^n)$ ein Untervektorraum von $\mathcal{L}^1(\mathbb{R}^n)$. Zwei Funktionen $f, g \in \mathcal{L}^1(\mathbb{R}^n)$ heißen *äquivalent* ($f \sim g$), wenn $f - g \in \mathcal{N}(\mathbb{R}^n)$, d.h. wenn f und g fast überall gleich sind. Die entsprechende Äquivalenzklasse einer Funktion $f \in \mathcal{L}^1(\mathbb{R}^n)$ wird mit $[f]$ oder $f + \mathcal{N}(\mathbb{R}^n)$ bezeichnet. Dann ist die Menge aller Äquivalenzklassen

$$L^1(\mathbb{R}^n) := \mathcal{L}^1(\mathbb{R}^n)/\mathcal{N}(\mathbb{R}^n) = \{[f] \ : \ f \in \mathcal{L}^1(\mathbb{R}^n)\}$$

ein Untervektorraum von $\mathcal{L}^1(\mathbb{R}^n)$. Dabei ist die Linearkombination von Klassen definiert als Klasse der Linearkombination: $c_1[f_1] + c_2[f_2] := [c_1 f_1 + c_2 f_2]$.

Durch die Äquivalenzklassenbildung wird das Problem mit der Normeigenschaft von $\| \ \|_1$ behoben: Dazu definieren wir $\|[f]\|_1 := \|f\|_1$. Diese Definition ist sinnvoll, denn aus $f \sim g$, also $[f] = [g]$, folgt

$$\begin{aligned} & \|f\|_1 = \|g + (f-g)\|_1 \leq \|g\|_1 + \|f-g\|_1 = \|g|_1 \\ \text{und} \qquad & \|g\|_1 = \|f + (g-f)\|_1 \leq \|f\|_1 + \|f-g\|_1 = \|f\|_1 \ , \end{aligned}$$

die Norm ist also unabhängig von der Wahl des Repräsentanten. Insbesondere gilt die Dreiechsungleichung sowie $\|c[f]\|_1 = |c|\|f\|_1$. Schließlich gilt nach Satz 3.18 $\|[f]\|_1 = 0$ genau dann, wenn $[f] = [0]$. Damit ist $\big(L^1(\mathbb{R}^n), \| \ \|_1\big)$ ein normierter Raum, nach dem Satz von Riesz-Fischer sogar ein Banach-Raum.

3.7 Der Satz von der monotonen Konvergenz

Wir verallgemeinern nun den Satz 3.8, der uns ein Integrierbarkeitskriterium für monotone Folgen von Treppenfunktionen mit beschränktem Integral geliefert hat, auf monotone Folgen integrierbarer Funktionen mit beschränktem Integral.

Satz 3.21 (von der monotonen Konvergenz (Beppo Levi))
Es sei $(f_k)_{k\in\mathbb{N}}$ eine monoton wachsende Folge integrierbarer Funktionen $f_k \in \mathcal{L}^1(\mathbb{R}^n)$. Die punktweise gebildete Grenzfunktion $f = \lim_{k\to\infty} f_k$ ist genau dann integrierbar, wenn die Folge der Integrale $\displaystyle\int_{\mathbb{R}^n} dx\ f_k(x)$ beschränkt ist. In diesem Fall gilt $\displaystyle\int_{\mathbb{R}^n} dx\ f(x) = \lim_{k\to\infty}\int_{\mathbb{R}^n} dx\ f_k(x)$.

Beweis. ($\Leftarrow$) Wegen $f_k \leq f$ ist $\displaystyle\int_{\mathbb{R}^n} dx\ f_k(x) \leq \int_{\mathbb{R}^n} dx\ f(x)$, so daß die Beschränktheit der Integrale notwendig ist.

($\Rightarrow$) Die Folge $\displaystyle\Big(\int_{\mathbb{R}^n} dx\ f_k(x)\Big)_{k\in\mathbb{N}}$ sei beschränkt (und monoton wachsend). Also konvergiert sie gegen einen Grenzwert in $\mathbb{R}$. Jede konvergente Folge ist eine Cauchy-Folge, so daß es zu jedem $\epsilon > 0$ ein $N \in \mathbb{N}$ gibt, so daß für alle $k \geq l \geq N$ gilt

$$\begin{aligned} \|f_k - f_l\|_1 &= \int_{\mathbb{R}^n} dx\ \big|f_k(x) - f_l(x)\big| = \int_{\mathbb{R}^n} dx\ \big(f_k(x) - f_l(x)\big) \\ &= \int_{\mathbb{R}^n} dx\ f_k(x) - \int_{\mathbb{R}^n} dx\ f_l(x) = \Big|\int_{\mathbb{R}^n} dx\ f_k(x) - \int_{\mathbb{R}^n} dx\ f_l(x)\Big| < \epsilon\ . \end{aligned}$$

Folglich ist $(f_k)_{k\in\mathbb{N}}$ eine L^1-Cauchyfolge, die nach dem Satz von Riesz-Fischer einen L^1-Grenzwert $\tilde{f}$ hat, so daß eine Teilfolge (f_{k_ν}) fast überall punktweise gegen $\tilde{f}$ konvergiert. Damit gilt $f = \lim_{\nu\to\infty} f_{k_\nu} = \tilde{f}$ fast überall punktweise. Nach dem Modifikationssatz ist dann auch f integrierbar mit

$$\int_{\mathbb{R}^n} dx\ f(x) = \int_{\mathbb{R}^n} dx\ \tilde{f}(x) = \lim_{\nu\to\infty} \int_{\mathbb{R}^n} dx\ f_{k_\nu}(x) = \lim_{k\to\infty} \int_{\mathbb{R}^n} dx\ f_k(x)\ . \qquad \square$$

Eine nützliche Methode, solche monoton wachsenden Folgen zu generieren, besteht darin, eine nichtnegative Funktion auf eine Folge verschachtelter Gebiete einzuschränken.

Satz 3.22 (Integration durch Ausschöpfung) *Für $A \subset \mathbb{R}^n$ sei die Familie $(A_k)_{k\in\mathbb{N}}$ eine Ausschöpfung, d.h. $A_0 \subset A_1 \subset A_2 \subset \dots$ und $\bigcup_{k\in\mathbb{N}} A_k = A$. Es sei f eine auf A definierte Funktion, so daß f über jede Teilmenge A_k integrierbar ist. Dann gilt: Die Funktion f ist genau dann über A integrierbar, wenn die Folge der Integrale $\int_{A_k} dx\ |f(x)|$ beschränkt ist. In diesem Fall gilt $\int_A dx\ f(x) = \lim_{k\to\infty} \int_{A_k} dx\ f_k(x)$.*

Beweis. ($\Leftarrow$) Mit f ist auch $|f|$ integrierbar, und wegen $\int_{A_k} dx\ \ |f(x)| \leq \int_A dx\ |f(x)|$ muß die Folge der Integrale über A_k beschränkt sein.

($\Rightarrow$) Es genügt, $0 \leq f < \infty$ zu betrachten. Dann ist f_A die Grenzfunktion der monoton wachsenden Folge f_{A_k} integrierbarer Funktionen. Ist die Folge der Integrale von f_{A_k} beschränkt, dann dann ist f_A nach dem Satz von Beppo Levi integrierbar mit

$$\int_A dx\ f(x) = \int_{\mathbb{R}^n} dx\ f_A(x) = \lim_{k\to\infty} \int_{\mathbb{R}^n} dx\ f_{A_k}(x) = \lim_{k\to\infty} \int_{A_k} dx\ f(x)\ . \qquad \square$$

Mittels Integration durch Ausschöpfung können wir nun rotationssymmetrische Funktionen integrieren. Für $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ sei $\|x\| = \sqrt{x_1^2 + \cdots + x_n^2}$ die Standardnorm. Eine Funktion $f : I \to \mathbb{R}$ sei auf einem Intervall $I \subset [0, \infty[$ definiert. Sei $]a, b[$ das Innere von I. Dann ist die durch $\tilde{f}(x) := f(\|x\|)$ definierte Funktion $\tilde{f} : K_{a,b} \to \mathbb{R}$ eine rotationssymmetrische Funktion auf der Kugelschale $K_{a,b} = \{x \in \mathbb{R}^n\ :\ a < \|x\| < b\}$.

Satz 3.23 *Es sei f eine Regelfunktion auf dem Intervall $]a, b[$. Die Funktion $\tilde{f}$ ist genau dann über die Kugelschale $K_{a,b}$ integrierbar, wenn die Funktion $|f(r)|r^{n-1}$ über $]a, b[$ integrierbar ist. In diesem Fall gilt*

$$\int_{K_{a,b}} dx\ f(\|x\|) = n\kappa_n \int_a^b dr\ r^{n-1}\, f(r)\ ,$$

wobei κ_n das Volumen der n-dimensionalen Einheitskugel ist. Da für $b \le M \ne \infty$ der Rand der Kugelschale eine Nullmenge ist, gilt in diesem Fall die Aussage auch für die abgeschlossene Kugelschale $\overline{K_{a,b}}$.

Beweis. 1) Sei zunächst $J = [\alpha, \beta]$ ein kompaktes Intervall und f Regelfunktion auf J. Es gibt eine Folge von Treppenfunktionen, die gleichmäßig gegen f konvergiert. Wir beweisen den Satz also zunächst für Treppenfunktionen. Wir können $f = \delta_{[\alpha',\beta']}$ anzunehmen, da die charakteristischen Funktionen von kompakten Intervallen ein Erzeugendensystem bilden. In Übungsaufgabe 1 wird bewiesen, daß wenn κ_n das Volumen der n-dimensionalen Einheitskugel ist, die n-dimensionale Kugel vom Radius R das Volumen $R^n\kappa_n$ hat. Also gilt

$$\int_{\overline{K_{\alpha',\beta'}}} dx\ \tilde{1}(x) = v_n(\overline{K_{\alpha',\beta'}}) = \kappa_n({\beta'}^n - {\alpha'}^n) = n\kappa_n \int_{\alpha'}^{\beta'} dr\ r^{n-1}\ .$$

2) Sei $f \ge 0$ Regelfunktion auf $J = [\alpha, \beta]$. Im Beweis von Satz 3.10 haben wir gezeigt, daß es in diesem Fall eine monoton wachsende Folge $(f_k)_{k\in\mathbb{N}}$ von Treppenfunktionen gibt, die punktweise gegen f konvergiert. Nach Definition des Regelintegrals ist $\lim\limits_{k\to\infty} \int_\alpha^\beta dr\ r^{n-1}\, f_k(r) = \int_\alpha^\beta dr\ r^{n-1}\, f(r)$. Insbesondere ist die Folge der Integrale $\Big(\int_\alpha^\beta dr\ r^{n-1}\, f_k(r)\Big)_{k\in\mathbb{N}}$ beschränkt.

Nach 1) gilt für Treppenfunktionen $\int_{\overline{K_{\alpha,\beta}}} dx\ \tilde{f}_k(x) = n\kappa_n \int_\alpha^\beta dr\ r^{n-1}\, f_k(r)$. Die Folge $(\tilde{f}_k)_{k\in\mathbb{N}}$ ist ebenfalls monoton wachsend und punktweise gegen $\tilde{f}$ konvergent. Die Folge der Integrale $\Big(\int_{\overline{K_{\alpha,\beta}}} dx\ \tilde{f}_k(x)\Big)_{k\in\mathbb{N}}$ ist durch $n\kappa_n \int_\alpha^\beta dr\ r^{n-1}\, f(r)$ beschränkt. Damit gilt nach dem Satz von Beppo Levi

$$\begin{aligned}\int_{\overline{K_{\alpha,\beta}}} dx\ \tilde{f}(x) &= \lim_{k\to\infty} \int_{\overline{K_{\alpha,\beta}}} dx\ \tilde{f}_k(x) = \lim_{k\to\infty} n\kappa_n \int_\alpha^\beta dr\ r^{n-1}\, f_k(r)\\ &= n\kappa_n \int_\alpha^\beta dr\ r^{n-1}\, f(r)\ .\end{aligned}$$

3) Sei jetzt $I \subset [0, \infty[$ ein beliebiges Intervall mit Randpunkten a, b. Wir können I durch kompakte Intervalle $J_l = [\alpha_l, \beta_l]$ mit $J_0 \subset J_1 \subset J_2 \subset \ldots$ ausschöpfen, $I = \bigcup_{l\in\mathbb{N}} J_l$. Die entsprechenden Kugelschalen $A_l := \overline{K_{\alpha_l,\beta_l}}$ bilden dann eine Ausschöpfung von $K_{a,b}$. Nach 2) ist $\tilde{f}$ über A_l integrierbar mit $\int_{A_l} dx\ |\tilde{f}(x)| = n\kappa_n \int_{\alpha_l}^{\beta_l} dr\ r^{n-1}\ |f(r)| =: F_l$. Dann gilt nach dem Ausschöpfungssatz

$$\int_{K_{a,b}} dx\ \tilde{f}(x) \text{ existiert } \Leftrightarrow\ (F_l)_{l\in\mathbb{N}} \text{ konvergent } \Leftrightarrow \int_a^b dr\ r^{n-1}|f(r)| \text{ existiert.}$$

Existieren diese Integrale, so gilt

$$\int_{K_{a,b}} dx\ \tilde{f}(x) = \lim_{l\to\infty} \int_{A_l} dx\ \tilde{f}(x) = n\kappa_n \int_{\alpha_l}^{\beta_l} dr\ r^{n-1}\ f(r)\ . \qquad \square$$

Beispiel 3.1 (Trägheitsmoment eines Kreiszylinders) Das Trägheitsmoment einer kompakten Menge K bezüglich einer Achse L ist definiert als

$$\Theta := \int_K dx\ \mu(x) \big(\mathrm{d}(x, L)\big)^2\ ,$$

wobei μ die Dichte und $\mathrm{d}(x, L)$ der Abstand eines Punktes $x \in K$ zur Drehachse L ist. Für einen homogenen (konstante Diche μ) geraden Kreiszylinder

$$K = \{(x, y, z) \in \mathbb{R}^3\ :\ x^2 + y^2 \leq R^2\ ,\ 0 \leq z \leq h\}$$

mit der z-Achse als Drehachse gilt nach Fubini

$$\Theta = \mu \int_0^h dz \int_{A_z} d(x, y)\ (x^2 + y^2)\ ,$$

wobei $A_z = \{(x, y) \in \mathbb{R}^2\ :\ x^2 + y^2 \leq R^2\}$ unabhängig von z ist. Übergang zu Zylinderkoordinaten mit $r = \sqrt{x^2 + y^2}$ liefert mit $\kappa_2 = \pi$

$$\Theta = \mu \int_0^h dz\ n\kappa_n \int_0^R dr\ r^{n-1} \cdot r^2\Big|_{n=2} = 2\pi\mu h \int_0^R dr\ r \cdot r^2 = 2\pi\mu h \frac{R^4}{4} = \frac{1}{2} m R^2\ ,$$

wobei $m = \mu v_3(K)$ die Gesamtmasse ist.

3.8 Lebesgue-Integral und Differentiation

Wir benötigen ein weiteres Integrierbarkeitskriterium:

Satz 3.24 (von der majorisierten Konvergenz (Satz von Lebesgue)) *Es sei $(f_k)_{k\in\mathbb{N}}$ eine Folge integrierbarer Funktionen auf $\mathbb{R}^n$, die fast überall punktweise gegen eine Funktion f konvergiert. Es gebe eine integrierbare Funktion F mit $|f_k| \leq F$ für alle $k \in \mathbb{N}$. Dann ist f integrierbar, und es gilt* $\int_{\mathbb{R}^n} dx\ f(x) = \lim_{k\to\infty} \int_{\mathbb{R}^n} dx\ f_k(x)$.

Beweisidee. Aus der beliebigen beschränkten Folge $(f_k))_{k\in\mathbb{N}}$ wird eine Folge $(g_k)_{k\in\mathbb{N}}$ monoton fallender Funktionen konstruiert durch $g_k(x) := \sup_{i\geq k} f_i(x)$ und für jedes feste k eine Folge $(g_{k,\nu})_{\nu\in\mathbb{N}}$ monoton wachsender Funktionen durch $g_{k,\nu}(x) := \max(f_k(x), f_{k+1}(x), \dots, f_{k+\nu}(x))$. Auf diese Folgen wird der Satz von Beppo Levi angewandt. $\square$

Eine wichtige Anwendung ist:

Satz 3.25 *Es sei f eine differenzierbare Funktion auf einem kompakten Intervall $[a,x]$, und die Ableitung f' sei beschränkt. Dann ist f' integrierbar über $[a,x]$, und es gilt $f(x)-f(a)=\displaystyle\int_{[a,x]} dx\ f'(x)$.*

Beweis. Wir betrachten die Folge $(f_k)_{k\geq 1}$ der Differenzenquotienten auf $[a,x]$,

$$f_k(t)=\begin{cases} k\Big(f(t+\frac{1}{k})-f(t)\Big) & \text{für } t\in[a,x-\frac{1}{k}]\ . \\ 0 & \text{sonst}\ . \end{cases}$$

Die Folge konvergiert auf $[a,x[$ punktweise gegen f'. Die Funktionen f_k sind als stückweise stetige Funktionen Regelfunktionen und nach dem Mittelwertsatz der Differentialrechnung beschränkt durch $\max|f'|\leq C$. Die konstante Funktion C ist über $[a,x]$ integrierbar. Nach dem Satz von der majorisierten Konvergenz ist f' integierbar mit

$$\int_{[a,x]} dt\ f'(t)=\lim_{k\to\infty}\int_{[a,x]} dt\ f_k(t)=\lim_{k\to\infty}\int_a^x dt\ f_k(t)\ .$$

Die letzte Gleichheit folgt aus Satz 3.7.

Für $x-a>\frac{2}{k}$ gilt

$$\begin{aligned}
\int_a^x dt\ f_k(t) &= k\Big(\int_a^{x-\frac{1}{k}} dt\ f(t+\tfrac{1}{k})-\int_a^{x-\frac{1}{k}} dt\ f(t)\Big)\\
&= k\Big(\int_{a+\frac{1}{k}}^{x} dt'\ f(t')-\int_a^{x-\frac{1}{k}} dt\ f(t)\Big)\\
&= k\Big(\Big(\int_{x-\frac{1}{k}}^{x} dt\ f(t)+\int_{a+\frac{1}{k}}^{x-\frac{1}{k}} dt\ f(t)\Big)-\Big(\int_a^{a+\frac{1}{k}} dt\ f(t)-\int_{a+\frac{1}{k}}^{x-\frac{1}{k}} dt\ f(t)\Big)\Big)\\
&= k\int_{x-\frac{1}{k}}^{x} dt\ f(t)-k\int_a^{a+\frac{1}{k}} dt\ f(t)\ .
\end{aligned}$$

Weiter folgt aus der Stetigkeit von f und der Tatsache, daß f auf einem kompakten Intervall das Maximum und Minimum annimmt,

$$\min_{t\in[x-\frac{1}{k},x]} f(t)\leq k\int_{x-\frac{1}{k}}^{x} dt\ f(t)\leq \max_{t\in[x-\frac{1}{k},x]} f(t)\ ,$$

$$\min_{t\in[a,a+\frac{1}{k}]} f(t)\leq k\int_a^{a+\frac{1}{k}} dt\ f(t)\leq \max_{t\in[a,a+\frac{1}{k}]} f(t)\ .$$

Damit gilt $\displaystyle\lim_{k\to\infty}\int_a^x dt\ f_k(t)=f(x)-f(a)$. □

Im folgenden sei für $X\subset\mathbb{R}^n$ und $T\subset\mathbb{R}^p$ eine Funktion $f:X\times T\to\mathbb{R}$ gegeben. Für festes $x\in X$ werde durch $f_x(t)=f(x,t)$ eine Funktion $f_x:T\to\mathbb{R}$ definiert und für festes $t\in T$ eine Funktion $f_t:X\to\mathbb{R}$ durch $f_t(x)=f(x,t)$.

Satz 3.26 *Für eine Funktion $f : X \times T \to \mathbb{R}$ gelte:*

i) *Für jedes $x \in X$ ist $f_x : T \to \mathbb{R}$ über T integrierbar.*

ii) *Für jedes $t \in T$ ist $f_t : X \to \mathbb{R}$ stetig auf X.*

iii) *Es gibt eine integrierbare Funktion $\Phi : T \to \mathbb{R}$ mit $|f(x,t)| \leq \Phi(t)$ für alle $(x,t) \in X \times T$.*

Dann liefert das Integral $F(x) := \displaystyle\int_T dt\ f_x(t)$ eine stetige Funktion $F : X \to \mathbb{R}$.

Beweis. Für eine gegen x konvergierende Folge $(x_k)_{k\in\mathbb{N}}$ von Punkten $x_k \in X$ sei $f_k(t) := f(x_k, t)$. Wegen ii) konvergiert die Folge (f_k) gegen f_x. Nach iii) ist $|f_k| \leq \Phi$. Dann liefert der Satz von der majorisierten Konvergenz

$$F(x) = \int_T dt\ f_x(t) = \lim_{k\to\infty} \int_T dt\ f_k(t) = \lim_{k\to\infty} F(x_k)\ ,$$

d.h. F ist stetig. □

Satz 3.27 *Für eine Funktion $f : X \times T \to \mathbb{R}$ mit X offen gelte:*

i) *Für jedes $x \in X$ ist $f_x : T \to \mathbb{R}$ über T integrierbar.*

ii) *Für jedes $t \in T$ ist $f_t : X \to \mathbb{R}$ stetig differenzierar auf X.*

iii) *Es gibt eine integrierbare Funktion $\Phi : T \to \mathbb{R}$ mit $|(\partial_i f)(x,t)| \leq \Phi(t)$ für alle $(x,t) \in X \times T$ und alle $1 \leq i \leq n$.*

Dann liefert das Integral $F(x) := \displaystyle\int_T dt\ f_x(t)$ eine stetig differenzierbare Funktion $F : X \to \mathbb{R}$. Außerdem ist für festes $x \in X$ die durch $(\partial_i f)_x(t) := (\partial_i f)(x,t)$ definierte Funktion $(\partial_i f)_x : T \to \mathbb{R}$ integrierbar, und es gilt

$$(\partial_i F)(x) = \int_T dt\ (\partial_i f)(x,t)\ .$$

Beweis. Sei $x \in X$ fixiert und $(h_k)_{k\in\mathbb{N}}$ eine gegen Null konvergierende Folge reeller Zahlen mit $h_k \neq 0$, so daß gilt $x_k := x + h_k e_i \in X$ für alle $k \in \mathbb{N}$, wobei $e_i \in \mathbb{R}^n$ der i-te Einheitsvektor ist. Wegen ii) konvergiert die durch $\phi_k(t) := \frac{1}{h_k}\big(f(x_k,t) - f(x,t)\big)$ definierte Folge (ϕ_k) von Funktionen gegen $(\partial_i f)_x$. Da die Differenzenquotienten ϕ_k wegen i) integrierbar sind und nach iii) und dem Mittelwertsatz der Differentialrechnung durch eine integrierbare Funktion beschränkt sind, liefert der Satz von der majorisierten Konvergenz, daß die Grenzfunktion $(\partial_i f)_x$ über T integrierbar ist mit

$$\lim_{k\to\infty} \int_T dt\ \phi_k(t) = \int_T dt\ (\partial_i f)(x,t)\ .$$

Andererseits ist $\int_T dt\ \phi_k(t) = \frac{1}{h_k}\big(F(x_k) - F(x)\big)$ und damit im Limes $k \to \infty$ die partielle Ableitung von F in i-ter Koordinatenrichtung. Aus der Stetigkeit von $\partial_i f$ folgt nach Satz 3.26 die Stetigkeit von $\partial_i F$, also die stetige Differenzierbarkeit von F. □

3.9 Integration über einen Produktraum

Es geht nun um die Verallgemeinerung des Satzes von Fubini von stetigen Funktionen auf beliebige integrierbare Funktionen. Zur Vereinfachung der Schreibweise sei $X = \mathbb{R}^p$ und $Y = \mathbb{R}^{n-p}$. Wie üblich entsteht aus $f(x,y)$ durch Festhalten von $y \in Y$ die Funktion f_y auf X und durch Festhalten von $x \in X$ die Funktion f_x auf Y.

Satz 3.28 (Fubini (allgemeinster Fall)) *Es sei $f : X \times Y \to \mathbb{R} \cup \{\infty\}$ eine integrierbare Funktion. Dann gilt:*

i) *Abgesehen von einer möglichen Nullmenge $N \subset Y$ ist für festes $y \in Y \setminus N$ die Funktion f_y über X integrierbar.*

ii) *Die durch $F(y) := \begin{cases} \int_X dx\ f(x,y) & \text{für } y \in Y \setminus N \\ 0 & \text{für } y \in N \end{cases}$ definierte Funktion $F : Y \to \mathbb{R}$ ist über Y integrierbar, und es gilt*

$$\int_{x\times Y} d(x,y)\ f(x,y) = \int_Y dy\ F(y) \equiv \int_Y dy \int_X dx\ f(x,y)\ .$$

Beweis. Zu f gibt es nach Satz 3.20 eine Folge $(f_k)_{k\in\mathbb{N}}$ von Treppenfunktionen und eine Nullmenge $A \subset X \times Y$, so daß für $(x,y) \notin A$ gilt $\lim_{k\to\infty} f_k(x,y) = f(x,y)$ und außerdem $\sum_{k=0}^{\infty} \|f_{k+1} - f_k\|_1 < \infty$. Im folgenden bezeichnen $\|\ \|_{1,X}$ und $\|\ \|_{1,Y}$ lie L^1-halbnormen auf X und Y.

Wir beweisen zunächst, daß es eine Nullmenge $N' \subset Y$ gibt, so daß für $y \in Y \setminus N'$ gilt, daß $A_y := \{x \in X\ :\ (x,y) \in A\} \subset X$ eine Nullmenge ist. Wegen $\|\delta_A\|_1 = 0$ gibt es für jedes $\epsilon > 0$ zu δ_A eine Hüllreihe $\sum_{i=0}^{\infty} \delta_{Q_i}$ mit $\sum_{i=0}^{\infty} v(Q_i) < \epsilon$. Die Quader zerlegen sich in $Q_i = Q_i' \times Q_i''$ mit $Q_i' \in X$ und $Q_i'' \in Y$. Für festes y ist (ohne Ausschluß einer Nullmenge) $\delta_{A_y} \leq \sum_{i=1}^{\infty} \delta_{Q_i'} \delta_{Q_i''}(y)$ und damit $a(y) := \int_X dx \delta_{A_y}(x) \leq \sum_{i=1}^{\infty} v(Q_i') \delta_{Q_i''}(y)$. Integration der so definierten Funktion a über Y liefert $\|a\|_{1,Y} = 0$, d.h. die Existenz einer Nullmenge $N' \subset Y$, so daß $0 = a(y) = \|\delta_{A_y}\|_{1,X} = 0$ für $y \notin N$. Das war zu zeigen.

Folglich konvergiert für festes $y \in Y \setminus N'$ die Folge $(f_{k,y})_{k\in\mathbb{N}}$ fast überall auf X (nämlich für $x \notin A_y$) gegen f_y.

Sei $H_k(y) := \int_X dx\ |f_{k+1,y}(x) - f_{k,y}(x)|$. Nach Satz 3.11 gilt $\int_Y dy\ H_k(y) = \int_{X\times Y} d(x,y)\ |f_{k+1}(x,y) - f_k(x,y)| = \|f_{k+1} - f_k\|_1$ und damit

$$\sum_{k=0}^{\infty} \int_Y dy\ H_k(y) < \infty\ . \tag{*}$$

Die Folge $(G_s)_{s\in\mathbb{N}}$ der Funktionen $G_s = \sum_{k=0}^{s} H_s$ ist monoton wachsend. Nach (*) gilt

$$I_s := \int_Y dy\ G_s(y) = \sum_{k=0}^{s} \int_Y dy\ H_k(y) \leq \sum_{k=0}^{\infty} \int_Y dy\ H_k(y) < \infty\ ,$$

d.h. die Folge $(I_s)_{s\in\mathbb{N}}$ der Integrale ist beschränkt. Nach dem Satz von Beppo Levi ist damit die Grenzfunktion $\sum_{k=0}^{\infty} H_k$ integrierbar. Nach Satz 3.16 gibt es höchstens eine Nullmenge N'', so daß $H_k(y) < \infty$ für alle $y \in Y \setminus N''$. Somit gilt für alle $y \in Y \setminus N$, mit $N = N' \cup N''$,

$$\sum_{k=0}^{\infty} \|f_{k+1,y} - f_{k,y}\|_{1,X} < \infty\ . \tag{**}$$

Damit ist $(f_{k,y})_{k\in\mathbb{N}}$ eine L^1-Cauchyfolge auf X, die nach dem Satz von Riesz-Fischer fast überall punktweise gegen eine über X integrierbare Funktion $\tilde{f}_y$ konvergiert. Damit ist nach dem Modifikationssatz auch f_y integrierbar, und für $y \in Y \setminus N$ gilt

$$\lim_{k\to\infty} \int_X dx\ f_k(x,y) = \int_X dx\ f(x,y) = F(y)\ .$$

Wir betrachten nun die Treppenfunktionen $F_k(y) := \int_X dx\ f_k(x,y)$. Die Folge $(F_k)_{k\in\mathbb{N}}$ konvergiert fast überall punktweise gegen F. Außerdem ist

$$\begin{aligned}\sum_{k=0}^{\infty} \|F_{k+1} - F_k\|_{1,Y} &= \sum_{k=0}^{\infty} \int_Y dy\ \Big| \int_X dx\ (f_{k+1,y}(x) - f_{k,y}(x)) \Big| \\ &\leq \sum_{k=0}^{\infty} \int_Y dy \int_X dx\ \big|f_{k+1,y}(x) - f_{k,y}(x)\big| = \sum_{k=0}^{\infty} \|f_{k+1} - f_k\|_1 < \infty\ .\end{aligned}$$

Somit ist $(F_k)_{k\in\mathbb{N}}$ eine L^1-Cauchyfolge, die nach dem Satz von Riesz-Fischer fast überall punktweise gegen eine integrierbare Funktion $\tilde{F}$ konvergiert. Damit ist auch F integrierbar, und es gilt

$$\begin{aligned}\int_Y dy\ F(y) &= \lim_{k\to\infty} \int_y dy\ F_k(y) = \lim_{k\to\infty} \int_y dy\ \int_X dx\ f_k(x,y) \\ &= \lim_{k\to\infty} \int_{X\times Y} d(x,y)\ f_k(x,y) = \int_{X\times Y} d(x,y)\ f(x,y)\ . \qquad \square\end{aligned}$$

Im Beweis können die Rollen von X, Y vertauscht werden, so daß für eine über $X \times Y$ integrierbare Funktion gilt

$$\int_{X\times Y} d(x,y)\ f(x,y) = \int_X dx \int_Y dy\ f(x,y) = \int_Y dy \int_X dx\ f(x,y)\ .$$

Die Voraussetzung der Integrierbarkeit über $X \times Y$ ist entscheidend. Es gibt Beispiele für Funktionen, für die die Integrale $\displaystyle\int_X dx \int_Y dy\ \ f(x,y)$ und $\displaystyle\int_Y dy \int_X dx\ f(x,y)$ existieren, ohne daß $f(x,y)$ integrierbar ist. Um die Umkehrung des Satzes von Fubini zu formulieren, benötigen wir fogende Bezeichnung:

Definition 3.11 Es sei A eine Vereinigung abzählbar vieler kompakter Mengen. Eine Funktion $f : A \to \mathbb{R} \cup \{\infty\}$ heißt *lokal-integrierbar*, wenn sie über jede kompakte Teilmenge $K \subset A$ integrierbar ist.

Satz 3.29 (Tonelli) *Eine lokal-integrierbare oder fast überall stetige Funktion $f : X \times Y \to \mathbb{R}$ ist genau dann über $X \times Y$ integrierbar, wenn wenigstens eines der iterierten Integrale $\displaystyle\int_X dx \int_Y dy\ |f(x,y)|$ oder $\displaystyle\int_Y dy \int_X dx\ |f(x,y)|$ existiert. Ist das der Fall, so gilt*

$$\int_{X\times Y} d(x,y)\ f(x,y) = \int_X dx \int_Y dy\ f(x,y) = \int_Y dy \int_X dx\ f(x,y)\ .$$

($\Leftarrow$) Mit f ist auch $|f|$ über $X \times Y$ integrierbar, und der Satz von Fubini liefert die Existenz und Eigenschaften der iterierten Integrale.

($\Rightarrow$) Wir beweisen nur den Fall, daß f lokal integrierbar ist. Für $k \geq 1$ sei $W_k := [-k,k]^n \subset \mathbb{R}^n$ der kompakte Würfel mit Kantenlänge $2k$ und $f_k := \min(|f|, k\delta_{W_k})$. Dann ist f_k integrierbar. Die Folge $(f_k)_{k\in\mathbb{N}}$ konvergiert monoton wachsend gegen $|f|$, und die Folge der Integrale ist nach dem Satz von Fubini für f_k beschränkt:

$$\begin{aligned}\int_{X\times Y} d(x,y)\ f_k(x,y) &= \int_Y dy \int_X dx f_k(x,y) \leq \int_Y dy \int_X dx|f_k(x,y)| \\ &\leq \int_Y dy \int_X dx|f(x,y)| < \infty\ .\end{aligned}$$

Nach dem Satz von Beppo Levi ist damit die Grenzfunktion $|f|$ über $X \times Y$ integrierbar. Andererseits ist f über $X \times Y$ lokal integrierbar, insbesondere ist $X \times Y$ die Vereinigung abzählbar vieler kompakter Teilmengen. Also gibt es eine Ausschöpfung $A_0 \subset A_1 \subset \ldots$ von $X \times Y = \bigcup_{i\in\mathbb{N}} A_i$ durch kompakte Teilmengen A_i, und f_{A_i} ist integrierbar. Da f_{A_i} durch die integrierbare Funktion $|f|$ beschränkt ist, ist f nach dem Satz von Lebesgue integrierbar. $\square$

3.10 Der Transformationssatz

Der Transformationssatz ist eine mächtige Methode zur Berechnung von Integralen.

Satz 3.30 *Seien $U, V \subset \mathbb{R}^n$ offene Teilmengen und sei $T : U \to V$ ein Diffeomorphismus. Eine Funktion $f : V \to \mathbb{R} \cup \{\infty\}$ ist genau dann über V integrierbar, wenn die Funktion $|\det(dT)| \cdot (f \circ T) : U \to \mathbb{R} \cup \{\infty\}$ über $U = T^{-1}(V)$ integrierbar ist. In diesem Fall gilt*

$$\int_U dx \, |\det(DT)(x)| \, f(T(x)) = \int_V dy \, f(y) \, .$$

Zur Erinnerung: Ein Diffeomorphismus T ist eine differenzierbare bijektive Abbildung mit differenzierbarem Inversen. Das Differential DT ist dann eine lineare Abbildung $DT : \mathbb{R}^n \to \mathbb{R}^n$, d.h. $DT \in \mathrm{End}(\mathbb{R}^n)$, so daß die Determinante korrekt definiert ist. Bei der Diskussion des Satzes über implizite Funktionen haben wir in Satz 2.35 gezeigt, daß das Differential einer bijektiven differenzierbare Abbildung invertierbar ist. Damit ist $|\det(DT)(x)| \neq 0$ für alle $x \in U$. Der Transformationssatz läßt sich aber auch verwenden, wenn T nur auf einer Nullmenge N kein Diffeomorphismus ist, da Nullmengen im Lebesgue-Integral keine Rolle spielen. In diesem Fall genügt es, über $U \setminus N$ bzw. $T^{-1}(U \setminus N)$ zu integrieren.

Bevor wir den Transformationssatz beweisen, diskutieren wir einige Folgerungen und Anwendungen.

- Im $\mathbb{R}^1$ reduziert sich der Transformationssatz auf folgende Form: Sei $T : [a, b] \to [\alpha, \beta]$ eine bijektive stetig differenzierbare Abbildung, dann gilt $\displaystyle\int_{[a,b]} dx \, |T'(x)| \, f(T(x)) = \int_{[\alpha,\beta]} dy \, f(y)$. Das Auftreten von $|T'|$ ist leicht zu verstehen: Mit $y = T(x) \in V$ folgt formell $dy = \frac{dT}{dx} dx = T'(x) \, dx$. Ist $T' < 0$, so ist $T^{-1}(\alpha) = b$ und $T^{-1}(\beta) = a$. Die dann notwendige Vertauschung von oberer und unterer Integrationsgrenze liefert ein weiteres Vorzeichen, so daß in die Transformationsformel nur $|T'(x)|$ eingeht.
- Im Spezialfall einer nichtausgearteten affinen Transformation $y = T(x) = Ax + b \in \mathbb{R}^n$ mit $\det A \neq 0$ ist $f : K \to \mathbb{R}^n \cup \{\infty\}$ genau dann über $K \subset \mathbb{R}^n$ integrierbar, wenn $f \circ T$ über $T^{-1}(K)$ integrierbar ist, und es gilt
$$\int_{T^{-1}(K)} dx \, f(Ax + b) = \frac{1}{|\det A|} \int_K dy \, f(y) \, .$$
- Beschreibt A eine Rotation oder Spiegelung (dann ist $|\det A| = 1$) und wählen wir für f die konstante Funktion $f = 1$, so folgt, daß K genau dann meßbar ist, wenn $T^{-1}(K)$ meßbar ist, und es gilt $v_n(K) = v_n(T^{-1}(K))$. Volumina bleiben also bei Kombinationen aus Verschiebung, Drehung und Spiegelung erhalten.

- Ist A eine Diagonalmatrix mit Diagonalelementen $a_{11}, \dots, a_{nn}$ und ist $b = 0$, so handelt es sich um eine richtungsabhängige Skalierung. Dadurch entsteht z.B. aus der dreidimensionalen Kugel $\overline{B_1(0)}$ ein Ellipsoid

$$E = \left\{(y_1, y_2, y_3) \in \mathbb{R}^3 \ : \ \frac{y_1^2}{a_{11}^2} + \frac{y_2^2}{a_{22}^2} + \frac{y_3^2}{a_{33}^2} \le 1\right\}$$

durch $T(E) = \overline{B_1(0)}$ bzw. $y = Ax$. Dann folgt für das Volumen des Ellipsoids

$$\begin{aligned} v_3(E) &= \int_E d(y_1, y_2, y_3)\ 1(y) = \int_{\overline{B_1(0)}} dx\ |\det A|\ 1(Ax) = |\det A|\ \kappa_3 \\ &= \frac{4}{3}\pi a_{11} a_{22} a_{33}\ . \end{aligned}$$

Sehr häufig treten Integrale über n-dimensionale Kugeln oder Kugelschalen auf. Solche Integrale lassen sich durch eine Transformation T zu Polarkoordinaten vereinfachen (und mit dem Satz von Fubini oft auch lösen). Polarkoordianten im $\mathbb{R}^n$ bestehen aus dem Radius r und $n-1$ Winkeln $\varphi, \vartheta_1, \dots, \vartheta_{n-2}$. Dann ist $T : (r, \varphi, \vartheta_1, \dots, \vartheta_{n-1}) \mapsto (y_1, \dots, y_n) = T_n(r, \varphi, \vartheta_1, \dots, \vartheta_{n-2})$ definiert durch $T_2(r, \varphi) = (r\cos\varphi, r\sin\varphi)$ und dann rekursiv $y_n = r\cos\vartheta_{n-2}$ und $(y_1, \dots, y_{n-1}) = T_{n-1}(r, \varphi, \vartheta_1, \dots, \vartheta_{n-3}) \cdot \sin\vartheta_{n-2}$. Konkret heißt das

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_{n-1} \\ y_n \end{pmatrix} = \begin{pmatrix} r\cos\varphi\sin\vartheta_1\cdots\sin\vartheta_{n-2} \\ r\sin\varphi\sin\vartheta_1\cdots\sin\vartheta_{n-2} \\ r\cos\vartheta_1\sin\vartheta_2\cdots\sin\vartheta_{n-2} \\ \vdots \\ r\cos\varphi_{n-3}\sin\vartheta_{n-2} \\ r\cos\vartheta_{n-2} \end{pmatrix} .$$

Damit die Transformation T bijektiv wird, ist (z.B.) $\vartheta_1, \dots, \vartheta_{n-2} \in \,]0, \pi[$ und $\varphi \in \,]0, 2\pi[$ zu wählen.

Satz 3.31 *Es gilt* $|\det(DT)(r, \varphi, \vartheta_1, \dots, \vartheta_{n-2})| = r^{n-1}\sin\vartheta_1^1\cdots\sin\vartheta_{n-2})^{n-2}$.

Beweis. Das Differential der Transformation ist

$$DT = \begin{pmatrix} \frac{\partial y_1}{\partial r} & \frac{\partial y_1}{\partial\varphi} & \frac{\partial y_1}{\partial\vartheta_1} & \cdots & \frac{\partial y_1}{\partial\vartheta_{n-2}} \\ \frac{\partial y_2}{\partial r} & \frac{\partial y_2}{\partial\varphi} & \frac{\partial y_2}{\partial\vartheta_1} & \cdots & \frac{\partial y_2}{\partial\vartheta_{n-2}} \\ \frac{\partial y_3}{\partial r} & \frac{\partial y_3}{\partial\varphi} & \frac{\partial y_3}{\partial\vartheta_1} & \cdots & \frac{\partial y_3}{\partial\vartheta_{n-2}} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_n}{\partial r} & \frac{\partial y_n}{\partial\varphi} & \frac{\partial y_n}{\partial\vartheta_1} & \cdots & \frac{\partial y_n}{\partial\vartheta_{n-2}} \end{pmatrix}$$

Für $n = 2$ ist $\partial_r y = (\cos\varphi, \sin\varphi)$ und $\partial_\varphi y = (-r\sin\varphi, -r\cos\varphi)$, so daß die Determinantenformel gilt. Im Schritt von n auf $n+1$ für $n \geq 2$ haben wir mit $\tilde{y} = (y_1, \dots, y_n) \in \mathbb{R}^n$ und $y = (\tilde{y}\sin\vartheta_{n-1}, r\cos\vartheta_{n-1}) \in \mathbb{R}^{n+1}$

$$\partial_r y = \begin{pmatrix} \partial_r\tilde{y}\cdot\sin\vartheta_{n-1} \\ \cos\vartheta_{n-1} \end{pmatrix}, \qquad \partial_\varphi y = \begin{pmatrix} \partial_\varphi\tilde{y}\cdot\sin\vartheta_{n-1} \\ 0 \end{pmatrix},$$
$$\partial_{\vartheta_i} y = \begin{pmatrix} \partial_{\vartheta_i}\tilde{y}\cdot\sin\vartheta_{n-1} \\ 0 \end{pmatrix} \text{ für } 1 \leq i \leq n-2\,, \qquad \partial_{\vartheta_{n-1}} y = \begin{pmatrix} \tilde{y}\cos\vartheta_{n-1} \\ -r\sin\vartheta_{n-1} \end{pmatrix}.$$

Nach Induktionsannahme gelte die Determinatenformel für $n \geq 2$. Im Schritt von n auf $n+1$ betrachten wir zunächst $\sin\vartheta_{n-1} = 0$. Dann ist gilt $\operatorname{rang}(dT) = 2$ und damit $\det dT = 0$. Sei also $\sin\vartheta_{n-1} \neq 0$. Dann addieren wir das $(-r\frac{\cos\vartheta_{n-1}}{\sin\vartheta_{n-1}})$-fache der ersten Spalte zur letzten. Wegen $r\partial_r\tilde{y} = \tilde{y}$ wird die neue letzte Spalte zu

$$\begin{pmatrix} -r\partial_r\tilde{y}\cdot\cos\vartheta_{n-1} \\ -r\frac{\cos^2\vartheta_{n+1}}{\sin\vartheta_{n-1}} \end{pmatrix} + \begin{pmatrix} \tilde{y}\cos\vartheta_{n-1} \\ -r\sin\vartheta_{n-1} \end{pmatrix} = \begin{pmatrix} 0 \\ \frac{-r}{\sin\vartheta_{n-1}} \end{pmatrix}.$$

Entwicklung nach der neuen letzten Spalte und Herausziehen des Faktors $\sin\vartheta_{n-1}$ aus jeder der ersten n Spalten der Unterdeterminante bestätigt die Determinatenformel. □

Sei nun $\Pi :=]0, 2\pi[\times(]0,\pi[)^{n-2}$ und $I \subset]0,\infty[$ ein offenes Intervall. Dann ist das Bild von $I \times \Pi$ unter T die offene Teilmenge $K(I) \setminus N \subset \mathbb{R}^n$, wobei $K[I] := \{x \in \mathbb{R}^n \,:\, \|x\| \in I\}$ die offene Kugelschale der Radien im Intervall I ist und N eine Nullmenge, die durch Aufschneiden der Kugel bei $y_2 = 0$ entlang der positiven y_1-Achse erhalten wird. Da für das Lebesgue-Integral Nullmengen (N und Ränder von I) keine Rolle spielen, erhalten wir

Satz 3.32 *Sei $I \subset [0,\infty[$ ein beliebiges Intervall und $K(I) \subset \mathbb{R}^n$ die entsprechende Kugelschale. Eine auf $K(I)$ definierte Funktion f ist genau dann über die Kugelschale $K(I)$ integrierbar, wenn die Funktion $f(T(r,\varphi,\vartheta_1,\dots,\vartheta_{n-2}))\cdot r^{n-1}C(\varphi,\vartheta_1,\dots,\vartheta_{n-2})$ über $I \times \Pi$ integrierbar ist. In diesem Fall gilt (wegen Fubini)*

$$\begin{aligned}&\int_{K(I)} dy\; f(y)\\ &= \int_I dr\; r^{n-1}\int_0^{2\pi} d\varphi \int_0^\pi d\vartheta_1\;\sin\vartheta_1\cdots\int_0^\pi d\vartheta_{n-2}\;\sin^{n-2}\vartheta_{n-2}\; f(T(r,\varphi,\vartheta_1,\dots,\vartheta_{n-2}))\,.\end{aligned}$$

Als Beispiel berechnen wir nochmals das Volumen der dreidimensionalen Kugel vom Radius R, d.h. $I = [0,R]$. Die Funktion 1 ist integrierbar, so daß gilt

$$v(K([0,R])) = \int_0^R dr\; r^2\int_0^{2\pi} d\varphi\int_0^\pi d\vartheta_1\;\sin\vartheta_1 = \frac{R^3}{3}\cdot 2\pi\cdot\big(-\cos\vartheta_1\big|_0^\pi\big) = \frac{4\pi}{3}R^3\,.$$

Wenn in Satz (3.32) die Funktion f nicht von den Winkeln abhängt, also rotationssymmetrisch ist, dann erhalten wir eine Verallgemeinerung von Satz 3.23:

Satz 3.33 *Es sei f eine Funktion auf dem Intervall $]a, b[$. Die Funktion $\tilde{f}$ auf $\mathbb{R}^n$ mit $\tilde{f}(x) = f(\|x\|)$ ist genau dann über die Kugelschale K_I integrierbar, wenn die Funktion $f(r)r^{n-1}$ über I integrierbar ist. In diesem Fall gilt*

$$\int_{K(I)} dx\ f(\|x\|) = n\kappa_n \int_I dr\ r^{n-1}\, f(r)\ .$$

Beweis. Unter Verwendung von Satz 3.32 ist nur zu zeigen, daß das Winkelintegral $\displaystyle\int_0^{2\pi} d\varphi \int_0^{\pi} d\vartheta_1\ \sin\vartheta_1 \cdots \int_0^{\pi} d\vartheta_{n-2}\ \sin^{n-2}\vartheta_{n-2} = n\kappa_n$ liefert. Das folgt aber sofort für das Volumen der Einheitskugel mit $f = 1$ und $I = [0, 1]$:

$$\begin{aligned}\kappa_n &= \int_{[0,1]} dr\ r^{n-1} \int_0^{2\pi} d\varphi \int_0^{\pi} d\vartheta_1\ \sin\vartheta_1 \cdots \int_0^{\pi} d\vartheta_{n-2}\ \sin^{n-2}\vartheta_{n-2} \\ &= \frac{1}{n} \int_0^{2\pi} d\varphi \int_0^{\pi} d\vartheta_1\ \sin\vartheta_1 \cdots \int_0^{\pi} d\vartheta_{n-2}\ \sin^{n-2}\vartheta_{n-2}\ .\end{aligned}$$

□

3.11 Beweis des Transformationssatzes

Wir unterteilen den Beweis in folgende Schritte:

i) Diskussion der Nullmengen

ii) Volumen eines affin transformierten Würfels

iii) Volumen eines diffeomorph transformierten Würfels

iv) Beweis für Treppenfunktionen

v) Beweis im allgemeinen Fall

In den Beweisen ist es vorteilhaft, auf $\mathbb{R}^n$ die *Maximumsnorm* $\|(x_1, \dots, x_i)\|_\infty := \max(x_1, \dots, x_n)$ einzuführen.

Lemma 3 *Ist $N \subset U$ eine Nullmenge, dann ist auch $T(N) \subset V$ eine Nullmenge.*

Beweis. Für $x, y \in U$ und $t \in [0, 1]$ sei $g(t) := T(x + t(y - x))$. Dann gibt es nach dem Mittelwertsatz ein $\theta \in [0, 1]$ mit

$$T(y) - T(x) = g(1) - g(0) = g'(\theta) = (DT)(x + \theta(y - x)) \cdot (y - x)\ ,$$

wobei im letzten Schritt die Kettenregel benutzt ist. Damit gilt nach Definition der Norm einer linearen Abbildung

$$\|T(y) - T(x)\|_\infty \leq \sup_{\theta\in[0,1]} \|(DT)(x + \theta(y - x))\| \cdot \|y - x\|_\infty\ .$$

Zu N gibt es für jedes $\epsilon > 0$ eine Überdeckung durch abzählbar viele kompakte Würfel $W_k \subset U$ mit $\sum_{k=0}^{\infty} v_n(W_k) < \epsilon$. Entsprechend ist $T(N) \subset \bigcup_{k=0}^{\infty} T(W_k)$.

Sei jetzt $x, y \in N \cap W_k$. Auf W_k ist $\|(DT)(x + \theta(y - x))\|$ als stetige Funktion auf einer kompakten Menge beschränkt, d.h. $\|T(y) - T(x)\|_\infty \leq L \cdot \|y - x\|_\infty$ für alle $x, y \in W_k$ und somit auch für alle $x, y \in N \cap W_k$. Also gilt $v_n(T(W_k)) \leq L^n v_n(W_k)$, d.h. $T(N \cap W_k)$ ist eine Nullmenge. Dann ist auch $T(N)$ als Vereinigung von abzählbar vielen Nullmengen eine Nullmenge. □

Lemma 4 *Es seien $a_1, \ldots, a_n \in \mathbb{R}^n$ und $P(a_1, \ldots, a_n) := \{x = t_1 a_1 + \cdots + t_n a_n \ : \ t_i \in [0, 1]\}$ das durch diese Vektoren aufgespannte Parallelotop. Dann gilt*

$$v_n(P(a_1, \ldots, a_n)) = |\det(a_1, \ldots, a_n)| \ ,$$

wobei a_i auf der rechten Seite die i-te Zeile einer $(n \times n)$-Matrix ist.

Beweis. Aus der Definition und dem Beweis der Eindeutigkeit der Determinante im letzten Semester folgt, daß der Betrag der Determinante eindeutig definiert ist durch

(D1) $|\det(\ldots, \lambda a_i, \ldots)| = |\lambda| |\det(\ldots, a_i, \ldots)|$

(D2) $|\det(\ldots, a_i, \ldots, a_j, \ldots)| = |\det(\ldots, a_i + a_j, \ldots, a_j, \ldots)|$

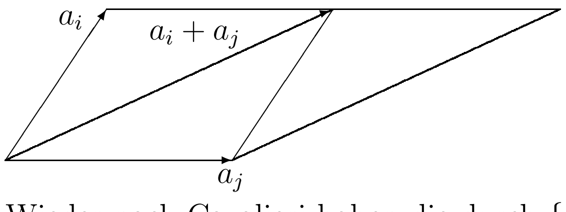
(D3) $|\det(e_1, \ldots, e_n)| = 1$

Die Punkte in (D1), (D2) bedeuten, daß die jeweiligen Zeilen der rechten und linken Seite identisch sind.

Wir beweisen, daß auch das Volumen diese Eigenschaften hat. (D3) ist klar.

(D1) Sei $P_\lambda := P(a_1, \ldots, a_{i-1}, \lambda a_i, a_{i+1}, \ldots, a_n)$. Die Parallelotope P_1 und P_{-1} sind nur gegeneinander verschoben und haben deshalb das gleiche Volumen (das war eine Übungsaufgabe). Wir können uns also auf $\lambda > 0$ beschränken. Für natürliche Zahlen $\lambda = l \in \mathbb{N} \setminus \{0\}$ gilt offenbar $v_n(P_l) = l v_n(P_1)$ nach Aneinandereihung von l Parallelotopen P_1 in i-ter Richtung. Sei $\lambda = \frac{p}{q}$ eine rationale Zahl mit $p, q \in \mathbb{N} \setminus \{0\}$. Dann gilt $v_n(P_{q \cdot \lambda}) = q v_n(P_\lambda) = v_n(P_p) = p v_n(P_1) = \frac{p}{q} v_n(P_q)$. Schließlich finden wir für $\lambda \in \mathbb{R}_+$ zu jedem $\epsilon > 0$ rationale Zahlen $r_1 \leq \lambda \leq r_2$ mit $|r_1 - r_2| \leq \frac{\epsilon}{v_n(P_1)}$. Das ergibt $v_n(P_{r_1}) \leq v_n(P_\lambda) \leq v_n(P_{r_2})$ und damit $|v_n(P_\lambda) - \lambda v_n(P_1)| \leq \epsilon$. Somit gilt $(D1)$ für alle $\lambda \in \mathbb{R}$.

(D2) Nach dem Prinzip von Cavalieri genügt es, die jeweiligen Flächen in der $\{i, j\}$-Ebene zu vergleichen:



Wieder nach Cavalieri haben die durch $\{a_i, a_j\}$ bzw. $\{a_i + a_j, a_j\}$ aufgespannten Parallelogramme die gleiche Fläche. Das beendet den Beweis. □

Sei nun $W = P(e_1, \ldots, e_n) \subset \mathbb{R}^n$ der Einheitswürfel und $T : \mathbb{R}^n \to \mathbb{R}^n$ mit $T : x \mapsto A \cdot x$ eine lineare Abbildung. Dann ist $A \cdot e_i = a_i$ die i-te Spalte von A bzw. die

i-te Zeile von A^t. Aus der Linearität von T folgt somit $T(W) = P(a_1, \ldots, a_n)$. Aus Lemma 4 und $\det A^t = \det A$ ergibt sich schließlich $v_n(T(W)) = |\det A| \cdot v_n(W)$.

Lemma 5 *Für jeden kompakten Würfel $W \subset U$ gilt* $v_n(T(W)) \leq \max_{x \in W} |\det(DT)(x)| \cdot v_n(W)$.

Beweis. Da jede kompakte Teilmenge meßbar ist und das Bild einer kompakten Teilmenge im $\mathbb{R}^n$ unter einer stetigen Abbildung wieder kompakt ist, sind $v_n(W)$ und $v_n(T(W))$ definiert. Wegen Lemma 3 genügt es, den Fall $v_n(W) > 0$ zu beweisen.

Wir setzen $\alpha := \frac{v_n(T(W))}{v_n(W)}$. Durch Halbierung sämtlicher Kanten zerlegen wir W in 2^n achsenparallele gleich große Teilwürfel. Dann gibt es einen Teilwürfel W_1 mit $v_n(T(W_1)) \geq \alpha v_n(W_1)$. Durch Wiederholung dieser Zerlegung gewinnt man eine Folge $W_1 \supset W_2 \supset \ldots$ von Würfeln mit $v_n(T(W_i)) \geq \alpha v_n(W_i)$. Wie im Beweis des Satzes von Heine-Borel (Satz 2.15) gibt es einen Punkt $a \in W$, der in allen W_i liegt. Sei $b := T(a)$ der Bildpunkt. Wir können das Koordinatensystem so verschieben, daß $a = b = 0$ gilt.

Ist m_k der Mittelpunkt des k-ten Würfels und hat der Ausgangswürfel W die Kantenlänge $2L$, dann ist $W_k = \{x \in U \ : \ \|x - m_k\|_\infty \leq \frac{L}{2^k}\}$. Nach Definition der Differenzierbarkeit von T im Nullpunkt gilt $T(x) = T(0) + (DT)(0) \cdot x + \phi(x)$ mit $\lim\limits_{x \to 0, x \neq 0} \dfrac{\phi(x)}{\|x\|} = 0$. Wir setzen $A := (DT)(0) \in GL(n, \mathbb{R})$. Wegen $T(0) = 0$ gilt dann $T(x) = A \cdot (x + \|x\|_\infty \cdot r(x))$, wobei $r(x) := \frac{1}{\|x\|_\infty} A^{-1} \cdot \phi(x)$ gegen 0 konvergiert für $x \neq 0$. Also gibt es zu jedem $\epsilon > 0$ ein $\delta > 0$, so daß $\|r(x)\|_\infty < \frac{\epsilon}{2}$ für alle $x \in \mathbb{R}^n$ mit $\|x\|_\infty < \delta$. Sei l ein Index, so daß für alle $x \in W_l$ gilt $\|x\|_\infty \leq 2 \cdot \frac{L}{2^l} < \delta$. Dann gilt

$$\|(x + \|x\|_\infty \cdot r(x)) - m_l\|_\infty \leq \|x - m_l\|_\infty + \|x\|_\infty \cdot \|r(x)\|_\infty \leq \frac{L}{2^l} + 2 \cdot \frac{L}{2^l} \cdot \frac{\epsilon}{2} = \frac{L}{2^l}(1 + \epsilon)$$

für alle $x \in W_l$. Somit ist die Menge $V_l := \{x + \|x\| \cdot r(x) \ : \ x \in W_l\}$ enthalten im Würfel $W_l^\epsilon := \{x \in \mathbb{R}^n \ : \ \|x - m_l\|_\infty \leq \frac{L}{2^k} \cdot (1 + \epsilon)\}$.

Also gilt $T(W_l) = A \cdot V_l \subset A \cdot W_l^\epsilon$ und weiter

$$v_n(T(W_l)) \leq v_n(A \cdot W_l^\epsilon) = |\det A|(1 + \epsilon)^n v_n(W_l) \ .$$

Angenommen, es gelte $\alpha > \max_{x \in W} |\det(DT)(x)| \geq |\det A|$. Dann finden wir ein $\epsilon > 0$, für das auch $\alpha > (1 + \epsilon)^n |\det A|$ gilt. Das bedeutet $v_n(T(W_l)) < \alpha v_n(W_l)$ im Widerspruch zur Konstruktion von W_l. $\square$

Lemma 6 *Sei $K \subset U$ eine kompakte Teilmenge, so daß der Rand ∂K eine Nullmenge ist. Dann gilt*

$$\min_{x \in K} |\det(DT)(x)| \cdot v_n(K) \leq v_n(T(K)) \leq \max_{x \in K} |\det(DT)(x)| \cdot v_n(K) \ .$$

Beweis. Die kompakte Menge K und damit ihr offenes Innere $K \setminus \partial K$ ist meßbar mit $v_n(K) = v_n(K \setminus \partial K)$. Zu $K \setminus \partial K$ gibt es eine Ausschöpfung $A_0 \subset A_1 \subset \dots$ mit $K \setminus \partial K = \bigcup_{k=0}^{\infty} A_k$, wobei die kompakten Teilmengen $A_k = \cup_{i_k=0}^{p_k} W_{i_k}$ durch Zusammenkleben von kompakten Würfeln W_{i_k} der Kantenlängen $1, \frac{1}{2}, \dots, \frac{1}{2^k}$ entlang ihrer Ränder gebildet werden. Dann ist $v_n(K \setminus \partial K) = \sum_{k=0}^{\infty} \sum_{i_k=0}^{p_k} v_n(W_{i_k})$. Wegen der Stetigkeit und Bijektivität von T gilt $T(K) \setminus \partial T(K) = T(K \setminus \partial K) = T(\bigcup_{k=0}^{\infty} \bigcup_{i_k=0}^{p_k} W_{i_k})$ und dann mit Lemma 3 und Lemma 5

$$\begin{aligned} v_n(T(K)) = v_n(T(K \setminus \partial K)) &= \sum_{k=0}^{\infty} \sum_{i_k=0}^{p_k} v_n(T(W_{i_k})) \\ &\leq \max_{x \in K} |\det(DT)(x)| \cdot \sum_{k=0}^{\infty} \sum_{i_k=0}^{p_k} v_n(W_{i_k}) = \max_{x \in K} |\det(DT)(x)| \cdot v_n(K) \; . \end{aligned}$$

Andererseits folgt daraus durch Vertauschung der Rollen von K und $T(K)$

$$v_n(K) = v_n(T^{-1}(T(K))) \leq \max_{y \in T(K)} |\det(DT^{-1})(y)| \cdot v_n(T(K)) \; .$$

Nach dem Satz über implizite Funktionen gilt $(DT^{-1})(y) = ((DT)(x))^{-1}$ mit $x := T^{-1}(y)$, also $|\det(DT^{-1})(y)| = \frac{1}{|\det(DT)(x)|}$. Nun ist $|\det(DT^{-1})(y)|$ dort maximal, wo $|\det(DT)(x)|$ minimal ist. Das bedeutet

$$v_n(T(K)) \geq \min_{x \in K} |\det(DT)(x)| \cdot v_n(K) \; . \qquad \square$$

Satz 3.34 *Der Transformationssatz gilt für jede Treppenfunktionen f auf V, deren Träger* $\mathrm{supp}(f) := \overline{\{y \in \mathbb{R}^n \; : \; f(y) \neq 0\}}$ *Teilmenge von V ist.*

Beweis. Wegen der Linearität des Integrals genügt es, den Transformationssatz für die charakteristische Funktion eines Quaders zu beweisen. Weiter brauchen wir nach Lemma 3 nur kompakte Quader $Q \in V$ zu betrachten, da der Rand eines Quaders eine Nullmenge ist. Die Integrierbarkeit von $|\det DT|\delta_Q \circ T$ ist klar, denn $\delta_Q \circ T$ verschwindet außerhalb der kompakten Menge $T^{-1}(Q) \subset U$, und $|\det DT|$ ist stetig auf $T^{-1}(Q)$. Zu zeigen bleibt

$$\int_Q dy = v_n(Q) = \int_{T^{-1}(Q)} dx \; |\det(DT)(x)| \; .$$

Da die stetige Funktion $|\det(DT^{-1})|^{-1}|$ auf der kompakten Menge Q gleichmäßig stetig ist (Satz 2.19), gibt es zu jedem $\epsilon > 0$ eine Zerlegung $Q = Q_1 \cup \dots \cup Q_p$ in kompakte Quader, die nur Randpunkte gemeinsam haben und so klein sind, daß $\max_{y \in Q_i} |\det(DT^{-1})(y)|^{-1} - \min_{y \in Q_i} |\det(DT^{-1})(y)|^{-1} \leq \epsilon$. Dann gilt im Urbild $T^{-1}(Q_i)$

$$\begin{aligned} \max_{x \in T^{-1}(Q_i)} |\det(DT)(x)| v_n(T^{-1}(Q_i)) - \min_{x \in T^{-1}(Q_i)} |\det(DT)(x)| v_n(T^{-1}(Q_i)) \\ \leq \epsilon v_n(T^{-1}(Q_i)) \; . \end{aligned}$$

Sowohl $\int_{T^{-1}(Q_i)} dx\ |\det(DT)(x)|$ als auch $v_n(Q_i)$ nach Lemma 6 sind enthalten im Intervall

$$\left[\min_{x\in T^{-1}(Q_i)} |\det(DT)(x)| v_n(T^{-1}(Q_i)), \max_{x\in T^{-1}(Q_i)} |\det(DT)(x)| v_n(T^{-1}(Q_i))\right] .$$

Also gilt

$$\Big|\int_{T^{-1}(Q_i)} dx\ |\det(DT)(x)| - v_n(Q_i)\Big| \leq \epsilon v_n(T^{-1}(Q_i)) .$$

Summation über alle Teilquader liefert

$$\begin{aligned}\Big|\int_{T^{-1}(Q)} dx\ |\det(DT)(x)| - v_n(Q)\Big| &\leq \sum_{i=1}^{p} \Big|\int_{T^{-1}(Q_i)} dx\ |\det(DT)(x)| - v_n(Q_i)\Big| \\ &\leq \epsilon v_n(T^{-1}(Q)) .\end{aligned}$$

Für $\epsilon \to 0$ ergibt sich die Behauptung. □

Beweis des Transformationssatzes. Nach Definition der Integrierbarkeit gibt es zu jeder über $V \subset \mathbb{R}^n$ integrierbaren Funktion f und jedem $\epsilon > 0$ eine auf einer beschränkten Teilmenge des $\mathbb{R}^n$ definierte Treppenfunktion f_ϵ mit $\|f - f_\epsilon\|_1 \leq \frac{\epsilon}{2}$. Wegen $|f_V - f_\epsilon\delta_V| \leq |f_V - f_\epsilon|$ gilt dann auch $\|f_V - f_\epsilon\delta_V\|_1 \leq \frac{\epsilon}{2}$. Sei $B \subset \mathbb{R}^n$ eine beschränkte offene Teilmenge mit $\mathrm{supp}(f_\epsilon) \subset B$ und $M = \max_{x\in\mathrm{supp}(f_\epsilon)} |f_\epsilon(x)|$. Dann gibt es zu der beschränkten offenen Teilmenge $V \cap B$ eine Vereinigung $A = Q_0 \cup \cdots \cup Q_k \subset V \cap B$ von endlich vielen kompakten Quadern Q_i mit $\big|v_n(V\cap B) - v_n(A)\big| < \frac{\epsilon}{2M}$. Damit ist $f_\epsilon\delta_A$ eine Treppenfunktion mit $\mathrm{supp}(f_\epsilon\delta_A) \subset V$, für die gilt

$$\|f_\epsilon\delta_A - f_\epsilon\delta_V\|_1 = \|f_\epsilon\delta_A - f_\epsilon\delta_{V\cap B}\|_1 \leq M|v_n(A) - v_n(V\cap B)| < \frac{\epsilon}{2} .$$

Somit gilt $\|f_V - f_\epsilon\delta_A\|_1 \leq \|f_V - f_\epsilon\delta_V\|_1 + \|f_\epsilon\delta_V - f_\epsilon\delta_A\|_1 < \epsilon$, d.h. wir können annehmen, daß die approximierenden Treppenfunktionen zu f_V ihren Träger in V haben. Nach Auwahl einer Teilfolge gemäß Satz 3.20 gibt es also zu f_V eine Familie $(f_k)_{k\in\mathbb{N}}$ von Treppenfunktionen mit Träger in V, die fast überall punktweise gegen f konvergieren und außerdem L^1-konvergent gegen f sind.

Wir betrachten die Folge der Funktionen $\tilde{f}_k := |\det(DT)|(f_k \circ T)$. Nach Satz 3.34 ist $\tilde{f}_k$ über U integrierbar, und es gilt

$$\|\tilde{f}_k - \tilde{f}_l\|_{1,U} = \int_U dx\ |\tilde{f}_k(x) - \tilde{f}_l(x)| = \int_V |f_k(y) - f_l(y)| = \|f_k - f_l\|_{1,V} .$$

Damit ist $(\tilde{f}_k)_{k\in\mathbb{R}^n}$ eine L^1-Cauchyfolge auf U, so daß eine Teilfolge fast überall punktweise gegen eine über U integrierbare Funktion $\tilde{f}$ konvergiert mit $\int_U dx\ \tilde{f}(x) = \lim_{k\to\infty}\int_U dx\ \tilde{f}_k(x)$. Andererseits konvergiert $\tilde{f}_k$ auch fast überall punktweise gegen die Funktion $|\det(DT)|(f\circ T)$. Nach dem Modifikationssatz ist dann auch $|\det(DT)|(f\circ T)$ über U integrierbar, und es gilt

$$\int_U dx\ |\det(DT)(x)|\ f(T(x)) = \lim_{k\to\infty}\int_U dx\ \tilde{f}_k(x) = \lim_{k\to\infty}\int_V dy\ f_k(y) = \int_V dy\ f(y)\ .$$

Ist umgekehrt $|\det(DT)|(f\circ T)$ über U integrierbar, dann folgt durch Vertauschen der Rollen von T und T^{-1}, daß $|\det(DT^{-1})|\big(|\det(DT)|(f\circ T)\big)\circ T^{-1} = f$ über V integrierbar ist. Damit ist der Transformationssatz bewiesen. $\square$

3.12 Integration über Teilmengen von $(\mathbb{R}_+)^n$

Häufig treten Integrationen auf, die auf das Standardsimplex

$$\Delta^n := \{(x_1,\dots,x_n)\in\mathbb{R}^n\ :\ x_i\geq 0\ ,\ x_1+\dots+x_n\leq 1\}\subset\mathbb{R}^n$$

zurückgeführt werden können, z.B. bei Funktionen auf $(\mathbb{R}_+)^n$, die entscheidend von der Summe $x_1+\dots+x_n$ abhängen. In diesem Fall ist eine auf Jacobi zurückgehende Transformation hilfreich. Dazu definiert man

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = J_2(u_1,u_2) := \begin{pmatrix} u_1(1-u_2) \\ u_1u_2 \end{pmatrix}$$

und dann rekursiv für $\tilde{x} = (x_1,\dots,x_n)^t$

$$\begin{pmatrix} \tilde{x} \\ x_{n+1} \end{pmatrix} = J_{n+1}(u_1,\dots,u_n,u_{n+1}) := \begin{pmatrix} J_n(u_1,\dots,u_n)\ (1-u_{n+1}) \\ u_1u_{n+1} \end{pmatrix}\ .$$

Wir zeigen, daß J_n einen Diffeomorphismus implementiert zwischen

- $\mathbb{R}_+\times(]0,1[)^{n-1}$ und $(\mathbb{R}_+)^n$,
- bzw. $(]0,1[)^n$ und $(\Delta^n)^o := \Delta^n\setminus\partial\Delta^n$.

Zunächst zur Bijektivität. Klar ist, daß das Bild Teilmenge von $(\mathbb{R}_+)^n$ ist. Es gilt $x_1+\dots+x_n = u_1$ zunächst für $n=2$ und dann rekursiv für alle n. Sei also $u_1>0$. Damit gilt $0<x_n<u_1$, es gibt also eine bijektive Zuordnung zwischen $u_n\in\,]0,1[$ und $x_n = u_1u_n$. Sei dann zusätzlich u_n fixiert, dann ist $x_1+\dots+x_{n-1} = u_1-x_n = u_1(1-u_n)$. Insbesondere folgt $0<x_{n-1}<u_1(1-u_n)$, damit eine bijektive Zuordnung zwischen $u_{n-1}\in\,]0,1[$ und $x_n = u_1u_{n-1}(1-u_n)$, usw.

Das Differential von J ist

$$\begin{aligned}
(DJ_2)(u_1,u_2) &= \begin{pmatrix} 1-u_2 & -u_1 \\ u_2 & u_1 \end{pmatrix}\ ,\\
(DJ_{n+1})(\tilde{u},u_{n+1}) &= \begin{pmatrix} (1-u_{n+1})(DJ_n)(\tilde{u}) & -J_n(\tilde{u}) \\ u_{n+1}\cdot e_1 & u_1 \end{pmatrix}\ ,
\end{aligned}$$

wobei $e_1 = (1, 0, \dots, 0) \in \mathbb{R}^n$ der erste Einheitsvektor ist. Also ist J differenzierbar. Es gilt $\det(DJ_2)(u_1, u_2) = u_1$. In der Rekursionsformel ist die erste Spalte gegeben durch $\left((1-u_{n+1})\frac{\partial J_n}{\partial u_1}, u_{n-1}\right)^t = \left(\frac{(1-u_{n+1})}{u_1} J_n, u_{n-1}\right)^t$, so daß Addition der $\frac{u_1}{1-u_{n+1}}$-fachen ersten Spalte zur letzten ergibt:

$$\begin{aligned}\det(DJ_{n+1})(\tilde{u}, u_{n+1}) &= \det \begin{pmatrix} (1-u_{n+1})(DJ_n)(\tilde{u}) & 0 \\ u_{n+1} \cdot e_1 & u_1 + \frac{u_1 u_{n+1}}{1-u_{n+1}} \end{pmatrix} \\ &= u_1 (1-u_{n+1})^{n-1} \det\left(\ (DJ_n)(\tilde{u}) \ \right) .\end{aligned}$$

Somit gilt $|\det(DJ_n)(u_1, \dots, u_n)| = u_1^{n-1}(1-u_3)(1-u_4)^2 \cdots (1-u_n)^{n-2}$, und J_n ist ein Diffeomorphismus. Damit haben wir den folgenden Satz bewiesen:

Satz 3.35 *Eine auf $(\mathbb{R}_+)^n$ bzw. auf $(\Delta^n)^o$ definierte Funktion f ist genau dann über $(\mathbb{R}_+)^n$ bzw. $(\Delta^n)^o$ integrierbar, wenn die Funktion $|\det(DJ_n)| f \circ J_n$ über $\mathbb{R}_+ \times W^{n-1}$ bzw. über W^n integrierbar ist, wobei $W^k := (]0,1[)^k$ der offene Würfel ist. In diesem Fall gilt*

$$\begin{aligned}&\int_{\substack{(\mathbb{R}_+)^n \\ bzw.\ \Delta^n}} dx\, f(x) \\ &= \int_{\substack{\mathbb{R}^+ \\ bzw.\ ]0,1[}} du_1\, u_1^{n-1} \int_{]0,1[} du_2 \int_{]0,1[} du_3\, (1-u_3) \dots \int_{]0,1[} du_n\, (1-u_n)^{n-2}\, f(J_n(u_1, \dots, u_n)).\end{aligned}$$

Speziell erhalten wir $v_n(\Delta^n) = \frac{1}{n!}$.

Beispiel 3.2 Mit der Jacobi-Abbildung lassen sich z.B. zweidimensionale Integrale des folgenden Typs lösen (dabei ist $p, q > 0$):

$$\begin{aligned}\int_{\Delta^2} d(x,y)\, x^{p-1} y^{p-2} f(x+y) &= \int_{]0,1[} du_1\, u_1^{p+q-1}\, f(u_1) \int_{]0,1[} du_2\, (1-u_2)^{p-1} u_2^{q-1} \\ &= B(p,q) \int_{]0,1[} du_1\, u_1^{p+q-1}\, f(u_1) .\end{aligned}$$

Hier ist $B(p,q) := \displaystyle\int_{]0,1[} dt\, (1-t)^{p-1} t^{q-1}$ die Beta-Funktion mit $B(p,q) = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}$ (Übungsaufgabe). Die obige Gleichung gilt, wenn eines der Integrale existiert. Statt über Δ^2 und $]0,1[$ kann auch über $(\mathbb{R}_+)^2$ und $\mathbb{R}_+$ integriert werden.

Die Integration über das Standardsimplex ist deshalb so wichtig, weil sich durch Potenzabbildungen viele Integrationsgebiete darauf zurückführen lassen. Dazu wird für $\alpha_i, a_i > 0$ folgende Transformation betrachtet:

$$(y_1, \dots, y_n) = T(x_1, \dots, x_n) := (a_1 x_1^{\frac{1}{\alpha_1}}, \dots, a_n x_n^{\frac{1}{\alpha_n}}) .$$

Die Transformation T bildet $(\mathbb{R}_+)^n$ diffeomorph auf sich selbst ab. Sie bildet andererseits das Innere des Standardsimplex Δ^n diffeomorph auf das Innere des verallgemeinerten Simplex

$$\Delta^{\alpha_1,\ldots,\alpha_n}_{a_1,\ldots,a_n} := \left\{ (y_1,\ldots,y_n) \in \mathbb{R}^n \ : \ y_i \geq 0 \ , \ \left(\frac{y_1}{a_1}\right)^{\alpha_1} + \ldots \left(\frac{y_n}{a_n}\right)^{\alpha_n} \leq 1 \right\}$$

ab. Das sind dann z.B. Viertelkreise ($n = 2$, $\alpha_1 = \alpha_2 = 2$, $a_1 = a_2 = r$) oder Kugeloktanten, ...

Die Determinante des Differentials ist offenbar

$$|\det(DT)(x_1,\ldots,x_n)| = \frac{a_1 \cdots a_n}{\alpha_1 \cdots \alpha_n} x_1^{\frac{1}{\alpha_1}-1} \cdots x_n^{\frac{1}{\alpha_n}-1} \ .$$

Satz 3.36 *Eine auf auf dem verallgemeinerten Simplex* ($\Delta^{\alpha_1,\ldots,\alpha_n}_{a_1,\ldots,a_n}$) *definierte Funktion f ist genau dann über dieses verallgemeinerte Simplex integrierbar, wenn die Funktion $f(a_1 x_1^{\frac{1}{\alpha_1}},\ldots,a_n x_n^{\frac{1}{\alpha_n}}) x_1^{\frac{1}{\alpha_1}-1} \cdots x_n^{\frac{1}{\alpha_n}-1}$ über das Standardsimplex integrierbar ist. In diesem Fall gilt*

$$\begin{aligned} &\int_{\Delta^{\alpha_1,\ldots,\alpha_n}_{a_1,\ldots,a_n}} dy \ f(y) \\ &= \frac{a_1 \cdots a_n}{\alpha_1 \cdots \alpha_n} \int_{\Delta^n} d(x_1,\ldots,x_n) \ x_1^{\frac{1}{\alpha_1}-1} \cdots x_n^{\frac{1}{\alpha_n}-1} f(a_1 x_1^{\frac{1}{\alpha_1}},\ldots,a_n x_n^{\frac{1}{\alpha_n}}) \ . \end{aligned}$$

Durch Kombination mit der Jacobi-Transformation entsteht so ein Diffeomorphismus $W^n \xrightarrow{J_n} \Delta^n \xrightarrow{T} \Delta^{\alpha_1,\ldots,\alpha_n}_{a_1,\ldots,a_n}$, mit dem wir Integrationen über ein verallgemeinertes Simplex auf Integrationen über den Würfel zurückführen können.

Beispiel 3.3 Wir berechnen das Volumen eines Kugeloktanten $KO := \Delta^{2,2,2}_{R,R,R}$ über die Jacobi-Transformation:

$$\begin{aligned} v_3(KO) &= \int_{\Delta^{2,2,2}_{R,R,R}} dy \\ &= \frac{R^3}{8} \int_{\Delta^3} d(x_1,x_2,x_3) \ x_1^{-\frac{1}{2}} x_2^{-\frac{1}{2}} x_3^{-\frac{1}{2}} \\ &= \frac{R^3}{8} \int_0^1 du_1 u_1^2 \int_0^1 du_2 \int_0^1 du_3 (1-u_3) \cdot u_1^{-\frac{3}{2}} u_2^{-\frac{1}{2}} (1-u_2)^{-\frac{1}{2}} u_3^{-\frac{1}{2}} (1-u_3)^{-1} \\ &= \frac{R^3}{8} \int_0^1 du_1 \ u_1^{\frac{1}{2}} \int_0^1 du_2 \ u_2^{-\frac{1}{2}} (1-u_2)^{-\frac{1}{2}} \int_0^1 du_3 \ u_3^{-\frac{1}{2}} \\ &= \frac{R^3}{8} \cdot \frac{2}{3} \cdot B(\tfrac{1}{2},\tfrac{1}{2}) \cdot 2 = \frac{1}{8} \cdot \frac{4R^3}{3} \frac{(\Gamma(\frac{1}{2}))^2}{\Gamma(1)} \ . \end{aligned}$$

Wegen $\Gamma(1) = 0! = 1$ und $v_n(KO) = \frac{R^3}{8}\kappa_3 = \frac{1}{8} \cdot \frac{4\pi R^3}{3}$ erhalten wir $\Gamma(\frac{1}{2}) = \sqrt{\pi}$.

3.13 Integration über Untermannigfaltigkeiten

Wir haben bisher die Methoden entwickelt, um Funktionen über Teilmengen $A \subset \mathbb{R}^n$ zu integrieren und z.B. Volumina solcher Teilmengen zu berechnen. Wir können damit aber noch nicht die Oberfläche des Randes von A berechnen. Die dazu notwendigen Ideen sollen nun kurz vorgestellt werden, wobei wir aus Zeitgründen keine Beweise angeben können.

Zur Erinnerung nochmals einige Definitionen und Eigenschaften aus 2.11:Untermannigfaltigkeiten.

Definition 3.12 Eine Teilmenge $M \subset \mathbb{R}^{n+k}$ heißt *n-dimensionale differenzierbare Untermannigfaltigkeit*, wenn zu jedem Punkt $a \in M$ eine offene Umgebung $U \subset \mathbb{R}^{n+k}$ von a und eine stetig differenzierbare Abbildung $f : U \to \mathbb{R}^k$ existieren, so daß

i) $U \cap M = f^{-1}(0)$

ii) für alle $x \in U$ mit $f(x) = 0 \in \mathbb{R}^k$ hat das Differential $(Df)(x) \in M(k \times (n+k), \mathbb{R})$ den maximalen Rang k.

Der Untervektorraum

$$T_a(M) := \ker((Df)(a)) = \{v \in \mathbb{R}^{n+k} \ : \ (Df)(a) \cdot v = 0\} \subset \mathbb{R}^{n+k}$$

heißt der *Tangentialraum* von M im Punkt $a \in M$. Sei orthogonales Komplement

$$N_a(M) := T_a(M)^{\perp} := \{w \in \mathbb{R}^{n+k} \ : \ \langle v, w \rangle = 0 \text{ für alle } v \in T_a(M)\}$$

heißt der *Normalenvektorraum* von M im Punkt a. Elemente $v \in T_a(M)$ bzw. $w \in N_a(M)$ heißen *Tangentialvektoren* bzw. *Normalenvektoren* an M im Punkt a.

Satz 3.37 *Eine Teilmenge $M \subset \mathbb{R}^{n+k}$ ist genau dann eine n-dimensionale differenzierbare Untermannigfaltigkeit, wenn es zu jedem Punkt $a \in M$ eine offene Umgebung $V \subset M$, eine offene Umgebung $T \subset \mathbb{R}^n$ und eine Immersion $\phi : T \to \mathbb{R}^{n+k}$ gibt, so daß T durch ϕ homöomorph auf V abgebildet wird.*

Bemerkungen. Zur Erinnerung: Immersion bedeutet, daß ϕ differenzierbar ist mit $\mathrm{rang}(D\phi)(t) = n$ für alle $t \in T$. Die Richtung $(\Rightarrow)$ hatten wir in Satz 2.36 bewiesen.

Insbesondere gibt es eine Überdeckung einer Untermannigfaltigkeit durch offene Mengen V_i. Dann heißt (V_i, ϕ_i) mit $\phi_i : T_i \to V_i$ eine *lokale Karte* von M. Für $V_{ij} := V_i \cap V_j \neq \varnothing$ gibt es zwei Homöomorphismen $\phi_i^{-1} : V_{ij} \to \phi_i^{-1}(V_{ij}) \subset T_i \subset \mathbb{R}^n$ und $\phi_j^{-1} : V_{ij} \to \phi_j^{-1}(V_ij) \subset T_j \subset \mathbb{R}^n$. Über die Konstruktion von ϕ im Satz 2.36 zeigt man, daß $\tau_{ij} := \phi_j^{-1} \circ \phi_i : \phi_i^{-1}(V_{ij}) \to \phi_j^{-1}(V_{ij})$ ein Diffeomorphismus ist zwischen Teilmengen des $\mathbb{R}^n$. Man sagt, die Kartenwechsel sind Diffeomorphismen.

Die Integration einer Funktion f über die Teilmenge $V \subset M$ wird nun über einen analogen Transformationssatz durch Integration der Funktion "$|\det D\phi|$"$(f \circ \phi)$ über T erklärt. Das Problem dabei ist, daß die Determinante der rechteckigen Matrix $D\phi$ so nicht existiert. Man zeigt, daß

$$\text{“}|\det D\phi|\text{”} := \sqrt{\det((D\phi)^t \cdot (D\phi))}$$

die richtigen Eigenschaften hat. Dabei ist $(D\phi)^t(D\phi)$ punktweise eine $n \times n$-Matrix. Entsprechend definiert man das Integral einer Funktion f über eine Karte (V, ϕ) von M mit $\phi(T) = V$ zu

$$\int_{(V,\phi)} dS\ f(x) := \int_T du\ \sqrt{\det((D\phi)^t(u) \cdot (D\phi)(u))}\ f(\phi(u))\ . \qquad (*)$$

Die Idee ist wieder zu beweisen, daß das durch die n Vektoren $a_1, \dots, a_n \subset \mathbb{R}^{n+k}$ aufgespannte Parallelotop das Volumen $\det(A^t \cdot A)$ hat, wobei a_i die Spalten von $A \in M((n+k) \times n, \mathbb{R})$ sind. Dann identifiziert man das Parallelotop mit dem Bild des n-dimensionalen Einheitswürfels im $\mathbb{R}^{n+k}$, dessen letzte k Komponenten identisch Null sind, unter einer affinen Transformation. Durch analoge Konvergenzbetrachtungen wie im Transformationssatz beweist man, daß durch (*) das Integral einer Funktion über $V \subset M$ sinnvoll definiert ist.

Beispiel 3.4 (Oberfläche der dreidimensionalen Kugel) Es sei

$$M := \{(x_1, x_2, x_3) \in \mathbb{R}^3 \ :\ x_1^2 + x_2^2 + x_3^2 = R^2\}$$

die Oberfläche der dreidimensionalen Kugel vom Radius R. Mittels Polarkoordinaten gewinnen wir die folgende Abbildung $\phi :]0, 2\pi[\times]0, \pi[\to V \subset M$:

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \phi(\varphi, \vartheta) := \begin{pmatrix} R\cos\varphi\sin\vartheta \\ R\sin\varphi\sin\vartheta \\ R\cos\vartheta \end{pmatrix} .$$

Das offene Rechteck $T :=]0, 2\pi[\times]0, \pi[$ wird durch ϕ homöomorph auf die Teilmenge $V := M \setminus HK$ abgebildet, d.h. aus der Kugeloberfläche wird der Halbkreis $HK := \{(x_1, x_2, x_3) \in \mathbb{R}^3 \ :\ x_2 = 0\ ,\ x_1 \geq 0\ ,\ x_1^2 + x_3^3 = R^2\}$ herausgeschnitten. Dann ist

$$(D\phi)(\varphi, \vartheta) = \begin{pmatrix} -R\sin\varphi\sin\vartheta & R\cos\varphi\cos\vartheta \\ R\cos\varphi\sin\vartheta & R\sin\varphi\cos\vartheta \\ 0 & -R\sin\vartheta \end{pmatrix} ,$$

$$(D\phi)^t(\varphi, \vartheta) \cdot (D\phi)(\varphi, \vartheta) = \begin{pmatrix} R^2\sin^2\vartheta & 0 \\ 0 & R^2 \end{pmatrix} ,$$

so daß wir erhalten:

$$\int_{(V,\phi)} dS\ f(x) = R^2 \int_0^{2\pi} d\varphi \int_0^{\pi} d\vartheta\ \sin\vartheta\ f(\phi(\varphi, \vartheta))\ .$$

Der Halbkreis HK ist eine Nullmenge. Man kann wieder zeigen, daß Nullmengen für die Integrationstheorie ignoriert werden können. Also stimmt das Integral mit dem Integral über ganz M überein. Insbesondere erhalten wir für $f = 1$ die Oberfläche der zweidimensionalen Sphäre vom Radius R zu $\displaystyle\int_M dS = 4\pi R^2$.

Eine wichige Konsequenz des Transformationssatzes ist, daß (*) unabhängig von der Wahl der Karte ist. Gibt es zu V zwei Karten (V, ϕ_1) und (V, ϕ_2) mit Immersionen $\phi_i : T_i \to \mathbb{R}^{n+k}$, so daß $\phi_i : T_i \to V_i$ Homöomorphismen sind, so gilt $\displaystyle\int_{(V,\phi_1)} dS\ f(x) = \int_{(V,\phi_2)} dS\ f(x)$. Zum Beweis verwendet man, daß $\phi_2 \circ \phi_1^{-1} : T_1 \to T_2$ ein Diffeomorphismus ist und den entsprechenden Transformationssatz, der $|\det D(\phi_2 \circ \phi_1^{-1})|$ beinhaltet.

Das nutzt man aus, um Integrationen über Untermannigfaltigkeiten zu definieren, die aus mehreren Karten zusammengesetzt werden müssen. Wir betrachten nur den einfachsten Fall, daß es endlich viele Karten $(V_1, \phi_1), \dots, (V_p, \phi_p)$ gibt, die $M = V_1 \cup \dots \cup V_p$ überdecken. Dann kann man immer eine Familie von Funktionen $f_i : M \to \mathbb{R}$ konstruieren mit

- $\mathrm{supp}(\alpha_i) \subset V_i$
- $\sum_{i=1}^p \alpha_i(x) = 1$ für alle $x \in M$.

Eine solche Familie heißt *Zerlegung der Eins*. Mittels Zerlegung der Eins erhalten wir:

$$\begin{aligned}\int_M dS\ f(x) &= \sum_{i=1}^p \int_M dS\ f(x)\alpha_i(x) = \sum_{i=1}^p \int_{U_i} dS\ (f\alpha_i)(x) \\ &= \sum_{i=1}^p \int_{T_i} du_i\ \sqrt{\det((D\phi_i)^t(u_i) \cdot (D\phi_i)(u_i))}\ (f\alpha_i)(\phi_i(u_i))\ .\end{aligned}$$

Die Eigenschaften der Zerlegung der Eins garantieren, daß diese Definition unabhängig von der Wahl der Überdeckung und der α_i ist. Die Konstruktion verallgemeinert sich sogar auf abzählbar viele Karten, wenn sich jeweils nur endlich viele schneiden und $|f\alpha_i|$ integrierbar ist:

Definition 3.13 Es sei $M \subset \mathbb{R}^{n+k}$ eine n-dimensionale Untermannigfaltigkeit, ausgestattet mit einem Atlas lokaler Karten (V_i, ϕ_i) entsprechend Satz 3.37, so daß $M = \bigcup_{i=0}^\infty V_i$ und jeder Punkt $x \in M$ nur in endlich vielen V_i enthalten ist.

Eine auf M definierte Funktion f heißt *über M integrierbar*, wenn es eine dem Atlas $(V_i, \phi_i)_{i\in\mathbb{N}}$ untergeordnete Zerlegung der Eins $(\alpha_i)_{i\in\mathbb{N}}$ gibt, so daß

i) Jede Funktion $f\alpha_i$ ist über V_i (damit über M) integrierbar

ii) $\displaystyle\sum_{i=0}^\infty \int_{(V_i,\phi_i)} dS\ |(f\alpha_i)(x)| < \infty$.

Dann ist das Integral von f über M (unabhängig von der Zerlegung der Eins) definiert durch

$$\int_M dS\ f(x) := \sum_{i=0}^{\infty} \int_{T_i} du_i\ \sqrt{\det((D\phi_i)^t(u_i) \cdot (D\phi_i)(u_i))}\ (f\alpha_i)(\phi_i(u_i))\ .$$

Wir sehen uns noch einige interessante Integrale über Karten an:

Beispiel 3.5 (Integrale entlang Kurven) Es sei $I \subset \mathbb{R}^n$ ein offenes Intervall und $\gamma : I \to \mathbb{R}^n$ eine differenzierbare Kurve. Unter der Annahme $(D\gamma)(t) = \gamma'(t) \neq 0$ handelt es sich um eine Immersion. Sei f jetzt eine Funktion auf der Kurve $\Gamma := f(I)$, dann ist das Kurvenintegral gegeben durch

$$\int_\Gamma dS\ f(x) = \int_I dt\ \|\gamma'(t)\|\ f(t)\ .$$

Insbesondere ist das eindimensionale Volumen der Kurve $\Gamma := f(I)$ gerade ihre Bogenlänge L,

$$L := v_1(\Gamma) = \int_\Gamma dS = \int_I dt\ \|f'(t)\|\ .$$

Die Berechnung von Determinanten des Typs $\det(A^t \cdot A)$ mit $A \in M((n+k) \times n, \mathbb{R})$ kann für große n, k sehr umständlich werden. Hier hilft das Determinanten-Multiplikationstheorem (Binet-Cauchy-Theorem) entscheidend weiter:

Theorem 3.2 (Binet-Cauchy) *Es seien $A = (a_1, \dots, a_{n+k}) \in M((n+k) \times n, \mathbb{R})$ und $B = (b_1, \dots, b_{n+k}) \in M((n+k) \times n, \mathbb{R})$ zwei rechteckige Matrizen, gebildet aus den Zeilenvektoren $a_i, b_i \in \mathbb{R}^n$. Für $1 \leq m_1 < m_2 < \dots < m_n \leq n+k$ seien quadratische Matrizen $A^{m_1 m_2 \dots m_n} := (a_{m_1}, a_{m_2}, \dots, a_{m_n}) \in M(n \times n, \mathbb{R})$ und $B^{m_1 m_2 \dots m_n} := (b_{m_1}, b_{m_2}, \dots, b_{m_n}) \in M(n \times n, \mathbb{R})$ definiert. Dann gilt*

$$\det(A^t \cdot B) = \sum_{1 \leq m_1 < m_2 \dots < m_n \leq n+k} (\det A^{m_1 m_2 \dots m_n}) \cdot (\det B^{m_1 m_2 \dots m_n})$$

Die Summe läuft über die $\binom{n+k}{n} = \frac{(n+k)!}{n!k!}$ verschiedenen Möglichkeiten, n der $n+k$ Zeilen der Matrizen auszuwählen.

Ein Beweis findet sich z.B. in G. Fischer: Lineare Algebra, Kapitel 3.3.

Beispiel 3.6 Es sei $T \subset \mathbb{R}^n$ offen und die Höhenfunktion $h : T \to \mathbb{R}$ differenzierbar. Dann ist die Abbildung $\phi : T \to \mathbb{R}^{n+1}$ mit $\phi(u) := (u, f(u))$ eine Immersion. Zur Berechnung von Integralen über den Graphen $\Gamma := \phi(T) \subset \mathbb{R}^n$ benötigen wir die Determinante der Matrix $G(u) = (D\phi)^t(u) \cdot (D\phi)(u)$. Dabei ist $(D\phi)(u) = \begin{pmatrix} E_n \\ (\operatorname{grad} h)(u) \end{pmatrix}$, wenn $(\operatorname{grad} h)(u) \in \mathbb{R}^{n-1}$ als Zeilenvektor betrachtet wird. Dann gilt mit den Bezeichnungen aus Theorem 3.2

$$\det((D\phi)^{1,2,\dots,n}) = 1\ , \quad \det((D\phi)^{1,\dots,i-1,i+1,\dots,n+1}) = \pm\partial_i h$$

und damit $\det(D\phi)^t(u)\cdot(D\phi)(u) = 1+\|(\operatorname{grad} h)(u)\|^2$. Somit erhalten wir das Integral einer Funktion f über den Graph $\Gamma := \phi(T)\subset\mathbb{R}^n$ (Höhenfäche) zu

$$\int_\Gamma dS\; f(x) = \int_T du\; \sqrt{1+\|(\operatorname{grad} h)(u)\|^2}\; f(\phi(u))\;.$$

Wir berechnen auf diese Weise noch einmal die Oberfläche der Halbkugel HK. Dazu sei $T := \{(x,y)\in\mathbb{R}^2\;:\;x^2+y^2<R^2\}$ und $h(x,y) := \sqrt{R^2-x^2-y^2}$. Dann ist $HK := (x,y,h(x,y))\subset\mathbb{R}^3$, und wir erhalten

$$v_2(HK) = \int_{HK} dS = \int_T d(x,y)\frac{R}{\sqrt{R^2-x^2-y^2}}\;.$$

Das Integral lösen wir in Polarkoordinaten $x = r\cos\varphi$, $y = r\sin\varphi$. Mit Satz 3.32 erhalten wir

$$v_2(HK) = \int_0^R dr\, r\int_0^{2\pi} d\varphi\; \frac{1}{\sqrt{1-\frac{r^2}{R^2}}} \overset{r=R\sin t}{=} 2\pi R^2\int_0^{\frac{\pi}{2}} dt\;\sin t = 2\pi R^2\;.$$

Beispiel 3.7 (Rotationsflächen im $\mathbb{R}^3$) Sei $I\subset\mathbb{R}$ offen und die Radiusfunktion $r: I\to\mathbb{R}_+$ differenzierbar. Sei $M := \{(x,y,z)\in\mathbb{R}^3\;:\;z\in I\;,\;x^2+y^2 = (r(z))^2\}$ die Rotationsfläche. Dann ist die Abbildung

$$\phi:\; I\times]0,2\pi[\;\to M\setminus N\;,\qquad \phi(\varphi,z) := \big(r(z)\cos\varphi,\; r(z)\sin\varphi,\; z\big)$$

eine Immersion, wobei der Nullmeridian $N = \{(x,y,z)\in\mathbb{R}^3\;:\;z\in I\;,\;y = 0\;,\;x^2 = (r(z))^2\}$ herausgeschnitten ist. Wir erhalten

$$(D\phi)(z,\varphi) = \begin{pmatrix} r'(z)\cos\varphi & -r(z)\sin\varphi\\ r'(z)\sin\varphi & r(z)\cos\varphi\\ 1 & 0\end{pmatrix}$$

und damit $(\det(D\phi)^t\cdot(D\phi))(z,\varphi) = r(z)^2(1+(r'(z))^2)$. Da N eine Nullmenge ist, erhalten wir das Integral einer Funktion f über die Rotationsfläche zu

$$\int_M dS\; f(x) = \int_I dz\int_0^{2\pi} d\varphi\; r(z)\sqrt{1+(r(z))^2}\, f\big(r(z)\cos\varphi,\; r(z)\sin\varphi,\; z\big)\;.$$

Für $I =]-R,R[$ und $r(z) = \sqrt{R^2-z^2}$ erhalten wir die Oberfläche der dreidimensionalen Kugel $M = \{(x,y,y)\in\mathbb{R}^3\;:\;x^2+y^2+z^2 = R^2\}$ zu

$$v_2(M) = \int_M dS = 2\pi\int_{-R}^R dz\;\sqrt{R^2-z^2}\sqrt{1+\Big(\frac{z}{\sqrt{R^2-z^2}}\Big)^2} = 2\pi R\int_{-R}^R dz = 4\pi R^2.$$

3.14 Der Gaußsche Integralsatz

Wir betrachten jetzt (differenzierbare) Hyperflächen im $\mathbb{R}^n$, d.h. $(n-1)$-dimensionale Untermannigfaltigkeiten $M \subset \mathbb{R}^n$. Lokal auf einer Teilmenge $U \subset M$ ist der Normalenvektorraum $N_a(U)$ ein eindimensionaler Untervektorraum des $\mathbb{R}^n$, gegeben durch Vielfache des Gradienten der Funktion $f : U \to \mathbb{R}^n$, die die Untermannigfaltigkeit beschreibt.

Definition 3.14 Ein *Einheitsnormalenfeld* auf einer Hyperfläche $M \subset \mathbb{R}^n$ ist ein stetiges Vektorfeld $\nu : M \to \mathbb{R}^n$, so daß in jedem Punkt $x \in M$ gilt

i) $\nu(x)$ steht senkrecht auf dem Tangentialraum $T_x(M)$

ii) $\|\nu(x)\| = 1$

Eine differenzierbare Hyperfläche heißt *orientierbar*, wenn es auf ihr ein Einheitsnormalenfeld gibt. Ein Paar (M, ν) mit festgelegtem Einheitsnormalenfeld ν heißt *orientierte Hyperfläche*.

Entweder es existieren zwei Einheitsnormalenfelder ν und $-\nu$, oder gar keines. Lokal in jeder Karte (V, ϕ) von M existiert immer ein Einheitsnormalenfeld $\nu = \frac{\operatorname{grad} f}{\|\operatorname{grad} f\|}$. Beim Zusammensetzen der Karten zu einer Überdeckung von M kann es aber das Problem geben, daß auf dem Durchschnitt $V_i \cap V_j$ sich die Einheitsnormalenfelder der Karten um das Vorzeichen unterscheiden. Bekanntestes Beispiel einer nichtorientierbaren Hyperfläche ist das Möbiusband.

Definition 3.15 Es sei (M, ν) eine orientierte differenzierbare Hyperfläche im $\mathbb{R}^n$. Ein Vektorfeld $F : M \to \mathbb{R}^n$ heißt *integrierbar über* M, wenn die Funktion $\langle F, \nu \rangle$ über M integrierbar ist. In diesem Fall setzt man

$$\int_{(M,\nu)} d\vec{S}\; F(x) := \int_M dS\; \langle F(x), \nu(x) \rangle\;.$$

Zur Formulierung des Gaußschen Integralsatzes benötigen wir den Begriff des $\mathcal{C}^1$-Polyeders:

Definition 3.16 Es sei $G \subset \mathbb{R}^n$ eine kompakte Teilmenge und ∂G der Rand von G. Ein Randpunkt $x \in \partial G$ heißt *regulärer Randpunkt*, wenn es eine Umgebung $U \subset \mathbb{R}^n$ von x und eine stetig differenzierbare Funktion $f : U \to \mathbb{R}$ gibt mit $(\operatorname{grad} f)(y) \neq 0$ für alle $y \in U$, so daß $G \cap U = \{y \in U \;:\; f(u) \leq 0\}$. Jeder nicht reguläre Randpunkt von ∂G heißt singulär. Die Menge der regulären Randpunkte heißt *regulärer Rand* $\partial_r G$. Die Menge der singulären Randpunkte heißt *singulärer Rand* $\partial_s G$. Die Menge G heißt *$\mathcal{C}^1$-Polyeder*, wenn $\partial_s G$ eine $n-1$-dimensionale Nullmenge ist.

Die Definition besagt, daß der reguläre Rand eine $(n-1)$-dimensionale Untermannigfaltigkeit (Hyperfläche) ist. Singuläre Randpunkte sind z.B. die Ecken und Kanten eines Quaders. Diese dürfen wir nicht ausschließen, da der Beweis des Gaußschen Integralsatzes auf den Fall der Quader zurückgeführt wird. Da es für $\partial_r G$ "innen" und "außen" gibt, ist $\partial_r G$ orientierbar. Das äußere Einheitsnormalenfeld ist dann dadurch ausgezeichnet, daß es für jeden Punkt $x \in \partial G \subset \mathbb{R}^n$ ein $\epsilon > 0$ gibt, so daß $x + t\nu(x) \notin G$ für alle $t \in]0, \epsilon[$.

In Vorbereitung des Gaußschen Integralsatzes sei an die Divergenz eines Vektorfeldes F auf einer offenen Teilmenge $U \subset \mathbb{R}^n$ erinnert: Ist $F = (F_1, \ldots, F_n)$ mit differenzierbaren Funktionen $F_i : U \to \mathbb{R}$, dann ist die Divergenz des Vektorfeldes die Funktion $(\mathrm{div} F) = \partial_1 F_1 + \cdots + \partial_n F_n$.

Theorem 3.3 (Gauß) *Es sei $G \subset \mathbb{R}^n$ ein $\mathcal{C}^1$-Polyeder, und $\partial_r G$ sei durch das äußere Einheitsnormalenfeld orientiert. Sei $F : U \to \mathbb{R}^n$ ein stetig differenzierbares Vektorfeld auf einer offenen Teilmenge $U \subset \mathbb{R}^n$ mit $A \subset U$. Ist die Divergenz des Vektorfeldes* div F *über $G \subset U$ integriebar und das Vektorfeld F über den regulären Rand $\partial_r G$ integrierbar, dann gilt*

$$\int_G dy\ (\mathrm{div}\, F)(y) = \int_{\partial_r G} d\vec{S}\ F(x)\ .$$

Der entscheidende Schritt im Beweis ist die Betrachtung der Situation für einen kompakten achsenparallelen Quader, der offenbar ein $\mathcal{C}^1$-Polyeder ist.

Lemma 7 *Es sei $Q \subset \mathbb{R}^n$ ein offener Quader und $F = (F_1, \ldots, F_n)$ ein stetig differenzierbares Vektorfeld auf einer Umgebung U von Q. Dann gilt*

$$\int_Q dy\ (\mathrm{div}\, F)(y) = \int_{\partial Q} d\vec{S}\ F(x)$$

Beweis. Es sei $\nu = (\nu_1, \ldots, \nu_n)$ das äußere Einheitsnormalenfeld auf ∂Q. Wegen Linearität der Integrale genügt es zu zeigen, daß für jede auf U stetig differenzierbare Funktion f und jede Komponente $i = 1, \ldots, n$ gilt

$$\int_Q dy\ (\partial_i f)(y) = \int_{\partial_r Q} (\nu_i f)(x)\ .$$

Durch Umnumerierung der Richtungen können wir $i = n$ annehmen. Dann ist $Q = Q' \times [a, b]$, wobei $Q' \subset \mathbb{R}^{n+1}$ wieder ein offener Quader ist. Entsprechend sei $y = (y', z)$ die Parametrisierung mit $y' \in \mathbb{R}^{n+1}$ und $z \in \mathbb{R}$. Der reguläre Rand von Q ist

$$\partial_r Q = \big((Q')^o \times \{a\}\big) \cup \big((Q')^o \times \{b\}\big) \cup \big(\partial_r Q' \times]a, b[\big)\ ,$$

wobei $(Q')^o$ das offene Innere von Q' ist. Für die n-te Komponente ν_n des äußeren Einheitsnormalenfeldes auf dem regulären Rand gilt dann

$$\nu_n(x) = \begin{cases} 1 & \text{auf } (Q')^o \times \{b\} \\ -1 & \text{auf } (Q')^o \times \{b\} \\ 0 & \text{auf } \partial Q' \times]a, b[\end{cases}$$

Also ist die Funktion $f\nu_n$ nur über die Randflächen $(Q')^o \times \{b\}$ und $(Q')^o \times \{a\}$ zu integrieren. Nach dem Hauptsatz der Differential- und Integralrechnung gilt

$$\begin{aligned}\int_{\partial_r Q} dx\ (\nu_n f)(x) &= \int_{(Q')^o} dy'\ f(y',b) - \int_{(Q')^o} dy'\ f(y',a) \\ &= \int_{(Q')^o} dy' \int_a^b dz\ (\partial_n f)(y',z) = \int_Q d(y',z)\ (\partial_n f)(y',z)\ . \quad \square\end{aligned}$$

Wir sehen uns einige Anwendungen des Gaußschen Integralsatzes an.

Beispiel 3.8 (Oberfläche der Einheitssphäre)
Es sei $G = K_n := \{x \in \mathbb{R}^n\ :\ \|x\| \leq 1\}$ die n-dimensionale Einheitskugel und $S^{n-1} := \partial G = \{x \in \mathbb{R}^n\ :\ \|x\| = 1\}$ die $(n-1)$-dimensionale Sphäre. Wir betrachten das Vektorfeld $F = x$ mit $(\operatorname{div} F)(x) = n$. Das äußere Einheitsnormalenfeld auf S^{n-1} ist $\nu(x) = x$. Dann gilt

$$\int_G dx\ (\operatorname{div} F)(x) = n\kappa_n = \int_{S^{n-1}} dS\ \langle x, x\rangle = \int_{S^{n-1}} dS =: \omega_n\ .$$

Die Oberfläche der S^{n-1} ist also $v_{n-1}(S^{n-1}) =: \omega_n = n\kappa_n$.

Beispiel 3.9 Es sei $G \subset \mathbb{R}^n$ ein $\mathcal{C}^1$-Polyeder, $a \in \mathbb{R}^n \setminus \partial G$ ein Punkt und $F(x) := \dfrac{x-a}{\|x-a\|^n}$. Wir beweisen

$$\int_{\partial_r G} d\vec{S}\ F(x) = \begin{cases} 0 & \text{für } a \notin G \\ \omega_n & \text{für } a \in G \end{cases} \qquad (*)$$

Zunächst gilt für $x \neq a$

$$\begin{aligned}(\operatorname{div} F)(x) &= \big\langle x-a, \operatorname{grad} \tfrac{1}{\|x-a\|^n}\big\rangle + \tfrac{1}{\|x-a\|^n}\operatorname{div}(x-a) \\ &= \big\langle x-a, \tfrac{-n}{\|x-a\|^{n+1}} \tfrac{x-a}{\|x-a\|}\big\rangle + \tfrac{n}{\|x-a\|^n} = 0\ .\end{aligned}$$

Ist $a \notin G$, dann liefert der Gaußsche Integralsatz sofort die Behauptung (*).

Ist andererseits $a \in G$, dann gibt es wegen $a \notin \partial G$ eine offene Kugel $K_r(a) \subset G$. Dann ist $G_a := G \setminus K_r(a)$ wieder ein $\mathcal{C}^1$-Polyeder, und $(\operatorname{div}\ F)(y) = 0$ für alle $y \in G_a$. Es gilt $\partial_r G_a = \partial_r G \cup \partial K_r(a)$. Das äußere Einheitsnormalenfeld ν auf $\partial K_r(a)$ aus Sicht von G_a ist das innere Einheitsnormalenfeld aus Sicht von $K_r(a)$, so daß gilt $\nu(x) = -\frac{x-a}{\|x-a\|} = -\frac{1}{r}(x-a)$. Das ergibt $\langle \nu(x), F(x)\rangle = -\frac{1}{r^{n-1}}$ für alle $x \in K_a(r)$ und damit

$$\int_{\partial_r G} d\vec{S}\ F(x) = -\int_{\partial K_r(a)} d\vec{S}\ F(x) = \frac{1}{r^{n-1}} \int_{\partial K_r(a)} dS = \omega_n\ .$$

Die Gleichung (*) verallgemeinert sich auf Linearkombinationen von Vektorfeldern F. Sei $G \subset \mathbb{R}^3$ ein $\mathcal{C}^1$-Polyeder und seien $q_1, \dots, q_k$ die Punktladungen in den Punkten $a_1, \dots, a_k \in \mathbb{R}^3 \setminus \partial G$, dann ist nach dem Coulombschen Gesetz die elektrische Feldstärke in einem Punkt $x \neq a_i$ gegeben durch

$$E = \sum_{i=1}^{k} \frac{q_i}{4\pi} \frac{x - a_i}{\|x - a_i\|^3} \ .$$

Der Fluß des elektrischen Feldstärke durch die Oberfläche ∂G ist dann gleich der Gesamtladung in G:

$$\int_{\partial G} d\vec{S}\ F(x) = \sum_{\{i\,:\ a_i \in G\}} q_i \ .$$

Satz 3.38 (Greensche Formeln) *Sei $G \subset \mathbb{R}^n$ ein $\mathcal{C}^1$-Polyeder und f, g zweimal stetig differenzierbare Funktionen auf einer offenen Umgebung von G. Dann gilt*

$$\begin{aligned} \int_G dy\ \big\langle \operatorname{grad} f, \operatorname{grad} g \big\rangle(y) &= \int_{\partial G} dS\ (f\, D_\nu g)(x) - \int_G dy\ (f\, \Delta g)(y)\ , \\ \int_G dy\ (f\, \Delta g - g\, \Delta f)(y) &= \int_{\partial G} dS\ (f\, D_\nu g - g\, D_\nu f)(x)\ , \end{aligned}$$

wobei $D_\nu f = \langle \nu, \operatorname{grad} f \rangle$ die Richtungsableitung in die äußere Normalenrichtung ist.

Beweis. Man wendet den Gaußschen Integralsatz auf das Vektorfeld $F = f \operatorname{grad} g$ an und benutzt die Leibnizregel. □

Die Greenschen Formeln spielen eine wichtige Rolle bei der Lösung wichtiger partieller Differentialgleichungen.

Definition 3.17 Sei $U \subset \mathbb{R}^n$ offen. Eine zweimal stetig differenzierbare Funktion f auf U heißt *harmonisch*, wenn $(\Delta f)(x) = 0$ für alle $x \in U$.

Die Newtonschen Potentiale $N_a : \mathbb{R}^n \to \mathbb{R}$ mit

$$N_a(x) := \begin{cases} -\dfrac{1}{(n-2)\omega_n} \dfrac{1}{\|x-a\|^{n-2}} & \text{für } n > 2 \\ \dfrac{1}{2\pi} \ln \|x - a\| & \text{für } n = 2 \end{cases}$$

sind auf $\mathbb{R}^n \setminus \{a\}$ harmonisch.

Satz 3.39 (Mittelwertsatz harmonischer Funktionen) *Es sei $f : U \to \mathbb{R}$ eine harmonische Funktion auf einer offenen Menge $U \subset \mathbb{R}^n$. Dann gilt für jede Kugel $K_r(a) \subset U$*

$$h(a) = \frac{1}{\omega_n r^{n-1}} \int_{\partial K_r(a)} dS\ h(x)\ .$$

Beweis. (für $n \geq 2$) Sei $G := \overline{K_r(a)} \setminus K_\rho(a)$ die Kugelschale mit innerem Radius ρ und äußerem Radius $r > \rho$, und sei $S_r := \partial K_r(a)$ und $S_\rho := \partial K_\rho(a)$. Dann sind h, N_a harmonisch auf G, so daß nach der 2. Greenschen Formel gilt

$$\int_{S_r} dS\ (hD_\nu N_a - N_a D_\nu h)(x) = \int_{S_\rho} dS\ (hD_\nu N_a - N_a D_\nu h)(x)\ .$$

Dabei ist ν jeweils das äußere Einheitsnormalenfeld auf den Sphären. Die 1. Greensche Formel für $G = \overline{K_R(a)}$ sowie $f \mapsto 1$ und $g \mapsto h$ liefert $\int_{S_R} dS\ (D_\nu h)(x) = 0$ für $R = \rho$ und $R = r$. Da N_a auf S_R konstant ist, folgt $\displaystyle\int_{S_r} dS\ (hD_\nu N_a)(x) = \int_{S_\rho} dS\ (hD_\nu N_a)(x)$. Für alle $x \in S_R$ gilt $(D_\nu N_a)(x) = \frac{1}{\omega_n} \frac{1}{\|x\|^{n-1}}$ und damit

$$\frac{1}{\omega_n r^{n-1}} \int_{S_r} dS\ h(x) = \frac{1}{\omega_n \rho^{n-1}} \int_{S_\rho} dS\ h(x)\ .$$

Für $\rho \to 0$ folgt aus der Stetigkeit von h die Behauptung. $\square$

Mit den Greenschen Formeln beweist man auch den folgenden Satz über eine Lösung der Potentialgleichung:

Satz 3.40 *Sei $\rho : \mathbb{R}^n \to \mathbb{R}$ eine zweimal stetig differenzierbare Funktion mit kompaktem Träger. Für $x \in \mathbb{R}^n$ sei*

$$\phi(x) := \int_{\mathbb{R}^n} dy\ N_y(x)\, \rho(y)\ .$$

Dann ist $\phi : \mathbb{R}^n \to \mathbb{R}$ zweimal stetig differenzierbar, und es gilt $\Delta\phi = \rho$.

Dabei kann man sich ρ als Ladungsdichte vorstellen und ϕ als elektrisches Potential. Auf diese Weise findet man das Coulombsche Gesetz als Lösung der statischen Maxwellschen Gleichungen.