

Analysis II für Physiker

Funktionen mehrerer Veränderlicher

Lutz Heindorf

Freie Universität Berlin
Wintersemester 2013/14

Inhaltsverzeichnis

1	Vorbereitungen	1
1.1	Die zwei Spielarten von \mathbb{R}^n	1
1.2	Topologie des \mathbb{R}^n	3
1.3	Zwischenspiel: das allgemeine Konzept des metrischen Raumes	6
1.4	Eigenschaften von Teilmengen metrischer Räume	9
1.5	Kompaktheit	12
1.6	Eigenschaften von Teilmengen von \mathbb{R}^n	13
2	Kurven	15
2.1	Begriffe	16
2.2	Die Ableitung einer Kurve; Differentiation und Integration von Vektorfunktionen	16
2.3	Die Länge einer Kurve	20
2.4	Kurvenintegrale von Skalarfeldern: $\int_{\gamma} f ds$	22
2.5	Skalare Kurvenintegrale von Vektorfeldern: $\int_{\gamma} \vec{f} \bullet d\vec{s}$	24
2.6	Äquivalente Kurven	26
3	Differentialrechnung in \mathbb{R}^n	28
3.1	Partielle Ableitungen	28
3.2	(Totale) Differenzierbarkeit	30
3.3	Kurven durch Skalarfelder; Gradienten	37
3.4	Implizite Funktionen I: der skalare Fall	41
3.5	Implizite Funktionen II: der allgemeine Fall	45
3.6	Notwendige Bedingungen für Extremwerte	48
3.7	Abschweifung; Extremwerte von quadratischen Formen und der Satz über die Hauptachsentransformation	55
3.8	Höhere partielle Ableitungen	58
3.9	Die TAYLORSche Formel für Skalarfelder	59
3.10	Hinreichende Bedingungen für Extremwerte	61
4	Integrale	64
4.1	Darstellung von Funktionen durch Integrale	64
4.2	Integrale über Rechtecke	68
4.3	Integration zweistelliger Funktionen über krummlinig begrenzte Gebiete	75
4.4	Zwei kleine Zusätze	80
4.5	Integration im \mathbb{R}^n	81
4.6	Die Transformationsformel	84
4.7	Beispiel: Integration über einen Torus	89
4.8	Zwischenspiel: Flächen in \mathbb{R}^3 ; Tangentialebene und Normalenvektor	91
4.9	Der Inhalt einer gekrümmten Oberfläche im Raum	96
4.10	Oberflächenintegrale	99
5	Integralsätze	102
5.1	Die GREENSche Formel (in der Ebene)	103
5.2	Der STOKESSche Integralsatz	106
5.3	Der Integralsatz von GAUSS	109
5.4	Die GREENSchen Formeln im Raum	114
6	Vektoranalysis	116
6.1	Anschaulich-physikalische Deutung von Divergenz und Rotation	116
6.2	Formeln	118
6.3	Potentialfelder	121
6.4	Quellenfreiheit und Vektorpotential	127
6.5	Der Satz von HELMHOLTZ	129

1 Vorbereitungen

Im letzten Semester hatten wir Funktionen untersucht, die jeder reellen Zahl aus ihrem Definitionsbereich wieder eine reelle Zahl zuordnen. Mathematisch gesprochen sind das Abbildungen $f : D \rightarrow \mathbb{R}$, für eine gewisse Teilmenge $D \subseteq \mathbb{R}$.

Jetzt geht es um Funktionen, die gewissen Paaren (x, y) , Tripeln (x, y, z) oder allgemein n -Tupeln (x_1, x_2, \dots, x_n) von reellen Zahlen wieder reelle Zahlen zuordnen. Mit Hilfe von mehreren derartigen Funktionen bekommt man dann auch Zuordnungen n -Tupel $\mapsto m$ -Tupel.

1.1 Die zwei Spielarten von \mathbb{R}^n

1.1.1 Punkte und Vektoren

Es erweist sich als günstig, eine geometrische Sprache zu benutzen und sich die Zahlentupel als (Koordinaten von) Punkte(n) bzw. Vektoren vorzustellen. Alle diese Tupel in ihrer Gesamtheit bilden den Raum \mathbb{R}^n .

Obwohl andere Autoren das nicht machen, werde ich versuchen, die beiden Aspekte *Punkt* und *Vektor* typographisch zu unterscheiden. Bei mir gibt es \mathbb{R}^n deshalb in zwei Spielarten:

- den Spaltenraum, den wir zur Unterscheidung mit $\vec{\mathbb{R}}^n$ bezeichnen. Seine Elemente heißen Vektoren und werden durch einen kleinen Pfeil gekennzeichnet¹:

$$\vec{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}, \vec{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}, \vec{c} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix}, \dots, \vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \vec{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \text{ u.ä.,}$$

- den Zeilenraum \mathbb{R}^n , dessen Elemente durch einen Querstrich gekennzeichnet werden

$$\bar{a} = (a_1, a_2, \dots, a_n), \quad \bar{b} = (b_1, b_2, \dots, b_n), \quad \dots, \quad \bar{z} = (z_1, z_2, \dots, z_n)$$

und Punkte heißen.

Den Spaltenraum $\vec{\mathbb{R}}^n$ betrachten wir als \mathbb{R} -Vektorraum mit den üblichen Operationen

$$\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} = \begin{pmatrix} a_1 + b_1 \\ a_2 + b_2 \\ \vdots \\ a_n + b_n \end{pmatrix} \quad \text{und} \quad c \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} c a_1 \\ c a_2 \\ \vdots \\ c a_n \end{pmatrix}.$$

Bei Bedarf kommt das kanonische Skalarprodukt dazu. Den Gepflogenheiten der Physiker folgend werde ich es durch einen Punkt notieren $\vec{a} \bullet \vec{b}$. Das Skalarprodukt induziert eine Norm, die als $\|\vec{a}\|$ geschrieben wird: $\|\vec{a}\| = \sqrt{a_1^2 + \dots + a_n^2}$.

Außerdem, und das ist neu, werden wir *Vektoren an Punkte antragen*. Anschaulich bedeutet das eine Verschiebung des Punktes in Richtung des (Vektor)Pfeils um dessen Länge. Die entsprechende Operation schreiben wir auch als $+$ und definieren offiziell

$$(a_1, a_2, \dots, a_n) + \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} = (a_1 + b_1, a_2 + b_2, \dots, a_n + b_n).$$

Bei uns wird der Punkt in Summen immer vorn stehen. Bei dreifachen Summen stellt sich das Problem der Klammerung: $\bar{a} + \vec{b} + \vec{c}$ kann aufgefaßt werden als $(\bar{a} + \vec{b}) + \vec{c}$ (der Punkt \bar{a}

¹Die griechischen Buchstaben zur Bezeichnung der Skalare geben wir wieder auf.

wird zweimal verschoben) oder als $\bar{a} + (\vec{b} + \vec{c})$ (der Punkt wird um die Summe aus zwei Vektoren verschoben). Wenn man die Koordinaten einsetzt, sieht man aber sofort, daß das Resultat dasselbe ist. Deshalb werden wir keine Klammern setzen.

Punkte werden wir nicht zueinander addieren oder mit Zahlen multiplizieren. Allerdings ist es bequem, eine *Subtraktion von Punkten* zu haben: $\vec{b} - \vec{a}$ ist der Vektor \vec{c} für den $\bar{a} + \vec{c} = \bar{b}$. In Koordinaten

$$(b_1, b_2, \dots, b_n) - (a_1, a_2, \dots, a_n) = \begin{pmatrix} b_1 - a_1 \\ b_2 - a_2 \\ \vdots \\ b_n - a_n \end{pmatrix}.$$

Suggestiver wäre wahrscheinlich statt $\bar{a} - \bar{c}$ so etwas wie $\overrightarrow{\bar{c}\bar{a}}$ zu schreiben, aber das ist mir typographisch zu aufwendig.

Die Operationen sind so einfach definiert, daß wir wohl keine Rechenregeln aufschreiben müssen. Ich fasse aber noch einmal kurz zusammen, was definiert ist und welches Ergebnis liefert:

Vektor	+	Vektor	=	Vektor
Punkt	+	Vektor	=	Punkt
Punkt	+	Punkt	=	nicht definiert
Punkt	-	Punkt	=	Vektor
Vektor	•	Vektor	=	Skalar
Skalar		Vektor	=	Vektor (kein Operationszeichen)
Skalar		Punkt	=	nicht definiert

Statt $\frac{1}{c}\vec{v}$ werden wir auch $\frac{\vec{v}}{c}$ schreiben. Die kanonischen Basisvektoren behalten ihre aus dem ersten Semester bewährte Bezeichnung \vec{e}_i . Bei \vec{e}_i sind alle Einträge Null, bis auf den i -ten, der gleich 1 ist. Beim *Nullpunkt* $\bar{0}$ und *Nullvektor* $\vec{0}$ sind alle Koordinaten Null.

Durch Transposition geht ein Punkt in einen Vektor über und umgekehrt. $\vec{a}^T = \bar{a} - \bar{0}$ heißt *Ortsvektor* des Punktes \bar{a} . Dagegen ist $\vec{a}^T = \bar{0} + \vec{a}$ der *Endpunkt* von \vec{a} .

Ich werde viele n -dimensionale Dinge für $n = 2$ oder $n = 3$ formulieren und/oder beweisen. Dann braucht man keine Indizes, sondern schreibt Punkte als (a, b) oder (a, b, c) und Vektoren als $\begin{pmatrix} a \\ b \\ c \end{pmatrix}$

bzw. $\begin{pmatrix} a \\ b \\ c \end{pmatrix}$.

Der Fall $n = 1$ bildet insofern eine Ausnahme, daß man die Spalte (a) nicht von der Zeile (a) unterscheiden kann. Obwohl die reellen Zahlen faktisch auch in drei verschiedenen Rollen auftreten (Skalare, (Zeit)Punkte, Vektoren) denen eigentlich die Bezeichnungen \mathbb{R} , \mathbb{R}^1 , $\vec{\mathbb{R}}^1$ entsprechen würden, werden wir diesen subtilen Unterschied nicht explizit machen, immer nur \mathbb{R} schreiben und weder Pfeile noch Querstriche setzen.

1.1.2 Metrik

Der Raum $\vec{\mathbb{R}}^n$ ist ein euklidischer Raum mit dazugehöriger Norm und Abstandsfunktion. Darüber wissen wir aus dem ersten Semester eine ganze Menge, das ich hier nicht wiederholen will. Die Abstandsfunktion von $\vec{\mathbb{R}}^n$ überträgt sich auf naheliegende Weise auf den Punktraum. Nach unserer Definition der Subtraktion von Punkten können wir sogar dieselbe Formel schreiben

$$d(\bar{a}, \bar{b}) = \|\vec{b} - \vec{a}\| = \sqrt{\sum_{i=1}^n (a_i - b_i)^2}.$$

Der Abstand zweier Punkte ist die Länge der von ihnen begrenzten Strecke. In diesem Semester werden wir unter anderem auch untersuchen, was die Länge einer gekrümmten Linie sein soll und wie man sie berechnet.

1.1.3 Volumen und Flächeninhalt

Auch darüber wissen wir schon einiges aus dem letzten Semester. Ist $\bar{a} \in \mathbb{R}^n$ ein Punkt und sind $\vec{b}_1, \dots, \vec{b}_k \in \mathbb{R}^n$ linear unabhängige Vektoren, so bestimmen diese Daten ein k -dimensionales Parallelepiped (einfacher auszusprechen einen k -Spat) im n -dimensionalen Raum:

$$\{\bar{a} + c_1\vec{b}_1 + \dots + c_k\vec{b}_k : 0 \leq c_1, \dots, c_k \leq 1\}.$$

Sein Volumen hängt nicht von \bar{a} ab. Für $n = k$ berechnet es sich als Betrag der Determinante: $|\det(\vec{b}_1, \dots, \vec{b}_n)|$. Ist $k < n$, so benutzt man die Wurzel der sogenannten GRAMSchen Determinante

$$\sqrt{\det(B^T B)} = \sqrt{\det(\vec{b}_i \bullet \vec{b}_j)_{i,j=1}^k},$$

wobei B die aus den nebeneinandergeschriebenen Spalten $\vec{b}_1, \dots, \vec{b}_k$ bestehende $n \times k$ -Matrix ist.

Sind nur zwei \vec{b} s vorhanden, so ist der 2-Spat ein Parallelogramm. Liegt dieses in \mathbb{R}^2 , so ist der Flächeninhalt $|\det(\vec{b}_1, \vec{b}_2)|$. Parallelogramme in höherdimensionalen Räumen haben den Flächeninhalt

$$\sqrt{\begin{vmatrix} \vec{b}_1 \bullet \vec{b}_1 & \vec{b}_1 \bullet \vec{b}_2 \\ \vec{b}_2 \bullet \vec{b}_1 & \vec{b}_2 \bullet \vec{b}_2 \end{vmatrix}} = \sqrt{\|\vec{b}_1\|^2 \cdot \|\vec{b}_2\|^2 - (\vec{b}_1 \bullet \vec{b}_2)^2}$$

Für Parallelogramme im dreidimensionalen Raum kann man auch das Vektorprodukt heranziehen. Der Flächeninhalt des von \vec{b}_1, \vec{b}_2 (von einem beliebigen Punkt aus) aufgespannten Parallelogramms ist $\|\vec{b}_1 \times \vec{b}_2\|$. Der halbe Parallelogramminhalt ist der Dreiecksinhalt.

Eine der Hauptaufgaben dieses Semesters betrifft die Klärung der Frage nach dem Volumen allgemeinerer Körper in \mathbb{R}^n . Speziell werden wir uns auch mit dem Inhalt gekrümmter Oberflächen in \mathbb{R}^3 beschäftigen.

1.1.4 Felder

Abbildungen, die auf einer Teilmenge des Punktraumes \mathbb{R}^n definiert sind und Vektoren aus \mathbb{R}^m als Werte annehmen, werden wir m -dimensionale *Vektorfelder* nennen. Zur Betonung bekommen die entsprechenden Funktionszeichen einen Pfeil: \vec{f}, \vec{g} , etc. Eigentlich ist der Fall $m = 1$ darin auch schon eingeschlossen. Trotzdem sprechen wir in diesem Fall lieber von n -stelligen Funktionen oder, oft suggestiver, *Skalarfeldern*. Manchmal brauchen wir auch Funktionen, die Punkte in Punkte umarbeiten, die heißen dann eventuell² *Punktfelder*.

Zwei Felder bekommen Standardbezeichnungen. Die Abbildung $\mathbb{R}^n \rightarrow \mathbb{R}^n$, die jedem Punkt seinen Ortsvektor zuordnet, heißt traditionell \vec{r} . Man kann sie so schreiben: $\vec{r}(\bar{x}) = \bar{x}^T = \bar{x} - \bar{0}$. Die Norm von \vec{r} , also der Abstand des gegebenen Punktes vom Nullpunkt, wird mit r bezeichnet: $r(\bar{x}) = \|\vec{r}(\bar{x})\| = \sqrt{x_1^2 + x_2^2 + \dots + x_n^2}$.

1.1.5 Kurven

Auf einem Intervall $I \subseteq \mathbb{R}$ definierte stetige einstellige Funktionen $I \rightarrow \mathbb{R}^n$ werden meistens *Kurven* genannt und mit griechischen Buchstaben bezeichnet. Ihnen wird später ein eigenes Kapitel gewidmet.

1.2 Topologie des \mathbb{R}^n

Mit Hilfe der Abstandsfunktion d können wir die sogenannten *topologischen* Konzepte der Konvergenz und Stetigkeit, die wir im letzten Semester einstellig kennengelernt hatten, in höhere

²Dieser Ausdruck ist nicht allgemein gebräuchlich. Man könnte auch von *Transformationen* reden, oder gar kein spezielles Wort einführen.

Dimensionen übertragen. Theoretisch müßte man alles doppelt erzählen, einmal für den Punkt- und einmal für den Vektorraum. Bis auf das Austauschen von Querstrichen gegen Vektorpfeile passiert dabei nichts. Ich beschränke mich daher meist auf den Punktraum.

1.2.1 Konvergenz von Folgen

Definitionsgemäß konvergiert eine Folge (\bar{a}_n) von Punkten gegen den Punkt \bar{b} , wenn die Zahlenfolge $d(a_n, b) = \|\bar{a}_n - \bar{b}\|$ eine Nullfolge ist. Also

$$\lim_{n \rightarrow \infty} \bar{a}_n = \bar{b} \quad : \iff \quad \lim_{n \rightarrow \infty} \|\bar{a}_n - \bar{b}\| = 0.$$

Analog für Vektorfolgen.

Beobachtung. Eine Punkt- oder Vektorfolge konvergiert genau dann, wenn ihre Koordinatenfolgen als Zahlenfolge (im Sinne des vorigen Semesters) konvergieren.

In \mathbb{R}^3 gilt also:

$$\lim_{n \rightarrow \infty} (a_n, b_n, c_n) = (p, q, r) \quad \iff \quad \lim_{n \rightarrow \infty} a_n = p \text{ und } \lim_{n \rightarrow \infty} b_n = q \text{ und } \lim_{n \rightarrow \infty} c_n = r.$$

Analog für Vektoren.

Der **Beweis** in der \leftarrow -Richtung ist eine Anwendung von Grenzwertsätzen aus dem vorigen Semester: aus $(a_n - p) \rightarrow 0$, $(b_n - q) \rightarrow 0$ und $(c_n - r) \rightarrow 0$ folgt

$$\sqrt{(a_n - p)^2 + (b_n - q)^2 + (c_n - r)^2} \rightarrow 0.$$

In der anderen Richtung, benutzt man

$$0 \leq |a_n - p|, |b_n - q|, |c_n - r| \leq \sqrt{(a_n - p)^2 + (b_n - q)^2 + (c_n - r)^2}$$

Wenn also die Wurzel gegen Null konvergiert, so auch die drei Beträge.

1.2.2 CAUCHY-Folgen und die Vollständigkeit von \mathbb{R}^n

Eine Punktfolge $(\bar{a}_n)_{n=0}^\infty$ heißt CAUCHY-Folge, wenn ihre Glieder beliebig dicht zusammenrücken. Formal, wenn für alle $\varepsilon > 0$ ein N derart existiert, daß $\|\bar{a}_n - \bar{a}_m\| < \varepsilon$ sobald $m, n \geq N$.

Dieselbe Betrachtung wie eben zeigt, daß eine Punktfolge in \mathbb{R}^n genau dann die CAUCHY-Eigenschaft hat, wenn alle ihre Koordinatenfolgen CAUCHY-Folgen in \mathbb{R} sind. Wegen der Vollständigkeit von \mathbb{R} ist das genau dann der Fall, wenn diese Koordinatenfolgen alle konvergieren. Das wiederum ist zur Konvergenz der Punktfolge äquivalent.

Fazit: In \mathbb{R}^n ist wie in \mathbb{R} die CAUCHY-Eigenschaft zur Konvergenz äquivalent: \mathbb{R}^n ist vollständig.

1.2.3 Grenzwerte von Funktionen

Ist \bar{a} ein Berührungspunkt des Definitionsbereiches $D \subseteq \mathbb{R}^n$ eines Skalarfeldes f bzw. eines m -dimensionalen Vektorfeldes \vec{f} , so soll

$$\lim_{\bar{x} \rightarrow \bar{a}} f(\bar{x}) = g \quad \text{bzw} \quad \lim_{\bar{x} \rightarrow \bar{a}} \vec{f}(\bar{x}) = \vec{g}$$

bedeuten, daß für alle $\varepsilon > 0$ ein $\delta > 0$ derart existiert, daß $|f(\bar{x}) - g| < \varepsilon$ bzw. $\|\vec{f}(\bar{x}) - \vec{g}\| < \varepsilon$ für alle $\bar{x} \in D$ mit $\|\bar{x} - \bar{a}\| < \delta$.

Bei konkreten Rechnungen stören die Wurzeln in der euklidischen Norm manchmal. Deshalb ist es ganz nützlich zu wissen, daß man statt der einen Ungleichung $\|\bar{x} - \bar{a}\| < \delta$ auch die n Ungleichungen $|x_1 - a_1| < \delta, \dots, |x_n - a_n| < \delta$ benutzen kann, ohne das Grenzwertkonzept zu ändern. Das liegt daran, daß

$$\max \left(|a_1 - x_1|, |a_2 - x_2|, \dots, |a_n - x_n| \right) \leq \|\bar{a} - \bar{x}\| \leq \sqrt{n} \cdot \max \left(|a_1 - x_1|, |a_2 - x_2|, \dots, |a_n - x_n| \right);$$

sind also alle Beträge klein, so auch die Norm und umgekehrt.

Analog kann bei Vektorfunktionen die Ungleichung $\|\vec{f}(\bar{x}) - \vec{g}\| < \varepsilon$ durch die m Ungleichungen $|f_i(\bar{x}) - g_i| < \varepsilon$ ersetzt werden.

Daraus folgt dann auch, daß man Grenzwerte ‘reinziehen’ kann:

$$\lim_{\bar{x} \rightarrow \bar{a}} \vec{f}(\bar{x}) = \begin{pmatrix} \lim_{\bar{x} \rightarrow \bar{a}} f_1(\bar{x}) \\ \lim_{\bar{x} \rightarrow \bar{a}} f_2(\bar{x}) \\ \vdots \\ \lim_{\bar{x} \rightarrow \bar{a}} f_m(\bar{x}) \end{pmatrix}.$$

1.2.4 Alle Wege müssen zum Grenzwert führen

Die folgenden Betrachtungen sind etwas lax, aber trotzdem ganz nützlich. Bei einstelligen Funktionen und $a \in \mathbb{R}$ müssen sich die Funktionswerte $f(x)$ dem Grenzwert g beliebig nähern, wenn sich x der Stelle a nähert. Für dieses Nähern gibt es unter anderem die Möglichkeiten von links und von rechts, was im vorigen Semester zu den Konzepten des linksseitigen ($\lim_{x \nearrow a}$) und rechtsseitigen ($\lim_{x \searrow a}$) Grenzwertes geführt hatte, die gleich sein müssen, damit ein Grenzwert an der Stelle a existiert.

Im mehrdimensionalen Fall gibt es nun viel mehr Möglichkeiten, sich einem Punkt zu nähern. Wenn ein Grenzwert der Funktion f im Punkt \bar{a} existieren soll, so müssen sich die Funktionswerte $f(\bar{x})$ immer diesem Grenzwert nähern, egal auf welchem Wege sich \bar{x} der Stelle \bar{a} nähert.

Diese Erkenntnis kann oft genutzt werden, um zu zeigen, daß *kein* Grenzwert existiert.

Als Beispiel betrachten wir die für $(x, y) \neq (0, 0)$ definierte Funktion $g(x, y) = \frac{x^2 y^2}{(x^2 + y^2)^2}$.

Auf beiden Koordinatenachsen ist sie konstant Null, nähert man sich dem Punkt $(0, 0)$ also achsenparallel, so kommt auch 0 als Grenzwert heraus. Auf der Winkelhalbierenden $y = x$ ist die Funktion aber konstant $g(x, x) = \frac{x^4}{(2x^2)^2} = \frac{1}{4}$. Auf anderen Wegen Richtung Nullpunkt werden auch noch andere Grenzwerte angestrebt, z.B. entlang der Geraden $y = 2x$ der Wert $\frac{4}{25}$. Also hat die Funktion g für $(x, y) \rightarrow (0, 0)$ keinen Grenzwert.

1.2.5 Stetigkeit

Mit Hilfe des Grenzwertkonzeptes wird die Stetigkeit definiert: der Grenzwert an der entsprechenden Stelle muß gleich dem Funktionswert sein. Dröseln man den Grenzwertbegriff auf, erhält man die ε - δ -Beschreibung: *Das Vektorfeld $\vec{f} : D \rightarrow \mathbb{R}^m$ ist stetig im Punkt $\bar{a} \in D \subseteq \mathbb{R}^n$, falls für alle $\varepsilon > 0$ ein $\delta > 0$ derart existiert, daß $\|\vec{f}(\bar{x}) - \vec{f}(\bar{a})\| < \varepsilon$ sobald $\bar{x} \in D$ und $\|\bar{x} - \bar{a}\| < \delta$.*

Die Möglichkeit Grenzwerte auf die Koordinatenfunktionen zu verteilen, liefert folgende **Beobachtung**. *Ist $D \subseteq \mathbb{R}^n$ und $\vec{f} : D \rightarrow \mathbb{R}^m$ oder $\vec{f} : D \rightarrow \mathbb{R}^m$ ein Punkt- oder Vektorfeld, dann ist die Stetigkeit von \vec{f} bzw. \vec{f} (im Punkt $\bar{a} \in D$) äquivalent zur Stetigkeit (im Punkt \bar{a}) der Koordinatenabbildungen $f_i : D \rightarrow \mathbb{R}$ für alle $i = 1, \dots, m$.*

Stetigkeitsuntersuchungen werden dadurch auf (eventuell mehrere) skalare Funktionen reduziert.

1.2.6 Wie erkennt man stetige Skalarfelder in der Praxis?

Normalerweise werden die auftretenden Funktionen durch Formeln beschrieben, die sagen, wie der Funktionswert $f(x_1, \dots, x_n)$ aus den Koordinaten x_1, \dots, x_n berechnet wird. Tauchen in diesen Formeln neben den Variablen nur Summen, Produkte, Differenzen, Quotienten sowie einstellige stetige Funktionen ($\sin, \sqrt{\quad}, \exp, \text{etc}$) auf, so sind die entsprechenden Funktionen überall dort stetig, wo sie definiert sind (keine Nenner Null, nichts Negatives unter der Wurzel, ...). Das liegt daran, daß die einzelnen Koordinaten x_i natürlich(?) stetig von den Punkten (x_1, \dots, x_n) abhängen, die Grundrechenarten stetige Abbildungen $\mathbb{R}^2 \rightarrow \mathbb{R}$ sind (für die Division nur $\mathbb{R}^2 \setminus y\text{-Achse} \rightarrow \mathbb{R}$) und die Hintereinanderausführung stetiger Abbildungen stetig bleibt.

Bei 'gestückelt' definierten Beispielen muß man die Ausnahmepunkte gesondert betrachten. Schauen wir uns etwa die beiden zweistelligen Funktionen

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

an. Die Formeln zeigen, daß beide in allen Punkten $(x, y) \neq (0, 0)$ stetig sind.

Um die Stetigkeit von f im Nullpunkt nachzuweisen, müssen wir einsehen, daß $|f(x, y) - f(0, 0)|$ beliebig klein wird, wenn nur $\sqrt{x^2 + y^2}$ klein genug ist. Nun ist aber stets $|x^2 - y^2| \leq |x^2 + y^2|$ und daher

$$|f(x, y) - f(0, 0)| = \left| y \cdot \frac{x^2 - y^2}{x^2 + y^2} \right| \leq |y| \leq \sqrt{x^2 + y^2}.$$

Dagegen ist g unstetig, denn wie wir oben schon gesehen hatten, existiert kein Grenzwert $\lim_{(x,y) \rightarrow (0,0)} g(x, y)$, der bei Stetigkeit Funktionswert sein müßte.

1.2.7 Beschränktheit und Stetigkeit linearer Abbildungen

Das folgende **Lemma** wird später öfter benutzt.

Zu jeder linearen Abbildung $\vec{L} : \mathbb{R}^n \rightarrow \mathbb{R}^m$ gibt es eine Zahl C , so daß $\|\vec{L}(\vec{v})\| \leq C \|\vec{v}\|$ für alle $\vec{v} \in \mathbb{R}^n$. Speziell sind alle linearen Abbildungen stetig.

Die Existenz der Zahl C ist Äquivalent zur *Beschränktheit* der Abbildungen (Bilder von beschränkten Mengen sind beschränkt) und charakteristisch für lineare Abbildungen **endlichdimensionaler** Räume. In der Quantenmechanik werden Sie das bedauern: die interessanten Operatoren der (unendlichdimensionalen) HILBERTräume sind fast alle unbeschränkt.

Beweis. Wie üblich sollen $\vec{e}_1, \dots, \vec{e}_n$ die kanonischen Basisvektoren von \mathbb{R}^n bezeichnen. Dann ist ein beliebiger Vektor \vec{v} als $v_1 \vec{e}_1 + \dots + v_n \vec{e}_n$ darstellbar. Wegen $|v_i| \leq \|\vec{v}\|$ folgt dann

$$\begin{aligned} \|\vec{L}(\vec{v})\| &= \left\| v_1 \vec{L}(\vec{e}_1) + \dots + v_n \vec{L}(\vec{e}_n) \right\| \\ &\leq |v_1| \|\vec{L}(\vec{e}_1)\| + \dots + |v_n| \|\vec{L}(\vec{e}_n)\| \\ &\leq \|\vec{v}\| \left(\|\vec{L}(\vec{e}_1)\| + \dots + \|\vec{L}(\vec{e}_n)\| \right) \\ &= \|\vec{v}\| \cdot \underbrace{\left(\|\vec{L}(\vec{e}_1)\| + \dots + \|\vec{L}(\vec{e}_n)\| \right)}_{=: C}. \end{aligned}$$

Damit ist die Existenz von C bewiesen. Für den Nachweis der (gleichmäßigen) Stetigkeit sei $\varepsilon > 0$ vorgegeben. Wir setzen $\delta := \frac{\varepsilon}{C}$ und haben $\|\vec{L}(\vec{v}) - \vec{L}(\vec{w})\| = \|\vec{L}(\vec{v} - \vec{w})\| \leq C \|\vec{v} - \vec{w}\| < \varepsilon$, sobald $\|\vec{v} - \vec{w}\| < \delta$.

1.3 Zwischenspiel: das allgemeine Konzept des metrischen Raumes

Sobald in der Mathematik dieselben Überlegungen mehrmals angestellt werden müssen, erfinden die Mathematiker ein abstraktes Konzept, das es (späteren Generationen) erlaubt, nur einmal zu denken und dann oft anzuwenden. Wir kennen das schon von Körpern und Vektorräumen. Ein natürlicher Rahmen für Stetigkeits- und Grenzwertbetrachtungen wird beschrieben durch den Begriff des metrischen Raumes.

1.3.1 Metrik, metrischer Raum

Eine auf einer nicht-leeren Menge X definierte zweistellige reellwertige Funktion d heißt *Abstandsfunktion* oder *Metrik*, falls sie folgende drei Bedingungen erfüllt:

$$(M1) \quad d(x, y) \geq 0 \quad \text{und} \quad d(x, y) = 0 \iff x = y \quad (\text{Positivität})$$

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

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

Eine Menge X mit einer fixierten Metrik heißt *metrischer Raum*. Um auf der sicheren Seite zu sein, definieren die Mathematiker einen metrischen Raum auch gern als Paar (X, d) , wobei X die Menge ist und d die Metrik. Die Elemente von X nennt man bei abstrakter Behandlung *Punkte* des metrischen Raumes. In konkreten Räumen können das Vektoren, Zahlen, Funktionen, etc. sein. Früher war der Buchstabe ρ zur Bezeichnung der Abstandsfunktion beliebt. Ich werde ihn manchmal benutzen, wenn wir einer zweiten Metrik einen Namen geben müssen.

Beispiele. Die drei Axiome (M1)–(M3) sind uns schon einmal begegnet, nämlich als Eigenschaften der auf euklidischen bzw. unitären Räumen definierten Abstandsfunktionen $d(x, y) = \|x - y\|$. Damit haben wir sofort viele Beispiele für metrische Räume.

Speziell sind die im vorigen Abschnitt behandelten Funktionen

$$d(\vec{x}, \vec{y}) := \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad \text{bzw.} \quad d(\vec{x}, \vec{y}) := \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

Metriken auf \mathbb{R}^n bzw. $\vec{\mathbb{R}}^n$. Daneben gibt es auf derselben Menge noch viele andere sinnvolle Abstandsbegriffe, etwa die Summenmetrik oder die Maximummetrik, die den Abstand nach den Formeln

$$\sum_{i=1}^n |x_i - y_i| \quad \text{bzw.} \quad \max(|x_1 - y_1|, \dots, |x_n - y_n|)$$

berechnen. Dieselben Formeln definieren jeweils auch Metriken in \mathbb{C}^n . Für $n = 1$, d.h. in \mathbb{R} bzw. \mathbb{C} ergibt sich in allen drei Fällen die Betragsfunktion $|x - y|$.

Drei analoge Metriken auf der Menge $C([a, b])$ der reell- oder komplexwertigen (die Formeln sind dieselben, nur das Betragszeichen hat eine andere Bedeutung) stetigen Funktionen auf dem abgeschlossenen Intervall $[a, b]$ werden gegeben durch die Vorschriften

$$\sqrt{\int_a^b |f(t) - g(t)|^2 dt}, \quad \int_a^b |f(t) - g(t)| dt \quad \text{und} \quad \sup\{|f(t) - g(t)| : t \in [a, b]\}$$

Die Überprüfung der Einzelheiten bleibt Ihnen überlassen.

Die folgenden eher exotischen **Beispiele** sollen uns davor bewahren, von Metriken Dinge zu erwarten, die sie nicht leisten. Auf der Menge der reellen Zahlen wird durch $d(x, y) = |\arctan x - \arctan y|$ eine Metrik definiert, bezüglich der beliebige zwei Punkte einen Abstand von weniger als π haben.

Auf einer beliebigen Menge $X \neq \emptyset$ können wir durch die Vorschrift $d(x, y) = \begin{cases} 0, & \text{falls } x = y \\ 1, & \text{falls } x \neq y \end{cases}$ eine Metrik definieren. Sie heißt naheliegenderweise 0-1-Metrik.

1.3.2 Konvergenz von Folgen

Im Rest dieses Abschnittes wollen wir den Grenzwert- und Stetigkeitsbegriff von \mathbb{R} auf beliebige metrische Räume übertragen. Das geht ganz 'kanonisch'; es ist nur in den alten Definitionen 'reelle Zahl' durch 'Punkt von X ' und $|x - y|$ durch $d(x, y)$ zu ersetzen. Ich halte die Darstellung entsprechend kurz, rate Ihnen aber, zur Wiederholung die Details selbst zu ergänzen, möglichst ohne dabei auf das Analysis-I-Skript zurückzugreifen.

Definition. Man sagt, daß eine Folge $(x_n)_{n=0}^{\infty}$ von Punkten eines metrischen Raumes gegen den Punkt a konvergiert und schreibt $x_n \rightarrow a$, falls für jedes $\varepsilon > 0$ ein $N \in \mathbb{N}$ derart existiert, daß $d(x_n, a) < \varepsilon$, sobald $n \geq N$.

Dasselbe kann man kurz auch so sagen:

$$x_n \rightarrow a \iff \lim_{n \rightarrow \infty} d(x_n, a) = 0,$$

wobei sich der Grenzwert auf die Zahlenfolge $d(x_n, a)$ bezieht.

1.3.3 Eindeutigkeit des Grenzwertes

Wie in Teil I beweist man, daß eine Folge nicht gleichzeitig gegen zwei verschiedene Punkte konvergieren kann. Das erlaubt wieder, für den eindeutig bestimmten Punkt, gegen den (x_n) gegebenenfalls konvergiert, die Bezeichnung $\lim_{n \rightarrow \infty} x_n$ einzuführen.

1.3.4 Stetige Abbildungen zwischen Teilmengen von metrischen Räumen

Sind (X, d) und (Y, ρ) metrische Räume, $D \subseteq X$, so nennt man eine Abbildung $f : D \rightarrow Y$ *stetig im Punkt* $a \in D$ falls für alle $\varepsilon > 0$ ein $\delta > 0$ derart existiert, daß für alle $x \in D$ mit $d(x, a) < \delta$ gilt $\rho(f(x), f(a)) < \varepsilon$.

Eine Abbildung $D \rightarrow Y$, die in allen Punkten von D stetig ist, nennt man *stetig*.

Praktisch derselbe Beweis wie im vorigen Semester zeigt, daß f *genau dann im Punkt* a *stetig ist, wenn für alle im Raum* X *gegen* a *konvergierenden Folgen* (x_n) *von Punkten aus* D *gilt, daß* $\lim_{n \rightarrow \infty} f(x_n) = f(a)$.

Einfach aber wichtig: *die Hintereinanderausführung zweier stetige Abbildungen liefert wieder eine stetige Abbildung.*

Auch das Konzept der gleichmäßigen Stetigkeit läßt sich leicht verallgemeinern: Sind D, X, Y wie oben, so nennt man $f : D \rightarrow Y$ *gleichmäßig stetig auf* D , wenn für alle $\varepsilon > 0$ ein $\delta > 0$ derart existiert, daß $\rho(f(x_1), f(x_2)) < \varepsilon$ sobald $x_1, x_2 \in D$ die Ungleichung $d(x_1, x_2) < \delta$ erfüllen.

1.3.5 CAUCHY-Folgen und Vollständigkeit

Eine Folge (x_n) von Punkten eines metrischen Raumes heißt **CAUCHY-Folge**, wenn für alle $\varepsilon > 0$ ein $N \in \mathbb{N}$ derart existiert, daß für alle $m, n \geq N$ gilt $d(x_n, x_m) < \varepsilon$.

Der Beweis dafür, daß alle konvergenten Folgen auch CAUCHY-Folgen sind, überträgt sich von den reellen Zahlen sofort auf beliebige metrische Räume. Die Umkehrung allerdings nicht, weil sie im allgemeinen falsch ist. Man nennt einen metrischen Raum *vollständig*, wenn alle CAUCHY-Folgen von Punkten dieses Raumes gegen einen Grenzwert aus diesem Raum konvergieren.

Beispiel. Wir wissen aus dem vorigen Semester, daß \mathbb{R} bezüglich der Betragsmetrik vollständig ist; das ist gewissermaßen Axiom. Betrachtet man die reellen Zahlen aber mit der oben angegebenen Metrik $d(x, y) = |\arctan x - \arctan y|$, so ist die Folge der natürlichen Zahlen eine CAUCHY-Folge, die aber nicht konvergiert (Beweis als **Übungsaufgabe**). Weiter können Sie zeigen, daß bezüglich dieser exotischen Metrik genau die gleichen Zahlenfolgen konvergieren, wie im üblichen Sinn. Das liegt an der Stetigkeit von Tangens und Arcustangens.

Der Raum $C([a, b])$ der auf $[a, b]$ stetigen Funktionen mit der sup-Metrik ist vollständig (das hatten wir im vorigen Semester bewiesen, ohne diese Terminologie zu benutzen (wo?)), jedoch nicht bezüglich der Integralmetrik $d(f, g) := \int_a^b |f - g|$. Eine nicht-konvergente CAUCHY-Folge in $C([-1, 1])$ wird beispielsweise gegeben durch

$$f_n(x) := \begin{cases} -1, & -1 \leq x \leq -\frac{1}{n} \\ nx, & -\frac{1}{n} \leq x \leq \frac{1}{n} \\ 1, & \frac{1}{n} \leq x \leq 1 \end{cases}$$

Malen Sie ein Bild, dann wird klar, was los ist. Ein formaler Beweis bleibt als Übungsaufgabe.

1.3.6 Kugeln

In \mathbb{R} haben wir zur Charakterisierung von Konvergenz und Stetigkeit manchmal ε -Umgebungen benutzt. Die damit verbundene geometrisch angehauchte Sprache ist auch in beliebigen metrischen Räumen nützlich. Als *offene Kugel* mit Mittelpunkt $x \in X$ und Radius $r > 0$ wird die Menge

$$K_x(r) := \{y \in X : d(x, y) < r\}$$

bezeichnet. Als abgeschlossene Kugel bzw. Sphäre bezeichnet man die Mengen

$$\bar{K}_x(r) = \{y \in X : d(x, y) \leq r\} \quad \text{bzw.} \quad S_x(r) = \{y \in X : d(x, y) = r\}.$$

In der Literatur sind auch andere Bezeichnungen für Kugeln verbreitet. Unser $K_x(r)$ könnte z.B. auch $K(x, r)$ oder $B(x, r)$ heißen.

Jede offene Kugel enthält zumindest den Mittelpunkt³. Eventuell aber auch nicht mehr, wie das Beispiel der 0-1-Metrik zeigt. Für diese ist $K_x(r) = \begin{cases} \{x\}, & \text{falls } r \leq 1 \\ X, & \text{falls } r > 1 \end{cases}$

In \mathbb{R} mit der Betragsmetrik sind offenen Kugeln offenen Intervalle, abgeschlossenen Kugeln abgeschlossenen Intervalle und Sphären zweipunktige Mengen:

$$K_x(r) =]x - r, x + r[, \quad \bar{K}_x(r) = [x - r, x + r] \quad \text{und} \quad S_x(r) = \{x \pm r\}.$$

Kugeln in \mathbb{R}^2 können offenen Vollkreise (euklidische Metrik) oder offene Quadrate (Summen- bzw. Maximummetrik) sein. Die entsprechenden Bilder malen Sie selbst. Für andere Metriken können auch andere geometrische Figuren Kugeln sein.

Als wesentliche Konsequenz der Dreiecksungleichung folgt die Disjunktheit zweier Kugeln, sobald die Summe der Radien den Abstand der Mittelpunkte nicht übersteigt (Übungsaufgabe):

$$r + s \leq d(a, b) \quad \implies \quad K_a(r) \cap K_b(s) = \emptyset.$$

Diese Eigenschaft ist der ‘geometrische Hintergrund’ der Eindeutigkeit des Grenzwertes konvergenter Folgen.

Ihre Umkehrung gilt nicht, wie das Beispiel der 0-1-Metrik wieder zeigt. Trotzdem sind auch hier beliebige zwei voneinander verschiedene Punkte Mittelpunkte disjunkter offener Kugeln (welcher?).

1.3.7 Umformulierungen

Mit Hilfe von Kugeln kann man einige der obigen Konzepte anders und eventuell anschaulicher beschreiben. Die Beweise der folgenden Aussagen bleiben als Übungsaufgaben.

Eine Folge konvergiert genau dann gegen den Punkt a , wenn jede offene Kugel mit dem Mittelpunkt a fast alle Folgenglieder enthält.

Die Abbildung $f : X \rightarrow Y$ ist genau dann stetig im Punkt a , wenn zu jeder in Y offenen Kugel L mit Mittelpunkt $f(a)$ eine in X offene Kugel K mit Mittelpunkt a gefunden werden kann, so daß $f[K] \subseteq L$.

1.4 Eigenschaften von Teilmengen metrischer Räume

In diesem Abschnitt ist (X, d) ein fixierter metrischer Raum. Alle vorkommenden Mengen sind Teilmengen von X .

³Wir setzen immer stillschweigend voraus, daß die Radien aller betrachteten Kugeln strikt positiv sind.

1.4.1 Klassifikation von Punkten hinsichtlich einer Menge

Mit Hilfe von offenen Kugeln werden die Konzepte innerer Punkt, äußerer Punkt und Randpunkt einer Menge definiert. Ist $M \subseteq X$ die fragliche Menge, so nennt man x einen *inneren Punkt* von M , falls es eine Kugel um x mit positivem Radius gibt, die ganz in M enthalten ist. Formaler: $K_x(\varepsilon) \subseteq M$ für ein $\varepsilon > 0$.

Gibt es eine analoge Kugel, die keinen Punkt von M enthält, so nennt man x *äußeren Punkt*. Formal $K_x(\varepsilon) \cap M = \emptyset$ für ein geeignetes $\varepsilon > 0$. Alle anderen Punkte sind Randpunkte der Menge M . Diese können zu M gehören, müssen aber nicht. Offensichtlich ist jeder Punkt genau eines: innerer Punkt, äußerer Punkt oder Randpunkt einer gegebenen Menge M .

In \mathbb{R}^n kann man von einem inneren Punkt \bar{a} einer Menge in Richtung jedes Einheitsvektors \vec{e} ein Stückchen gehen, ohne die Menge zu verlassen: $\bar{a} + t\vec{e} \in M$ für alle $t < \varepsilon$. Tatsächlich gibt es sogar ein $\varepsilon > 0$, das in jeder Richtung funktioniert.

1.4.2 Umgebungen

eines Punktes x nennt man alle Mengen U , von denen x innerer Punkt ist; anders gesagt: U ist Umgebung von x falls es ein $\varepsilon > 0$ gibt, so daß $K_x(\varepsilon) \subseteq U$. Die Kugel selbst werden wir manchmal ε -Umgebung nennen.

1.4.3 Offene und abgeschlossene Mengen

Eine Menge heißt *offen*, wenn jeder ihrer Punkte innerer Punkt ist. Sie heißt *abgeschlossen*, wenn jeder nicht zu ihr gehörende Punkt äußerer Punkt ist. Anders gesagt, alle Randpunkte müssen auch zur abgeschlossenen Menge M gehören.

Offensichtlich ist eine Menge A genau dann offen/abgeschlossen, wenn ihr Komplement $X \setminus A$ abgeschlossen/offen ist. Achtung: die meisten Mengen sind weder noch.

Die folgende Selbstverständlichkeit ist keine:

Offene/abgeschlossene Kugeln sind offene/abgeschlossene Mengen. Sphären sind abgeschlossen.

Ich zeige nur, daß $\bar{K}_x(r)$ abgeschlossen ist. Dazu muß zu einem $y \notin \bar{K}_x(r)$ eine offene Kugel $K_y(\varepsilon)$ gefunden werde, für die

$$\bar{K}_x(r) \cap K_y(\varepsilon) = \emptyset.$$

Da der Mittelpunkt y festliegt, kommt es nur auf den Radius ε an. Wir wählen ihn so, daß $0 < \varepsilon < d(y, x) - r$ (Warum geht das?).

Wäre dann, im Widerspruch zu unserem angepeilten Ziel, ein Punkt z im Durchschnitt der beiden Kugeln $\bar{K}_x(r)$ und $K_y(\varepsilon)$, so ergäbe sich nach der Dreiecksungleichung

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

Das geht nicht.

1.4.4 Was es mit der ‘Abgeschlossenheit’ auf sich hat

Satz. *Eine Teilmenge $A \subseteq X$ ist genau dann abgeschlossen, wenn der Grenzwert jeder (in X) konvergenten Folge von Punkten aus A wieder zu A gehört.*

Beweis. Angenommen A ist abgeschlossen und $x_n \rightarrow y$ für eine Folge von Punkten aus A . Wäre $y \notin A$, so wäre y äußerer Punkt und es gäbe eine Kugel $K_y(\varepsilon)$, die A nicht schneidet. Diese Kugel enthielte keine Glieder der Folge, wo sie doch wegen $x_n \rightarrow y$ fast alle enthalten müßte. Der Widerspruch zeigt, daß $y \notin A$ unmöglich ist.

Für die andere Richtung setzen wir voraus, daß A unter Grenzwertbildung abgeschlossen ist, und zeigen, daß alle nicht zu A gehörenden Punkte äußere Punkte sind. Sei also $y \notin A$. Wir müssen eine offene Kugel $K_y(r)$ finden, die keine Punkte aus A enthält. Da der Mittelpunkt der gesuchten Kugel festliegt, kommt es nur auf den Radius an. Für $n = 1, 2, 3, \dots$ probieren wir die

Radien $\frac{1}{n}$. Entweder $K_y(\frac{1}{n})$ enthält keinen Punkt aus A , dann sind wir fertig. Oder wir können ein $x_n \in A \cap K_y(\frac{1}{n})$ wählen. Wenn kein n funktionieren würde, entstünde eine unendliche Folge $(x_n)_{n=1}^\infty$ von Punkten aus A , so daß $d(x_n, y) < \frac{1}{n}$. Also konvergiert diese Folge gegen y . Wegen der vorausgesetzten Abgeschlossenheit gegen Grenzwerte wäre $y \in A$. Da das nicht sein kann, muß die Konstruktion der x_n abbrechen und ein passender Radius existieren.

1.4.5 Der Abschluß einer Menge

Ist $A \subseteq X$, so nennt man die kleinste abgeschlossene Teilmenge von X , die A enthält den *Abschluß von A* oder auch die *abgeschlossene Hülle von A* . Populäre Bezeichnungen für den Abschluß sind \bar{A} oder $cl(A)$ auch $Cl(A)$.

Jede Teilmenge eines metrischen Raumes besitzt einen Abschluß. Er entsteht, indem man zu A alle Randpunkte hinzunimmt.

Alternativ entsteht \bar{A} indem man zu A die Grenzwerte aller in X konvergenten Folgen mit Gliedern aus A hinzunimmt.

1.4.6 Beschränktheit

Eine Teilmenge von X heißt *beschränkt*, wenn sie in einer (genügend großen) Kugel enthalten ist. In \mathbb{R}^n ist das dazu äquivalent, daß es eine feste Zahl R gibt, die die Beträge aller Koordinaten aller Punkte der Menge übertrifft.

Man zeigt leicht, daß die Vereinigung zweier (und dann auch endlich vieler) beschränkter Mengen wieder beschränkt ist. Auch sind Teilmengen beschränkter Mengen wieder beschränkt.

1.4.7 Zusammenhang

Eine Teilmenge $M \subseteq X$ heißt *zusammenhängend*⁴, wenn es für je zwei Punkte $a, b \in M$ eine stetige 'Kurve' $\gamma : [0, 1] \rightarrow M$ derart gibt, daß $\gamma(0) = a$ und $\gamma(1) = b$.

Wichtig ist, daß die Kurve ganz innerhalb von M verläuft und nicht nur im Raum X .

Der Hauptgrund für die Nützlichkeit des Begriffes ist der **Zwischenwertsatz**. *Jede stetige auf der zusammenhängenden Menge M definierte reellwertige Funktion nimmt mit je zwei Werten auch alle Zwischenwerte an. Kurz: Ist $D \subseteq X$ zusammenhängend und $f : D \rightarrow \mathbb{R}$ stetig, so ist $f[D]$ ein Intervall.*

Beweis. Seien $a, b \in D$ und $f(a) \leq y \leq f(b)$. Wir suchen ein $x \in D$ mit $f(x) = y$. Wegen des Zusammenhangs gibt es eine stetige Kurve $\gamma : [0, 1] \rightarrow D$ mit $a = \gamma(0)$ und $b = \gamma(1)$. Dann ist $f \circ \gamma : [0, 1] \rightarrow \mathbb{R}$ stetig, also gibt es nach dem Zwischenwertsatz aus dem letzten Semester ein $t \in [0, 1]$ mit $f(\gamma(t)) = y$. Da die Kurve ganz in D verläuft wird y von f auf D angenommen.

Das typische Beispiel einer nicht zusammenhängenden Menge wird beschrieben durch das folgende **Lemma**. *Sind A und B nichtleere und offenen Teilmengen von X , die sich nicht überschneiden, so ist $A \cup B$ nicht zusammenhängend.*

Beweis. Wir definieren eine Funktion $f : A \cup B \rightarrow \mathbb{R}$ durch die Festlegung $f(a) = 0$, falls $a \in A$ und $f(b) = 1$, falls $b \in B$. Da sich A und B nicht überschneiden, ist das eine korrekte Definition. Diese Funktion ist stetig. Um das einzusehen, sei $x \in A \cup B$ beliebig und $\varepsilon > 0$ vorgegeben. Gesucht ist $\delta > 0$, so daß $|f(x) - f(x')| < \varepsilon$, sobald $d(x, x') < \delta$. Sei zunächst $x \in A$. Da diese Menge offen ist, gibt es $\delta > 0$ so daß $K_x(\delta) \subseteq A$. Dann liegen aber alle x' mit $d(x, x') < \delta$ in A und daher ist $|f(x) - f(x')| = |0 - 0| = 0 < \varepsilon$, wie verlangt. Der Fall $x \in B$ geht analog. Da der Zwischenwert $\frac{1}{2}$ von der stetigen Funktion f nicht angenommen wird, kann der Definitionsbereich von f nicht zusammenhängend sein.

⁴Es gibt auch andere Zusammenhangsbegriffe, die bei uns jedoch nicht vorkommen werden. Korrekt müßte man von *Wegzusammenhang* sprechen.

1.5 Kompaktheit

In Analysis I hatten wir unter anderem bewiesen, daß stetige Funktionen so richtig gut erst auf abgeschlossenen Intervallen sind. Sie nehmen dann einen größten und einen kleinsten Funktionswert an und sind gleichmäßig stetig.

Statt ‘abgeschlossenes Intervall’ hatten wir manchmal auch ‘kompaktes Intervall’ gesagt. Hier will ich nun ein allgemeineres Konzept von Kompaktheit liefern und zeigen, daß zumindest die beiden genannten angenehmen Eigenschaften stetiger Funktionen sich tatsächlich an dieser Kompaktheit festmachen lassen. Wir befinden uns immer noch in einem festen metrischen Raum (X, d) .

1.5.1 Definition

Eine Teilmenge C eines metrischen Raumes heißt *kompakt*⁵, wenn jede Folge $(c_m)_{m=0}^\infty$ von Elementen aus C eine Teilfolge enthält, die gegen einen Grenzwert aus C konvergiert. Mit anderen Worten: Jede Folge aus C hat einen Häufungspunkt in C .

Erinnerung. Teilfolgen entstehen aus Folgen, indem Glieder gestrichen werden (eventuell auch unendlich viele) und die restlichen unendlich vielen Glieder in ihrer alten Reihenfolge zusammenrücken. Formal schreibt man eine Teilfolge als $(c_{m_k})_{k=0}^\infty$, wobei stillschweigend $m_0 < m_1 < m_2 < \dots$ vorausgesetzt wird. ‘Ausdünnen’ ist ein ganz suggestives Wort, um den Übergang zu Teilfolgen zu bezeichnen. Häufungspunkte waren einfach als Grenzwerte von Teilfolgen definiert. Es sollte klar sein, daß Teilfolgen von Teilfolgen wieder Teilfolgen sind. Für Zahlenfolgen hatten wir im letzten Semester (wo sie das alles noch einmal nachschlagen sollten) schon festgestellt, daß jede Teilfolge einer konvergenten Folge selbst konvergiert und denselben Grenzwert wie die ganze Folge hat. Das überträgt sich sofort auf Folgen in beliebigen metrischen Räumen.

1.5.2 Satz vom Maximum.

Ist C eine kompakte Teilmenge eines metrischen Raumes und $f : C \rightarrow \mathbb{R}$ stetig, so nimmt f einen größten und einen kleinsten Funktionswert an.

Der **Beweis** aus dem vorigen Semester überträgt sich ohne Schwierigkeiten von abgeschlossenen Intervallen auf beliebige kompakte Mengen. Ich schreibe ihn trotzdem noch einmal auf. Wir werden nur beweisen, daß das Supremum angenommen wird, und beginnen mit einer

Vorüberlegung/Erinnerung. *Ist $Y \neq \emptyset$ eine nach oben beschränkte (bzw. unbeschränkte) Teilmenge von \mathbb{R} , so gibt es eine Folge (y_k) von Elementen aus Y , die gegen $\sup Y$ konvergiert (bzw. bestimmt gegen ∞ divergiert).*

Ist nämlich Y beschränkt, also $\sup Y$ endlich, so kann man für jedes $k \geq 1$ ein $y_k \in Y$ finden, mit $\sup Y - \frac{1}{k} < y_k \leq \sup Y$. Ist Y dagegen unbeschränkt, so gibt es $y_k \in Y$ mit $y_k \geq k$. Die so gewählten Folgen sind wie verlangt.

Nun zum **Beweis der Annahme des Supremums**. Sei $s = \sup\{f(c) : c \in C\}$. Dann liefert die Vorüberlegung eine Folge y_k von Funktionswerten, die gegen s konvergiert (bzw. bestimmt divergiert): $\lim y_k = s$.

Seiner Herkunft nach läßt sich jedes y_k als $y_k = f(c_k)$ mit $c_k \in C$ schreiben. Wegen der Kompaktheit besitzt (c_k) eine konvergente Teilfolge etwa $\lim_{m \rightarrow \infty} c_{k_m} = p \in C$. Die entsprechende Folge der Funktionswerte $(f(c_{k_m}))_{m=0}^\infty$ konvergiert dann wegen der Stetigkeit gegen $f(p)$ (kann also nicht bestimmt divergieren). Andererseits hat (y_{k_m}) als Teilfolge denselben Grenzwert wie die ganze Folge (y_k) , nämlich s . Folglich $f(p) = s$, und das Maximum ist gefunden.

1.5.3 Der Satz über die gleichmäßige Stetigkeit

(X, d) und (Y, ρ) seien metrische Räume, $C \subseteq X$ sei kompakt und $f : C \rightarrow Y$ eine stetige Abbildung. Dann gibt es zu jedem $\varepsilon > 0$ ein $\delta > 0$, mit $\rho(f(a), f(b)) < \varepsilon$ sobald $a, b \in C$ einen Abstand $< \delta$ voneinander haben.

⁵Wenn man es genau nimmt ‘folgenkompakt’, aber andere Kompaktheitsbegriffe werden bei uns nicht vorkommen.

Beweis. Sei $\varepsilon > 0$ vorgegeben. Wenn es kein passendes δ gäbe, so sind speziell $\delta = 1, \frac{1}{2}, \frac{1}{3}, \dots, \frac{1}{m}, \dots$ unbrauchbar. Das heißt, daß es jeweils Punkte $a_m, b_m \in C$ geben muß für die $\rho(f(a_m), f(b_m)) \geq \varepsilon$ ist, obwohl $d(a_m, b_m) < \frac{1}{m}$. Wegen der Kompaktheit von C gibt es eine konvergente Teilfolge (a_{m_k}) mit Grenzwert $a \in C$. Aus

$$d(b_{m_k}, a) \leq d(b_{m_k}, a_{m_k}) + d(a_{m_k}, a) < \frac{1}{m_k} + d(a_{m_k}, a) \rightarrow 0$$

folgt, daß auch $b_{m_k} \rightarrow a$. Geht man nun in der Ungleichung

$$\varepsilon \leq \rho(f(a_{m_k}), f(b_{m_k})) \leq \rho(f(a_{m_k}), f(a)) + \rho(f(b_{m_k}), f(a))$$

für $k \rightarrow \infty$ zum Grenzwert über, so ergibt sich wegen der Stetigkeit von f , daß $\varepsilon \leq 0$. Dieser Widerspruch zeigt, daß eine der Zahlen $\frac{1}{m}$ die Rolle des gesuchten δ spielen kann.

1.5.4 Abgeleitete Eigenschaften kompakter Mengen

- (1) *Kompakte Teilmengen metrischer Räume sind abgeschlossen.*
- (2) *Kompakte Teilmengen metrischer Räume sind beschränkt.*
- (3) *Kompaktheit vererbt sich auf abgeschlossenen Teilmengen.*

Beweis. Sei $C \subseteq X$ kompakt. Um die Abgeschlossenheit nachzuweisen, betrachten wir eine konvergente Folge (c_m) von Punkten aus C . Wir müssen zeigen, daß ihr Grenzwert, etwa a auch zu C gehört. Wegen der Kompaktheit gibt es eine Teilfolge (c_{m_k}) , deren Grenzwert zu C gehört. Dieser Grenzwert muß aber mit dem Grenzwert der ganzen Folge übereinstimmen: $a = \lim_m(c_m) = \lim_k(c_{m_k}) \in C$.

Um die Beschränktheit nachzuweisen, fixieren wir einen Punkt $a \in X$ und stellen zunächst fest, daß die reellwertige Funktion $x \mapsto d(a, x)$ auf X (gleichmäßig) stetig ist. Das ist auch in andern Argumenten nützlich und folgt sofort aus der Dreiecksungleichung

$$\left. \begin{array}{l} d(a, y) \leq d(a, x) + d(x, y) \\ d(a, x) \leq d(a, y) + d(y, x) \end{array} \right\} \implies |d(a, x) - d(a, y)| \leq d(x, y).$$

Nach dem Satz über die Annahme des Maximums, gibt es auf C einen größten Funktionswert dieser Funktion, etwa R . Dann ist also $d(a, c) \leq R$ für alle c in C , bzw. $C \subseteq \bar{K}_a(R)$.

Sei schließlich $A \subseteq C$ eine abgeschlossene Teilmenge. Um die Kompaktheit von A einzusehen, betrachten wir eine beliebige Folge (a_n) von Punkten aus A . Klar gehören alle Folgenglieder auch zu C . Daher gibt es eine konvergente Teilfolge (a_{n_k}) mit Grenzwert in C . Da aber A abgeschlossen ist und die Teilfolge nur Glieder aus A enthält, gehört der Grenzwert dieser Teilfolge sogar zu A .

1.6 Eigenschaften von Teilmengen von \mathbb{R}^n

Jetzt kehren wir zurück in den n -dimensionalen Raum \mathbb{R}^n bzw. $\vec{\mathbb{R}}^n$ mit der euklidischen Metrik. Alle eben definierten Konzepte bleiben in Kraft. Im Gegensatz zur abstrakten Situation des metrischen Raumes gibt es aber hier noch eine 'Geometrie' (Strecken, Richtungen, etc.), mit deren Hilfe weitere wichtige Eigenschaften von Mengen beschrieben werden können. Außerdem wollen wir einige der oben diskutierten Begriffe noch etwas detaillierter ansehen.

1.6.1 Konvexität

Man nennt eine Teilmenge $A \subseteq \mathbb{R}^n$ *konvex*, wenn sie mit je zwei Punkten \bar{a} und \bar{b} die ganze Strecke von \bar{a} nach \bar{b} enthält, also die Menge

$$\{\bar{a} + t(\bar{b} - \bar{a}) : 0 \leq t \leq 1\}.$$

Fakt. *Offene und abgeschlossene Kugeln sind konvex.*

Um das etwa für offene Kugeln einzusehen, betrachten wir \bar{a} und \bar{b} in $K_{\bar{c}}(r)$. Sei $0 \leq t \leq 1$ und $\bar{x} = \bar{a} + t(\bar{b} - \bar{a})$. Wir müssen $\bar{x} \in K_{\bar{c}}(r)$, also $\|\bar{x} - \bar{c}\| < r$ nachweisen. Das geht mit ein bißchen Rechnen⁶

$$\begin{aligned} \|\bar{x} - \bar{c}\| &= \|\bar{a} + t(\bar{b} - \bar{a}) - \bar{c}\| = \|(1-t)(\bar{a} - \bar{c}) + t(\bar{b} - \bar{c})\| \\ &\leq (1-t)\|\bar{a} - \bar{c}\| + t\|\bar{b} - \bar{c}\| < (1-t)r + tr = r. \end{aligned}$$

Wo wurde in diesem Argument $0 \leq t \leq 1$ benutzt?

1.6.2 Umwege sind möglich

Gehört \bar{a} zur offenen Kugel $K_{\bar{c}}(r)$ so ist nicht nur die Strecke vom Mittelpunkt \bar{c} nach \bar{a} ganz in der Kugel, sondern auch alle *achsenparallelen Umwege* von \bar{c} nach \bar{a} verlaufen innerhalb der Kugel. In \mathbb{R}^3 führt ein solcher Weg zum Beispiel

$$\text{von } \bar{c} = (c_1, c_2, c_3) \text{ über } (c_1, a_2, c_3) \text{ und } (c_1, a_2, a_3) \text{ nach } (a_1, a_2, a_3) = \bar{a}.$$

Die Eckpunkte eines solchen Streckenzuges haben einige Koordinaten c_i und die anderen Koordinaten a_j . Deshalb liegen Sie alle in der Kugel. Wegen deren Konvexität gilt das dann auch für alle Verbindungsstrecken.

1.6.3 Polygonaler Zusammenhang

Konvexe Mengen sind offensichtlich zusammenhängend. Die Umkehrung gilt nicht, wie das Beispiel von (mindestens eindimensionalen) Sphären zeigt. Zwischen Konvexität und Zusammenhang ist folgender Begriff angesiedelt.

Eine Menge M wird *polygonal zusammenhängend* genannt, wenn es zu je zwei Punkten \bar{a}, \bar{b} aus M einen Polygonzug gibt, der sie verbindet und ganz in M verläuft. Exakter: wenn es eine endliche Folge $\bar{c}_0, \bar{c}_1, \bar{c}_2, \dots, \bar{c}_k$ von Punkten aus M derart gibt, daß $\bar{a} = \bar{c}_0$, $\bar{b} = \bar{c}_k$ und jede der k Strecken von \bar{c}_i nach \bar{c}_{i+1} ganz innerhalb M verläuft. Aus formalen Gründen erlauben wir auch einpunktige Strecken (z. B. damit jede einpunktige Menge polygonal zusammenhängend wird).

Um einzusehen, daß polygonal zusammenhängende Mengen zusammenhängend sind, braucht man nur eine stetige Kurve anzugeben, die den Polygonzug durchläuft. Hier wäre die Formel:

$$\bar{\gamma}(t) = \begin{cases} \bar{c}_0 + kt(\bar{c}_1 - \bar{c}_0), & \text{für } 0 \leq t \leq \frac{1}{k} \\ \bar{c}_1 + (kt - 1)(\bar{c}_2 - \bar{c}_1), & \text{für } \frac{1}{k} \leq t \leq \frac{2}{k} \\ \vdots & \vdots \\ \bar{c}_{k-1} + (kt - k + 1)(\bar{c}_k - \bar{c}_{k-1}), & \text{für } \frac{k-1}{k} \leq t \leq 1 \end{cases}$$

Die Umkehrung gilt nicht, wie die Sphären wieder zeigen. Aber

Satz. *Offene zusammenhängende Teilmengen von \mathbb{R}^n sind polygonal zusammenhängend.*

Im **Beweis** benutzen wir das Lemma aus 1.4.7. Sei M die gegebene offene und zusammenhängende Menge und sei \bar{a} ein beliebiger Punkt aus M . Wir setzen

$$A := \{\bar{x} \in M : \bar{a} \text{ und } \bar{x} \text{ lassen sich in } M \text{ durch einen Polygonzug verbinden}\} \text{ und } B := M \setminus A.$$

A ist offen. Sei nämlich $\bar{x} \in A$. Dann ist speziell $\bar{x} \in M$ und, weil M offen ist, $K_{\bar{x}}(\varepsilon) \subseteq M$ für ein passendes $\varepsilon > 0$. Dann ist aber die ganze Kugel $K_{\bar{x}}(\varepsilon)$ auch in A enthalten, denn man kommt von \bar{a} zu \bar{x} mit einem Polygonzug und von \bar{x} aus erreicht man mit einer weiteren Strecke jeden Punkt der Kugel.

Aus einem ähnlichen Grund ist auch die Menge B offen. Ist $\bar{x} \in B \subseteq M$, so gibt es $K_{\bar{x}}(\varepsilon) \subseteq M$ und kein Punkt dieser Kugel kann von \bar{a} aus polygonal erreichbar sein, weil sonst auch \bar{x} erreichbar wäre. Mit anderen Worten, $K_{\bar{x}}(\varepsilon) \subseteq B$.

Gäbe es nun einen Punkt $\bar{b} \in M$, der sich nicht mit \bar{a} durch einen Polygonzug verbinden ließe, so hätten wir $\bar{a} \in A$ und $\bar{b} \in B$. Beide Mengen wären also nicht leer. Beide sind offen und

⁶Wobei ich eine der nötigen Regeln der ‘Punkt-Vektor-Rechnung’ nicht explizit begründet habe; sie werden das trotzdem verstehen.

überschneiden sich nicht. Nach dem Lemma, wäre dann $A \cup B = M$ nicht zusammenhängend, Widerspruch.

Mit einer kleinen Zusatzüberlegung auf der Basis von 1.6.2 sieht man, daß je zwei Punkte einer offenen zusammenhängenden Teilmenge von \mathbb{R}^n durch einen Streckenzug verbunden werden können, deren Teilstrecken alle achsenparallel sind.

1.6.4 Quader

sind Mengen der Art $[a_1, b_1] \times [a_2, b_2] \times \dots \times [a_n, b_n]$

$$= \{ \bar{x} = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n : a_1 \leq x_1 \leq b_1, a_2 \leq x_2 \leq b_2, \dots, a_n \leq x_n \leq b_n \}.$$

Wir werden diesen Quader als $[\bar{a}, \bar{b}]$ notieren und die Version ohne Randpunkte (also alle Ungleichungen strikt) als $] \bar{a}, \bar{b} [$.

1.6.5 Charakterisierung der kompakten Teilmengen von \mathbb{R}^n

Beobachtung. Jeder Quader

$$[\bar{a}, \bar{b}] = \{ \bar{c} \in \mathbb{R}^n : a_1 \leq c_1 \leq b_1, a_2 \leq c_2 \leq b_2, \dots, a_n \leq c_n \leq b_n \}$$

ist eine kompakte Teilmenge von \mathbb{R}^n .

Ich beweise nur die Fälle $n = 1, 2$ und deute an, wie es weitergeht. Der eindimensionale Fall behauptet die Kompaktheit von abgeschlossenen Intervallen. Nach dem Satz von BOLZANO-WEIERSTRASS hat jede in $[a, b]$ enthaltene Folge (c_m) eine in \mathbb{R} konvergente Teilfolge (c_{m_k}) . Da man aber in den Ungleichungen $a \leq c_{m_k} \leq b$ zum Grenzwert übergehen darf, liegt dieser auch in $[a, b]$, wie verlangt.

Für das Rechteck haben wir eine Folge $(\bar{c}_m) = (c_m^1, c_m^2)_{m=0}^\infty$ zu betrachten, wobei $a_1 \leq c_m^1 \leq b_1$ und $a_2 \leq c_m^2 \leq b_2$ für alle m . Wie gerade gesehen, läßt sich die Zahlenfolge (c_m^1) so ausdünnen, daß $(c_{m_k}^1)$ einen Grenzwert zwischen a_1 und b_1 hat. Dasselbe Argument läßt sich auf $(c_{m_k}^2)$ anwenden und liefert eine Teilteilstfolge $(c_{m_{k_l}}^2)_{l=0}^\infty$ die einen Grenzwert zwischen a_2 und b_2 hat. Klar konvergiert dann die Punktfolge $(\bar{c}_{m_{k_l}})_{l=0}^\infty = (c_{m_{k_l}}^1, c_{m_{k_l}}^2)_{l=0}^\infty$ gegen einen Punkt aus dem Rechteck $[(a_1, a_2), (b_1, b_2)]$.

Im dreidimensionalen Fall muß man entsprechend dreimal ausdünnen. Das wird typographisch schon schwierig. Die allgemeine Aussage kann man induktiv nach der Dimension beweisen.

Charakterisierungssatz. Eine Teilmenge $C \subseteq \mathbb{R}^n$ ist genau dann kompakt, wenn sie abgeschlossen und beschränkt ist. Daraus folgt unter anderem, daß abgeschlossene Kugeln und Sphären kompakt sind.

Eine Richtung hatten wir schon für beliebige metrische Räume bewiesen. Daher bleibt nur die Umkehrung einzusehen. Angenommen $C \subseteq \mathbb{R}^n$ ist abgeschlossen und beschränkt. Dann können wir C in einen großen Quader einschließen: $C \subseteq [\bar{a}, \bar{b}]$. Dessen Kompaktheit überträgt sich dann nach 1.5.4 auf die abgeschlossene Teilmenge C .

2 Kurven

Sowohl in der Umgangssprache als auch in der (Schul)Geometrie sind Kurven gewisse Punktmen- gen, die auf verschiedene Arten beschrieben werden können. Ein Kreis ist z.B.

- (1) die Menge (in der Sprache der alten Griechen: der geometrische Ort) aller Punkte einer Ebene, die von einem festen Punkt O einen festen Abstand r haben.

Nach Fixierung eines rechtwinkligen Koordinatensystems mit O als Ursprung ist ein Kreis auch

- (2) die Menge aller Punkte, deren Koordinaten die Gleichung $x^2 + y^2 = r^2$ erfüllen, bzw.

- (3) die Menge der Punkte mit den Koordinaten $(r \cos t, r \sin t)$, wenn t das Intervall $[0, 2\pi]$ (oder ein längeres) durchläuft.

Es gibt verschiedene Arten zu präzisieren, welche Eigenschaften eine Punktmenge haben muß, um als Kurve (im Gegensatz etwa zu einer Fläche) durchzugehen. Dafür haben wir leider keine Zeit. Wir wollen Analysis betreiben und werden deshalb nur solche Kurven betrachten, die sich parametrisieren also durch Funktionen beschreiben lassen. Um allen für uns nebensächlichen Schwierigkeiten von vornherein aus dem Weg zu gehen, benutzen wir das Wort Kurve zumindest offiziell⁷ nur für Parametrisierungen. Damit wir sie typographisch sofort als solche erkennen, werden derartige Kurven normalerweise mit kleinen griechischen Buchstaben bezeichnet, vorzugsweise $\bar{\gamma}$. Der Querstrich drückt aus, daß wir uns im Punktraum bewegen.

Manchmal kommen auch einstellige Funktionen $I \rightarrow \mathbb{R}^n$ vor. Sie verhalten sich im Prinzip wie Kurven, nur daß ihre Koordinatenfunktionen untereinander stehen. Alles, was wir in diesem Abschnitt beweisen werden, überträgt sich sinngemäß. Wir sprechen zur Betonung und Unterscheidung von ‘Kurven im Vektorraum’ oder einstelligen Vektorfunktionen.

2.1 Begriffe

Definition. Eine *Kurve* ist eine **stetige** Abbildung $\bar{\gamma} = (\gamma_1, \gamma_2, \dots, \gamma_n) : I \rightarrow \mathbb{R}^n$ eines in \mathbb{R} enthaltenen Intervalls in den n -dimensionalen Punktraum. Die von der Kurve durchlaufene Punktmenge $\bar{\gamma}[I] = \{\bar{\gamma}(t) : t \in I\}$ werden wir zur Unterscheidung die *Bahn* von $\bar{\gamma}$ nennen.

Oft ist es günstig, sich die Kurve als ‘Bahnkurve’ eines sich bewegenden Punktes vorzustellen. Der Definitionsbereich $I \subset \mathbb{R}$ ist dann ein Zeitintervall und die Variable wird meist als t geschrieben. Ich will betonen, daß die *Stetigkeit* Bestandteil der Kurvendefinition ist und auch, daß die Bewegung in einem (eventuell freilich unbeschränkten) *Zeitintervall* verläuft. Dagegen wird nicht verlangt, daß sich wirklich etwas bewegt: auch eine konstante Funktion ist eine Kurve in unserem Sinne.

Noch einige weitere **Vokabeln**. Ist $I = [a, b]$ ein abgeschlossenes Intervall, so nennt man $\bar{\gamma}(a)$ den *Anfangs-* und $\bar{\gamma}(b)$ den *Endpunkt* der Kurve. Stimmen beide überein, nennt man die Kurve *geschlossen*. Man spricht von einer *JORDAN-Kurve*, wenn es außer eventuell Anfangs- und Endpunkt keine Doppelpunkte gibt (kein Punkt zweimal durchlaufen wird, $\bar{\gamma}$ injektiv ist).

Sind $\bar{\alpha} : [a, b] \rightarrow \mathbb{R}^n$ und $\bar{\beta} : [c, d] \rightarrow \mathbb{R}^n$ zwei Kurven, bei denen der Endpunkt von $\bar{\alpha}$ gleich dem Anfangspunkt von $\bar{\beta}$ ist: $\bar{\alpha}(b) = \bar{\beta}(c)$, dann kann man sie hintereinander durchlaufen. Offiziell nennen wir das *Zusammenstücken* und bezeichnen die neue Kurve mit $\bar{\alpha}\bar{\beta} : [a, b + d - c] \rightarrow \mathbb{R}^n$. Sie wird definiert (prüfen sie Korrektheit und Stetigkeit) durch

$$\bar{\alpha}\bar{\beta}(t) = \begin{cases} \alpha(t), & a \leq t \leq b \\ \beta(c + t - b), & b \leq t \leq b + d - c. \end{cases}$$

Die in 1.6.3 schon benutzten Polygonzüge sind z.B. die Kurven, die entstehen, wenn endlich viele Strecken zusammengestückt werden.

Durchläuft man die Kurve $\bar{\gamma}$ im gleichen Zeitintervall $[a, b]$ rückwärts, so nennt man das Resultat die inverse Kurve $\bar{\gamma}^-$. Als Formel: $\bar{\gamma}^-(t) = \bar{\gamma}(a + b - t)$.

2.2 Die Ableitung einer Kurve; Differentiation und Integration von Vektorfunktionen

2.2.1 Der Geschwindigkeitsvektor

Bereits in der Schule lernt man die Augenblicksgeschwindigkeit einer durch $\bar{\gamma}$ beschriebenen Bewegung zum Zeitpunkt t kennen als

$$\lim_{h \rightarrow 0} \frac{\bar{\gamma}(t+h) - \bar{\gamma}(t)}{h}.$$

⁷Wenn ich später von einfachen Kurven wie Strecken, Geraden, Polygonzügen oder Kreisen rede, ohne eine Parametrisierung anzugeben, so ist diese doch immer mitgemeint und (von Ihnen) bei Bedarf ohne Schwierigkeit zu ergänzen.

Die Physiker schreiben gern einen Punkt für die Ableitung nach der Zeit: $\dot{\bar{\gamma}}(t)$. Wir werden das alternativ zu $\bar{\gamma}'(t)$ auch häufig tun, wenn der Parameter t heißt. Manchmal ist auch $\frac{d}{dt}\bar{\gamma}(t)$ eine praktische Notation. Wie auch immer dieses Objekt bezeichnet wird: **in jedem Fall ist die Geschwindigkeit ein Vektor** (denn die Differenz der beiden Punkte $\bar{\gamma}(t+h) - \bar{\gamma}(t)$ ist ein Vektor), wie es sich gehört.

Koordinatenweise aufgeschrieben, ergibt sich, weil man den Grenzwert ‘reinziehen’ darf

$$\frac{d}{dt}\bar{\gamma}(t) = \lim_{h \rightarrow 0} \begin{pmatrix} \frac{\gamma_1(t+h) - \gamma_1(t)}{h} \\ \frac{\gamma_2(t+h) - \gamma_2(t)}{h} \\ \vdots \\ \frac{\gamma_n(t+h) - \gamma_n(t)}{h} \end{pmatrix} = \begin{pmatrix} \frac{d}{dt}\gamma_1(t) \\ \frac{d}{dt}\gamma_2(t) \\ \vdots \\ \frac{d}{dt}\gamma_n(t) \end{pmatrix} \quad \text{oder} \quad \dot{\bar{\gamma}}(t) = \begin{pmatrix} \dot{\gamma}_1(t) \\ \dot{\gamma}_2(t) \\ \vdots \\ \dot{\gamma}_n(t) \end{pmatrix}.$$

Aus der koordinatenweisen Anwendung der üblichen Ableitung läßt sich alles leicht herleiten, was man allgemein über Geschwindigkeiten von Kurven wissen muß. Z.B. die Kettenregel in der Form $\frac{d}{dt}\bar{\gamma}(\varphi(t)) = \dot{\varphi}(t) \dot{\bar{\gamma}}(\varphi(t))$ ($\dot{\varphi}$ steht vorn, weil nach unserer Konvention Skalare immer vor Vektoren stehen).

2.2.2 Kurventangenten

Die Differenzenquotienten $\frac{\bar{\gamma}(t+h) - \bar{\gamma}(t)}{h}$ sind offensichtlich Richtungsvektoren der Sekanten durch die Kurvenpunkte $\bar{\gamma}(t)$ und $\bar{\gamma}(t+h)$ und im Grenzfall gehen die Sekanten in die Tangente über. Um nicht geometrisch erklären zu müssen, was Tangenten sein sollen, definieren wir einfach: Die Gerade durch den Punkt $\bar{\gamma}(t_0)$ in Richtung $\dot{\bar{\gamma}}(t_0) \neq \vec{0}$ heißt Tangente. Als Kurve wird sie durch $t \mapsto \bar{\gamma}(t_0) + (t - t_0)\dot{\bar{\gamma}}(t_0)$ beschrieben. Sie ist in einer Umgebung von $\bar{\gamma}(t_0)$ die **beste lineare Approximation an die Kurve**.

Ist \bar{p} ein Doppelpunkt der Kurve, d.h. $\bar{\gamma}(t_1) = \bar{p} = \bar{\gamma}(t_2)$ für $t_1 \neq t_2$, so können $\dot{\bar{\gamma}}(t_1)$ und $\dot{\bar{\gamma}}(t_2)$ verschieden sein. In Doppelpunkten gibt es eventuell verschiedene Tangenten. Ein Beispiel liefert die Kurve $(t^2 - 1, t^3 - t)$; sie hat im Nullpunkt, d.h. für $t = \pm 1$ zwei Tangenten (nachrechnen).

Bei sogenannten *singulären* Parameterwerten, d.h. wenn $\dot{\bar{\gamma}}(t) = \vec{0}$ ist, gibt es keine Tangenten (in dem eben definierten Sinn; es kann andere Parametrisierungen derselben Bahn geben, bei denen im entsprechenden Raumpunkt sehr wohl eine Tangente existiert).

Gehen zwei Kurven $\bar{\alpha}$ und $\bar{\gamma}$ durch denselben Raumpunkt \bar{p} , so definiert man den Schnittwinkel als Winkel zwischen den Kurventangenten in diesem Punkt. Je nach Parametrisierung kann \bar{p} zu verschiedenen Zeitpunkten erreicht werden, etwa $\bar{\alpha}(s) = \bar{p} = \bar{\gamma}(t)$. Dann ist der Kosinus des Schnittwinkels

$$\frac{\dot{\bar{\alpha}}(s) \bullet \dot{\bar{\gamma}}(t)}{\|\dot{\bar{\alpha}}(s)\| \cdot \|\dot{\bar{\gamma}}(t)\|}.$$

2.2.3 Differenzierbarkeit und Glattheit; Regularität

Die Grenzwerte, die in der Ableitungsdefinition vorkommen, müssen freilich nicht immer existieren. Wenn sie zu jedem Zeitpunkt aus I (in den Randpunkten im einseitigen Sinn) existieren, so nennt man die Kurve *differenzierbar*. Sind die Funktionen $\dot{\gamma}_i$ zudem stetig (bzw. ist $\dot{\bar{\gamma}}$ eine stetige Vektorfunktion), so werden wir die Kurve *glatt*⁸ nennen. In praktischen Anwendungen kommen oft Kurven vor (z.B. Polygonzüge), die aus glatten Teilen gestückt sind. Man spricht dann von *stückweise differenzierbaren* bzw. *stückweise glatten* Kurven: Das ganze Intervall $I = [a, b]$ läßt sich durch $a = c_1 < c_2 < \dots < c_k = b$ in endlich viele Teilintervalle zerlegen, so daß die Kurve auf

⁸Achtung: manche Autoren sagen hier nur stetig differenzierbar und reservieren das Wort ‘glatt’ für solche Kurven, die unendlich oft differenzierbar sind.

jedem $[c_i, c_{i+1}]$ differenzierbar oder glatt ist (jeweils mit Einschluß der Randpunkte c_i , in denen müssen also die einseitigen Ableitungen existieren, brauchen aber nicht übereinzustimmen). Eine glatte Kurve heißt *regulär*, wenn ihre Ableitung nie verschwindet. Reguläre Kurven haben zu jedem Zeitpunkt eine Tangente, die sich stetig ändert.

2.2.4 Eine Kuriosität am Rande

Durchläuft man einen Polygonzug mit Einheitsgeschwindigkeit, so ist die entsprechende Kurve nur im Inneren der beteiligten Strecken stetig differenzierbar. An den Eckpunkten stimmen die beiden einseitigen Ableitungen nicht überein. Allerdings kann man jeden Polygonzug auch als Bahn einer durchgängig stetig differenzierbaren Kurve bekommen. Die Idee besteht darin, die Geschwindigkeit in jedem Eckpunkt auf Null herunterzufahren. Der (Geschwindigkeits)Nullvektor merkt dann nichts vom Richtungswechsel. Beispielsweise werden die beiden Strecken von $(-1, 1)$ nach $(0, 0)$ und weiter nach $(1, 1)$, d.h. der Graph der Betragsfunktion auf $[-1, 1]$, von der Kurve

$$x(t) = t^3 \quad y(t) = \begin{cases} -t^3, & t < 0 \\ t^3, & t \geq 0 \end{cases} \quad t \in [-1, 1]$$

durchlaufen, die durchgängig stetig differenzierbar ist. Die Differenzierbarkeit von y zum Zeitpunkt $t = 0$ folgt aus der Gleichheit der linksseitigen und rechtsseitigen Ableitung.

Als **Übungsaufgabe** können Sie beweisen, daß Polygonzüge **nicht von regulären** Kurven durchlaufen werden können. Will man glatt um eine Ecke kommen, muß man auf Null abbremsen.

2.2.5 Regulärer Zusammenhang

Wir können jetzt ein Resultat des vorigen Abschnitts verbessern: *Je zwei Punkte einer offenen zusammenhängenden Teilmenge von \mathbb{R}^n lassen sich durch eine reguläre Kurve verbinden.*

Wir wissen schon, daß es einen Polygonzug gibt, der die gegebenen beiden Punkte verbindet. Bis auf die endlich vielen Ecken, in denen die Strecken des Polygonzuges zusammenstoßen, ist diese Kurve auch glatt. Die Idee, die ich nicht im Detail ausführen werde (weil das Hinschreiben der Formeln keine Einsicht bringt), besteht darin, diese Ecken abzurunden. Da jeder Eckpunkt in einer ganz in der offenen Menge enthaltenen Kugel liegt, ist dazu genug Platz.

2.2.6 Differentiation von Vektorfunktionen

Im Rest dieses Abschnittes betrachten wir eine auf einem Intervall $I \subseteq \mathbb{R}$ definierte Funktion \vec{f} mit Werten in \mathbb{R}^n . Das ist im Prinzip dasselbe wie eine Kurve, nur daß die Koordinaten jetzt untereinander und nicht nebeneinander geschrieben werden.

Ist t ein innerer (Zeit)Punkt des Definitionsbereiches, so definiert man die Ableitung von \vec{f} an der Stelle t durch

$$\vec{f}'(t) := \lim_{h \rightarrow 0} \frac{\vec{f}(t+h) - \vec{f}(t)}{h}.$$

In Randpunkten des Definitionsintervalls I benutzt man einseitige Grenzwerte $h \searrow 0$ bzw. $h \nearrow 0$). Der Grenzwert muß natürlich existieren und wenn dem so ist, dann nennt man \vec{f} an der Stelle (zum Zeitpunkt) t differenzierbar. Alles für Kurven gesagte überträgt sich. Speziell läßt sich der Grenzwert reinziehen:

$$\vec{f}'(t) = \lim_{h \rightarrow 0} \begin{pmatrix} \frac{f_1(t+h) - f_1(t)}{h} \\ \frac{f_2(t+h) - f_2(t)}{h} \\ \vdots \\ \frac{f_n(t+h) - f_n(t)}{h} \end{pmatrix} = \begin{pmatrix} f'_1(t) \\ f'_2(t) \\ \vdots \\ f'_n(t) \end{pmatrix}$$

Nach dieser Erkenntnis ist es Routine (also Übungsaufgabe), die folgenden

2.2.7 Ableitungsregeln nachzuprüfen:

(1) **Linearität.** $(c\vec{f} + d\vec{g})' = c\vec{f}' + d\vec{g}'$, für konstante c, d .

(2) **Produktregeln.**

$$(g\vec{f})' = g'\vec{f} + g\vec{f}', \quad \frac{d}{dt}(\vec{f} \bullet \vec{g}) = \vec{f}' \bullet \vec{g} + \vec{f} \bullet \vec{g}', \quad \text{und} \quad \frac{d}{dt}(\vec{f} \times \vec{g}) = (\vec{f}' \times \vec{g}) + (\vec{f} \times \vec{g}').$$

(3) **Ableitung einer Determinante.** Sind $\vec{f}, \vec{g}, \vec{h} : I \rightarrow \mathbb{R}^3$ differenzierbare Vektorfunktionen, so ist $\det(\vec{f}, \vec{g}, \vec{h})$ eine differenzierbare Funktion $I \rightarrow \mathbb{R}$, für deren Ableitung gilt:

$$\frac{d}{dt} \det(\vec{f}, \vec{g}, \vec{h}) = \det(\vec{f}', \vec{g}, \vec{h}) + \det(\vec{f}, \vec{g}', \vec{h}) + \det(\vec{f}, \vec{g}, \vec{h}').$$

Analoge Formeln (die Ableitung wandert durch) gelten in höheren Dimensionen.

2.2.8 Integrale von Vektorfunktionen

Ist $\vec{f} = \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_n \end{pmatrix}$ eine auf dem kompakten Intervall $[a, b]$ definierte einstellige Vektorfunktion mit

Werten in \mathbb{R}^n , deren Koordinatenfunktionen integrierbar sind, so ist es sinnvoll, ihr Integral als

$$\int_a^b \vec{f}(t) dt := \begin{pmatrix} \int_a^b f_1(t) dt \\ \int_a^b f_2(t) dt \\ \vdots \\ \int_a^b f_n(t) dt \end{pmatrix}$$

zu definieren. Die üblichen Eigenschaften des RIEMANN-Integrals einstelliger Funktionen lassen sich ohne Mühe auf diesen Fall übertragen. Speziell sind alle stückweise stetigen Vektorfunktionen integrierbar. Die NEWTON-LEIBNIZ-Formel nimmt folgende Gestalt an:

$$\int_a^b \vec{f}'(t) dt = \vec{f}(b) - \vec{f}(a) \quad \text{bzw. für Kurven} \quad \int_a^b \dot{\vec{\gamma}}(t) dt = \vec{\gamma}(b) - \vec{\gamma}(a).$$

Außerdem ist mit \vec{f} auch die skalare Funktion $\|\vec{f}\|$ integrierbar und es gilt

$$\left\| \int_a^b \vec{f}(t) dt \right\| \leq \int_a^b \|\vec{f}(t)\| dt.$$

Die Integrierbarkeit von $\|\vec{f}\| = \sqrt{f_1^2 + f_2^2 + \dots + f_n^2}$ ist (nach Analysis I) klar. Der Beweis der Ungleichung braucht einen Trick, der durchsichtiger wird, wenn wir $\int_a^b f_i(t) dt$ mit y_i bezeichnen. Dann ist $\int_a^b \vec{f}(t) dt = \vec{y}$ und es gilt (alle Summen von $i = 1$ bis n):

$$\|\vec{y}\|^2 = \sum y_i^2 = \sum y_i \int_a^b f_i(t) dt = \int_a^b \left(\sum y_i \cdot f_i(t) \right) dt = \int_a^b \vec{y} \bullet \vec{f}(t) dt$$

weiter mit CAUCHY-SCHWARZ

$$\leq \int_a^b \|\vec{y}\| \cdot \|\vec{f}(t)\| dt = \|\vec{y}\| \cdot \int_a^b \|\vec{f}(t)\| dt.$$

Ist $\|\vec{y}\| \neq 0$ so kann man kürzen und erhält

$$\left\| \int_a^b \vec{f}(t) dt \right\| = \|\vec{y}\| \leq \int_a^b \|\vec{f}(t)\| dt,$$

wie verlangt. Ist aber \vec{y} der Nullvektor, so ist dieselbe Ungleichung trivialerweise erfüllt.

2.2.9 Eine schwache Version des Mittelwertsatzes

Der Mittelwertsatz der Differentialrechnung (aus dem vorigen Semester) betrifft differenzierbare Funktionen $\varphi : [a, b] \rightarrow \mathbb{R}$ und liefert eine Stelle $c \in]a, b[$, für die $\varphi(b) - \varphi(a) = \varphi'(c) \cdot (b - a)$. Seine unmittelbare Übertragung auf Kurven bzw. einstellige Vektorfunktionen ist falsch (außer bei geradliniger Bewegung). Folgende abgeschwächte Form gilt aber und ist oft nützlich.

Satz. Ist $\vec{f} : [a, b] \rightarrow \mathbb{R}^n$ eine stetig differenzierbare Vektorfunktion, so gibt es ein $c \in]a, b[$ mit

$$\|\vec{f}(b) - \vec{f}(a)\| \leq \|\vec{f}'(c)\| \cdot (b - a).$$

Tatsächlich genügt die stückweise stetige Differenzierbarkeit, wie sie sich schnell selbst überlegen können. Der **Beweis** besteht nur aus einer Zeile und benutzt den Mittelwertsatz der Integralrechnung angewandt auf die stetige(!) Funktion $t \mapsto \|\vec{f}'(t)\|$.

$$\|\vec{f}(b) - \vec{f}(a)\| = \left\| \int_a^b \vec{f}'(t) dt \right\| \leq \int_a^b \|\vec{f}'(t)\| dt = \|\vec{f}'(c)\| \cdot (b - a).$$

Hier ist die Version für Kurven (gleicher Beweis): Ist $\bar{\gamma} : [a, b] \rightarrow \mathbb{R}^n$ eine glatte Kurve, so gibt es einen Zeitpunkt $c \in]a, b[$ mit

$$\|\bar{\gamma}(b) - \bar{\gamma}(a)\| \leq \|\dot{\bar{\gamma}}(c)\| \cdot (b - a).$$

Dem Physiker sollte die letzte Aussage völlig klar sein, sobald er sich $\bar{\gamma}$ als Beschreibung einer Bewegung vorstellt. Ist S der während der ganzen Zeit zurückgelegte Gesamtweg, so ist $S \geq \|\bar{\gamma}(b) - \bar{\gamma}(a)\|$, also der Quotient $\frac{\|\bar{\gamma}(b) - \bar{\gamma}(a)\|}{b - a}$ höchstens so groß wie die Durchschnittsgeschwindigkeit $\frac{S}{b - a}$. Diese muß aber irgendwann mal als Augenblicksgeschwindigkeit erreicht oder übertroffen werden, also $\leq \|\dot{\bar{\gamma}}(c)\|$ für ein c sein.

2.3 Die Länge einer Kurve

Gegeben sei eine Kurve K im geometrisch naiven Sinn (man denke an ein Stück Spirale oder eine Ellipse), die von \bar{a} nach \bar{b} führt und deren Länge bestimmt werden soll. Wenn es sich zufällig um eine Strecke handelt, so ist die Länge gleich dem Abstand von Anfangs- und Endpunkt: $\|\bar{b} - \bar{a}\|$. Ein Polygonzug besteht aus endlich vielen Strecken, deren aufaddierte Längen die Länge der ganzen Kurve ergeben. Im allgemeinen Fall besteht die Idee darin, die Kurve durch Polygonzüge zu approximieren und deren Längen als Approximationen an die gesuchte Länge zu benutzen. Für parametrisierte Kurven kann man das folgendermaßen exakt machen.

2.3.1 Definition der Länge

Gegeben sei eine Kurve $\bar{\gamma} : [a, b] \rightarrow \mathbb{R}^n$. Der Definitionsbereich ist also ein kompaktes Intervall; nur solche Kurven sind im Moment zugelassen. Statt der Eckpunkte des Polygonzuges auf der Bahn wählt man die Zeitpunkte, zu denen sie erreicht werden. Man unterteilt das Zeitintervall durch Zwischenpunkte $a = t_0 < t_1 < t_2 < \dots < t_k = b$ und benutzt die Länge $\sum_{i=0}^{k-1} \|\bar{\gamma}(t_{i+1}) - \bar{\gamma}(t_i)\|$ des Polygonzuges von $\bar{\gamma}(t_0)$ nach $\bar{\gamma}(t_1)$ weiter nach $\bar{\gamma}(t_2)$ weiter nach $\bar{\gamma}(t_3)$ weiter nach \dots schließlich nach $\bar{\gamma}(t_k)$ als Näherung für die Länge der Kurve. Diese sollte umso besser sein, je feiner die Unterteilung ist. Je feiner man die Unterteilung aber wählt, desto länger wird der Polygonzug (Dreiecksungleichung in jedem zusätzlich unterteilten Teilintervall). Daher ist

$$L(\bar{\gamma}) := \sup \sum_{i=0}^{k-1} \|\bar{\gamma}(t_{i+1}) - \bar{\gamma}(t_i)\|$$

eine sinnvolle Längendefinition, wobei das Supremum über alle Unterteilungen von $[a, b]$ genommen wird. Kurven bei denen das klappt (d.h. ein endliches Supremum herauskommt), nennt man *rektifizierbar*⁹.

⁹Früher auf deutsch auch *streckbar*.

2.3.2 Eigenschaften des Längenfunctionals

Die Beweise der folgenden drei **Beobachtungen** bleiben als Übungsaufgaben:
 $\bar{\alpha}$ und $\bar{\beta}$ seien Kurven, so daß der Endpunkt von $\bar{\alpha}$ gleich dem Anfangspunkt von $\bar{\beta}$ ist.

- (1) Die Zusammenstückung $\bar{\alpha}\bar{\beta}$ ist genau dann rektifizierbar, wenn beide Teilstücke rektifizierbar sind. In diesem Fall gilt $L(\bar{\alpha}\bar{\beta}) = L(\bar{\alpha}) + L(\bar{\beta})$.
- (2) Speziell ist jedes Teilstück einer rektifizierbaren Kurve rektifizierbar.
- (3) Mit $\bar{\alpha}$ ist auch $\bar{\alpha}^-$ rektifizierbar und beide haben dieselbe Länge.

2.3.3 Die Rektifizierbarkeit stückweise glatter Kurven

folgt aus der Möglichkeit zu stückeln und dem folgenden

Lemma. *Glatte Kurven sind rektifizierbar und haben die Länge*

$$\int_a^b \|\dot{\gamma}(t)\| dt = \int_a^b \sqrt{\dot{\gamma}_1(t)^2 + \dot{\gamma}_2(t)^2 + \dots + \dot{\gamma}_n(t)^2} dt.$$

Beweis. Sei $a = t_0 < t_1 < \dots < t_k = b$ eine Zerlegung des Definitionsbereiches. Dann ist die Länge des Polygonzuges (nach 2.2.8)

$$\sum \|\bar{\gamma}(t_{i+1}) - \bar{\gamma}(t_i)\| = \sum \left\| \int_{t_i}^{t_{i+1}} \dot{\gamma}(u) du \right\| \leq \sum \int_{t_i}^{t_{i+1}} \|\dot{\gamma}(u)\| du = \int_a^b \|\dot{\gamma}(u)\| du$$

höchstens so groß wie das Integral. Damit existiert das Supremum über alle Polygonzüge und wir erhalten

$$L(\bar{\gamma}) \leq \int_a^b \|\dot{\gamma}(u)\| du.$$

Für die Umkehrung sei ein beliebiges $\varepsilon > 0$ vorgegeben. Da die stetige Vektorfunktion $\dot{\gamma}$ auf dem kompakten Intervall $[a, b]$ gleichmäßig stetig ist, schwankt sie auf genügend kurzen Teilintervallen um weniger als ε . Ist also die Zerlegung $a = t_0 < t_1 < \dots < t_k = b$ genügend fein, so gilt für alle s zwischen t_i und t_{i+1} daß $\|\dot{\gamma}(s) - \dot{\gamma}(t_i)\| < \varepsilon$. Daraus folgt nach Dreiecksungleichung $\|\dot{\gamma}(s)\| \leq \|\dot{\gamma}(t_i)\| + \varepsilon$. Jetzt integrieren wir diese Ungleichungen zwischen t_i und t_{i+1} :

$$\begin{aligned} \int_{t_i}^{t_{i+1}} \|\dot{\gamma}(u)\| du &\leq \|\dot{\gamma}(t_i)\| \cdot (t_{i+1} - t_i) && + \varepsilon \cdot (t_{i+1} - t_i) \\ &= \left\| \int_{t_i}^{t_{i+1}} \dot{\gamma}(t_i) du \right\| && + \varepsilon \cdot (t_{i+1} - t_i) \\ &= \left\| \int_{t_i}^{t_{i+1}} [\dot{\gamma}(u) + \dot{\gamma}(t_i) - \dot{\gamma}(u)] du \right\| && + \varepsilon \cdot (t_{i+1} - t_i) \\ &\leq \left\| \int_{t_i}^{t_{i+1}} \dot{\gamma}(u) du \right\| + \left\| \int_{t_i}^{t_{i+1}} [\dot{\gamma}(t_i) - \dot{\gamma}(u)] du \right\| && + \varepsilon \cdot (t_{i+1} - t_i) \\ &\leq \left\| \int_{t_i}^{t_{i+1}} \dot{\gamma}(u) du \right\| + \int_{t_i}^{t_{i+1}} \|\dot{\gamma}(t_i) - \dot{\gamma}(u)\| du && + \varepsilon \cdot (t_{i+1} - t_i) \\ &\leq \|\gamma(t_{i+1}) - \gamma(t_i)\| + \varepsilon \cdot (t_{i+1} - t_i) && + \varepsilon \cdot (t_{i+1} - t_i) \\ &= \|\gamma(t_{i+1}) - \gamma(t_i)\| && + 2\varepsilon \cdot (t_{i+1} - t_i). \end{aligned}$$

Addiert man diese Ungleichungen für alle i , so ergibt sich

$$\int_a^b \|\dot{\gamma}(u)\| du \leq \sum_{i=0}^{k-1} \|\gamma(t_{i+1}) - \gamma(t_i)\| + 2\varepsilon \cdot (b - a).$$

Die Summe auf der rechten Seite ist höchstens so groß wie das Supremum aller Summen, also

$$\int_a^b \|\dot{\gamma}(u)\| du \leq L(\bar{\gamma}) + 2\varepsilon(b - a).$$

Nun läßt man noch $\varepsilon \searrow 0$ gehen.

2.3.4 Beispiele

Als dreidimensionale **Schraubenlinie** bezeichnet man die Kurve

$$t \mapsto (r \cos t, r \sin t, at) \quad \text{mit } -\infty < t < \infty.$$

Dabei ist r der Radius des Zylinders um den die Linie sich wickelt und $2\pi a$ ist die ‘Ganghöhe’. Wenn wir die Länge einer Windung bestimmen wollen, so führt das auf das Integral

$$\int_0^{2\pi} \sqrt{\left(\frac{d}{dt}r \cos t\right)^2 + \left(\frac{d}{dt}r \sin t\right)^2 + \left(\frac{d}{dt}at\right)^2} dt$$
$$\int_0^{2\pi} \sqrt{(-r \sin t)^2 + (r \cos t)^2 + a^2} dt = \int_0^{2\pi} \sqrt{r^2 + a^2} dt = 2\pi \sqrt{r^2 + a^2},$$

wie auch der gesunde Menschenverstand vorhergesagt hätte.

Eine ebene **Schneckenlinie** oder **Spirale** hat die Parametrisierung $t \mapsto (r(t) \cos t, r(t) \sin t)$, wobei $r(t)$ je nach dem, um welchen Spiraltyp es sich handelt, eine positive streng wachsende oder fallende stetige Funktion von t ist. Bei der sogenannten *Archimedischen Spirale* nimmt der Abstand vom Nullpunkt bei jeder Windung um den gleichen Betrag zu: $r(t) = at$ für ein positives a . Will man etwa die Länge der k -ten Windung bestimmen, so führt das auf das Integral

$$\int_{(2k-2)\pi}^{2k\pi} \sqrt{(a \cos t - at \sin t)^2 + (a \sin t + at \cos t)^2} dt = a \int_{(2k-2)\pi}^{2k\pi} \sqrt{1 + t^2} dt.$$

Die Stammfunktion kann man bei BRONSTEIN nachgucken und dann einsetzen. Das Resultat ist nicht besonders hübsch, ich spare mir die weitere Arbeit.

2.4 Kurvenintegrale von Skalarfeldern: $\int_{\bar{\gamma}} f ds$

Wir nehmen jetzt an, unsere Kurve $\bar{\gamma} : [a, b] \rightarrow \mathbb{R}^n$ verläuft im Definitionsbereich eines Skalarfeldes $f : D \rightarrow \mathbb{R}$. Analog zum RIEMANNschen Integral über einem Intervall (Strecke) soll jetzt das Integral von f entlang der Kurve erklärt werden. Statt die Strecke in kleine Teilstrecken einzuteilen, wird jetzt die Kurve in immer mehr und immer kleinere Teilstücke zerlegt.

2.4.1 Integrale entlang rektifizierbarer Kurven

Das zu definierende Kurvenintegral $\int_{\bar{\gamma}} f(\bar{x}) ds$ wird als Grenzwert bei zunehmender Verfeinerung der Unterteilungen $a = t_1 < t_2 < \dots < t_k = b$ von Integralsummen der Art

$$\sum_{i=1}^{k-1} f(\bar{\gamma}(\tau_i)) L(\bar{\gamma}, t_i, t_{i+1})$$

definiert. Dabei liegt τ_i zwischen t_i und t_{i+1} und $L(\bar{\gamma}, u, v)$ bezeichnet provisorisch die Länge des Kurvenstücks zwischen $\bar{\gamma}(u)$ und $\bar{\gamma}(v)$. Damit die Kurvenstückchen Längen haben, muß $\bar{\gamma}$ als rektifizierbar vorausgesetzt werden. Um es von anderen Arten von Integralen zu unterscheiden, schreibt man für Kurvenintegrale besonders bei geschlossenen Kurven gern \oint .

2.4.2 Berechnung

Bei glatten Parametrisierungen und stetigen Skalarfeldern kann man das Kurvenintegral nach der folgenden Formel ausrechnen:

$$\int_{\bar{\gamma}} f ds = \int_a^b f(\bar{\gamma}(t)) \cdot \|\dot{\bar{\gamma}}(t)\| dt,$$

wobei rechts ein ganz gewöhnliches Integral aus dem vorigen Semester steht. Um das einzusehen, betrachten wir eine Unterteilung von $[a, b]$. Dann zerfällt das Integral $\int_a^b f(\bar{\gamma}(t)) \|\dot{\bar{\gamma}}(t)\| dt$ in eine Summe

$$\sum_{i=1}^{k-1} \int_{t_i}^{t_{i+1}} f(\bar{\gamma}(t)) \|\dot{\bar{\gamma}}(t)\| dt = \sum_{i=1}^{k-1} f(\bar{\gamma}(\tau_i)) \cdot \int_{t_i}^{t_{i+1}} \|\dot{\bar{\gamma}}(t)\| dt = \sum_{i=1}^{k-1} f(\bar{\gamma}(\tau_i)) L(\bar{\gamma}, t_i, t_{i+1}),$$

wobei die τ_i zwischen t_i und t_{i+1} liegen und vom erweiterten Mittelwertsatz der Integralrechnung (letztes Semester) geliefert werden. Das Integral $\int_a^b f(\bar{\gamma}(t)) \|\dot{\bar{\gamma}}(t)\| dt$ selbst ist also *eine* der zur gegebenen Zerlegung (t_i) gehörenden Integralsummen. Ist $\sum_{i=1}^{k-1} f(\bar{\gamma}(\sigma_i)) L(\bar{\gamma}, t_i, t_{i+1})$ eine zweite, so unterscheidet sie sich vom Integral um

$$\left| \sum_{i=1}^{k-1} \left(f(\bar{\gamma}(\tau_i)) - f(\bar{\gamma}(\sigma_i)) \right) L(\bar{\gamma}, t_i, t_{i+1}) \right| \leq \sum_{i=1}^{k-1} |f(\bar{\gamma}(\tau_i)) - f(\bar{\gamma}(\sigma_i))| L(\bar{\gamma}, t_i, t_{i+1}) \leq S \cdot L(\bar{\gamma}),$$

wobei mit S die maximale Schwankung von $f \circ \bar{\gamma}$ auf den Teilintervallen $[t_i, t_{i+1}]$ gemeint ist. Wegen der gleichmäßigen Stetigkeit von $f \circ \bar{\gamma}$ auf dem (kompakten!) Intervall $[a, b]$ wird S beliebig klein, wenn nur alle Teilintervalle kurz genug sind. Bei genügend feiner Zerlegung wird also der Unterschied zwischen $\sum_{i=1}^{k-1} f(\bar{\gamma}(\sigma_i)) L(\bar{\gamma}, t_i, t_{i+1})$ und $\int_a^b f(\bar{\gamma}(t)) \|\dot{\bar{\gamma}}(t)\| dt$ beliebig klein.

2.4.3 Berechnung der Kurvenlänge als Kurvenintegral

Die Formel ist selbstverständlich, sollte aber irgendwo mal stehen:

$$L(\bar{\gamma}) = \int_{\bar{\gamma}} 1 ds.$$

2.4.4 Das Bogenelement und seine Darstellungen in Spezialfällen

Der formale Ausdruck ds wird *Bogenelement* (auch *Linielement*) genannt. Die Umrechnung

$$ds = \|\dot{\bar{\gamma}}(t)\| dt = \sqrt{\dot{\gamma}_1^2 + \dots + \dot{\gamma}_n^2} dt$$

wird klassisch auch

$$ds = \sqrt{(\dot{\gamma}_1 dt)^2 + \dots + (\dot{\gamma}_n dt)^2} = \sqrt{(dx_1)^2 + \dots + (dx_n)^2}$$

geschrieben. Man nennt sie dann die Darstellung des Bogenelements durch rechtwinklige Koordinaten.

Nicht immer ist es günstig, sich den Kurvenparameter als Zeit vorzustellen. Daneben gibt es noch andere populäre und/weil nützliche Möglichkeiten, vorallem in der Ebene.

Eine Koordinate als Parameter. In der Ebene ist die Kurve oft ein Funktionsgraph $y = y(x)$, der zwischen den Grenzen $x = a$ und $x = b$ betrachtet wird. Dann bietet sich x als Parameter an: $\bar{\gamma}(x) = (x, y(x))$. Das Bogenelement schreibt sich in diesem Fall als

$$ds = \sqrt{\left(\frac{d}{dx}x\right)^2 + \left(\frac{d}{dx}y\right)^2} dx = \sqrt{1 + (y')^2} dx$$

und die Formel für das Kurvenintegral wird zu

$$\int_a^b f(x, y(x)) \cdot \sqrt{1 + y'(x)^2} dx.$$

Polarkoordinaten mit dem Winkel als Parameter. Ebene Kurven werden auch oft in Polarkoordinaten beschrieben als $r = r(\varphi)$. Übersetzt man das in die oben verwendete Sprache, so ist eine Kurve

$$\bar{\gamma} : [\varphi_0, \varphi_1] \rightarrow \mathbb{R}^2 \quad \text{mit} \quad \bar{\gamma}(\varphi) = (r(\varphi) \cos \varphi, r(\varphi) \sin \varphi)$$

gemeint. Ihr Bogenelement berechnet sich als $ds = \sqrt{\left(\frac{d}{d\varphi} r(\varphi) \cos \varphi\right)^2 + \left(\frac{d}{d\varphi} r(\varphi) \sin \varphi\right)^2} d\varphi$

$$= \sqrt{(r'(\varphi) \cos \varphi - r(\varphi) \sin \varphi)^2 + (r'(\varphi) \sin \varphi + r(\varphi) \cos \varphi)^2} d\varphi = \sqrt{r'(\varphi)^2 + r(\varphi)^2} d\varphi.$$

Ist auch die Funktion in Polarkoordinaten gegeben $f(r, \varphi)$, so wird das Kurvenintegral entsprechend zu

$$\int_{\varphi_0}^{\varphi_1} f(r(\varphi), \varphi) \cdot \sqrt{r'(\varphi)^2 + r(\varphi)^2} d\varphi.$$

2.4.5 Eigenschaften des Integrals

Für eine feste Kurve $\bar{\gamma}$ und verschiedene Felder gelten (Existenz der Integrale vorausgesetzt) Linearität und Monotonie:

$$\int_{\bar{\gamma}} [uf(\bar{x}) + vg(\bar{x})] ds = u \int_{\bar{\gamma}} f(\bar{x}) ds + v \int_{\bar{\gamma}} g(\bar{x}) ds \quad f(\bar{x}) \leq g(\bar{x}) \Rightarrow \int_{\bar{\gamma}} f(\bar{x}) ds \leq \int_{\bar{\gamma}} g(\bar{x}) ds.$$

Für ein festes Skalarfeld und zusammenstückbare rektifizierbare Kurven gelten Additivität und Invarianz bei Zeitumkehr:

$$\int_{\bar{\alpha}\bar{\beta}} f(\bar{x}) ds = \int_{\bar{\alpha}} f(\bar{x}) ds + \int_{\bar{\beta}} f(\bar{x}) ds \quad \int_{\bar{\gamma}^-} f(\bar{x}) ds = \int_{\bar{\gamma}} f(\bar{x}) ds.$$

2.4.6 Alternative Definition des Kurvenintegrals

Wenn die Rektifizierbarkeit von $\bar{\gamma}$ nicht vorausgesetzt wird, hat die oben gegebene Integraldefinition keinen Sinn, weil dann die Längen $L(\bar{\gamma}, t_i, t_{i+1})$ eventuell gar nicht definiert sind. Man kann dann auf die Polygonzugidee zurückgehen und Integralsummen der Art

$$\sum_{i=0}^{k-1} f(\bar{\gamma}(\tau_i)) \|\bar{\gamma}(t_{i+1}) - \bar{\gamma}(t_i)\|$$

betrachten. Läßt man die Feinheit der Unterteilung immer kleiner werden, kann man hoffen, daß die Summen sich einem Grenzwert nähern, der dann als $\int_{\bar{\gamma}} f(\bar{x}) ds$ bezeichnet wird. Im Gegensatz zur ersten Variante ist hier die Rückführung auf

$$\int_a^b f(\bar{\gamma}(t)) \cdot \|\dot{\bar{\gamma}}(t)\| dt$$

etwas mühsamer, gelingt aber für glatte Kurven und sagen wir stetiges f . Im wesentlichen (nämlich für rektifizierbare Kurven) liefert die zweite Definition dasselbe wie die erste, in pathologischen Fällen (etwa f konstant Null, $\bar{\gamma}$ nicht rektifizierbar) liefert die zweite Definition eventuell einen Wert, wogegen die erste versagt.

2.5 Skalare Kurvenintegrale von Vektorfeldern: $\int_{\bar{\gamma}} \vec{f} \bullet d\vec{s}$

2.5.1 Motivierendes Beispiel

Wenn der sich auf der Kurve bewegendende Punkt eine Masse hat und sich in einem Kraftfeld bewegt, verrichtet er Arbeit. Sei $\bar{\gamma} : [a, b] \rightarrow \mathbb{R}^n$ die entsprechende Kurve durch den Definitionsbereich des Kraftfeldes $\vec{f} : D \rightarrow \mathbb{R}^n$. Man nähert die gesuchte Arbeit dadurch an, daß das ganze Zeitintervall

in viele kleine Teilintervalle $[t_i, t_{i+1}]$ zerlegt wird, auf denen die Kraft näherungsweise konstant als $\vec{f}(\bar{\gamma}(\tau_i))$ und die Kurve durch die Strecke von $\bar{\gamma}(t_i)$ nach $\bar{\gamma}(t_{i+1})$ ersetzt wird. Die auf dem Teilstück geleistete Arbeit ist näherungsweise gleich Streckenlänge mal Betrag der Kraft mal Kosinus des Winkels zwischen Kraftrichtung und Wegrichtung, also wissenschaftlich Skalarprodukt aus Kraft und Wegzuwachs. Summiert man über alle Teilstrecken auf, ergibt sich

$$\sum_{i=0}^{k-1} \vec{f}(\bar{\gamma}(\tau_i)) \bullet [\bar{\gamma}(t_{i+1}) - \bar{\gamma}(t_i)] = \sum_{i=0}^{k-1} \sum_{p=1}^n f_p(\bar{\gamma}(\tau_i)) \cdot [\gamma_p(t_{i+1}) - \gamma_p(t_i)].$$

Wenn man Glück hat (und bei stetigen Kraftfeldern und (stückweise) glatten Kurven ist das so), dann streben diese Summen bei fortschreitender Verfeinerung der Unterteilung gegen einen Grenzwert, eben die Arbeit des Feldes \vec{f} entlang des Weges $\bar{\gamma}$.

2.5.2 Allgemeine Definition

Derartige Grenzwerte von Summen von Skalarprodukten kann man nun für beliebige Vektorfelder und Kurven durch ihren Definitionsbereich betrachten. Man spricht vom Kurvenintegral des Feldes \vec{f} entlang der Kurve $\bar{\gamma}$. Das Ergebnis ist ein Skalar, der von mir mit $\int_{\bar{\gamma}} \vec{f} \bullet d\vec{s}$ bezeichnet werden soll und sich bei einer glatten Kurve auf das gewöhnliche Integral

$$\int_a^b \vec{f}(\bar{\gamma}(t)) \bullet \dot{\bar{\gamma}}(t) dt = \int_a^b [f_1(\bar{\gamma}(t)) \cdot \dot{\gamma}_1(t) + \dots + f_n(\bar{\gamma}(t)) \cdot \dot{\gamma}_n(t)] dt$$

zurückführen läßt. Die Begründung ist der für den anderen Integraltyp gegebenen sehr ähnlich und soll hier entfallen. Bei stückweise glatten Kurven wird das Integral für jedes glatte Teilstück ausgerechnet und dann addiert. Die klassische Schreibweise für diese Art Kurvenintegral ist

$$\int_{\bar{\gamma}} f_1(\bar{x}) dx_1 + f_2(\bar{x}) dx_2 + \dots + f_n(\bar{x}) dx_n$$

Im dreidimensionalen heißen die Koordinaten von \vec{f} traditionsgemäß P, Q, R und das Integral wird als $\int_{\bar{\gamma}} Pdx + Qdy + Rdz$ notiert. Wenn zum Beispiel P und R konstant Null sind, werden die entsprechenden Summanden nicht mitgeschrieben und es entstehen Schreibweisen der Art: $\int_{\bar{\gamma}} Q dy$. Statt $d\vec{s}$ kommt in Büchern auch $d\vec{x}$ oder, besonders bei Physikern, $d\vec{r}$ vor.

Bei Integration über geschlossenen Kurven verwendet man gern das Zeichen \oint und spricht bisweilen von der *Zirkulation* des Feldes \vec{f} entlang $\bar{\gamma}$.

Man kann diesen zweiten Typ Kurvenintegral in gewisser Weise als Spezialfall des ersten auffassen. Bezeichnet nämlich \vec{t} den im jeweiligen Punkt ausgerechneten Tangenteneinheitsvektor an die Kurve, so ist $\int_{\bar{\gamma}} \vec{f} \bullet d\vec{s} = \int_{\bar{\gamma}} (\vec{f} \bullet \vec{t}) ds$. Das ist allerdings deshalb nicht ganz korrekt, weil $\vec{f} \bullet \vec{t}$ kein a priori vorhandenes Skalarfeld ist, sondern gewissermaßen erst mit der Kurve entsteht.

2.5.3 Beispiel

Wir wollen das Vektorfeld $\begin{pmatrix} yz \\ zx \\ xy \end{pmatrix}$ entlang der Schraubenlinie $t \mapsto (r \cos t, r \sin t, at)$ integrieren

und zwar vom Punkt $(r, 0, 0)$ zum Punkt $(r, 0, 2\pi a)$ (eine Windung).

Nach dem oben angegebenen Rezept erhalten wir $\int_{\bar{\gamma}} yz dx + zx dy + xy dz$

$$\begin{aligned} &= \int_0^{2\pi} \left[(r \sin t) \cdot (at) \frac{d}{dt}(r \cos t) + (at) \cdot (r \cos t) \cdot \frac{d}{dt}(r \sin t) + (r \cos t) \cdot (r \sin t) \cdot \frac{d}{dt}(at) \right] dt \\ &= ar^2 \int_0^{2\pi} \left[-t \sin^2 t + t \cos^2 t + \cos t \cdot \sin t \right] dt = ar^2 \int_0^{2\pi} \frac{d}{dt} [t \cos t \sin t] dt = ar^2 [t \cos t \sin t] \Big|_0^{2\pi} = 0. \end{aligned}$$

2.5.4 Eigenschaften

zähle ich nur auf. Die Beweise sind mit unseren bisherigen Methoden leicht zu führen.

- Für eine feste Kurve $\bar{\gamma}$ und verschiedene Felder ist das Integral (Existenz vorausgesetzt) linear:

$$\int_{\bar{\gamma}} [u\vec{f}(\bar{x}) + v\vec{g}(\bar{x})] \bullet d\vec{s} = u \int_{\bar{\gamma}} \vec{f}(\bar{x}) \bullet d\vec{s} + v \int_{\bar{\gamma}} \vec{g}(\bar{x}) \bullet d\vec{s}$$

- Für ein festes Vektorfeld und zusammenstückbare rektifizierbare Kurven ist das Integral additiv

$$\int_{\bar{\alpha}\bar{\beta}} \vec{f}(\bar{x}) \bullet d\vec{s} = \int_{\bar{\alpha}} \vec{f}(\bar{x}) \bullet d\vec{s} + \int_{\bar{\beta}} \vec{f}(\bar{x}) \bullet d\vec{s}$$

- Bei ‘Zeitumkehr’ wechselt das Integral sein Vorzeichen¹⁰:

$$\int_{\bar{\gamma}^-} \vec{f}(\bar{x}) \bullet d\vec{s} = - \int_{\bar{\gamma}} \vec{f}(\bar{x}) \bullet d\vec{s}.$$

Versuchen Sie, den Vorzeichenwechsel mathematisch zu begründen. Wieso war das bei Skalarfeldern nicht so?

2.6 Äquivalente Kurven

In der Praxis wird die Kurve, entlang der integriert wird, oft nur ungenau angegeben. Es kann zum Beispiel sein, daß nur die Bahn beschrieben wird. Dabei ist in jedem Fall darauf zu achten, daß die ‘Orientierung’ klar ist: welcher Punkt ist als Anfangspunkt, welcher als Endpunkt gemeint. Davon hängt, wie gerade gesehen, das Vorzeichen des Integrals ab. Bei geschlossenen Kurven muß der Umlaufsinn angegeben werden. Wenn nichts gesagt ist, meint man in \mathbb{R}^2 immer den Umlaufsinn bei dem das Innere des umrandeten Gebietes links von der Kurve liegt. Außerdem ist stillschweigend immer gemeint, daß die Bahn vom Anfangs- zum Endpunkt auf (stückweise) reguläre Weise, d.h. stetig differenzierbar ohne Stillstand und auch ohne zwischenzeitliche Umkehr durchlaufen wird.

Mit diesen unvollständigen Angaben sind normalerweise mehrere Parametrisierungen vereinbar. In die Formeln zur Integralberechnung geht jedoch eine konkrete Parametrisierung ein. Wenn die nicht genau angegeben wird, könnte es doch passieren, daß sich verschiedene Menschen für verschiedene Parametrisierungen entscheiden und dann verschiedene Integrale herausbekommen. In diesem Abschnitt wollen wir begründen, daß das nicht passiert.

2.6.1 Beispiel

Wir wollen sehen, wie sich verschiedene Parametrisierungen auf ein Integral $\int_{\bar{\gamma}} Pdx + Qdy$ auswirken, wenn $\bar{\gamma}$ der von $(1, 0)$ nach $(0, 1)$ durchlaufene Viertelkreisbogen ist.

- Die ‘übliche’ Parametrisierung $x(t) = \cos t$, $y(t) = \sin t$, $t \in [0, \frac{\pi}{2}]$ liefert

$$\int_{\bar{\gamma}} Pdx + Qdy = \int_0^{\frac{\pi}{2}} [P(\cos t, \sin t) \cdot (-\sin t) + Q(\cos t, \sin t) \cdot \cos t] dt.$$

- $y \in [0, 1]$ als Parameter liefert $x = \sqrt{1 - y^2}$ und

$$\int_{\bar{\gamma}} Pdx + Qdy = \int_0^1 \left[P(\sqrt{1 - y^2}, y) \cdot \left(\frac{-y}{\sqrt{1 - y^2}} \right) + Q(\sqrt{1 - y^2}, y) \right] dy.$$

¹⁰Das ist dem Physiker sofort klar; wenn in der einen Richtung Arbeit geleistet wird, dann wird in der anderen Richtung dieselbe Menge Energie wieder frei.

- Wählt man $x \in [0, 1]$ als Parameter und $y = \sqrt{1-x^2}$, so wird der Viertelkreis im falschen Sinn durchlaufen. Das muß durch ein Minuszeichen ausgeglichen werden:

$$\int_{\bar{\gamma}} Pdx + Qdy = - \int_0^1 \left[P(x, \sqrt{1-x^2}) + Q(x, \sqrt{1-x^2}) \left(\frac{-x}{\sqrt{1-x^2}} \right) \right] dx.$$

- Schließlich kann man auch noch $\int_{\bar{\gamma}} Pdx + Qdy = \int_{\bar{\gamma}} Pdx + \int_{\bar{\gamma}} Qdy$ aufspalten und für beide Summanden verschiedene Parametrisierungen benutzen. Für den ersten bietet sich x für den zweiten y als Parameter an:

$$\int_{\bar{\gamma}} Pdx + Qdy = - \int_0^1 P(x, \sqrt{1-x^2})dx + \int_0^1 Q(\sqrt{1-y^2}, y)dy.$$

Ein scharfer Blick und ein Moment des Nachdenkens wird Sie davon überzeugen, daß trotz ihres unterschiedlichen Aussehens bei allen diesen Formeln dasselbe herauskommen muß; die Integrale gehen durch Substitutionen auseinander hervor.

2.6.2 Heuristik

Nehmen wir an, zwei Kurven $\bar{\alpha} : I \rightarrow \mathbb{R}^n$ und $\bar{\beta} : J \rightarrow \mathbb{R}^n$ durchlaufen dieselbe Bahn mit dem gleichen Anfangs- und Endpunkt (also im selben Sinn) ohne zu stoppen oder zwischendurch umzukehren. Dann gehört zu jedem Zeitpunkt $t \in I$ ein Raumpunkt $\bar{\alpha}(t)$, der aber auch von der Kurve $\bar{\beta}$ zu genau einer Zeit $u \in J$ erreicht wird: $\bar{\alpha}(t) = \bar{\beta}(u)$. Die Zuordnung $t \mapsto u$, wird von einer Funktion $g : I \rightarrow J$ hergestellt. Diese wird monoton sein, weil die Punkte der Bahn in derselben Reihenfolge (früher/später) durchlaufen werden. Da es keinen Stillstand geben soll, wird g sogar streng monoton sein. Außerdem muß g surjektiv sein, denn beide Kurven durchlaufen dieselbe Bahn und dazu wird jeweils das ganze Zeitintervall gebraucht. Diese Betrachtungen motivieren folgende

2.6.3 Definition

Zwei Kurven $\bar{\alpha} : I \rightarrow \mathbb{R}^n$ und $\bar{\beta} : J \rightarrow \mathbb{R}^n$ heißen *äquivalent*, wenn es eine surjektive streng monoton wachsende Funktion $g : I \rightarrow J$ derart gibt, daß $\bar{\alpha} = \bar{\beta} \circ g$. In vielen Zusammenhängen ist es sinnvoll, von der umrechnenden Funktion g noch stetige Differenzierbarkeit zu fordern. Ich bleibe mit dem Begriff etwas schwammig.

2.6.4 Weitere Bemerkungen

Die Äquivalenz von Kurven hat die drei typischen Eigenschaften

- *Reflexivität*: Jede Kurve ist zu sich selbst äquivalent,
- *Symmetrie*: Wenn $\bar{\alpha}$ äquivalent zu $\bar{\beta}$, so auch umgekehrt (weil g bijektiv sein muß und g^{-1} , wie im vorigen Semester gelernt, auch differenzierbar mit strikt positiver Ableitung ist) und
- *Transitivität*: Sind $\bar{\alpha}$ und $\bar{\beta}$ sowie $\bar{\beta}$ und $\bar{\gamma}$ jeweils äquivalent, so auch $\bar{\alpha}$ und $\bar{\gamma}$ (man benutzt einfach $g \circ h$ zur Umrechnung.).

2.6.5 Äquivalente Kurven haben dieselbe Länge.

Sind nämlich $\bar{\alpha} : [a, b] \rightarrow \mathbb{R}^n$ und $\bar{\beta} : [c, d] \rightarrow \mathbb{R}^n$ mittels $g : [a, b] \rightarrow [c, d]$ äquivalent, so läßt sich jeder Polygonzug sowohl durch Teilungspunkte t_i des Intervalls $[a, b]$ als auch durch Teilungspunkte $u_i = g(t_i)$ des Intervalls $[c, d]$ realisieren. Die beiden Mengen, deren Suprema die Kurvenlängen jeweils sind, sind also identisch. Dann auch ihre Suprema.

Bemerkung. Für äquivalente glatte Kurven (und eine stetig differenzierbare Umrechnungsfunktion g) folgt die Längengleichheit auch nach der Formel 2.3.3 und der Substitutionsregel für Integrale:

$$\int_a^b \|\dot{\alpha}(t)\| dt = \int_a^b \left\| \frac{d}{dt} \bar{\beta}(g(t)) \right\| dt = \int_a^b \left\| \dot{\bar{\beta}}(g(t)) \right\| \cdot g'(t) dt = \int_{g(a)}^{g(b)} \left\| \dot{\bar{\beta}}(u) \right\| du = \int_c^d \left\| \dot{\bar{\beta}}(u) \right\| du.$$

2.6.6 Kurvenintegrale

Offensichtlich tauchen bei äquivalenten Kurven dieselben Integralsummen auf (nur entsprechen sie jeweils anderen Unterteilungen des Zeitintervalls). Daher hängt die Existenz des Grenzwertes und sein Wert nicht davon ab, welche von zwei äquivalenten Kurven benutzt wird.

3 Differentialrechnung in \mathbb{R}^n

Das Ableiten von Kurven fand schon in \mathbb{R}^n statt. Aber dabei konnte alles sofort auf die einstelligen Koordinatenfunktionen heruntergespielt werden. Jetzt wird es mit dem Ableiten im n -Dimensionalen aber ernst.

3.1 Partielle Ableitungen

In diesem ganzen Abschnitt betrachten wir ein Skalarfeld $f : D \rightarrow \mathbb{R}$, das in der Nähe eines inneren Punktes \bar{a} seines Definitionsbereiches $D \subset \mathbb{R}^n$ untersucht wird.

Für dieses lokale Verhalten von f bei \bar{a} spielen die Werte von f außerhalb einer kleinen Kugel mit Zentrum \bar{a} keine Rolle. Erst wenn wir das lokale Verhalten von f in der Nähe verschiedener Punkte vergleichen wollen, müssen wir alle diese als innere Punkte des Definitionsbereiches voraussetzen. Wir machen uns die Arbeit dadurch etwas leichter, daß wir gleich alle Punkte von D als innere Punkte, die ganze Menge D also als offen annehmen. In der Praxis sind die 'natürlichen Definitionsbereiche' von Funktionen (d.h. die Punkte für die die definierenden Formeln sinnvoll gelesen werden können) meist¹¹ automatisch offen.

3.1.1 Definition der partiellen Ableitung

Wenn wir bis auf die i -te alle Koordinaten von \bar{a} festhalten, die i -te aber variieren, so entsteht eine einstellige Funktion

$$t \mapsto f(a_1, \dots, a_{i-1}, t, a_{i+1}, \dots, a_n)$$

die auf einem Intervall $]a_i - \delta, a_i + \delta[$ definiert ist. Die Ableitung (im Sinne des vorigen Semesters) dieser einstelligen Funktion an der Stelle $t = a_i$ heißt *partielle Ableitung der Funktion f an der Stelle \bar{a} nach der i -ten Koordinate* und wird je nachdem mit $\frac{\partial f}{\partial x_i}(\bar{a})$ oder $\frac{\partial}{\partial x_i} f(\bar{a})$ oder $f_{x_i}(\bar{a})$ oder $D_i f(\bar{a})$ bezeichnet. Bei Funktionen $f(x, y)$ oder $f(x, y, z)$ von zwei oder drei Veränderlichen schreibt man $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}$ oder f_x, f_y, f_z .

Als Grenzwert kann man die partielle Ableitung so definieren.

$$\frac{\partial f}{\partial x_i}(a_1, a_2, \dots, a_n) = \lim_{h \rightarrow 0} \frac{f(a_1, \dots, a_i + h, \dots, a_n) - f(a_1, \dots, a_i, \dots, a_n)}{h}$$

oder in kompakter Vektorschreibweise

$$\frac{\partial f}{\partial x_i}(\bar{a}) = \lim_{h \rightarrow 0} \frac{f(\bar{a} + h \vec{e}_i) - f(\bar{a})}{h}.$$

Das Hinschreiben dieser Formeln bedeutet natürlich nicht, daß die Grenzwerte auch existieren. Wenn das aber so ist, dann heißt f im Punkt/an der Stelle \bar{a} *partiell nach x_i differenzierbar*. Wenn f an allen Punkten des Definitionsbereiches partiell nach x_i differenzierbar ist, so entsteht eine neue Funktion $\frac{\partial f}{\partial x_i} : D \rightarrow \mathbb{R}$, die *partielle Ableitung von f nach x_i* heißt. Andere Schreibweisen dafür sind f_{x_i} und $D_i f$. Bei kleinen Dimensionen auch $\frac{\partial f}{\partial x}$ oder f_x , etc.

¹¹Aber nicht immer; denken Sie an \sqrt{t} !

3.1.2 Die praktische Berechnung

von partiellen Ableitungen macht normalerweise keine Schwierigkeiten. Man betrachtet die übrigen Variablen als (konstante) Parameter und leitet nach der interessierenden Variablen ab, wie im vorigen Semester gelernt. Die Differentiationsregeln (Summen-, Produkt-, Quotienten- und Kettenregel) bleiben erhalten.

Beispiele.

$$\begin{aligned} \frac{\partial}{\partial x} (x^2 + 5xy^2 - y^6) &= 2x + 5y^2 & \frac{\partial}{\partial y} (x^2 + 5xy^2 - y^6) &= 10xy - 6y^5. \\ \frac{\partial}{\partial x} \arctan \frac{x}{y} &= \frac{1}{1 + \left(\frac{x}{y}\right)^2} \cdot \frac{1}{y} = \frac{y}{x^2 + y^2} & \frac{\partial}{\partial y} \arctan \frac{x}{y} &= \frac{1}{1 + \left(\frac{x}{y}\right)^2} \cdot \frac{-x}{y^2} = \frac{-x}{x^2 + y^2}. \end{aligned}$$

Allerdings muß man dann aufpassen, wenn eine der Variablen einen Wert annimmt, bei dem die gewöhnlichen Ableitungsregeln nicht funktionieren. Im folgenden **Beispiel** soll $\sqrt[3]{t}$ entgegen unseren sonstigen Gepflogenheiten auch für negatives t definiert sein und diejenige Zahl liefern, deren dritte Potenz t ist. Diese Funktion ist, außer bei $t = 0$, überall differenzierbar mit der Ableitung $\frac{1}{3}t^{-2/3} = \frac{1}{3\sqrt[3]{t^2}}$. Die zweistellige Funktion $f(x, y) = \sqrt[3]{xy}$ ist dann auch überall definiert und ihre partielle Ableitung nach x ergibt sich nach den formalen Regeln als

$$\frac{\partial}{\partial x} \sqrt[3]{xy} = \frac{y}{3\sqrt[3]{(xy)^2}} = \frac{1}{3} \sqrt[3]{\frac{y}{x^2}}.$$

Diese dürfen allerdings nur angewendet werden, solange $xy \neq 0$, weil $\sqrt[3]{}$ sonst nicht differenzierbar ist. Trotzdem existieren die partiellen Ableitungen $\frac{\partial f}{\partial x}(a, 0) = \lim_{h \rightarrow 0} \frac{f(a+h, 0) - f(a, 0)}{h} = \lim \frac{0-0}{h} = 0$, sogar für $a = 0$. Für $b \neq 0$ existiert $\frac{\partial f}{\partial x}(0, b)$ allerdings nicht.

3.1.3 Partielle Differenzierbarkeit und Stetigkeit

Im Gegensatz zum eindimensionalen Fall, wo die Stetigkeit eine notwendige Voraussetzung der Differenzierbarkeit war, können auch unstetige Funktionen partielle Ableitungen besitzen.

Das ist so verwunderlich nicht: partielle Differenzierbarkeit impliziert 'achsenparalleles Wohlverhalten' aber zwischen den Achsen ist noch genug Platz für Unartigkeiten. Ein Beispiel ist die Funktion $(x, y) \mapsto \operatorname{sgn}(xy)$. Sie gibt das Vorzeichen (+1, -1 oder 0) von xy heraus und ist auf den Koordinatenachsen konstant 0. Daher existieren beide partielle Ableitungen in $(0, 0)$ und sind 0, während die Funktion in diesem Punkt nicht stetig ist.

Satz. *Setzt man dagegen voraus, daß alle partiellen Ableitungen von f in allen Punkten einer Umgebung (etwa einer offenen Kugel mit Mittelpunkt \bar{a}) existieren und die Funktionen $\frac{\partial f}{\partial x_i}$ in dieser Umgebung alle beschränkt sind, dann ist die Funktion f stetig in \bar{a} . Die Bedingung ist immer dann erfüllt, wenn alle $\frac{\partial f}{\partial x_i}$ im Punkt \bar{a} stetig sind.*

Um uns nicht durch Indizes zu verwirren, führe ich den **Beweis** für eine dreistellige Funktion $f(x, y, z)$. Angenommen, die partiellen Ableitungen $\frac{\partial f}{\partial x}$, $\frac{\partial f}{\partial y}$ und $\frac{\partial f}{\partial z}$ existieren für alle $(x, y, z) \in K = K_{(a,b,c)}(r)$ und sind dort betragsmäßig kleiner als C . Wir müssen zeigen, daß $|f(x, y, z) - f(a, b, c)|$ beliebig klein wird, sobald $\sqrt{(x-a)^2 + (y-b)^2 + (z-c)^2}$ genügend klein ist. Die Idee, die hier und in vielen anderen Beweisen zum Ziel führt, benutzt einen 'achsenparallelen Umweg', z.B.

$$f(x, y, z) - f(a, b, c) = (f(x, y, z) - f(a, y, z)) + (f(a, y, z) - f(a, b, z)) + (f(a, b, z) - f(a, b, c)).$$

Jede der drei Differenzen kann nach dem Mittelwertsatz der (eindimensionalen!) Differentialrechnung umgeschrieben werden. Um das einmal ganz ausführlich zu machen, betrachten wir die mittlere Differenz $f(a, y, z) - f(a, b, z)$. Sie läßt sich schreiben als $\varphi(y) - \varphi(b)$ für die Hilfsfunktion $\varphi(t) = f(a, t, z)$. Nach Voraussetzung ist φ differenzierbar. Also liefert der Mittelwertsatz ein η

zwischen b und y mit $\varphi(y) - \varphi(b) = \varphi'(\eta) \cdot (y - b)$. Es ist aber $\varphi'(\eta)$ nichts anderes als $\frac{\partial f}{\partial y}(a, \eta, z)$, also $f(a, y, z) - f(a, b, z) = \frac{\partial f}{\partial y}(a, \eta, z) \cdot (y - b)$. Auf analoge Weise erhält man passende ξ und ζ mit

$$f(x, y, z) - f(a, y, z) = \frac{\partial f}{\partial x}(\xi, y, z) \cdot (x - a) \quad \text{und} \quad f(a, b, z) - f(a, b, c) = \frac{\partial f}{\partial z}(a, b, \zeta) \cdot (z - c).$$

Wenn $(x, y, z) \in K$, so liegen auch die drei Zwischenpunkte (ξ, y, z) , (a, η, z) und (a, b, ζ) in K . Also sind die Beträge der drei Ableitungen höchstens C und wir erhalten

$$\begin{aligned} |f(x, y, z) - f(a, b, c)| &\leq |f(x, y, z) - f(a, y, z)| + |f(a, y, z) - f(a, b, z)| + |f(a, b, z) - f(a, b, c)| \\ &= \left| \frac{\partial f}{\partial x}(\xi, y, z) \cdot (x - a) \right| + \left| \frac{\partial f}{\partial y}(a, \eta, z) \cdot (y - b) \right| + \left| \frac{\partial f}{\partial z}(a, b, \zeta) \cdot (z - c) \right| \\ &\leq C \cdot (|x - a| + |y - b| + |z - c|) \\ &\leq 3C \cdot \sqrt{(x - a)^2 + (y - b)^2 + (z - c)^2}. \end{aligned}$$

Nun sollte klar sein, daß die rechte Seite mit dem Abstand von (x, y, z) zu (a, b, c) beliebig klein wird.

3.1.4 Partielle Ableitungen von Punkt- und Vektorfeldern

lassen sich entweder in Analogie zu den Skalarfeldern einführen oder ohne weiteres auf die partiellen Ableitungen der Koordinatenfunktionen zurückführen. Allerdings bieten die Punktfelder eine kleine Überraschung: sie werden zu Vektorfeldern.

$$\frac{\partial \vec{f}}{\partial x_i}(\vec{a}) = \lim_{h \rightarrow 0} \frac{\vec{f}(\vec{a} + h\vec{e}_i) - \vec{f}(\vec{a})}{h} = \begin{pmatrix} \lim_{h \rightarrow 0} \frac{f_1(\vec{a} + h\vec{e}_i) - f_1(\vec{a})}{h} \\ \lim_{h \rightarrow 0} \frac{f_2(\vec{a} + h\vec{e}_i) - f_2(\vec{a})}{h} \\ \vdots \\ \lim_{h \rightarrow 0} \frac{f_n(\vec{a} + h\vec{e}_i) - f_n(\vec{a})}{h} \end{pmatrix} = \begin{pmatrix} \frac{\partial f_1}{\partial x_i}(\vec{a}) \\ \frac{\partial f_2}{\partial x_i}(\vec{a}) \\ \vdots \\ \frac{\partial f_n}{\partial x_i}(\vec{a}) \end{pmatrix}.$$

So überraschend ist das allerdings auch nicht mehr; wir hatten ja den Effekt schon bei Kurven erwähnt ihre Werte liegen im Punktraum aber die Ableitungen sind Geschwindigkeitsvektoren.

Bei Vektorfeldern passiert nichts unvorhergesehenes:

$$\frac{\partial \vec{f}}{\partial x_i} = \frac{\partial}{\partial x_i} \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_n \end{pmatrix} = \begin{pmatrix} \frac{\partial f_1}{\partial x_i} \\ \frac{\partial f_2}{\partial x_i} \\ \vdots \\ \frac{\partial f_n}{\partial x_i} \end{pmatrix}.$$

3.2 (Totale) Differenzierbarkeit

Im Eindimensionalen war die Existenz der Ableitung einer Funktion in einem Punkt äquivalent zur linearen Approximierbarkeit dieser Funktion in der Nähe dieses Punktes. Das mehrdimensionale Analogon dieses zweiten Konzeptes nennt man Differenzierbarkeit. Zur Betonung manchmal totale Differenzierbarkeit.

3.2.1 Definition für Skalarfelder

Wieder betrachten wir zunächst ein Skalarfeld $f : D \rightarrow \mathbb{R}$, wobei D eine offene Teilmenge von \mathbb{R}^n sein soll. f heißt *an der Stelle \vec{a} (total) differenzierbar*, wenn es **eine** lineare Abbildung¹²

¹²Hier spielen die reellen Zahlen mal die Rolle von eindimensionalen Vektoren: eigentlich gehen L und ρ von \mathbb{R}^n nach \mathbb{R}^1 . Wir wollten den Unterschied nicht machen und halten uns auch daran. Einzig die Reihenfolge $\|\vec{v}\| \cdot \rho(\vec{v})$ (Skalar mal Vektor) deutet vielleicht noch darauf hin.

$L_{\bar{a}} : \mathbb{R}^n \rightarrow \mathbb{R}$ und eine Funktion $\rho : \mathbb{R}^n \rightarrow \mathbb{R}$ derart gibt, daß, wann immer die linke Seite definiert ist,

$$f(\bar{a} + \vec{v}) = f(\bar{a}) + L_{\bar{a}}(\vec{v}) + \|\vec{v}\| \cdot \rho(\vec{v}) \quad \text{und} \quad \lim_{\vec{v} \rightarrow \vec{0}} \rho(\vec{v}) = 0.$$

Alternative Beschreibungen: Der Zuwachs $\Delta_{\bar{a}} f(\vec{v}) = f(\bar{a} + \vec{v}) - f(\bar{a})$ besteht aus dem linearen Hauptbestandteil $L(\vec{v})$ und einem Fehler, der für $\vec{v} \rightarrow \vec{0}$ nicht nur gegen Null geht, sondern sogar von höherer als erster Ordnung. Letzteres meint: selbst wenn man den Fehler durch $\|\vec{v}\|$ dividiert, geht der Quotient noch gegen 0.

Obwohl die Version mit dem Vektor \vec{v} die eigentlich angemessenere ist, kann man die Definition der Differenzierbarkeit auch nur mit Punkten schreiben (der Vektor versteckt sich in der Differenz $\bar{b} - \bar{a}$):

$$f(\bar{b}) = f(\bar{a}) + L_{\bar{a}}(\bar{b} - \bar{a}) + d(\bar{a}, \bar{b}) \cdot \sigma(\bar{b}), \quad \text{wobei} \quad \lim_{\bar{x} \rightarrow \bar{a}} \sigma(\bar{x}) = 0.$$

Die **lineare Abbildung** $L_{\bar{a}}$ hat bezüglich der kanonischen Basis eine darstellende Matrix, die in diesem und allgemeineren Fällen JACOBI-Matrix von f an der Stelle \bar{a} genannt und mit $J_f(\bar{a})$ bezeichnet wird. Die lineare Abbildung selbst wird von manchen Autoren (*totale*) *Ableitung* von f an der Stelle \bar{a} genannt und folgerichtig mit $f'(\bar{a})$ bezeichnet. Auch der Name (*totales*) *Differential* ist gebräuchlich. Dann schreibt man gern $d_{\bar{a}} f$.

3.2.2 Berechnung der JACOBI-Matrix

Im Falle eines Skalarfeldes ist das totale Differential eine Linearform auf \mathbb{R}^n , ihre darstellende Matrix also eine Zeile. Schreiben wir sie erst mal provisorisch als $J = (l_1, l_2, \dots, l_n)$ und die Koordinaten von \vec{v} als v_i , so ergibt sich die ausführlichere Variante der obigen Gleichung:

$$\begin{aligned} & f(a_1 + v_1, a_2 + v_2, \dots, a_n + v_n) \\ &= f(a_1, a_2, \dots, a_n) + l_1 v_1 + l_2 v_2 + \dots + l_n v_n + \sqrt{v_1^2 + v_2^2 + \dots + v_n^2} \cdot \rho \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}. \end{aligned}$$

Diese Gleichung muß für alle \vec{v} gelten, für die $f(\bar{a} + \vec{v})$ definiert ist, damit mindestens für alle hinreichend kleinen Vektoren \vec{v} . Um die l_i als neue Bekannte zu identifizieren, fixieren wir ein i und betrachten $\vec{v} = h\vec{e}_i$. Dann wird die Gleichung zu

$$f(\bar{a} + h\vec{e}_i) = f(\bar{a}) + l_i h + |h| \cdot \rho(h\vec{e}_i) \quad \text{oder} \quad \frac{f(\bar{a} + h\vec{e}_i) - f(\bar{a})}{h} = l_i \pm \rho(h\vec{e}_i)$$

Läßt man $h \rightarrow 0$ gehen, wird klar, daß $l_i = \frac{\partial f}{\partial x_i}(\bar{a})$ sein muss. Die JACOBI-Matrix eines Skalarfeldes f besteht also aus den nebeneinander geschriebenen partiellen Ableitungen.

$$J_f(\bar{a}) = \left(\frac{\partial f}{\partial x_1}(\bar{a}), \frac{\partial f}{\partial x_2}(\bar{a}), \dots, \frac{\partial f}{\partial x_n}(\bar{a}) \right). \quad \text{Begriff des Gradienten?}$$

Speziell folgt aus der Differenzierbarkeit im Punkt \bar{a} , daß alle partiellen Ableitungen im Punkt \bar{a} existieren müssen. Das Differential läßt sich in Koordinatenschreibweise so notieren:

$$d_{\bar{a}} f(\vec{v}) = \frac{\partial f}{\partial x_1}(\bar{a}) \cdot v_1 + \frac{\partial f}{\partial x_2}(\bar{a}) \cdot v_2 + \dots + \frac{\partial f}{\partial x_n}(\bar{a}) \cdot v_n.$$

In der klassischen Schreibweise betrachtet man $\vec{v} = \begin{pmatrix} dx_1 \\ dx_2 \\ \vdots \\ dx_n \end{pmatrix}$ und läßt es als auf der linken Seite

weg:

$$d_{\bar{a}} f = \frac{\partial f}{\partial x_1}(\bar{a}) dx_1 + \frac{\partial f}{\partial x_2}(\bar{a}) dx_2 + \dots + \frac{\partial f}{\partial x_n}(\bar{a}) dx_n.$$

Bemerkung. Die Definition der Differenzierbarkeit fordert die Existenz *einer* linearen Abbildung, schließt aber a priori nicht aus, daß es mehrere gibt. Oben hatten wir aber schon den bestimmten Artikel benutzt und von *der* Ableitung, *dem* Differential etc. gesprochen. Das war eigentlich etwas voreilig, ist aber im nachhinein gerechtfertigt. Die darstellende Matrix von *irgend einem* $L_{\bar{a}}$ hatten wir als *die* aus den partiellen Ableitungen gebildete und somit eindeutig bestimmte JACOBI-Matrix gefunden.

3.2.3 Differenzierbarkeit und Stetigkeit

Jede im Punkt \bar{a} total differenzierbare Funktion ist dort auch stetig.

Da nach 1.2.7 lineare Abbildungen $\bar{\mathbb{R}}^n \rightarrow \mathbb{R}$ immer stetig sind, folgt aus $\bar{b} \rightarrow \bar{a}$ erst $\bar{b} - \bar{a} \rightarrow \vec{0}$ und dann $L_{\bar{a}}(\bar{b} - \bar{a}) \rightarrow 0$. Daher

$$\lim_{\bar{b} \rightarrow \bar{a}} f(\bar{b}) = \lim_{\bar{b} \rightarrow \bar{a}} [f(\bar{a}) + L_{\bar{a}}(\bar{b} - \bar{a}) + d(\bar{b}, \bar{a}) \cdot \sigma(\bar{b})] = f(\bar{a}).$$

3.2.4 Partielle und totale Differenzierbarkeit

Wir haben schon erwähnt (3.1.3), daß aus der alleinigen Existenz der partiellen Ableitungen nicht die Stetigkeit folgt. Also kann erst recht nicht die totale Differenzierbarkeit folgen.

Existieren aber alle partiellen Ableitungen von f in allen Punkten einer gewissen Umgebung des Punktes \bar{a} und sind die Funktionen $\frac{\partial f}{\partial x_i}$ alle stetig im Punkt \bar{a} , so ist f in \bar{a} total differenzierbar.

Ich **beweise** das im zweidimensionalen Fall. Seien $\frac{\partial f}{\partial x}$ und $\frac{\partial f}{\partial y}$ beide in einer Kugel mit Mittelpunkt (a, b) definiert und in (a, b) stetig. Wir definieren $\rho: \bar{\mathbb{R}}^2 \rightarrow \mathbb{R}$ für alle kleinen Vektoren $\begin{pmatrix} h \\ k \end{pmatrix}$ durch die Festsetzungen¹³ $\rho\begin{pmatrix} 0 \\ 0 \end{pmatrix} := 0$ und für $\begin{pmatrix} h \\ k \end{pmatrix} \neq \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

$$\rho\begin{pmatrix} h \\ k \end{pmatrix} := \frac{f(a+h, b+k) - f(a, b) - \frac{\partial f}{\partial x}(a, b) \cdot h - \frac{\partial f}{\partial y}(a, b) \cdot k}{\sqrt{h^2 + k^2}}.$$

Dann gilt jedenfalls

$$f(a+h, b+k) = f(a, b) + \frac{\partial f}{\partial x}(a, b) \cdot h + \frac{\partial f}{\partial y}(a, b) \cdot k + \sqrt{h^2 + k^2} \cdot \rho\begin{pmatrix} h \\ k \end{pmatrix}$$

und es bleibt zu zeigen, daß $\rho\begin{pmatrix} h \\ k \end{pmatrix} \rightarrow 0$ falls $\begin{pmatrix} h \\ k \end{pmatrix} \rightarrow \begin{pmatrix} 0 \\ 0 \end{pmatrix}$. Dazu genügt es, zu jedem $\varepsilon > 0$ ein derartiges $\delta > 0$ zu finden, daß

$$\left| f(a+h, b+k) - f(a, b) - \frac{\partial f}{\partial x}(a, b)h - \frac{\partial f}{\partial y}(a, b)k \right| < \varepsilon \sqrt{h^2 + k^2},$$

sobald $\sqrt{h^2 + k^2} < \delta$. Die Stetigkeit der beiden partiellen Ableitungen in (a, b) erlaubt es, ein $\delta > 0$ derart zu fixieren, daß

$$\left| \frac{\partial f}{\partial x}(a+\xi, b+\eta) - \frac{\partial f}{\partial x}(a, b) \right| < \frac{\varepsilon}{2} \quad \text{und} \quad \left| \frac{\partial f}{\partial y}(a+\xi, b+\eta) - \frac{\partial f}{\partial y}(a, b) \right| < \frac{\varepsilon}{2}$$

sobald $\sqrt{\xi^2 + \eta^2} < \delta$. Dieses δ leistet das verlangte. Sei nämlich $\sqrt{h^2 + k^2} < \delta$. Dann gilt

$$\left| f(a+h, b+k) - f(a, b) - \frac{\partial f}{\partial x}(a, b)h - \frac{\partial f}{\partial y}(a, b)k \right|$$

¹³Formal muß ρ für alle $\vec{v} \in \bar{\mathbb{R}}^2$ definiert sein. Wichtig sind aber nur die kleinen Vektoren, da nur der Grenzwert für $\vec{v} \rightarrow \vec{0}$ eine Rolle spielt. Wo die folgende Formel nichts liefert, kann man etwa $\rho = 100$ setzen.

$$\begin{aligned}
&\leq \left| f(a+h, b+k) - f(a, b+k) - \frac{\partial f}{\partial x}(a, b)h \right| + \left| f(a, b+k) - f(a, b) - \frac{\partial f}{\partial y}(a, b)k \right| \\
&\quad \text{Mittelwertsatz liefert } \xi \text{ und } \eta \text{ zwischen } 0 \text{ und } h \text{ bzw. } k \text{ mit} \\
&= \left| \frac{\partial f}{\partial x}(a+\xi, b+k) - \frac{\partial f}{\partial x}(a, b) \right| |h| + \left| \frac{\partial f}{\partial y}(a, b+\eta) - \frac{\partial f}{\partial y}(a, b) \right| |k| \\
&< \frac{\varepsilon}{2} (|h| + |k|) < \varepsilon \sqrt{h^2 + k^2}.
\end{aligned}$$

3.2.5 C^1 -Funktionen

Die Stetigkeit der partiellen Ableitungen hat schon zweimal eine rettende Rolle gespielt. Das ist Anlaß genug für folgende

Definition. Eine auf einer offenen Teilmenge von \mathbb{R}^n definierte Funktion heißt C^1 -Funktion, wenn sie in jedem Punkt ihres Definitionsbereiches alle partiellen Ableitungen besitzt und diese als Funktionen stetig sind.

Mit dieser Terminologie können wir die Aussagen 3.1.3 und 3.2.4 jetzt so formulieren: C^1 -Abbildungen sind stetig und total differenzierbar. Die Umkehrung ist übrigens falsch. Man erhält ein Gegenbeispiel, indem man eine einstellige differenzierbare Funktion $g: \mathbb{R} \rightarrow \mathbb{R}$ mit unstetiger Ableitung benutzt. $F(x, y) := g(x) + g(y)$ ist dann in jedem Punkt total differenzierbar aber nicht C^1 . Die Details bleiben als **Übungsaufgabe**.

stimmt das?

3.2.6 Der geometrische Aspekt: Tangentialebenen an Niveaulflächen

Später in dieser Vorlesung werden wir uns noch einmal und dann etwas strenger mit Tangentialebenen beschäftigen. Hier möchte ich nur einige heuristische Betrachtungen anstellen, die der Anschaulichkeit halber in \mathbb{R}^3 spielen.

Betrachten wir eine skalare Funktion $F(x, y, z)$. Durch den Punkt $\bar{p}_0 = (x_0, y_0, z_0)$ geht dann die ‘Niveaulfläche’

$$\Gamma = \{(x, y, z) : F(x, y, z) = c\}$$

wobei zur Abkürzung $c := F(x_0, y_0, z_0)$ gesetzt wurde. Wir wollen die Tangentialebene an Γ im Punkt \bar{p}_0 bestimmen. Das ist die Ebene, die durch \bar{p}_0 geht und sich bestmöglich an die Fläche anschmiegt. Letzteres ist geometrisch gar nicht so einfach (ehrlich gesagt: gar nicht) zu beschreiben. Deshalb nehmen wir Zuflucht zur Gleichung. Jede Ebene durch \bar{p}_0 hat eine Gleichung der Form $A \cdot (x - x_0) + B \cdot (y - y_0) + C \cdot (z - z_0) = 0$. Für die Tangentialebene sollen die Koeffizienten A, B, C so bestimmt werden, daß diese Ebenengleichung möglichst gut mit der Flächengleichung $F(x, y, z) = c$ bzw. $F(x, y, z) - F(x_0, y_0, z_0) = 0$ von Γ übereinstimmt.

Genau dieser Forderung ist aber die Definition der totalen Differenzierbarkeit angepaßt. Man ersetzt F durch die Approximation in \bar{p}_0 und läßt den Fehlerterm weg:

$$\frac{\partial F}{\partial x}(x_0, y_0, z_0) \cdot (x - x_0) + \frac{\partial F}{\partial y}(x_0, y_0, z_0) \cdot (y - y_0) + \frac{\partial F}{\partial z}(x_0, y_0, z_0) \cdot (z - z_0) = 0.$$

Das ist sinnvoll, wenn F in (x_0, y_0, z_0) total differenzierbar ist und nicht alle drei partiellen Ableitungen verschwinden (sonst handelt es sich nicht um eine Ebenengleichung). Letztere Bedingung kann man, im Moment etwas hergeholt klingend aber im Hinblick auf Verallgemeinerungen das Wesen der Sache genauer treffend, auch so formulieren: Die JACOBI-Matrix $J_F(x_0, y_0, z_0) = \left(\frac{\partial F}{\partial x}, \frac{\partial F}{\partial y}, \frac{\partial F}{\partial z} \right)$ von F muß im gegebenen Punkt vollen Rang haben. Die Wichtigkeit dieser Bedingung sieht man an der Funktion $F(x, y, z) = xyz$. Die Funktion ist C^1 aber die Gleichung $xyz = 0$ definiert die Vereinigung der drei Koordinatenebenen. Klar hat diese Fläche weder in $(0, 0, 0)$ noch in einem Punkt auf einer der Koordinatenachsen eine Tangentialebene.

Dieselbe Terminologie Niveaulfläche/Tangentialebene wird auch in Dimensionen > 3 benutzt. In \mathbb{R}^2 spricht man dagegen von Niveaulinien oder Höhenlinien (des von $F(x, y)$ beschriebenen

Gebirges) und Tangenten. Die Tangente in (x_0, y_0) an die durch $F(x, y) = c$ definierte Kurve hat also die Gleichung

$$\frac{\partial F}{\partial x}(x_0, y_0)(x - x_0) + \frac{\partial F}{\partial y}(x_0, y_0)(y - y_0) = 0.$$

3.2.7 Tangentialebenen an Funktionsgraphen

Sei f eine zweistellige Funktion mit dem Definitionsbereich $D \subseteq \mathbb{R}^2$. Ihr Graph ist die Fläche

$$\Gamma_f := \{(x, y, z) \in \mathbb{R}^3 : (x, y) \in D \text{ und } z = f(x, y)\}.$$

Ich finde die Vorstellung von einem ‘Gebirge’ ganz hilfreich. Wir wollen die im Punkt $\vec{p}_0 = (x_0, y_0, z_0) \in \Gamma_f$ angelegte Tangentialebene bestimmen. Dazu interpretieren wir Γ_f als Niveaufläche $\{F = 0\}$ der auf $D \times \mathbb{R}$ definierten Funktion $F(x, y, z) := f(x, y) - z$. Ihre partiellen Ableitungen in \vec{p}_0 sind

$$\frac{\partial F}{\partial x}(\vec{p}_0) = \frac{\partial f}{\partial x}(x_0, y_0), \quad \frac{\partial F}{\partial y}(\vec{p}_0) = \frac{\partial f}{\partial y}(x_0, y_0) \quad \text{und} \quad \frac{\partial F}{\partial z}(\vec{p}_0) = -1.$$

Deshalb hat die Tangentialebene die Gleichung

$$\frac{\partial f}{\partial x}(x_0, y_0)(x - x_0) + \frac{\partial f}{\partial y}(x_0, y_0)(y - y_0) - (z - z_0) = 0.$$

Stellt man diese noch nach z um, so ergibt sich, wenig verwunderlich, die rechte Seite der Näherungsgleichung, die entsteht, wenn in der Beziehung für totale Differenzierbarkeit der Fehlerterm weggelassen wird.

3.2.8 Differenzierbarkeit von Vektorfeldern

Die Definition unterscheidet sich nur durch ein paar kleine Pfeile von der für Skalarfelder gegebenen. Wir betrachten jetzt ein Vektorfeld $\vec{f} : D \rightarrow \mathbb{R}^m$, wobei D eine offene Teilmenge von \mathbb{R}^n sein soll. \vec{f} heißt *an der Stelle \vec{a} (total) differenzierbar*, wenn es eine lineare Abbildung $L_{\vec{a}} : \mathbb{R}^n \rightarrow \mathbb{R}^m$ und eine Funktion $\vec{\rho} : \mathbb{R}^n \rightarrow \mathbb{R}^m$ derart gibt, daß wann immer die linke Seite definiert ist

$$\vec{f}(\vec{a} + \vec{v}) = \vec{f}(\vec{a}) + L_{\vec{a}}(\vec{v}) + \|\vec{v}\| \cdot \vec{\rho}(\vec{v}) \quad \text{und} \quad \lim_{\vec{v} \rightarrow \vec{0}} \vec{\rho}(\vec{v}) = \vec{0}.$$

Alternative Beschreibungen: Der Zuwachs $\vec{\Delta}_{\vec{a}} \vec{f}(\vec{v}) = \vec{f}(\vec{a} + \vec{v}) - \vec{f}(\vec{a})$ besteht aus dem linearen Hauptbestandteil $L(\vec{v})$ und einem Fehler, der für $\vec{v} \rightarrow \vec{0}$ nicht nur gegen den Nullvektor aus \mathbb{R}^m geht, sondern sogar von höherer als erster Ordnung. Das heißt, selbst wenn man den Fehler durch $\|\vec{v}\|$ dividiert, geht der Quotient noch gegen $\vec{0}$.

Wieder kann man die entscheidende Gleichung auch nur mit Punkten schreiben (der Vektor \vec{v} versteckt sich in der Differenz $\vec{b} - \vec{a}$):

$$\vec{f}(\vec{b}) = \vec{f}(\vec{a}) + L_{\vec{a}}(\vec{b} - \vec{a}) + d(\vec{a}, \vec{b}) \cdot \vec{\sigma}(\vec{b}), \quad \text{wobei} \quad \lim_{\vec{x} \rightarrow \vec{a}} \vec{\sigma}(\vec{x}) = \vec{0}.$$

Die lineare Abbildung $L_{\vec{a}}$ geht jetzt von \mathbb{R}^n in \mathbb{R}^m . Sie wird von manchen Autoren *Ableitung* von \vec{f} an der Stelle \vec{a} genannt und folgerichtig mit $\vec{f}'(\vec{a})$ bezeichnet. Auch der Name *(total) Differential* ist gebräuchlich. Dann schreibt man gern $d_{\vec{a}} \vec{f}$. Ihre darstellende Matrix, die wieder JACOBI-Matrix heißt und mit $J_{\vec{f}}(\vec{a})$ bezeichnet wird, hat das Format $m \times n$. Wie im Falle eines Skalarfeldes leitet man her, daß die Einträge der JACOBI-Matrix die partiellen Ableitungen sind.

$$\text{Für } \vec{f} = \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_m \end{pmatrix} \quad \text{ergibt sich} \quad J_{\vec{f}}(\vec{a}) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(\vec{a}) & \frac{\partial f_1}{\partial x_2}(\vec{a}) & \cdots & \frac{\partial f_1}{\partial x_n}(\vec{a}) \\ \frac{\partial f_2}{\partial x_1}(\vec{a}) & \frac{\partial f_2}{\partial x_2}(\vec{a}) & \cdots & \frac{\partial f_2}{\partial x_n}(\vec{a}) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(\vec{a}) & \frac{\partial f_m}{\partial x_2}(\vec{a}) & \cdots & \frac{\partial f_m}{\partial x_n}(\vec{a}) \end{pmatrix}.$$

Man kann das (spaltenweise!) kompakt schreiben als

$$J_{\vec{f}}(\bar{a}) = \left(\frac{\partial \vec{f}}{\partial x_1}(\bar{a}), \quad \frac{\partial \vec{f}}{\partial x_2}(\bar{a}), \quad \dots, \quad \frac{\partial \vec{f}}{\partial x_n}(\bar{a}) \right).$$

Dagegen sind die Zeilen von $J_{\vec{f}}$ gerade die JACOBI-Matrizen der Koordinatenfunktionen f_1, f_2, \dots, f_m .

In Matrixschreibweise sieht die charakteristische Gleichung so aus:

$$\vec{f}(\bar{a} + \vec{v}) = \vec{f}(\bar{a}) + J_{\vec{f}}(\bar{a}) \cdot \vec{v} + \text{Fehlerterm.}$$

Für die einzelnen (skalaren) Koordinatenfunktionen ergibt sich

$$f_j(a_1 + v_1, \dots, a_n + v_n) = f_j(a_1, \dots, a_n) + \sum_{i=1}^n \frac{\partial f_j}{\partial x_i}(a_1, \dots, a_n) \cdot v_i + \text{Fehlerterm.}$$

Man sieht daran, daß ein Vektorfeld genau dann differenzierbar ist, wenn alle seine (skalaren) Koordinatenfunktionen differenzierbar sind. Das ist immer dann der Fall, wenn alle partiellen Ableitungen aller Koordinatenfunktionen stetig sind. In diesem Fall spricht man von einem C^1 -Vektorfeld.

3.2.9 Differenzierbarkeit von Punktfeldern

Unsere Unterscheidung von Punkt- und Vektorraum hat den Nachteil, daß man viele (fast) gleichlautende Dinge mehrmals erzählen muß. Eigentlich könnten Sie den Rest dieses Abschnittes ganz allein aufschreiben, aber ich deute wenigstens an.

Wir betrachten jetzt ein Punktfeld $\vec{f} : D \rightarrow \mathbb{R}^m$, wobei D wieder eine offene Teilmenge von \mathbb{R}^n sein soll. \vec{f} heißt *an der Stelle* \bar{a} (total) differenzierbar, wenn es eine lineare Abbildung $L_{\bar{a}} : \mathbb{R}^n \rightarrow \mathbb{R}^m$ und eine Funktion $\vec{\rho} : \mathbb{R}^n \rightarrow \mathbb{R}^m$ derart gibt, daß, wann immer die linke Seite definiert ist,

$$\vec{f}(\bar{a} + \vec{v}) = \vec{f}(\bar{a}) + L_{\bar{a}}(\vec{v}) + \|\vec{v}\| \cdot \vec{\rho}(\vec{v}) \quad \text{und} \quad \lim_{\vec{v} \rightarrow \vec{0}} \vec{\rho}(\vec{v}) = \vec{0}.$$

Obwohl wir die Ableitung eines Punktfeldes vor uns haben rechnet die lineare Abbildung Vektoren in Vektoren um! Sie wird als Ableitung oder Differential der Funktion f bezeichnet. Ihre darstellende Matrix $J_{\vec{f}}(\bar{a}) \in M(m \times n; \mathbb{R})$ hat wieder die partiellen Ableitungen der Koordinatenfunktionen als Einträge

$$\text{Für } \vec{f} = (f_1, f_2, \dots, f_m) \quad \text{ergibt sich} \quad J_{\vec{f}}(\bar{a}) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(\bar{a}) & \frac{\partial f_1}{\partial x_2}(\bar{a}) & \dots & \frac{\partial f_1}{\partial x_n}(\bar{a}) \\ \frac{\partial f_2}{\partial x_1}(\bar{a}) & \frac{\partial f_2}{\partial x_2}(\bar{a}) & \dots & \frac{\partial f_2}{\partial x_n}(\bar{a}) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(\bar{a}) & \frac{\partial f_m}{\partial x_2}(\bar{a}) & \dots & \frac{\partial f_m}{\partial x_n}(\bar{a}) \end{pmatrix}.$$

Man kann das (wieder spaltenweise!) kompakt schreiben als

$$J_{\vec{f}}(\bar{a}) = \left(\frac{\partial \vec{f}}{\partial x_1}(\bar{a}), \quad \frac{\partial \vec{f}}{\partial x_2}(\bar{a}), \quad \dots, \quad \frac{\partial \vec{f}}{\partial x_n}(\bar{a}) \right).$$

3.2.10 Wie sich Kurven einreihen

Kurven kann man als Punktfelder ansehen, die auf einem eindimensionalen Intervall definiert sind. Folgt man dem bisherigen Schema, so ist eine Kurve $\vec{\gamma} : I \rightarrow \mathbb{R}^n$ zum Zeitpunkt t differenzierbar, wenn es eine lineare Abbildung $L_t : \mathbb{R} \rightarrow \mathbb{R}^n$ und eine Vektorfunktion $\vec{\rho} : \mathbb{R} \rightarrow \mathbb{R}^n$ derart gibt, daß für alle kleinen h gilt

$$\vec{\gamma}(t + h) = \vec{\gamma}(t) + L_t(h) + |h| \cdot \vec{\rho}(h) \quad \text{wobei} \quad \lim_{h \rightarrow 0} \vec{\rho}(h) = \vec{0}.$$

Die reelle Zahl h spielt die Rolle des eindimensionalen Vektors. Wie sehen lineare Abbildungen $L : \mathbb{R} \rightarrow \mathbb{R}^n$ aus? Ist $\vec{a} = L(1)$, so gilt für alle h , daß $L(h) = L(h \cdot 1) = hL(1) = h\vec{a}$. Die Definition der Differenzierbarkeit fordert also die Existenz eines Vektors \vec{a} , so daß für alle hinreichend kleinen h gilt

$$\bar{\gamma}(t+h) = \bar{\gamma}(t) + h\vec{a} + \text{Fehlerterm.}$$

Klar muß dann \vec{a} die im vorigen Kapitel definierte Ableitung $\dot{\bar{\gamma}}(t)$ der Kurve im Punkt t sein und ihre Existenz ist der Differenzierbarkeit äquivalent. Die JACOBI-Matrix ist mit dem Geschwindigkeitsvektor identisch.

3.2.11 Der eindimensionale Fall

interessiert uns in diesem Semester weniger. Aber manchmal tauchen auch bei mehrdimensionalen Problemen auf natürliche Weise einstellige Funktionen auf, z.B. wenn man ein Skalarfeld längs einer Kurve untersucht. Dann müssen wir keine Panik bekommen; die mehrdimensionalen Konzepte spezialisieren sich auf natürliche Weise. (Man sollte versuchen, die drei Rollen auseinanderzuhalten, die die reellen Zahlen spielen können.) Die totale Differenzierbarkeit einer Funktion $f : I \rightarrow \mathbb{R}$ in einem Punkt a erweist sich als die übliche Differenzierbarkeit aus dem vorigen Semester. Die JACOBI-Matrix hat das Format 1×1 und hat $f'(a)$ als einzigen Eintrag.

3.2.12 Differenzierbarkeit zusammengesetzter Abbildungen; Kettenregel

Wir betrachten offene Mengen $D \subseteq \mathbb{R}^n$ und $E \subseteq \mathbb{R}^m$ und darauf definierte Felder $\bar{f} : D \rightarrow E$ und $\bar{g} : E \rightarrow \mathbb{R}^p$. Dann ist $\bar{g} \circ \bar{f} : D \rightarrow \mathbb{R}^p$ ein auf D definiertes Vektorfeld, nach dessen Differenzierbarkeit man fragen kann. Die Antwort gibt die

Kettenregel. *Ist \bar{f} im Punkt $\bar{a} \in D$ differenzierbar und \bar{g} im Punkt $\bar{f}(\bar{a}) \in E$ differenzierbar, so ist auch $\bar{g} \circ \bar{f}$ im Punkt \bar{a} differenzierbar und für die JACOBI-Matrizen gilt*

$$J_{\bar{g} \circ \bar{f}}(\bar{a}) = J_{\bar{g}}(\bar{f}(\bar{a})) \cdot J_{\bar{f}}(\bar{a}).$$

Klassischerweise nimmt man $\bar{f} = (f_1, f_2, \dots, f_m)$ als Funktion von $\bar{x} = (x_1, x_2, \dots, x_n)$ und $\bar{g} = \begin{pmatrix} g_1 \\ g_2 \\ \vdots \\ g_p \end{pmatrix}$ als Funktion von $\bar{y} = (y_1, \dots, y_m)$ an (um die partiellen Ableitungen unterscheiden zu können). Dann läßt sich das Matrizenprodukt ausschreiben:

$$\text{für } i = 1, 2, \dots, p \text{ und } j = 1, 2, \dots, n \text{ gilt } \frac{\partial g_i(f_1, \dots, f_m)}{\partial x_j}(\bar{a}) = \sum_{k=1}^m \frac{\partial g_i}{\partial y_k}(\bar{f}(\bar{a})) \cdot \frac{\partial f_k}{\partial x_j}(\bar{a}).$$

Achtung. Für die Gültigkeit dieser Formel ist es **nicht ausreichend**, daß alle beteiligten partiellen Ableitungen existieren. Die Abbildungen müssen tatsächlich ^{total} differenzierbar sein. Man ist auf der sicheren Seite, wenn man C^1 -Abbildungen voraussetzt: alle vorkommenden partiellen Ableitungen sind nicht nur vorhanden sondern sogar stetig.

Praktisch dieselbe Aussage gilt, wenn g kein Vektorfeld sondern ein Skalar- oder Punktfeld ist. Ich überlasse Ihnen die genaue Formulierung des Resultats.

Übrigens ist diese Kettenregel im Gegensatz zur eindimensionalen Variante, mit der man ständig rechnet, nicht primär dazu da, partielle Ableitungen und JACOBI-Matrizen konkreter Funktionen zu bestimmen. Für die Theorie ist sie aber ganz wichtig, wie wir später sehen werden.

Der **Beweis** der Kettenregel wird umso einfacher, je abstrakter man ihn führt. Wegen der vorausgesetzten Differenzierbarkeiten gibt es lineare Abbildungen $L : \mathbb{R}^n \rightarrow \mathbb{R}^m$ und $K : \mathbb{R}^m \rightarrow \mathbb{R}^p$, so daß

$$\bar{f}(\bar{a} + \vec{v}) = \bar{f}(\bar{a}) + L(\vec{v}) + \|\vec{v}\| \cdot \vec{\rho}(\vec{v}) \quad \text{und} \quad \bar{g}(\bar{f}(\bar{a}) + \vec{w}) = \bar{g}(\bar{f}(\bar{a})) + K(\vec{w}) + \|\vec{w}\| \cdot \vec{\sigma}(\vec{w})$$

jeweils für alle (kleinen) $\vec{v} \in \mathbb{R}^n$ und $\vec{w} \in \mathbb{R}^m$, wobei die Funktionen $\vec{\rho}$ und $\vec{\sigma}$ mit \vec{v} bzw. \vec{w} gegen den (jeweiligen) Nullvektor konvergieren.

Diese beiden Gleichungen setzen wir ineinander ein, indem wir $L(\vec{v}) + \|\vec{v}\| \cdot \vec{\rho}(\vec{v})$ als \vec{w} interpretieren.

$$\begin{aligned} \vec{g}(\vec{f}(\vec{a} + \vec{v})) &= \vec{g}\left(\vec{f}(\vec{a}) + L(\vec{v}) + \|\vec{v}\| \cdot \vec{\rho}(\vec{v})\right) \\ &= \vec{g}(\vec{f}(\vec{a})) + K\left(L(\vec{v}) + \|\vec{v}\| \cdot \vec{\rho}(\vec{v})\right) + \left\|L(\vec{v}) + \vec{\rho}(\vec{v}) \cdot \|\vec{v}\|\right\| \cdot \vec{\sigma}\left(L(\vec{v}) + \|\vec{v}\| \cdot \vec{\rho}(\vec{v})\right). \end{aligned}$$

Nun kann man die Linearität von K benutzen und dann (für $\vec{v} \neq \vec{0}$) sortieren:

$$\vec{g}(\vec{f}(\vec{a} + \vec{v})) = \vec{g}(\vec{f}(\vec{a})) + K(L(\vec{v})) + \|\vec{v}\| \cdot \underbrace{\left[K\left(\vec{\rho}(\vec{v})\right) + \left\| \frac{L(\vec{v})}{\|\vec{v}\|} + \vec{\rho}(\vec{v}) \right\| \right]}_{\cdot \vec{\sigma}\left(L(\vec{v}) + \|\vec{v}\| \cdot \vec{\rho}(\vec{v})\right)}.$$

Da die Hintereinanderausführung von K und L linear ist, bleibt zu sehen, daß der unterklammerte Ausdruck mit \vec{v} gegen den Nullvektor konvergiert. Einzig fraglich ist aber das Verhalten des Faktors $\left\| \frac{L(\vec{v})}{\|\vec{v}\|} + \vec{\rho}(\vec{v}) \right\|$. Um seine Harmlosigkeit einzusehen, erinnern wir uns an die in 1.2.7 bewiesene Beschränktheit linearer Abbildungen. Es ist also $\|L(\vec{v})\| \leq C \|\vec{v}\|$ für einen geeigneten Faktor C . Damit bleibt der Quotient beschränkt und stört die Nullkonvergenz des Restes nicht. Was beweist diese Gleichung? Wir haben $\vec{g}(\vec{f}(\vec{a} + \vec{v})) - \vec{g}(\vec{f}(\vec{a}))$ in den linearen Bestandteil $K(L(\vec{v}))$ und einen genügend schnell gegen Null strebenden Rest zerlegt. Das zeigt die Differenzierbarkeit von $\vec{g} \circ \vec{f}$. Die lineare Abbildung ergibt sich als Hintereinanderausführung der linearen Abbildungen K und L . Also ist die darstellende Matrix gleich dem Produkt der darstellenden Matrizen. Genau das war aber die Aussage.

3.2.13 Kettenregel, bodenständig aufgeschrieben

Es schadet sicher nichts, wenn wir noch eine Version der Kettenregel aufschreiben, die etwas bodenständiger aussieht. Meist will man eine zusammengesetzte skalare Funktion (partiell, wenn es mehrere sind) nach einer der vorkommenden Variablen differenzieren, also etwa $\frac{\partial F}{\partial v}$ berechnen, wobei $F(u, v) = h(u, f(u, v), g(u, v))$ aus drei C^1 -Funktionen zusammengesetzt ist. Die dreistellige Funktion h stellen wir uns als $h(x, y, z)$ vor. Dann gilt (der formal fehlende erste Summand $\frac{\partial h}{\partial x} \cdot \frac{\partial u}{\partial v}$ ist nämlich Null)

$$\frac{\partial F}{\partial v}(u, v) = \frac{\partial h}{\partial y}(u, f(u, v), g(u, v)) \cdot \frac{\partial f}{\partial v}(u, v) + \frac{\partial h}{\partial z}(u, f(u, v), g(u, v)) \cdot \frac{\partial g}{\partial v}(u, v).$$

Diese Formel ist gewissermaßen ein Ausschnitt der vollen Kettenregel, die sich in diesem Spezialfall als Matrizengleichung

$$\begin{pmatrix} \frac{\partial F}{\partial u} & \frac{\partial F}{\partial v} \end{pmatrix} = \begin{pmatrix} \frac{\partial h}{\partial x} & \frac{\partial h}{\partial y} & \frac{\partial h}{\partial z} \end{pmatrix} \cdot \begin{pmatrix} \frac{\partial u}{\partial u} & \frac{\partial u}{\partial v} \\ \frac{\partial f}{\partial u} & \frac{\partial f}{\partial v} \\ \frac{\partial g}{\partial u} & \frac{\partial g}{\partial v} \end{pmatrix}$$

darstellt.

3.3 Kurven durch Skalarfelder; Gradienten

3.3.1 Die Kettenregel für Funktionen auf Kurven

Ich will noch einen Spezialfall der Kettenregel diskutieren, der insofern etwas aus dem Rahmen fällt, als daß normale Ableitungen berechnet werden.

Gegeben sei eine Kurve $\vec{\gamma} = (\gamma_1, \gamma_2, \dots, \gamma_n) : I \rightarrow \mathbb{R}^n$ für ein Intervall $I \subset \mathbb{R}$. Ist dann $f : D \rightarrow \mathbb{R}$ ein Skalarfeld, durch dessen Definitionsbereich die Kurve verläuft, so ergibt sich $f \circ \vec{\gamma} : I \rightarrow \mathbb{R}$, eine ganz normale einstellige Funktion. Ist die Kurve zum Zeitpunkt t differenzierbar und ist

das Skalarfeld f im Raumpunkt $\bar{\gamma}(t)$ differenzierbar, so ist die Funktion $f \circ \bar{\gamma}$ zum Zeitpunkt t differenzierbar und hat die Ableitung

$$(*) \quad \frac{d}{dt} f(\bar{\gamma}(t)) = \frac{\partial f}{\partial x_1}(\bar{\gamma}(t)) \cdot \dot{\gamma}_1(t) + \frac{\partial f}{\partial x_2}(\bar{\gamma}(t)) \cdot \dot{\gamma}_2(t) + \dots + \frac{\partial f}{\partial x_n}(\bar{\gamma}(t)) \cdot \dot{\gamma}_n(t).$$

Im dreidimensionalen Fall, schreibt man (besonders als Praktiker) die Kurve auch als $(x(t), y(t), z(t))$ und die Kettenregel wird

$$\frac{d}{dt} f(x(t), y(t), z(t)) = \frac{\partial f}{\partial x} \cdot \dot{x} + \frac{\partial f}{\partial y} \cdot \dot{y} + \frac{\partial f}{\partial z} \cdot \dot{z}.$$

Zu beweisen brauchen wir nichts. Es handelt sich um einen Spezialfall der allgemeinen Kettenregel. Die JACOBI-Matrizen werden multipliziert:

$$\left((f \circ \bar{\gamma})'(t) \right) = J_{f \circ \bar{\gamma}}(t) = J_f(\bar{\gamma}(t)) \cdot J_{\bar{\gamma}}(t) = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \dots, \frac{\partial f}{\partial x_n} \right) \cdot \begin{pmatrix} \dot{\gamma}_1 \\ \dot{\gamma}_2 \\ \vdots \\ \dot{\gamma}_n \end{pmatrix}.$$

Dann entsteht (*) durch Vergleich der jeweils einzigen Einträge in den entstehenden 1×1 -Matrizen.

Übung. Ich empfehle Ihnen, sich die Formel noch einmal unabhängig von der allgemeinen Kettenregel durch Betrachtung der Differenz $f(x(t+h), y(t+h), z(t+h)) - f(x(t), y(t), z(t))$ klar zu machen. Der Trick mit dem achsenparallelen Umweg und Mittelwertsatz liefert

$$\begin{aligned} & \frac{f(x(t+h), y(t+h), z(t+h)) - f(x(t), y(t), z(t))}{h} \\ &= \frac{1}{h} [f(x(t+h), y(t+h), z(t+h)) - f(x(t), y(t+h), z(t+h))] + \dots \\ &= \frac{1}{h} \left[\frac{\partial f}{\partial x}(x(t) + \xi, y(t+h), z(t+h)) \cdot [x(t+h) - x(t)] + \dots \right] \\ &= \frac{\partial f}{\partial x}(x(t) + \xi, y(t+h), z(t+h)) \cdot \frac{x(t+h) - x(t)}{h} + \dots \end{aligned}$$

Setzen Sie die Existenz und Stetigkeit aller vorkommenden (partiellen) Ableitungen voraus.

3.3.2 Der Gradient

Statt als Produkt der JACOBI-Matrizen interpretiert man die rechte Seite

$$\frac{\partial f}{\partial x_1}(\bar{\gamma}(t)) \cdot \dot{\gamma}_1(t) + \frac{\partial f}{\partial x_2}(\bar{\gamma}(t)) \cdot \dot{\gamma}_2(t) + \dots + \frac{\partial f}{\partial x_n}(\bar{\gamma}(t)) \cdot \dot{\gamma}_n(t).$$

der obigen Gleichung (*) auch gern als Skalarprodukt. Ein Faktor ist $\dot{\bar{\gamma}}(t)$, der Geschwindigkeitsvektor der Kurve. Er hängt nur von der Kurve und nicht von dem Feld ab. Der andere Faktor ist technisch gesehen die transponierte JACOBI-Matrix von f und hängt nur von dem Verhalten des Feldes ab. Er wird üblicherweise *Gradient* des Skalarfeldes genannt und mit $\vec{\text{grad}} f$ bezeichnet¹⁴:

$$\vec{\text{grad}} f(\bar{a}) = J_f(\bar{a})^T = \begin{pmatrix} \frac{\partial f}{\partial x_1}(\bar{a}) \\ \frac{\partial f}{\partial x_2}(\bar{a}) \\ \vdots \\ \frac{\partial f}{\partial x_n}(\bar{a}) \end{pmatrix}.$$

Die Formel (*) liest sich mit der neuen Bezeichnung so:

$$\frac{d}{dt} f(\bar{\gamma}(t)) = \vec{\text{grad}} f(\bar{\gamma}(t)) \bullet \dot{\bar{\gamma}}(t).$$

¹⁴Der Vektorpfeil ist meine Extravaganz und nicht unbedingt üblich.

3.3.3 Eine Anwendung

Ist der Definitionsbereich $D \subset \mathbb{R}^n$ des Skalarfeldes f offen und zusammenhängend und sind alle partiellen Ableitungen $\frac{\partial f}{\partial x_i}$ in allen Punkten von D gleich Null, ist also $\text{grad } f = \vec{0}$, so muß f konstant sein.

Beweis. Die Idee ist, Kurven durch das Gebiet zu schicken. Wegen des Zusammenhangs kommt man überall hin.

Um das auszuführen, sei $\bar{a} \in D$ fest gewählt und \bar{b} ein beliebiger Punkt aus D . Wir müssen $f(\bar{b}) = f(\bar{a})$ nachweisen. Dazu nehmen wir eine Kurve $\bar{\gamma} : [0, 1] \rightarrow D$ mit $\bar{\gamma}(0) = \bar{a}$ und $\bar{\gamma}(1) = \bar{b}$. Auf die Funktion $\varphi = f \circ \bar{\gamma}$ wird der Mittelwertsatz der Differentialrechnung angewendet $\varphi(1) - \varphi(0) = \varphi'(\tau)$. Ausgeschrieben und nach Kettenregel differenziert:

$$f(\bar{b}) - f(\bar{a}) = \sum_{i=1}^n \frac{\partial f}{\partial x_i}(\bar{\gamma}(\tau)) \cdot \dot{\gamma}_i(\tau) = \sum_{i=1}^n 0 \cdot \dot{\gamma}_i(\tau) = 0.$$

Also ist $f(\bar{b}) = f(\bar{a})$.

Das ist der Beweis, wie ihn wahrscheinlich jeder Physiker (sobald er aus der Mathematik-Vorlesung heraus ist) akzeptieren würde. Genau genommen, d.h. für Mathematiker hat er einen Haken. Der Zusammenhang von D garantiert uns zwar die Existenz einer Kurve, aber es ist nicht ohne weiteres klar, daß diese differenzierbar ist. Glücklicherweise hatten wir uns aber früher schon mal überlegt (2.2.5), daß man bei offenen zusammenhängenden Mengen immer eine sogar reguläre Verbindungskurve findet.

3.3.4 Der Mittelwertsatz für Skalarfelder

$f : D \rightarrow \mathbb{R}$ sei ein auf der offenen Menge D definiertes C^1 -Skalarfeld. Liegt die Verbindungsstrecke $\{\bar{a} + t(\bar{b} - \bar{a}) : 0 \leq t \leq 1\}$ der Punkte \bar{a} und \bar{b} ganz in D , so gibt es einen Punkt \bar{c} auf (sogar im Inneren) dieser Strecke für den gilt $f(\bar{b}) - f(\bar{a}) = \sum_{i=1}^n \frac{\partial f}{\partial x_i}(\bar{c})(b_i - a_i)$.

Hier haben wir mit der Differenzierbarkeit kein Problem. Wir betrachten die geradlinig gleichförmige Bewegung von \bar{a} nach \bar{b} , also die Kurve $\bar{\gamma}(t) = \bar{a} + t(\bar{b} - \bar{a})$. In Koordinaten: $\gamma_1(t) = a_1 + t(b_1 - a_1), \dots, \gamma_n(t) = a_n + t(b_n - a_n)$. Nun wenden wir auf die Funktion $\varphi(t) = f(\gamma_1(t), \dots, \gamma_n(t))$ den Mittelwertsatz der Differentialrechnung aus dem letzten Semester an:

$$(*) \quad \varphi(1) - \varphi(0) = \varphi'(\tau)$$

für einen geeigneten Zeitpunkt $\tau \in]0, 1[$. Setzt man $\bar{c} = \bar{\gamma}(\tau) = \bar{a} + \tau(\bar{b} - \bar{a})$, so verwandelt sich (*) in die behauptete Gleichung.

Manchmal ist folgende Variante des Mittelwertsatzes passender:

$$f(\bar{a} + \vec{v}) - f(\bar{a}) = \sum_{i=1}^n \frac{\partial f}{\partial x_i}(\bar{a} + \tau\vec{v})v_i = \vec{\text{grad}} f(\bar{a} + \tau\vec{v}) \bullet \vec{v}$$

für ein $\tau \in [0, 1]$. Sie entsteht aus der vorigen, indem $\bar{b} - \bar{a} = \vec{v}$ gesetzt wird.

3.3.5 Unabhängigkeit von Variablen

Ist $f : D \rightarrow \mathbb{R}$ ein auf der offenen konvexen Menge D definiertes C^1 -Skalarfeld, dessen partielle Ableitung nach x_i überall verschwindet, so hängt f nicht von der i -ten Koordinate ab. Dasselbe gilt für Vektorfelder.

Zum **Beweis** betrachten wir beliebige Punkte $\bar{a}, \bar{b} \in D$, die sich nur in der i -ten Koordinate unterscheiden. Dann folgt aber $f(\bar{b}) = f(\bar{a})$ sofort aus dem Mittelwertsatz. Nach Voraussetzung

ist die Strecke von \bar{a} nach \bar{b} ganz in D enthalten, bis auf die i -te sind alle Differenzen $b_j - a_j = 0$ und $\frac{\partial f}{\partial x_i}(\bar{c}) = 0$, also

$$f(\bar{b}) - f(\bar{a}) = \sum_{j=1}^n \frac{\partial f}{\partial x_j}(\bar{c})(b_j - a_j) = 0.$$

Kurioserweise gilt die Aussage nicht mehr für beliebige offene zusammenhängende Definitionsbereiche.

3.3.6 Die Richtungsableitung eines Skalarfeldes

Wir betrachten wieder einen inneren Punkt \bar{a} des Definitionsbereiches $D \subset \mathbb{R}^n$ eines Skalarfeldes f . Dazu wird in beliebiger Richtung noch ein Einheitsvektor \vec{e} fixiert. Wegen der Offenheit kann man von \bar{a} aus in Richtung $\pm\vec{e}$ jeweils ein kleines Stück gehen, ohne den Definitionsbereich von f zu verlassen. Ist also $\delta > 0$ genügend klein, so ist die Funktion $t \mapsto f(\bar{a} + t\vec{e})$ für $t \in]-\delta, \delta[$ definiert. Ihre Ableitung im Nullpunkt wird Richtungsableitung von f im Punkt \bar{a} und in Richtung \vec{e} genannt und mit $\frac{\partial f}{\partial \vec{e}}(\bar{a})$ bezeichnet. Als Formel

$$\frac{\partial f}{\partial \vec{e}}(\bar{a}) = \lim_{h \rightarrow 0} \frac{f(\bar{a} + h\vec{e}) - f(\bar{a})}{h}.$$

Offensichtlich sind die oben diskutierten partiellen Ableitungen auch Richtungsableitungen und zwar gerade die Ableitungen in Richtung der Koordinatenachsen, sprich: der kanonischen Basisvektoren. $\frac{\partial f}{\partial x_i} = \frac{\partial f}{\partial \vec{e}_i}$.

Um einem **möglichen Mißverständnis** vorzubeugen: Der Grenzwert $\lim_{h \rightarrow 0}$ bei der Berechnung der Richtungsableitung ist beidseitig gemeint (positive und negative h). Als Richtung darf man sich nicht nur einen Strahl vorstellen, sondern die ganze Gerade. Allerdings mit einer Orientierung ('Durchlaufsin') versehen. Folglich ändert die Richtungsableitung auch nur das Vorzeichen, wenn der Vektor \vec{e} gegen $-\vec{e}$ ausgetauscht wird.

Beobachtung. Wenn die Funktion f im Punkt \bar{a} total differenzierbar ist, so existieren alle Richtungsableitungen. Sie lassen sich aus den partiellen Ableitungen und den Koordinaten des Richtungsvektors nach der Formel

$$\frac{\partial f}{\partial \vec{e}}(\bar{a}) = \frac{\partial f}{\partial x_1}(\bar{a}) \cdot e_1 + \frac{\partial f}{\partial x_2}(\bar{a}) \cdot e_2 + \dots + \frac{\partial f}{\partial x_n}(\bar{a}) \cdot e_n = \vec{\text{grad}} f(\bar{a}) \bullet \vec{e}$$

berechnen.

Um das einzusehen betrachten wir die Kurve $\bar{\gamma}(t) = \bar{a} + t\vec{e}$ mit dem konstanten Geschwindigkeitsvektor \vec{e} . Dann wird offensichtlich, daß

$$\frac{\partial f}{\partial \vec{e}}(\bar{a}) = \frac{d}{dt} f(\bar{\gamma}(0)) = \vec{\text{grad}} f(\bar{\gamma}(0)) \bullet \dot{\bar{\gamma}}(0) = \vec{\text{grad}} f(\bar{a}) \bullet \vec{e},$$

wie behauptet.

Achtung. Ohne die Voraussetzung der totalen Differenzierbarkeit gilt diese Aussage nicht. $\vec{\text{grad}} f$ kann existieren, ohne daß auch andere Richtungsableitungen existieren müssen.

Er meint $\partial f / \partial \vec{e}$

Bemerkung. Die gefundene Formel kann auch gewissermaßen rückwärts gelesen werden. Dann charakterisiert sie den Gradienten von f im Punkt \bar{a} als denjenigen Vektor, dessen Komponente in Richtung des Einheitsvektors \vec{e} die Richtungsableitung ist.

3.3.7 Die Extremaleigenschaft des Gradienten.

f sei ein C^1 -Skalarfeld. Ist $\vec{\text{grad}} f(\bar{a}) \neq \vec{0}$, so zeigt der Gradient in die Richtung des stärksten Wachstums der Funktion.

Genauer: Ist $\vec{g} = \frac{\vec{\text{grad}} f}{\|\vec{\text{grad}} f\|}$ der Einheitsvektor in Gradientenrichtung und \vec{e} ein beliebiger Einheitsvektor, so gilt

$$\left| \frac{\partial f}{\partial \vec{e}}(\bar{a}) \right| \leq \left\| \vec{\text{grad}} f(\bar{a}) \right\| = \frac{\partial f}{\partial \vec{g}}(\bar{a}).$$

Das folgt aus der obigen Charakterisierung der Richtungsableitung und der CAUCHY-SCHWARZschen Ungleichung ($\|\vec{e}\| = 1$):

$$\left| \frac{\partial f}{\partial \vec{e}} \right| = \left| \text{grad } f \cdot \vec{e} \right| \leq \left\| \text{grad } f \right\| \cdot \|\vec{e}\| = \left\| \text{grad } f \right\| = \text{grad } f \cdot \frac{\text{grad } f}{\left\| \text{grad } f \right\|} = \text{grad } f \cdot \vec{g} = \frac{\partial f}{\partial \vec{g}}.$$

Warnung. Auch wenn man im Gebirge (Funktionsgraph) auf einer Kammlinie, einem Gipfel oder einem Sattelpunkt steht, gibt es meist eine Richtung, in der es am stärksten bergauf/bergab geht. Diese wird aber *nicht* vom Gradienten angezeigt, da dieser als Nullvektor dann keine Richtung hat.

3.3.8 Geometrische Interpretation der Gradientenrichtung

Ich betrachte ein in \mathbb{R}^3 definiertes C^1 -Skalarfeld $F(x, y, z)$ und setze $c := F(x_0, y_0, z_0)$. In 3.2.7 hatten wir die Gleichung der Tangentialebene an die Niveaufäche $\Gamma = \{(x, y, z) : F(x, y, z) = c\}$ im Punkt $\bar{p}_0 = (x_0, y_0, z_0)$ als

$$\frac{\partial F}{\partial x}(x_0, y_0, z_0) \cdot (x - x_0) + \frac{\partial F}{\partial y}(x_0, y_0, z_0) \cdot (y - y_0) + \frac{\partial F}{\partial z}(x_0, y_0, z_0) \cdot (z - z_0) = 0$$

gefunden. Bedingung war, daß nicht alle partiellen Ableitungen verschwinden. Diese Gleichung lesen wir jetzt als

$$0 = \begin{pmatrix} \frac{\partial F}{\partial x}(x_0, y_0, z_0) \\ \frac{\partial F}{\partial y}(x_0, y_0, z_0) \\ \frac{\partial F}{\partial z}(x_0, y_0, z_0) \end{pmatrix} \cdot \begin{pmatrix} x - x_0 \\ y - y_0 \\ z - z_0 \end{pmatrix} = \vec{\text{grad}} F(\bar{p}_0) \cdot (\bar{p} - \bar{p}_0).$$

Das gilt für alle \bar{p} aus der Tangentialebene und läßt sich so interpretieren: der Gradient steht auf der Tangentialebene an die Niveaufäche senkrecht. Meist kürzt man das ab und sagt, daß der Gradient auf der Niveaufäche selbst senkrecht steht. Die Aussage gilt auch, wenn der Gradient verschwindet; nur hat sie dann keinen Nährwert.

Dasselbe gilt natürlich auch in höheren Dimensionen ($n > 3$) und für Niveaulinien ($n = 2$).

3.4 Implizite Funktionen I: der skalare Fall

3.4.1 Heuristische Beschreibung des Problems

Um zu verstehen, worum es in diesem Abschnitt eigentlich geht, erinnern wir uns an den Abbildungsbegriff. Jedem Element aus dem Definitionsbereich wird genau ein Element aus dem Wertebereich zugeordnet. Die Elemente können dabei auch unter- oder nebeneinandergeschriebene Paare, Tripel, etc. von Zahlen sein. Mehrstellige Funktionen, Vektorfelder etc. fallen also auch unter diesen Begriff.

Angenommen F ist eine Funktion, die jedem Zahlenpaar $(x, y) \in D \subseteq \mathbb{R}^2$ eine reelle Zahl zuordnet. Dann kann man folgende Vorschrift formulieren

(*) ordne einem gegebenen x dasjenige y zu, für das $F(x, y) = 0$ gilt

und hoffen, daß auf diese Weise eine neue jetzt einstellige Funktion entsteht. Wenn das klappt, spricht man von der durch die Gleichung $F(x, y) = 0$ definierten impliziten Funktion. In den meisten Fällen klappt es jedoch nicht, weil zu manchen gegebenen x gar kein oder mehrere y existieren. Man kann die Chancen von (*) verbessern, indem man nicht alle denkbaren x -Werte sondern nur solche aus einem (kleinen) Intervall zuläßt und auch von den y -Werten zusätzlich verlangt, daß sie in einem kleinen Intervall liegen. Wenn das funktioniert, so wird y durch die Gleichung $F(x, y) = 0$ lokal als implizite Funktion von x definiert.

Bevor ich zu einer exakten Definition komme, will ich dasselbe Problem noch geometrisch beschreiben. Die Gleichung $F(x, y) = 0$ definiert in der Ebene eine Niveaulinie (zumindest bei

anständigen Funktionen F). Man kann fragen, ob dieselbe Linie auch als Funktionsgraph aufgefaßt werden kann. Selbst bei sehr anständigen Funktionen klappt das nicht. Man mache sich das Problem am Beispiel der Kreisgleichung $x^2 + y^2 = 1$ klar. Der Einheitskreis ist kein Funktionsgraph. Pickt man sich aber einen Punkt auf der Kreislinie heraus, so kann man fast immer ein kleines offenes Rechteck um diesen Punkt legen, innerhalb dessen die Kreislinie zum Funktionsgraphen wird. Ausnahmen bilden die beiden Punkte $(-1, 0)$ und $(1, 0)$. Jedes Rechteck, das einen dieser Punkte im Inneren enthält, muß zu gewissen x -Werten sowohl positive als auch negative y -Werte einschließen, deshalb ist die Kreislinie in der Nähe dieser Punkte kein Funktionsgraph. Die schlechten Punkte erkennt man daran, daß in ihnen die Tangenten an den Kreis parallel zur y -Achse verlaufen.

Das ganze noch einmal rechnerisch gesehen: Es ist global keine (eindeutige!) Auflösung $y = \dots$ der Kreisgleichung $x^2 + y^2 = 1$ möglich. Um fast alle Punkte (a, b) mit $a^2 + b^2 = 1$ läßt sich aber ein Rechteck legen, in dem die Kreisgleichung $x^2 + y^2 = 1$ nach y aufgelöst werden kann. Entweder $y = \sqrt{1 - x^2}$ falls das Rechteck oberhalb der x -Achse liegt, oder $y = -\sqrt{1 - x^2}$, falls das Rechteck unterhalb der x -Achse liegt. Nur in der Nähe der Punkte $(\pm 1, 0)$ geht das nicht. Leider sind nicht alle praktischen Beispiele so einfach zu durchschauen: uneingeweiht erkennt man sicher nicht, ob die Gleichung $x^2 + y^2 + y^5 = 3$ in der Nähe des Punktes $(1, 1)$ eine Funktion definiert.

3.4.2 Definition

Als **Generalvoraussetzung** für den ganzen Abschnitt nehmen wir an, daß die zweistellige Funktion $F(x, y)$ in einer offenen Teilmenge von \mathbb{R}^2 definiert ist und im gesamten Definitionsbereich stetige partielle Ableitungen besitzt. Weiter sei $F(a, b) = 0$.

Wir sagen, daß die Funktion $F(x, y)$ in der Nähe des Punktes (a, b) die Variable y lokal als (implizite) Funktion von x definiert, wenn es offene Intervalle $I \ni a$ und $J \ni b$ und eine Funktion $f : I \rightarrow \mathbb{R}$ derart gibt, daß für $x \in I, y \in J$

$$F(x, y) = 0 \quad \iff \quad y = f(x).$$

Man kann dasselbe auch so sagen: es gibt offene Intervalle $I \ni a$ und $J \ni b$ sowie eine eindeutig bestimmte Funktion $f : I \rightarrow J$ mit $f(a) = b$ und $F(x, f(x)) = 0$ für alle $x \in I$.

3.4.3 Der eindimensionale Satz über implizite Funktionen

F sei eine C^1 -Funktion und $F(a, b) = 0$. Wenn $\frac{\partial F}{\partial y}(a, b) \neq 0$, so definiert F die Variable y in der Nähe von (a, b) als implizite Funktion von x . Die entsprechende Funktion $f : I \rightarrow \mathbb{R}$ ist stetig und differenzierbar (im genügend klein gewählten offenen Intervall I) und ihre Ableitung berechnet sich als $f'(x) = -\frac{\frac{\partial F}{\partial x}(x, f(x))}{\frac{\partial F}{\partial y}(x, f(x))}$. Speziell ist $f'(a) = -\frac{\frac{\partial F}{\partial x}(a, b)}{\frac{\partial F}{\partial y}(a, b)}$.

Der **Beweis** verläuft in mehreren Schritten (machen Sie beim Lesen unbedingt Bilder!). Zunächst finden wir I und J , so daß tatsächlich in I eine Funktion definiert wird, d.h. zu jedem $x \in I$ genau ein $y \in J$ derart existiert, daß $F(x, y) = 0$.

OBdA nehmen wir $\frac{\partial F}{\partial y}(a, b) > 0$ an (sonst wird mit -1 multipliziert). Wegen der vorausgesetzten Stetigkeit der partiellen Ableitung gibt es dann ein Rechteck $s < x < t, \quad u < y < v$, das (a, b) enthält und in dem $\frac{\partial F}{\partial y}$ positiv ist.

Bei jedem festen $x \in]s, t[$ ist die einstellige Funktion $y \mapsto F(x, y)$ auf $]u, v[$ streng monoton wachsend, denn ihre Ableitung ist strikt positiv. Das gilt speziell auch für die Funktion $F(a, y)$. Da diese im Punkt b Null ist, muß sie im Punkt u echt kleiner und im Punkt v echt größer Null sein: $F(a, u) < 0$ und $F(a, v) > 0$.

Jetzt halten wir u fest und betrachten $F(x, u)$ als Funktion von x . Sie ist stetig in $]s, t[$ und im Punkt a negativ. Dann ist sie auch in einer Umgebung von a negativ. Indem wir das Intervall $]s, t[$ notfalls nachträglich verkleinern, können wir annehmen, daß $F(x, u) < 0$ für alle $x \in]s, t[$. Dasselbe Argument angewendet auf $F(x, v)$ erlaubt uns nach eventuellem nochmaligem Verkleinern von

$]s, t[$ anzunehmen, daß $F(x, v) > 0$ für alle $x \in]s, t[$. Ich behaupte, daß $I =]s, t[$ und $J =]u, v[$ die verlangten Intervalle sind.

Tatsächlich ist für jedes $x \in]s, t[$ die Funktion $y \mapsto F(x, y)$ streng wachsend auf $]u, v[$, negativ im Anfangspunkt u und positiv im Endpunkt v . Also gibt es genau eine Nullstelle, d.h. ein $y \in]u, v[$ mit $F(x, y) = 0$.

Als nächstes wollen wir die Stetigkeit der impliziten Funktion einsehen. Da auf alle Punkte $(x, f(x))$ mit $x \in I$ dieselbe Voraussetzung (nämlich $\frac{\partial F}{\partial y} \neq 0$) wie auf (a, b) zutrifft, genügt es, die Stetigkeit von f im Punkt a nachzuweisen. Die folgt aus den bereits angestellten Überlegungen; wir müssen uns das bloß noch klar machen. Sei $\varepsilon > 0$ gegeben. Indem wir ε notfalls verkleinern, können wir $u < b - \varepsilon < b < b + \varepsilon < v$ annehmen. Wegen der strengen Monotonie von $F(a, y)$ und $F(a, b) = 0$, muß $F(a, b - \varepsilon) < 0 < F(a, b + \varepsilon)$ gelten. Die Stetigkeit von $F(x, b \pm \varepsilon)$ liefert dann ein $\delta > 0$, so daß $F(x, b \pm \varepsilon) \lesseqgtr 0$ für $x \in [a - \delta, a + \delta[$. Für diese x -Werte liegt also die Nullstelle $f(x)$ von $F(x, y)$ zwischen $b - \varepsilon$ und $b + \varepsilon$. Mit anderen Worten: aus $|x - a| < \delta$ folgt $|f(x) - b| < \varepsilon$, wie von der Stetigkeit verlangt.

Schließlich wollen wir noch zeigen, daß die implizite Funktion f eine Ableitung besitzt, die sich nach der angegebenen Formel berechnet. Wieder genügt es, den Punkt a zu betrachten, weil auf die anderen Punkte $(x, f(x))$ aus $I \times J$ dieselben Voraussetzungen wie auf (a, b) zutreffen.

Wir lassen h eine kleine Zahl $\neq 0$ sein und schreiben für $f(a + h) - f(a)$ kurzzeitig Δf . Es geht um den Grenzwert des Differenzenquotienten $\frac{\Delta f}{h}$ für $h \rightarrow 0$.

Nach Mittelwertsatz 3.3.4 gibt es für alle kleinen h und k , ein passendes τ zwischen 0 und 1, so daß

$$F(a + h, b + k) - F(a, b) = \frac{\partial F}{\partial x}(a + \tau h, b + \tau k) \cdot h + \frac{\partial F}{\partial y}(a + \tau h, b + \tau k) \cdot k.$$

Setzt man hier $k = \Delta f$ und berücksichtigt $f(a) = b$, so hat man

$$F(a + h, b + \Delta f) = F(a + h, f(a + h)) = 0 = F(a, b), \text{ also}$$

$$0 = \frac{\partial F}{\partial x}(a + \tau h, b + \tau \Delta f) \cdot h + \frac{\partial F}{\partial y}(a + \tau h, b + \tau \Delta f) \cdot \Delta f \quad \text{oder} \quad \frac{\Delta f}{h} = - \frac{\frac{\partial F}{\partial x}(a + \tau h, b + \tau \Delta f)}{\frac{\partial F}{\partial y}(a + \tau h, b + \tau \Delta f)}.$$

(Warum ist der letzte Nenner für alle kleinen h von Null verschieden?) Jetzt läßt man $h \rightarrow 0$ gehen. Dann geht Δf gegen Null (Stetigkeit von f schon bewiesen) und wegen der Stetigkeit der partiellen Ableitungen von F , ergibt sich als Grenzwert tatsächlich

$$f'(a) = - \frac{\frac{\partial F}{\partial x}(a, b)}{\frac{\partial F}{\partial y}(a, b)}.$$

Bemerkung. Die Bedingung $\frac{\partial F}{\partial y} \neq 0$ ist hinreichend für die Existenz der impliziten Funktion. Sie ist aber nicht notwendig: Die Gleichung $x^3 - y^3 = 0$ definiert überall die identische Funktion. Aber im Nullpunkt sind beide partiellen Ableitungen Null.

Bemerkung. Die Formel für die Ableitung kann man sich immer wieder schnell herleiten, indem die Gleichung $F(x, f(x)) = 0$ nach der Kettenregel differenziert wird

$$\frac{\partial F}{\partial x}(x, f(x)) + \frac{\partial F}{\partial y}(x, f(x)) \cdot f'(x) = 0$$

und dann nach $f'(x)$ umgestellt. Auch praktisch wird man meist so rechnen. Diese Rechnung ersetzt den obigen Beweis allerdings nicht, denn bevor man so rechnen kann, muß man wissen, daß f differenzierbar ist.

3.4.4 Beispiele

Die Gleichung $x^2 + y^2 + y^5 = 3$ wird vom Punkt $(1, 1)$ erfüllt. In der Nähe dieses Punktes handelt es sich bei dieser Niveaulinie um einen Funktionsgraphen, denn

$$\frac{\partial[x^2 + y^2 + y^5 - 3]}{\partial y}(1, 1) = 2y + 5y^4|_{(1,1)} = 7 \neq 0.$$

Die Ableitung der entsprechenden Funktion ergibt sich durch Differenzieren der Gleichung (nach x) und anschließendes Umstellen:

$$2x + 2yy' + 5y^4y' = 0 \rightsquigarrow y' = \frac{-2x}{2y + 5y^4}, \quad \text{speziell } y'(1) = \frac{-2}{2+5} = -\frac{2}{7}.$$

Zweites Beispiel. Die Kurve¹⁵ mit der Gleichung $x^3 + y^3 = 3axy$ nennt man *Kartesisches Blatt*. Dabei ist a ein positiver Parameter. Wir wollen Sie im Bereich $x \geq 0, y \geq 0$ untersuchen. Dort handelt es sich um eine geschlossene Kurve, bei der ein maximaler y -Wert auftreten muß, den wir bestimmen wollen.

Nehmen wir erst einmal an, das passiert in einem Punkt (x, y) , in dem $F(x, y) = x^3 + y^3 - 3axy$ die Variable y als Funktion von x definiert. Dann kann die Ableitung y' dieser Funktion im Punkt x ausgerechnet werden indem die Gleichung $x^3 + y^3 = 3axy$ nach x differenziert wird und dabei y als Funktion von x angesehen.

$$3x^2 + 3y^2y' = 3ay + 3axy'.$$

Da es sich um einen Extrempunkt handelt, muß $y' = 0$ sein, das liefert die Gleichung $3x^2 = 3ay$. Das kann man nach y umstellen und in die Gleichung der Kurve einsetzen:

$$x^3 + \left(\frac{x^2}{a}\right)^3 = 3ax\frac{x^2}{a} \quad \text{oder } x^6 = 2x^3a^3.$$

Da wir nur an Lösungen ≥ 0 interessiert sind, ergeben sich 0 und $a\sqrt[3]{2}$. Aus $x = 0$ folgt $y = 0$. Der zweite Wert liefert $y = \frac{x^2}{a} = a\sqrt[3]{4}$. Dieser Wert ist heißer Kandidat für das gesuchte Maximum. Allerdings müssen wir noch die Kurvenpunkte untersuchen, in denen y nicht als Funktion von x definiert wird. Das sind die Punkte wo $\frac{\partial F}{\partial y} = 0$, also

$$3y^2 - 3ax = 0.$$

Diese Gleichung kann man nun nach x umstellen, einsetzen wie oben usw. Es ergibt sich einfach das symmetrische Ergebnis

$$x = a\sqrt[3]{4} \quad y = a\sqrt[3]{2} \quad \text{bzw. } x = y = 0.$$

Da $\sqrt[3]{2} < \sqrt[3]{4}$, ist der zuletzt erhaltene y -Wert kleiner, also das Maximum $y = a\sqrt[3]{4}$ bestätigt.

3.4.5 Implizite skalare Funktionen mehrerer Veränderlicher

Jetzt geht es darum, eine Gleichung der Form

$$F(x_1, x_2, \dots, x_n, y) = 0$$

in der Nähe eines Punktes $(a_1, a_2, \dots, a_n, b)$ nach y umzustellen. Es ist im gegebenen Zusammenhang bequem, Punkte von \mathbb{R}^{n+1} als (\bar{x}, y) zu schreiben. Das Ergebnis und sein Beweis ergeben sich als einfache Modifikationen von 3.4.3; man braucht nur einige Querstriche anzubringen, $\|\bar{x} - \bar{a}\|$ statt $|x - a|$ zu schreiben und das offene Intervall I durch eine offene Kugel K zu ersetzen. Bei

¹⁵Korrektur: Niveaulinie.

der Berechnung der nun partiellen Ableitungen wird jetzt $\Delta f := f(\bar{a} + h\bar{e}_i) - f(\bar{a})$ betrachtet. Ich empfehle Ihnen, den obigen Beweis noch einmal durchzugehen und die entsprechenden Modifikationen tatsächlich vorzunehmen. Hier ist jedenfalls das Ergebnis:

F sei eine auf einer offenen Teilmenge von \mathbb{R}^{n+1} definierte C^1 -Funktion und $F(\bar{a}, b) = 0$. Wenn $\frac{\partial F}{\partial y}(\bar{a}, b) \neq 0$, so definiert F die Variable y in der Nähe von (\bar{a}, b) als implizite Funktion von \bar{x} . Die entsprechende Funktion $f : K \rightarrow J$ ist stetig und differenzierbar (in einer genügend klein gewählten offenen Kugel $K = K_{\bar{a}}(r)$) und ihre partiellen Ableitungen berechnen sich als

$$\frac{\partial f}{\partial x_i}(\bar{x}) = -\frac{\frac{\partial F}{\partial x_i}(\bar{x}, f(\bar{x}))}{\frac{\partial F}{\partial y}(\bar{x}, f(\bar{x}))}.$$

3.4.6 Vorgriff: höhere Ableitungen der impliziten Funktion

Offiziell haben wir noch nicht über höhere partielle Ableitungen von Funktionen gesprochen, aber spätestens beim Wiederlesen werden Sie diese Passage verstehen.

Alle Voraussetzungen und Bezeichnungen von oben bleiben in Kraft. *Wenn F eine C^k -Funktion ist, so ist auch die implizite Funktion f eine C^k -Funktion.* Das sieht man induktiv an der Formel

$$\frac{\partial f}{\partial x_i}(\bar{x}) = -\frac{\frac{\partial F}{\partial x_i}(\bar{x}, f(\bar{x}))}{\frac{\partial F}{\partial y}(\bar{x}, f(\bar{x}))}.$$

Da auf der rechten Seite eine stetige Funktion steht (Verkettung stetiger Funktionen und Quotient mit Nenner $\neq 0$), sind die partiellen Ableitungen von f stetig, d.h. f ist mindestens C^1 . Dann, frischer Blick, steht aber auf der rechten Seite eine C^1 -Funktion. Also sind alle $\frac{\partial f}{\partial x_i}$ stetig differenzierbar. Damit wird f zur C^2 -Funktion usw. Das Argument klappt solange, wie die rechte Seite aus soundsooft stetig differenzierbaren Funktionen zusammengesetzt ist. Das hängt aber nur von der Differenzierbarkeit von F ab. Durch Anwendung der Quotientenregel entstehen zwar Nenner, aber das sind alles nur Potenzen von $\frac{\partial F}{\partial y}(\bar{x}, f(\bar{x})) \neq 0$.

3.5 Implizite Funktionen II: der allgemeine Fall

Jetzt wollen wir die Erkenntnisse des vorigen Abschnitts auf Gleichungssysteme (in denen wieder oBdA alle rechten Seiten Null sind)

$$(*) \quad \begin{array}{rcl} F_1(x_1, \dots, x_n, y_1, \dots, y_m) & = & 0 \\ F_2(x_1, \dots, x_n, y_1, \dots, y_m) & = & 0 \\ & \vdots & \vdots \\ F_m(x_1, \dots, x_n, y_1, \dots, y_m) & = & 0 \end{array}$$

verallgemeinern, die nach den y_j aufgelöst werden sollen. Es handelt sich um m Gleichungen mit m Unbekannten, alles andere wäre wohl auch kaum sinnvoll. Kompakt kann man das System schreiben als $\vec{F}(\bar{x}, \bar{y}) = \vec{0}$ wobei $\vec{F} : \mathbb{R}^{n+m} \rightarrow \mathbb{R}^m$ ein Vektorfeld ist (das wieder C^1 vorausgesetzt wird).

Im Allgemeinen wird eine globale Auflösung nicht gelingen. Wir können aber hoffen, unter gewissen Bedingungen in der Nähe eines gegebenen Punktes \bar{a}, \bar{b} mit $\vec{F}(\bar{a}, \bar{b}) = \vec{0}$ eine Funktion $\bar{f} : \mathbb{R}^n \rightarrow \mathbb{R}^m$ derart zu finden, daß $\vec{F}(\bar{x}, \bar{y}) = \vec{0}$ und $\bar{y} = \bar{f}(\bar{x})$ gleichbedeutend sind. (nämlich für alle $\bar{x} \in K_{\bar{a}}(\varepsilon)$ und $\bar{y} \in K_{\bar{b}}(\delta)$ für passende positive ε und δ .)

Die Funktionen \bar{f} werden auch noch stetig und sogar differenzierbar sein.

3.5.1 Was ist die richtige Bedingung?

Um das herauszufinden versuchen wir die im einfachsten Fall erfolgreiche Bedingung $\frac{\partial F}{\partial y} \neq 0$ so umzuf formulieren, daß das Ergebnis auch im höherdimensionalen Fall sinnvoll ist.

F war als C^1 Funktion vorausgesetzt. Daraus folgt die totale Differenzierbarkeit im Punkt (a, b) . Es läßt sich also $F(x, y)$ in der Nähe von (a, b) gut durch eine lineare Funktion approximieren. Die Gleichung $F(x, y) = 0$ kann in der Nähe von (a, b) näherungsweise durch die lineare Gleichung

$$F(x, y) \approx \underbrace{F(a, b)}_{=0} + \frac{\partial F}{\partial x}(a, b)(x - a) + \frac{\partial F}{\partial y}(a, b)(y - b) = 0$$

ersetzt werden. Die Bedingung $\frac{\partial F}{\partial y}(a, b) \neq 0$ ist dann äquivalent dazu, daß sich die approximierende lineare Gleichung nach y umstellen läßt.

Die entsprechende lineare Approximation des Gleichungssystems (*) sieht so aus (alle partiellen Ableitungen im Punkt (\bar{a}, \bar{b}) genommen)

$$\begin{array}{ccccccc} \frac{\partial F_1}{\partial x_1}(x_1 - a_1) + \dots + \frac{\partial F_1}{\partial x_n}(x_n - a_n) & + & \frac{\partial F_1}{\partial y_1}(y_1 - b_1) + \dots + \frac{\partial F_1}{\partial y_m}(y_m - b_m) & = & 0 \\ \frac{\partial F_2}{\partial x_1}(x_1 - a_1) + \dots + \frac{\partial F_2}{\partial x_n}(x_n - a_n) & + & \frac{\partial F_2}{\partial y_1}(y_1 - b_1) + \dots + \frac{\partial F_2}{\partial y_m}(y_m - b_m) & = & 0 \\ \vdots & & \vdots & & \vdots \\ \frac{\partial F_m}{\partial x_1}(x_1 - a_1) + \dots + \frac{\partial F_m}{\partial x_n}(x_n - a_n) & + & \frac{\partial F_m}{\partial y_1}(y_1 - b_1) + \dots + \frac{\partial F_m}{\partial y_m}(y_m - b_m) & = & 0 \end{array}$$

Wenn man hierbei die x_i als gegeben und die y_j als gesucht annimmt, so ist eine hinreichende Lösbarkeitsbedingung für das LGS, daß die Koeffizientenmatrix vollen Rang hat, bzw., daß ihre Determinante nicht verschwindet:

$$\det \left(\frac{\partial F_i}{\partial y_j} \right) = \begin{vmatrix} \frac{\partial F_1}{\partial y_1}(\bar{a}, \bar{b}) & \frac{\partial F_1}{\partial y_2}(\bar{a}, \bar{b}) & \dots & \frac{\partial F_1}{\partial y_m}(\bar{a}, \bar{b}) \\ \frac{\partial F_2}{\partial y_1}(\bar{a}, \bar{b}) & \frac{\partial F_2}{\partial y_2}(\bar{a}, \bar{b}) & \dots & \frac{\partial F_2}{\partial y_m}(\bar{a}, \bar{b}) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial F_m}{\partial y_1}(\bar{a}, \bar{b}) & \frac{\partial F_m}{\partial y_2}(\bar{a}, \bar{b}) & \dots & \frac{\partial F_m}{\partial y_m}(\bar{a}, \bar{b}) \end{vmatrix} \neq 0.$$

Als kompakte Schreibweise benutzt man für die Koeffizientenmatrix auch $\frac{\partial \vec{F}}{\partial \vec{y}}(\bar{a}, \bar{b})$. Es handelt sich um den hinteren $m \times m$ -Teil der JACOBI-Matrix $J_{\vec{F}}(\bar{a}, \bar{b})$. Man spricht auch von einer partiellen JACOBI-Matrix. Determinanten von derartigen Matizen werden als *Funktionaldeterminanten* bezeichnet. Sie werden oft auch als $\frac{\partial(F_1, F_2, \dots, F_m)}{\partial(y_1, y_2, \dots, y_m)}$ notiert. Ihr Nichtverschwinden ist übrigens dazu äquivalent, daß die im Punkt (\bar{a}, \bar{b}) berechneten partiellen Ableitungen $\frac{\partial \vec{F}}{\partial y_1}, \frac{\partial \vec{F}}{\partial y_2}, \dots, \frac{\partial \vec{F}}{\partial y_m}$ linear unabhängig in $\vec{\mathbb{R}}^m$ sind.

3.5.2 Der allgemeine Satz über die impliziten Funktionen

Ist $\vec{F} : \mathbb{R}^{n+m} \rightarrow \vec{\mathbb{R}}^m$ ein C^1 -Vektorfeld, $\vec{F}(\bar{a}, \bar{b}) = \vec{0}$ und $\det \frac{\partial \vec{F}}{\partial \vec{y}}(\bar{a}, \bar{b}) \neq 0$, dann gibt es offene Kugeln K und L mit den Mittelpunkten \bar{a} bzw. \bar{b} und eine Funktion $\vec{f} : K \rightarrow \mathbb{R}^m$, so daß für alle $\bar{x} \in K$ und $\bar{y} \in L$

$$\vec{F}(\bar{x}, \bar{y}) = \vec{0} \iff \bar{y} = \vec{f}(\bar{x}).$$

Die Funktion \vec{f} ist stetig und differenzierbar in K und für ihre JACOBI-Matrix gilt

$$J_{\vec{f}}(\bar{x}) = - \left(\frac{\partial \vec{F}}{\partial \vec{y}}(\bar{x}, \vec{f}(\bar{x})) \right)^{-1} \cdot \frac{\partial \vec{F}}{\partial \bar{x}}(\bar{x}, \vec{f}(\bar{x})) \quad \text{speziell} \quad J_{\vec{f}}(\bar{a}) = - \left(\frac{\partial \vec{F}}{\partial \vec{y}}(\bar{a}, \bar{b}) \right)^{-1} \cdot \frac{\partial \vec{F}}{\partial \bar{x}}(\bar{a}, \bar{b}).$$

Wir verzichten auf den **Beweis**. Um Ihnen den Satz schmackhafter zu machen, gehe ich aber noch auf einen speziellen Fall ein.

3.5.3 Raumkurven als Durchschnitte von Oberflächen

Schneidet man zwei Sphären, so entsteht ein Kreis (im entarteten Fall auch ein Punkt), schneidet man einen Kegel mit einer Ebene, so entsteht (von entarteten Fällen abgesehen) je nachdem eine Ellipse, Parabel oder Hyperbel. Allgemein sollte beim Durchschnitt zweier im Raum gelegener Oberflächen eine Kurve (im naiven Sinn) entstehen. Der Satz über implizite Funktionen bestätigt das und garantiert sogar die differenzierbare Parametrisierbarkeit der entstehenden Kurven, allerdings nur stückweise.

Dabei muß man sogenannte Berührungspunkte der beiden Flächen ausschließen, d.h. gemeinsame Punkte mit gemeinsamer Tangentialebene (und deshalb linear abhängigen Gradienten).

Die beiden Flächen seien durch Gleichungen $F(x, y, z) = 0$ und $G(x, y, z) = 0$ definiert. Der Punkt (a, b, c) liege auf beiden Flächen, wobei $\text{grad} F(a, b, c)$ und $\text{grad} G(a, b, c)$ linear unabhängig sein sollen. Dann ist eine der drei möglichen im Punkt (a, b, c) berechneten Funktionaldeterminanten ungleich Null, etwa

$$\frac{\partial(F, G)}{\partial(x, z)} = \begin{vmatrix} \frac{\partial F}{\partial x} & \frac{\partial F}{\partial z} \\ \frac{\partial G}{\partial x} & \frac{\partial G}{\partial z} \end{vmatrix} \neq 0.$$

Nach dem Satz über implizite Funktionen kann man daher x und z in der Nähe von a als differenzierbare Funktionen von y finden. Es gibt also stetig differenzierbare Funktionen $f, g : I \rightarrow \mathbb{R}$, die auf einem b enthaltenden offenen Intervall I definiert sind und die $F(f(y), y, g(y)) = G(f(y), y, g(y)) = 0$ für alle $y \in I$ erfüllen. Also ist der Durchschnitt der beiden Flächen in der Nähe von (a, b, c) gerade die Bahn der differenzierbaren Kurve $\bar{\gamma}(y) := (f(y), y, g(y))$ mit $y \in I$.

Der Tangentenvektor im Punkt (a, b, c) ist $\bar{\gamma}'(b) = \begin{pmatrix} f'(b) \\ 1 \\ g'(b) \end{pmatrix}$, wobei die beiden Ableitungen

aus dem Gleichungssystem
$$\begin{aligned} \frac{\partial F}{\partial x}(a, b, c)f'(b) + \frac{\partial F}{\partial y}(a, b, c) + \frac{\partial F}{\partial z}(a, b, c)g'(b) &= 0 \\ \frac{\partial G}{\partial x}(a, b, c)f'(b) + \frac{\partial G}{\partial y}(a, b, c) + \frac{\partial G}{\partial z}(a, b, c)g'(b) &= 0 \end{aligned}$$
 bestimmt werden können. Das ist oft die schnellere Alternative zur orthodoxen Berechnung von

$$\begin{pmatrix} f'(b) \\ g'(b) \end{pmatrix} = - \begin{pmatrix} \frac{\partial F}{\partial x} & \frac{\partial F}{\partial z} \\ \frac{\partial G}{\partial x} & \frac{\partial G}{\partial z} \end{pmatrix}^{-1} \cdot \begin{pmatrix} \frac{\partial F}{\partial y} \\ \frac{\partial G}{\partial y} \end{pmatrix} \quad \text{rechte Seite in } (a, b, c) \text{ ausgerechnet.}$$

3.5.4 Die lokale Invertierbarkeit

Definition. Ein Punktfeld $\bar{f} : D \rightarrow \mathbb{R}^n$ mit offenem n -dimensionalen Definitionsbereich heißt bei $\bar{a} \in D$ *lokal invertierbar*, wenn es eine offene Menge U mit $\bar{a} \in U \subset D$ derartig gibt, daß $V = \bar{f}(U)$ offen in \mathbb{R}^n ist und die Einschränkung $\bar{f}|U : U \rightarrow V$ bijektiv.

Dann gibt es nämlich $\bar{g} : V \rightarrow U$, so daß $\bar{f}(\bar{g}(\bar{v})) = \bar{v}$ für alle $\bar{v} \in V$ und $\bar{g}(\bar{f}(\bar{u})) = \bar{u}$ für alle $\bar{u} \in U$ gilt.

Umkehrsatz. Eine C^1 -Abbildung $f : D \rightarrow \mathbb{R}^n$ ist bei $\bar{a} \in D$ bereits dann lokal invertierbar, wenn ihre JACOBI-Matrix $J_{\bar{f}}(\bar{a})$ invertierbar ist, wenn also die Determinante $\det J_{\bar{f}}(\bar{a})$ nicht Null wird. In diesem Fall ist auch die lokale Inverse $\bar{g} := (\bar{f}|U)^{-1} : V \rightarrow U$ eine C^1 -Abbildung und für alle $\bar{u} \in U$ gilt

$$J_{\bar{g}}(\bar{f}(u)) = (J_{\bar{f}}(u))^{-1}.$$

Diese Aussage läßt sich relativ leicht aus dem Satz über implizite Funktionen herleiten, nur darf man sich nicht davon beirren lassen, daß \bar{x} und \bar{y} die Rollen tauschen.

Sei $\bar{f}(\bar{a}) = \bar{b}$. Die gesuchte Umkehrfunktion muß ja das Gleichungssystem $\bar{f}(\bar{x}) = \bar{y}$ in der Nähe von (\bar{a}, \bar{b}) nach \bar{x} umstellen. Das paßt in das Schema des Satzes über implizite Funktionen, wenn wir die Hilfsfunktion $\bar{F}(\bar{x}, \bar{y}) = \bar{f}(\bar{x}) - \bar{y}$ benutzen. Damit das Gleichungssystem $\bar{F}(\bar{x}, \bar{y}) = \vec{0}$ bei (\bar{a}, \bar{b}) die Variablen \bar{x} als Funktion von \bar{y} definiert, muß die partielle JACOBI-Matrix $\frac{\partial \bar{F}}{\partial \bar{x}}(\bar{a}, \bar{b})$ invertierbar sein. Nun ist aber $\frac{\partial F_j}{\partial x_i}(\bar{a}, \bar{b}) = \frac{\partial [f_j(\bar{x}) - y_j]}{\partial x_i}(\bar{a}, \bar{b}) = \frac{\partial f_j}{\partial x_i}(\bar{a})$. Also ist die Auflösbarkeitsbedingung $\det \frac{\partial \bar{F}}{\partial \bar{x}}(\bar{a}) \neq 0$ oder $\det J_{\bar{f}}(\bar{a}) \neq 0$.

Die Formel für die JACOBI-Matrix der Umkehrfunktion, nennen wir sie kurz \bar{g} , folgt leicht aus der Kettenregel. $f(\bar{g}(\bar{x})) = \bar{x}$ also ist $J_{\bar{f}}(\bar{g}(\bar{x})) \cdot J_{\bar{g}}(\bar{x})$ die Einheitsmatrix $J_{id}(\bar{x})$.

3.5.5 Beispiel

Die Transformation $(x, y) \mapsto (e^x \cos y, e^x \sin y)$ hat die JACOBI-Matrix

$$\begin{pmatrix} e^x \cos y & -e^x \sin y \\ e^x \sin y & e^x \cos y \end{pmatrix}, \quad \text{also} \quad \det J = e^{2x}$$

überall positiv. Die Abbildung ist also überall lokal invertierbar. Global ist die Abbildung nicht invertierbar, wegen der Periodizität in y -Richtung. Jeder zur x -Achse parallele Streifen einer Breite $\geq 2\pi$ wird auf ganz \mathbb{R}^2 abgebildet.

3.6 Notwendige Bedingungen für Extremwerte

Auf die Definitionen von lokalem/globalem Maximum/Minimum brauchen wir wohl kein Papier zu verschwenden. Wir werden notwendige Bedingungen für das Vorliegen eines Extremwertes angeben. Mehr ist im Allgemeinen mit ersten Ableitungen nicht zu machen. Wenn wir später auch die zweiten Ableitungen zur Verfügung haben, wird eine hinreichende Bedingung nachgeliefert.

Generalvoraussetzung. Alle in diesem Abschnitt vorkommenden Funktionen sollen C^1 sein, auch wenn das nicht immer ausdrücklich gesagt wird.

3.6.1 Innere Extrempunkte des Definitionsbereiches

Nimmt das C^1 -Skalarfeld $f : D \rightarrow \mathbb{R}$ in einem inneren Punkt \bar{a} von $D \subset \mathbb{R}^n$ ein relatives Maximum oder Minimum an, so verschwinden alle partiellen Ableitungen in diesem Punkt: $\frac{\partial f}{\partial x_i}(\bar{a}) = 0$. Mit anderen Worten: der Gradient ist der Nullvektor.

Beweis. Das ist klar, denn alle partiellen Funktionen $t \mapsto f(a_1, \dots, a_{i-1}, t, a_{i+1}, \dots, a_n)$ nehmen im Punkt $t = a_i$ einen relativen Extremwert an, also verschwinden ihre Ableitungen.

Warnung. Wie im eindimensionalen Fall ist die Bedingung nicht hinreichend. Neben dem x^3 -Effekt, gibt es von der zweiten Dimension an noch die Möglichkeit, daß beim Durchgang durch \bar{a} in der einen achsenparallelen Richtung ein Minimum und in einer anderen achsenparallelen Richtung ein Maximum durchschritten wird. Man spricht dann von Sattelpunkten. Das populärste Beispiel für dieses Phänomen ist die Funktion $f(x, y) = x^2 - y^2$, deren Graph die sogenannte Sattelfläche darstellt.

3.6.2 Extrema unter Nebenbedingungen.

Wieder ist ein Skalarfeld $f : D \rightarrow \mathbb{R}$ mit offenem Definitionsbereich $D \subset \mathbb{R}^n$ gegeben, das im Zusammenhang mit der Suche nach Extremwerten gern *Zielfunktion* genannt wird. Wieder suchen wir relative und eventuell auch absolute Extremwerte, aber diesmal sind nicht alle $\bar{x} \in D$ zur Konkurrenz zugelassen, sondern nur solche, die gewissen sogenannten Nebenbedingungen genügen. Darunter wollen wir endlich viele Gleichungen $g_1(\bar{x}) = 0, g_2(\bar{x}) = 0, \dots, g_k(\bar{x}) = 0$ verstehen. Die kann man wieder zu einer vektoriellen Gleichung zusammenfassen $\vec{g}(\bar{x}) = \vec{0}$. Sowohl f als auch \vec{g} werden C^1 vorausgesetzt. Die Menge

$$\Gamma = \{ \bar{x} \in \mathbb{R}^n : \vec{g}(\bar{x}) = \vec{0} \}$$

aller Punkte, die der Nebenbedingung genügen, sollte man sich als eine $(n - k)$ -dimensionale Fläche in \mathbb{R}^n vorstellen¹⁶ (eventuell auch eine Kurve) und gesucht werden die Extrempunkte von f auf dieser Fläche. \bar{a} muß zu Γ gehören und $f(\bar{a})$ muß größer (kleiner) sein als $f(\bar{x})$ für alle $\bar{x} \in \Gamma$, die dicht bei \bar{a} liegen.

¹⁶Das ist nicht immer der Fall, aber normalerweise reduziert jede Gleichung die Dimension der Lösungsmenge um 1. Vgl. die Dimensionsformel aus der Linearen Algebra.

3.6.3 Beispiel

In der Ebene suchen wir den minimalen Abstand des Nullpunktes von einem Punkt des Kreises mit Mittelpunkt $(1, 3)$ und Radius 1. Die anschauliche Lösung ist ganz klar: Der Schnittpunkt der Strecke von $(0, 0)$ zum Mittelpunkt $(1, 3)$ realisiert das Minimum $\sqrt{10} - 1 \approx 2,162$.

Wie könnte man aber rechnen? Die Zielfunktion ist $\sqrt{x^2 + y^2}$ und die Nebenbedingung $\sqrt{(x-1)^2 + (y-3)^2} = 1$.

Sofort können wir uns das Leben erleichtern, indem wir bemerken, daß der Abstand kleinstmöglich wird, wenn sein Quadrat kleinstmöglich ist und die Nebenbedingung auch quadriert werden kann. Die Zielfunktion $f(x, y) = x^2 + y^2$ und die Nebenbedingung $g(x, y) = (x-1)^2 + (y-3)^2 - 1 = 0$ sehen schon deutlich sympathischer aus.

Nullsetzen der partiellen Ableitungen bringt gar nichts: $\frac{\partial f}{\partial x} = 2x = 0$ und $\frac{\partial f}{\partial y} = 2y = 0$ gilt nur für $x = y = 0$, aber der Punkt $(0, 0)$ liegt nicht auf der Kreislinie, kommt also als Lösung nicht infrage.

Cleverer ist es, die Nebenbedingung umzuschreiben. Im Fall des Kreises ist das einfach. Die Kurve läßt sich parametrisieren $x = 1 + \cos t, y = 3 + \sin t$. Das setzt man in f ein und sucht das Minimum der einstelligen Funktion $(1 + \cos t)^2 + (3 + \sin t)^2 = 11 + 2 \cos t + 6 \sin t$.

Dazu leitet man ab, setzt Null und löst die Gleichung $0 = -2 \sin t + 6 \cos t$. Da $\cos t = 0$ offensichtlich(?) keine Lösung liefert, dürfen wir dividieren und erhalten

$$\tan t = 3 \quad \text{also} \quad t = \arctan 3 \approx 1,25 \quad \text{oder} \quad t = \pi + \arctan 3 \approx 4,39.$$

Für diese t kann man nun Sinus und Cosinus berechnen usw. Laut Taschenrechner:

$$\sqrt{11 + 2 \cos(1,25) + 6 \sin(1,25)} \approx 4,162 \quad \sqrt{11 + 2 \cos(4,39) + 6 \sin(4,39)} \approx 2,162$$

Der zweite Wert stimmt mit $\sqrt{10} - 1$ überein, der erste ist $\sqrt{10} + 1$ und entspricht dem maximalen Abstand.

3.6.4 Problem

Leider findet man nicht immer und schon gar keine einfachen Parametrisierungen der Fläche oder Kurve Γ als ganzer. Wenn man lokale Extremwerte sucht, genügt es aber, diese Parametrisierungen lokal zu finden, eventuell verschiedene für verschiedene Teile der Fläche. Der Satz über implizite Funktionen sagt, daß das geht, solange die C^1 -Nebenbedingung $\vec{g}(\vec{x}) = \vec{0}$ eine JACOBI-Matrix hat, die in jedem (eigentlich nur in jedem Extrem-) Punkt von Γ vollen Rang (nämlich k) besitzt. Das geniale an der gleich herzuleitenden Methode besteht darin, daß man diese Parametrisierungen gar nicht explizit zu kennen braucht. Um auf die richtige Idee zu kommen, zunächst eine

3.6.5 heuristische Betrachtung

Die Zielfunktion $f(x, y)$ habe zwei Veränderliche und es gebe nur ein Nebenbedingung $g(x, y) = 0$. Dann ist $\Gamma = \{(x, y) : g(x, y) = 0\}$ eine Niveaulinie. Sei $\bar{p} = (x_0, y_0)$ ein relativer Extrempunkt von f auf Γ und $c := f(\bar{p})$ der extreme Funktionswert. Dann ist $L := \{(x, y) : f(x, y) = c\}$ eine Niveaulinie von f , die mit Γ den Punkt \bar{p} gemeinsam hat. Nehmen wir an, Γ würde L schneiden. Wenn wir dann auf Γ durch den Punkt \bar{p} hindurchgehen, überschreiten wir¹⁷ die Grenze L zwischen dem Gebiet $f < c$ und $f > c$. Das geht nicht, wenn c Extremwert ist. Also können sich die beiden Niveaulinien Γ und L in \bar{p} nicht wirklich schneiden sondern nur berühren. Sie müssen also eine gemeinsame Tangente haben. Und da die Gradienten von f und g jeweils senkrecht auf der Tangente stehen, liegen auch die Gradienten auf einer Geraden. Anders gesagt: im relativen Extrempunkt gilt $\text{grad } f = \lambda \text{grad } g$ für ein passendes λ .

Machen Sie sich das gegebene Argument noch einmal an dem obigen Beispiel klar. Die Niveaulinien der Zielfunktion f sind dort Kreise.

¹⁷Normalerweise ist das so. Sollte L eine 'Kammlinie' sein oder nur aus einem Punkt bestehen, ist der Gradient Null und das Endergebnis stimmt trotzdem (mit $\lambda = 0$).

3.6.6 Die Rechnung im Falle einer Nebenbedingung

Angenommen das C^1 -Feld f nimmt im Punkt \bar{p} einen relativen Extremwert auf der Fläche $\Gamma = \{\bar{x} \in \mathbb{R}^n : g(\bar{x}) = 0\}$ an, wobei der Gradient der C^1 -Funktion g im Punkt \bar{p} nicht verschwindet. Dann gibt es eine reelle Zahl λ mit $\vec{\text{grad}} f(\bar{p}) = \lambda \vec{\text{grad}} g(\bar{p})$.

Beweis. Wegen $\vec{\text{grad}} g(\bar{p}) \neq \vec{0}$ gibt es eine partielle Ableitung von g , die in \bar{p} nicht Null ist. Um die Bezeichnungen einfach zu haben, sei das $\frac{\partial g}{\partial x_n}$. Dann sagt der Satz über implizite Funktionen, daß sich die Gleichung $g(\bar{x}) = 0$ in der Nähe von \bar{p} mit Hilfe einer differenzierbaren Funktion nach x_n umstellen läßt:

$$\bar{x} \in \Gamma \iff x_n = \varphi(x_1, \dots, x_{n-1}), \quad \text{wobei} \quad \frac{\partial \varphi}{\partial x_i}(p_1, \dots, p_{n-1}) = -\frac{\frac{\partial g}{\partial x_i}(\bar{p})}{\frac{\partial g}{\partial x_n}(\bar{p})}.$$

φ liefert uns also eine lokale Parametrisierung der Fläche Γ . Setzt man diese in die Zielfunktion f ein, so hat die entstehende Hilfsfunktion $h(x_1, \dots, x_{n-1}) = f(x_1, \dots, x_{n-1}, \varphi(x_1, \dots, x_{n-1}))$ einen relativen Extremwert im nunmehr inneren Punkt (p_1, \dots, p_{n-1}) ihres Definitionsbereiches. Also verschwinden dort alle ihre partiellen Ableitungen. Die kann man aber mit Hilfe der Kettenregel ausrechnen. Für $i = 1, \dots, n-1$ ergibt sich

$$0 = \frac{\partial h}{\partial x_i}(p_1, \dots, p_{n-1}) = \frac{\partial f}{\partial x_i}(\bar{p}) + \frac{\partial f}{\partial x_n}(\bar{p}) \cdot \frac{\partial \varphi}{\partial x_i}(p_1, \dots, p_{n-1}) = \frac{\partial f}{\partial x_i}(\bar{p}) - \frac{\frac{\partial g}{\partial x_i}(\bar{p})}{\frac{\partial g}{\partial x_n}(\bar{p})} \cdot \frac{\partial f}{\partial x_n}(\bar{p}).$$

Setzt man $\lambda := \frac{\frac{\partial f}{\partial x_n}(\bar{p})}{\frac{\partial g}{\partial x_n}(\bar{p})}$ so wird daraus für $i = 1, \dots, n-1$

$$\frac{\partial f}{\partial x_i}(\bar{p}) = \lambda \frac{\partial g}{\partial x_i}(\bar{p}).$$

Trivialerweise gilt das aber auch für $i = n$. Also lassen sich die n Gleichungen zu der behaupteten Vektorgleichung $\vec{\text{grad}} f(\bar{p}) = \lambda \vec{\text{grad}} g(\bar{p})$ zusammenfassen.

Die so trickreich in das Problem eingeführt neue Unbestimmte λ heißt nach ihrem Erfinder LAGRANGEScher Multiplikator. Er ist mit den gesuchten Koordinaten p_1, \dots, p_n des Extrempunktes durch die n skalaren Gleichungen $\frac{\partial f}{\partial x_i}(\bar{p}) = \lambda \frac{\partial g}{\partial x_i}(\bar{p})$ verknüpft. Die Koordinaten erfüllen noch die Gleichung $g(\bar{p}) = 0$. Man hat also $n+1$ Gleichungen für $n+1$ Unbestimmte. Mit einigem Glück (und schlimmstenfalls der Hilfe von Computeralgebra) kann man das System lösen.

Achtung. Die Bedingung $\vec{\text{grad}} g(\bar{p}) \neq \vec{0}$ darf nicht übersehen werden. Punkte von Γ , in denen sie verletzt ist, müssen extra untersucht werden.

3.6.7 Verdächtiger = Täter?

Ich will nochmals darauf hinweisen, daß wir nur eine notwendige Bedingung für Extremwerte vor uns haben. Ob in verdächtigen Punkten tatsächlich ein Extremwert (unter Nebenbedingungen) vorliegt, muß jeweils extra untersucht werden.

Bei vielen Aufgaben wird nach absoluten Extremwerten auf Γ gefragt. Wenn man weiß, daß es ein absolutes Maximum bzw. Minimum gibt, so muß dieses in einem der verdächtigen Punkte angenommen werden. Ob es aber Maximum und/oder Minimum wirklich gibt, bedarf einer Extrabetrachtung. Ist Γ z.B. kompakt, so weiß man daß die Zielfunktion Maximum und Minimum annimmt, die entsprechenden Punkte also unter den verdächtigen sein müssen. Im nicht kompakten Fall kann man manchmal folgendermaßen argumentieren. Ist $\Gamma \subseteq \mathbb{R}^n$ abgeschlossen, f mindestens auf Γ definiert und stetig und gilt $\lim_{\bar{x} \rightarrow \infty} f(\bar{x}) = \infty$, so nimmt f auf Γ einen kleinsten Wert an. Dieser muß dann in einem verdächtigen Punkt angenommen werden. Dabei ist $\lim_{\bar{x} \rightarrow \infty} f(\bar{x}) = \infty$, Kurzschreibweise für:

$$\text{für alle } C > 0 \text{ gibt es } R > 0 \text{ so daß } f(\bar{x}) > C \text{ falls } \bar{x} \notin \bar{K}_R(0).$$

Beweis. Sei $\bar{a} \in \Gamma$ beliebig. Wähle R so daß für $\bar{x} \notin \bar{K}_0(R)$ gilt $f(\bar{x}) > f(\bar{a})$. Die Menge $\Gamma_R := \Gamma \cap \bar{K}_0(R)$ ist abgeschlossen (als Durchschnitt zweier abgeschlossener Mengen) und beschränkt (weil in der Kugel enthalten), also kompakt. Außerhalb von Γ_R werden von f nur Werte $> f(\bar{a})$ angenommen. Deshalb muß \bar{a} zu Γ_R gehören. Wegen der Kompaktheit von Γ_R gibt es ein $\bar{b} \in \Gamma_R$, so daß $f(\bar{b}) \leq f(\bar{x})$ für alle $\bar{x} \in \Gamma_R$. Dann gilt aber für ein beliebiges $\bar{x} \in \Gamma$

- entweder $\bar{x} \in \Gamma_R$, dann $f(\bar{b}) \leq f(\bar{x})$
- oder $\bar{x} \notin \Gamma_R$, dann $f(\bar{x}) > f(\bar{a}) \geq f(\bar{b})$.

Es ist also $f(\bar{b})$ der versprochene kleinste Funktionswert auf Γ .

3.6.8 Beispiel

Wir greifen das Beispiel von oben noch einmal auf. Die Funktion $f(x, y) = x^2 + y^2$ ist unter der Nebenbedingung $g(x, y) = (x - 1)^2 + (y - 3)^2 - 1 = 0$ zu minimieren.

Es gibt einen LAGRANGE-Multiplikator λ und die drei Gleichungen lauten:

$$\begin{array}{rcll} \frac{\partial f}{\partial x} & = & \lambda \frac{\partial g}{\partial x} & 2x & = & \lambda 2(x - 1) \\ \frac{\partial f}{\partial y} & = & \lambda \frac{\partial g}{\partial y} & 2y & = & \lambda 2(y - 3) \\ g(x, y) & = & 0 & (x - 1)^2 + (y - 3)^2 & = & 1 \end{array} \quad \text{also}$$

Dieses System kann man mit wenig Mühe auflösen. Zunächst liefern die beiden ersten Gleichungen $x - 1 = \frac{x}{\lambda}$, $y - 3 = \frac{y}{\lambda}$. Setzt man das in die dritte Gleichung ein, ergibt sich $x^2 + y^2 = \lambda^2$. Damit ist $|\lambda|$ der gesuchte minimale Abstand; wir brauchen x und y gar nicht unbedingt zu bestimmen. Weiter kann man die erste und zweite Gleichung nach x bzw. y umstellen und das Resultat in $x^2 + y^2 = \lambda^2$ einsetzen:

$$x = \frac{-\lambda}{1 - \lambda} \quad y = \frac{-3\lambda}{1 - \lambda} \quad \text{liefert} \quad \frac{\lambda^2 + 9\lambda^2}{(1 - \lambda)^2} = \lambda^2$$

also $(1 - \lambda)^2 = 10$ oder $\lambda = 1 \pm \sqrt{10}$, damit $|\lambda| = \sqrt{10} \pm 1$. Der Wert $\sqrt{10} + 1$ ist der maximale Abstand eines Kreispunktes vom Koordinatenursprung, $\sqrt{10} - 1$ der gesuchte minimale. Halt! Die Gradientenbedingung muß noch geprüft werden. Es wird aber $\text{grad } g = \begin{pmatrix} 2(x-1) \\ 2(y-3) \end{pmatrix}$ nur im Punkt $(1, 3)$ Null und der liegt nicht auf Γ .

Eigentlich liefert die Methode nur 'extremwertverdächtige' Punkte auf dem Kreis. Da aber die stetige Funktion auf der kompakten Menge Γ ein Maximum und ein Minimum haben muß, müssen die Verdächtigen auch die Täter sein.

3.6.9 Anwendung: Die A-G-Ungleichung

Wir suchen zunächst die minimale Summe $s + t + u + v$ von 4 positiven Zahlen, deren Produkt 1 ist. Zielfunktion ist $s + t + u + v$. Sie wird untersucht unter der Nebenbedingung $stuv - 1 = 0$ (auf Positivität achten wir extra). Es gibt einen Multiplikator λ und das Gleichungssystem wird zu

$$\begin{array}{rcll} \frac{\partial}{\partial s} : & 1 & = & \lambda \cdot tuv & s & = & \lambda stuv = \lambda \\ \frac{\partial}{\partial t} : & 1 & = & \lambda \cdot suv & t & = & \lambda stuv = \lambda \\ \frac{\partial}{\partial u} : & 1 & = & \lambda \cdot stv & u & = & \lambda stuv = \lambda \\ \frac{\partial}{\partial v} : & 1 & = & \lambda \cdot stu & v & = & \lambda stuv = \lambda \\ & & & & stuv & = & 1 \end{array} \quad \text{also}$$

Es müssen also $s = t = u = v = \lambda$ sein und weiter $stuv = \lambda^4 = 1$. Positivität erlaubt nur $\lambda = 1$. Also $s = t = u = v = 1$. Der entsprechende Funktionswert ist $1 + 1 + 1 + 1 = 4$.

Damit haben wir einen einzigen verdächtigen Punkt, in dem ein Extremwert vorliegen könnte. Der Gradient von g im Punkt $(1, 1, 1, 1)$ besteht auch aus lauter Einsen, ist also sicher nicht Null. Überhaupt ist $\text{grad } g = \vec{0}$ nur möglich, wenn eine Koordinate Null ist; das passiert nicht auf Γ , d.h. neben dem gefundenen gibt es keine weiteren verdächtigen Punkte.

Um zu sehen, daß es sich um das Minimum der Zielfunktion handelt, schneiden wir den Definitionsbereich in Gedanken ab. Auf der kompakten Menge

$$\{(s, t, u, v) \in \mathbb{R}^4 : stuv = 1; 0 \leq s \leq 5, 0 \leq t \leq 5, 0 \leq u \leq 5, 0 \leq v \leq 5\}$$

hat $s + t + u + v$ ein absolutes Minimum, das ≤ 4 sein muß.

Außerhalb der abgeschnittenen Menge ist wenigstens eine Koordinate ≥ 5 und daher $s + t + u + v \geq 5$. Das heißt, das Minimum auf der abgeschnittenen Menge ist auch globales Minimum. Da es aber nur einen verdächtigen Punkt gibt, muß dieser auch der Täter sein.

Folgerung. *Das geometrische Mittel $G = \sqrt[4]{abcd}$ von 4 positiven Zahlen ist höchstens so groß wie das arithmetische Mittel $A = \frac{1}{4}(a + b + c + d)$. Gleichheit $A = G$ besteht genau dann, wenn $a = b = c = d$.*

Zum Beweis setzen wir $s = \frac{a}{G}$, $t = \frac{b}{G}$, $u = \frac{c}{G}$ und $v = \frac{d}{G}$. Dann ist $stuv = \frac{abcd}{G^4} = 1$, also $s + t + u + v \geq 4$, was gerade $\frac{a+b+c+d}{G} \geq 4$ oder $G \leq \frac{a+b+c+d}{4} = A$ bedeutet.

Bemerkung. Offenbar kann man denselben Beweis auch für 5, 6, 7, ... positive Zahlen führen; die A-G-Ungleichung gilt für beliebig viele positive Zahlen.

Noch eine Bemerkung. Die Nebenbedingung $stuv = 1$ läßt sich sogar global nach v umstellen und dann in die Zielfunktion einsetzen. Dann ist nur noch die dreistellige Funktion $s + t + u + \frac{1}{stu}$ ohne Nebenbedingungen zu minimieren. Das geht durch Nullsetzen der partiellen Ableitungen, wird aber eher schwerer als die LAGRANGE-Methode.

3.6.10 Mehrere Nebenbedingungen

Hier gebe ich nur das Rezept an. In den Ergänzungen gibt es für Kenner und Liebhaber einen Beweis in einem hinreichend allgemeinen Fall. Die beteiligten Funktionen sind (stillschweigend) C^1 -vorausgesetzt. *Angenommen, die Zielfunktion $f : D \rightarrow \mathbb{R}$ nimmt im Punkt \bar{a} unter den Nebenbedingungen $g_1(\bar{x}) = 0, \dots, g_k(\bar{x}) = 0$ einen relativen Extremwert an und die JACOBI-Matrix $\left(\frac{\partial g_i}{\partial x_j}(\bar{a})\right)$ hat vollen Rang. Dann gibt es Skalare $\lambda_1, \dots, \lambda_k$ mit*

$$\vec{\text{grad}} f(\bar{a}) = \lambda_1 \vec{\text{grad}} g_1(\bar{a}) + \lambda_2 \vec{\text{grad}} g_2(\bar{a}) + \dots + \lambda_k \vec{\text{grad}} g_k(\bar{a}),$$

in Koordinaten:

$$\frac{\partial f}{\partial x_i}(\bar{a}) = \lambda_1 \frac{\partial g_1}{\partial x_i}(\bar{a}) + \lambda_2 \frac{\partial g_2}{\partial x_i}(\bar{a}) + \dots + \lambda_k \frac{\partial g_k}{\partial x_i}(\bar{a}) \quad i = 1, \dots, n.$$

Zusammen mit den k Gleichungen $g_j(\bar{a}) = 0$ liefern diese n Gleichungen ein System, aus dem man (mit etwas Glück) die $k + n$ Unbestimmten $\lambda_1, \dots, \lambda_k, a_1, \dots, a_n$ bestimmen kann. Die auftauchenden λ_i nennt man wieder LAGRANGE-Multiplikatoren.

Im Hinblick auf eine hinreichende Bedingung halten wir gleich noch fest

Zusatz. *Hat man die λ_i und \bar{a} nach obigem Rezept bestimmt und erweist es sich, daß die Hilfsfunktion $h(\bar{x}) = f(\bar{x}) - \lambda_1 g_1(\bar{x}) - \dots - \lambda_k g_k(\bar{x})$ im inneren Punkt \bar{a} ihres Definitionsbereiches einen relativen Extremwert annimmt, so besitzt auch f im Punkt \bar{a} einen relativen Extremwert auf der Fläche Γ .*

Um den Zusatz zu begründen, nehmen wir an, daß $h(\bar{a}) \leq h(\bar{x})$ für alle \bar{x} mit $d(\bar{a}, \bar{x}) < \delta$. Sei jetzt $d(\bar{a}, \bar{x}) < \delta$ und $\bar{x} \in \Gamma$. Da auf Γ die g_j alle Null sind, haben wir dann

$$f(\bar{a}) = h(\bar{a}) \leq h(\bar{x}) = f(\bar{x}),$$

wie von einem lokalen Minimum auf Γ verlangt.

3.6.11 Ein exotischer Fall

Nehmen wir an, daß die JACOBI-Matrix $\left(\frac{\partial g_i}{\partial x_j}(\bar{a})\right) = \left(\frac{\partial \vec{g}}{\partial x_1}(\bar{a}), \frac{\partial \vec{g}}{\partial x_2}(\bar{a}), \dots, \frac{\partial \vec{g}}{\partial x_n}(\bar{a})\right)$ vollen Rang hat. Dann hat auch ihre Transponierte $\left(\vec{\text{grad}} g_1(\bar{a}), \vec{\text{grad}} g_2(\bar{a}), \dots, \vec{\text{grad}} g_k(\bar{a})\right)$ vollen Rang. Solange $k \leq n$ ist, bedeutet das, daß die Gradienten $\vec{\text{grad}} g_i(\bar{a})$ linear unabhängig sind. Das ist ganz gut zu wissen. Ist $k \geq n$ so befindet sich unter den $\vec{\text{grad}} g_i(\bar{a})$ eine Basis von \mathbb{R}^n aus ihr läßt sich jeder Vektor, also auch $\vec{\text{grad}} f$ linear kombinieren. Die notwendige Bedingung ist also auf jeden Fall erfüllt (unabhängig von f !). Es liegt auch immer ein relativer Extremwert vor, denn (ohne Beweis) \bar{a} ist in diesem Fall isolierter Punkt von Γ ; es gibt in genügender Nähe von \bar{a} gar keine weiteren Punkte auf Γ deren Funktionswerte zu über- bzw. unterbieten wären. Die obige Aussage gilt also auch dann, ist aber wertlos.

3.6.12 Beispiel

Gesucht sind Maximum und Minimum der Zielfunktion $f(x, y, z) = 2x + y + z$ unter den zwei Nebenbedingungen $x^2 + 4y^2 + 4z^2 = 4$ und $(x-1)^2 + y^2 + z^2 = 1$. Die beiden Gleichungen definieren eine kompakte Menge (warum?), also existieren die gesuchten Extremwerte. Die LAGRANGE-Gleichungen lauten:

$$\begin{aligned} 2 &= 2\lambda x + 2\mu(x-1) \\ 1 &= 8\lambda y + 2\mu y &= (8\lambda + 2\mu)y \\ 1 &= 8\lambda z + 2\mu z &= (8\lambda + 2\mu)z \end{aligned}$$

Die letzten beiden Gleichungen zeigen $(8\lambda + 2\mu) \neq 0$ und (daher) $y = z$ sowie $y \neq 0 \neq z$. Setzt man $y = z$ in die Nebenbedingungen ein, erhält man $x^2 + 8y^2 = 4$ und $(x-1)^2 + 2y^2 = 1$. Daraus $x^2 = 4(x-1)^2$ oder $\pm x = 2(x-1)$, also $x = 2$ oder $x = \frac{2}{3}$. Aus $x = 2$ würde aber wegen $2^2 + 8y^2 = 4$ folgen $y = 0$, das geht nicht. Also ist $x = \frac{2}{3}$ und $y = z = \pm \frac{2}{3}$. In den beiden verdächtigen Punkten $(\frac{2}{3}, \frac{2}{3}, \frac{2}{3})$ und $(\frac{2}{3}, -\frac{2}{3}, -\frac{2}{3})$ nimmt f die Werte $\frac{8}{3}$ und 0 an. Wer an dieser Stelle aufhört, hat jedenfalls das Maximum der Funktion f verfehlt, denn $(2, 0, 0)$ erfüllt beide Nebenbedingungen und $f(2, 0, 0) = 4 > \frac{8}{3}$.

Wo werden die beiden Gradienten $\begin{pmatrix} 2x \\ 8y \\ 8z \end{pmatrix}$ und $\begin{pmatrix} 2(x-1) \\ 2y \\ 2z \end{pmatrix}$ linear abhängig? Entweder einer von ihnen ist der Nullvektor, das passiert in $(0, 0, 0)$ bzw. $(1, 0, 0)$, beide Punkte erfüllen die Nebenbedingungen nicht. Oder einer ist Vielfacher des anderen. Falls y oder z nicht Null sind, muß der Faktor 4 sein, woraus $2x = 8(x-1)$ oder $x = \frac{4}{3}$ folgt. Setzt man das in die beiden Nebenbedingungen ein, so ergibt sich $4(y^2 + z^2) = 4 - (\frac{4}{3})^2 = \frac{20}{9}$ und $y^2 + z^2 = \frac{5}{9}$, das geht nicht. Bleibt der Fall $y = z = 0$. Dann kann x beliebig sein und die Nebenbedingungen liefern den oben schon aufgeführten Punkt $(2, 0, 0)$, in dem also das Maximum angenommen wird.

3.6.13 Noch ein Beispiel mit zwei Nebenbedingungen

Gesucht sind Maximum und Minimum der Funktion $f(x, y, z) = xy + yz$ auf der Ellipse, die sich beim Schnitt des Zylinders $x^2 + y^2 = 2$ mit der Ebene $y + z = 2$ ergibt. Hier haben wir die Nebenbedingungen $g(x, y, z) = x^2 + y^2 - 2 = 0$ und $h(x, y, z) = y + z - 2 = 0$. Also gibt es zwei LAGRANGEMultiplikatoren, die im Extrempunkt die Gleichungen $\vec{\text{grad}} f = \lambda \vec{\text{grad}} g + \mu \vec{\text{grad}} h$ erfüllen. Damit haben wir das Gleichungssystem:

$$\begin{aligned} y &= 2\lambda x + \mu 0 & x^2 + y^2 &= 2 \\ x + z &= 2\lambda y + \mu 1 & y + z &= 2 \\ y &= \lambda 0 + \mu 1 \end{aligned}$$

Aus der letzten Gleichung sehen wir $y = \mu$. Das gestattet μ zu eliminieren. Statt λ führen wir noch die Unbestimmte $\alpha := 2\lambda$ ein. Die drei Koordinaten des Extrempunktes lassen sich dann durch x und α ausdrücken:

$$x = x \quad y = \alpha x \quad z = 2 - \alpha x.$$

Das wird in die beiden restlichen Gleichungen substituiert:

$$\begin{aligned} x + z &= \alpha y + y \rightsquigarrow x + (2 - \alpha x) = \alpha^2 x + \alpha x & \text{oder} & \quad x = 2/(\alpha^2 + 2\alpha - 1), \\ x^2 + y^2 &= 2 \rightsquigarrow x^2 + \alpha^2 x^2 = 2 & \text{oder} & \quad x^2 = 2/(1 + \alpha^2). \end{aligned}$$

Wenn man die erste der beiden erhaltenen Gleichungen quadriert und dann beide gleichsetzt, so erhält man

$$\frac{4}{(\alpha^2 + 2\alpha - 1)^2} = \frac{2}{1 + \alpha^2} \text{ daraus } (\alpha^2 + 2\alpha - 1)^2 = 2(1 + \alpha^2) \text{ also } \alpha^4 + 4\alpha^3 - 4\alpha - 1 = 0.$$

Glücklicherweise sieht man die beiden Lösungen $\alpha = \pm 1$ mit bloßem Auge. Nach Division durch $\alpha^2 - 1$ bleibt für die restlichen Lösungen die Gleichung $\alpha^2 + 4\alpha + 1 = 0$ daraus $\alpha = -2 \pm \sqrt{3}$. Statt jetzt zu den vier gefundenen α s die Koordinaten der 'mutmaßlichen' Extrempunkte zu bestimmen (diese sind nicht gefragt), versuchen wir die Funktionswerte in diesen Punkten direkt aus α zu berechnen. Dazu benutzen wir die zweite LAGRANGE-Gleichung:

$$f(x, y, z) = y(x + z) = y(\alpha y + y) = (1 + \alpha)y^2 = (1 + \alpha)\alpha^2 x^2 = (1 + \alpha)\alpha^2 \frac{2}{1 + \alpha^2}.$$

Setzt man die vier potentiellen Werte für α ein, so ergeben sich

$$\alpha = 1 : f = 2 \quad \alpha = -1 : f = 0 \quad \alpha = \sqrt{3} - 2 : f = \frac{3}{2}\sqrt{3} - \frac{5}{2} \quad \alpha = -\sqrt{3} - 2 : f = -\frac{3}{2}\sqrt{3} - \frac{5}{2}.$$

Da die Funktion auf der Ellipse Maximum und Minimum annimmt, müssen diese unter den gefundenen Werten sein. Daher Maximum = 2 und Minimum = $-\frac{3}{2}\sqrt{3} - \frac{5}{2}$.

Damit der letzte Schluß stimmt, muß allerdings noch geprüft werden, daß die beiden Gradienten von g und h auf der Ellipse nicht linear abhängig werden (sonst hätten wir eventuell Punkte übersehen). Das sieht man aber mit bloßem Auge:

$$\vec{\text{grad}} g = \begin{pmatrix} 2x \\ 2y \\ 0 \end{pmatrix} \neq \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \quad \text{und} \quad \vec{\text{grad}} h = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}.$$

Alternative Lösung. Wir Parametrisieren die Ellipse und drücken die Zielfunktion durch den Parameter aus. Die Bedingung $x^2 + y^2 = 2$ wird von $x = \sqrt{2} \cos t$ und $y = \sqrt{2} \sin t$ befriedigt. Damit dann noch $y + z = 2$ wird, muß $z = 2 - \sqrt{2} \sin t$ gesetzt werden. Dann ergibt sich

$$f = 2 \sin t \cos t + \sqrt{2} \sin t (2 - \sqrt{2} \sin t) = \sin 2t - 2 \sin^2 t + 2\sqrt{2} \sin t.$$

Diese Funktion auf Extremwerte zu untersuchen ist, jedenfalls per Hand auch nicht ganz einfach. Ein programmierbarer Taschenrechner kann sie aber plotten und dann sieht man im Intervall $[0, 2\pi]$ vier relative Extremwerte, die unseren vier α s entsprechen.

3.6.14 Was tun, wenn die Bedingungen nicht erfüllt sind?

Angenommen, wir sollen das Maximum einer Funktion f auf der abgeschlossenen Kugel $\bar{K}_a(r)$ bestimmen. Um überhaupt Differentialrechnung verwenden zu können, wird von f vorausgesetzt, daß es sich um eine C^1 -Funktion handelt, die in einer offenen, die Kugel umfassenden Menge definiert sein soll. Jedenfalls folgt aus der Kompaktheit von \bar{K} , daß es ein absolutes Maximum gibt. Es zu finden, ist eine Extremwertaufgabe mit Nebenbedingung, die allerdings nicht in unser bisheriges Schema paßt. Der Ausweg besteht darin, die Aufgabe in zwei zu teilen. Man sucht erstens

nach relativen Extremwerten in der offenen Kugel $K_{\bar{a}}(r)$ und zweitens nach relativen Extremwerten auf dem Rand. Der läßt sich nämlich durch eine Nebenbedingung, der oben behandelten Art beschreiben: $\sum(x_i - a_i)^2 - r^2 = 0$.

Danach muß man unter den (hoffentlich endlich vielen) gefundenen verdächtigen Punkten noch nach dem größten Funktionswert Ausschau halten.

Schlimmer sieht es aus, wenn statt der Kugel ein abgeschlossener Würfel zu untersuchen ist. Dann ist nämlich der Rand nicht mehr als ganzes durch eine Nebenbedingung $\vec{g} = 0$ zu beschreiben, sondern jeweils nur ein Stück des Randes. Man muß das Innere des Würfels und jede Seitenfläche extra behandeln. Das führt im dreidimensionalen Fall bereits auf 1+6 Teilaufgaben, die hinterher wieder zusammengeführt werden müssen. Schlimmer noch: die Seitenflächen bestehen selbst wieder aus einem 'inneren Teil' und ihrem Rand. Das macht dann bald wirklich Mühe. Hier ist noch ein eher harmloses

3.6.15 Beispiel

Angenommen es ist das Maximum und das Minimum der Funktion $f(x, y) = xy$ im (schiefe) Quadrat $Q = \{(x, y) : |x| + |y| \leq 1\}$ zu bestimmen.

Die Extremwerte können angenommen werden (a) im Inneren des Quadrates oder auf dem Rand. Im zweiten Fall können Sie (b) im Inneren der Quadratseiten liegen oder (c) Eckpunkte sein.

(a) Wird der Extremwert im Inneren von Q angenommen, so handelt es sich sicher auch um einen relativen Extrempunkt von f (bezüglich der ganzen Ebene). Diese findet man als Lösungen des Gleichungssystems

$$\frac{\partial f}{\partial x} = 0, \quad \frac{\partial f}{\partial y} = 0, \quad \text{in unserem Fall also } y = 0, \quad x = 0.$$

Die einzige Lösung $(0, 0)$ liegt auch wirklich im Inneren des Quadrates.

(b) Liegt der Extrempunkt im Inneren einer Quadratseite, so wird in ihm ein relativer Extremwert der Zielfunktion auf der entsprechenden Geraden realisiert. Diese Punkte werden gefunden, indem man f jeweils unter der Nebenbedingung (4 Teilaufgaben) $y = \pm x \pm 1$ untersucht und nachsieht, ob die gefundenen Lösungen im inneren der Strecken liegen. Für die linke obere Quadratseite ($y = x + 1$ bzw. $y - x - 1 = 0$) ergibt sich das Gleichungssystem (1 LAGRANGEMultiplikator)

$$\begin{array}{l} y = -\lambda \\ x = \lambda \\ y = x + 1 \end{array} \quad \text{daraus} \quad \lambda = \frac{1}{2} \text{ und } x = -\frac{1}{2}, y = \frac{1}{2}, \text{ der Seitenmittelpunkt.}$$

Analog ergeben sich die anderen drei Seitenmittelpunkte $(\pm\frac{1}{2}, \pm\frac{1}{2})$ als verdächtig.

(c) Bisher noch nicht berücksichtigt sind die vier Eckpunkte $(\pm 1, 0)$ und $(0, \pm 1)$.

Damit haben wir insgesamt neun Punkte gefunden, in denen die Extremwerte angenommen werden könnten. Von diesen liefern $(\frac{1}{2}, \frac{1}{2})$ und $(-\frac{1}{2}, -\frac{1}{2})$ den Funktionswert $\frac{1}{4}$, die beiden anderen Seitenmitten den Funktionswert $-\frac{1}{4}$ und die fünf anderen Punkte 0. Damit sind $\pm\frac{1}{4}$ als größter und kleinster Funktionswert gefunden.

3.7 Abschweifung: Extremwerte von quadratischen Formen und der Satz über die Hauptachsentransformation

Wir werden beweisen, daß jede symmetrische Matrix mit reellen Einträgen eine Orthonormalbasis aus Eigenvektoren besitzt. Gleichzeitig wird beschrieben, wie die Eigenwerte und Eigenvektoren (im Prinzip) gefunden werden können.

3.7.1 Zur Erinnerung

Für beliebige \vec{x} und \vec{y} aus \mathbb{R}^n und jede quadratische Matrix $A \in M(n \times n; \mathbb{R})$ gilt $A\vec{x} \bullet \vec{y} = \vec{x} \bullet A^T \vec{y}$. Dabei bezeichnet A^T die transponierte Matrix. Ist A symmetrisch, also $A = A^T$, so gilt stets $A\vec{x} \bullet \vec{y} = \vec{x} \bullet A\vec{y}$.

3.7.2 Nebenrechnungen

Wir betrachten zunächst irgendeine Matrix $A = (a_{ij})_{i,j=1}^n \in M(n \times n; \mathbb{R})$ und die zugehörige quadratische Form $Q : \mathbb{R}^n \rightarrow \mathbb{R}$ mit $Q(\vec{x}) := A\vec{x} \bullet \vec{x}$. Dabei soll $\vec{x} = \vec{x}^T$ der Ortsvektor des Punktes \vec{x} sein (den wir sonst auch mit $\vec{r}(\vec{x})$ bezeichnen). Bodenständig aufgeschrieben:

$$Q(\vec{x}) = \sum_{i,j=1}^n a_{ij} x_i x_j.$$

Weiter sei $s : \mathbb{R}^n \rightarrow \mathbb{R}$ definiert durch $s(\vec{x}) = x_1^2 + x_2^2 + \dots + x_n^2 - 1$, so daß die Nebenbedingung $s(\vec{x}) = 0$ die Einheitssphäre von \mathbb{R}^n beschreibt. Schließlich betrachten wir für festes $\vec{b} \in \mathbb{R}^n$ das durch $g_{\vec{b}}(\vec{x}) := \vec{b} \bullet \vec{x}$ definierte Skalarfeld.

Wir wollen die Gradienten dieser Funktionen ausrechnen. Zwei überblickt man sofort:

$$\frac{\partial s}{\partial x_i} = 2x_i \quad \text{und} \quad \frac{\partial g_{\vec{b}}}{\partial x_i} = b_i \quad \text{daher} \quad \vec{\text{grad}} s(\vec{x}) = 2\vec{x} \quad \text{und} \quad \vec{\text{grad}} g_{\vec{b}}(\vec{x}) = \vec{b} \quad (\text{konstant}).$$

Bei der Berechnung von $\frac{\partial Q}{\partial x_i}$ braucht man nur die Summanden zu berücksichtigen, in denen x_i wirklich vorkommt. Das sind $a_{ii}x_i^2$ sowie $a_{ij}x_i x_j$ und $a_{ji}x_j x_i$, jeweils für alle $j \neq i$. Daher ergibt sich

$$\frac{\partial Q}{\partial x_i}(\vec{x}) = 2a_{ii}x_i + \sum_{j \neq i} a_{ij}x_j + \sum_{j \neq i} a_{ji}x_j = \sum_{j=1}^n a_{ij}x_j + \sum_{j=1}^n a_{ji}x_j.$$

Untereinandergeschrieben ergeben diese partiellen Ableitungen den Gradienten:

$$\vec{\text{grad}} Q(\vec{x}) = \begin{pmatrix} \sum_{j=1}^n (a_{1j} + a_{j1})x_j \\ \sum_{j=1}^n (a_{2j} + a_{j2})x_j \\ \vdots \\ \sum_{j=1}^n (a_{nj} + a_{jn})x_j \end{pmatrix} = (A + A^T) \cdot \vec{x}.$$

Bei symmetrischem A ergibt sich also $\vec{\text{grad}} Q(\vec{x}) = 2A\vec{x}$.

3.7.3 Der erste Eigenwert

Von nun an sei A eine symmetrische Matrix. Die zugehörige quadratische Form Q ist stetig und hat daher auf der Einheitssphäre $S := \{\vec{x} \in \mathbb{R}^n : s(\vec{x}) = 0\}$ ein Maximum, das im Punkt \vec{b}_1 angenommen werden möge.

Wegen $\text{grad } s(\vec{b}_1) = 2\vec{b}_1 \neq \vec{0}$ gibt es einen LAGRANGE-Multiplikator λ_1 , so daß

$$\vec{\text{grad}} Q(\vec{b}_1) = \lambda_1 \vec{\text{grad}} s(\vec{b}_1) \quad \text{also} \quad 2A\vec{b}_1 = \lambda_1(2\vec{b}_1).$$

Da man durch 2 kürzen darf, ist λ_1 Eigenwert und \vec{b}_1 Eigenvektor von A . Außerdem ist

$$Q(\vec{b}_1) = A\vec{b}_1 \bullet \vec{b}_1 = \lambda_1 \vec{b}_1 \bullet \vec{b}_1 = \lambda_1 \|\vec{b}_1\|^2 = \lambda_1.$$

Daher ist λ_1 gleich dem Maximum der Form Q auf S . Analog zeigt man übrigens, daß auch das Minimum von Q auf S Eigenwert ist, daran sind wir aber im Moment nicht interessiert.

3.7.4 Der zweite Eigenwert

Als nächstes lassen wir nur noch solche Punkte von S zur Konkurrenz zu, deren Ortsvektoren senkrecht auf \vec{b}_1 stehen.

Diese bilden auch eine kompakte Menge auf der Q ein Maximum (das notwendig $\leq \lambda_1$ sein muß) hat. Sei \vec{b}_2 ein Punkt in dem dieses Maximum angenommen wird und \vec{b}_2 der zugehörige Ortsvektor. Genauer hat Q im Punkt \vec{b}_2 ein relatives Maximum unter den Nebenbedingungen $s(\vec{x}) = 0$ und $\vec{b}_1 \bullet \vec{x} = 0$ bzw. $g_{\vec{b}_1}(\vec{x}) = 0$.

Da $\vec{\text{grad}} s(\vec{b}_2) = 2\vec{b}_2$ und $\vec{\text{grad}} g_{\vec{b}_1}(\vec{b}_2) = \vec{b}_1$ orthogonal also linear unabhängig sind, gibt es zwei LAGRANGE-Multiplikatoren, die ich mit λ_2 und μ bezeichne:

$$\vec{\text{grad}} Q(\vec{b}_2) = \lambda_2 \vec{\text{grad}} s(\vec{b}_2) + \mu \vec{\text{grad}} g_{\vec{b}_1}(\vec{b}_2) \quad \text{also} \quad 2A\vec{b}_2 = \lambda_2(2\vec{b}_2) + \mu\vec{b}_1.$$

Um einzusehen, daß $\mu = 0$ sein muß, multiplizieren wir die entstandene Gleichung skalar mit \vec{b}_1 . Dann ergibt sich links

$$2A\vec{b}_2 \bullet \vec{b}_1 = 2\vec{b}_2 \bullet A\vec{b}_1 = 2\vec{b}_2 \bullet \lambda_1 \vec{b}_1 = 2\lambda_1 \vec{b}_2 \bullet \vec{b}_1 = 0$$

(wegen der Symmetrie, darf man A rüberwerfen) und rechts

$$2\lambda_2 \vec{b}_2 \bullet \vec{b}_1 + \mu \vec{b}_1 \bullet \vec{b}_1 = 2\lambda_2 \cdot 0 + \mu \cdot 1 = \mu.$$

Kürzt man $2A\vec{b}_2 = \lambda_2(2\vec{b}_2) + 0\vec{b}_1$ noch durch 2, so folgt $A\vec{b}_2 = \lambda_2 \vec{b}_2$. Es ist also λ_2 Eigenwert und \vec{b}_2 zugehöriger Eigenvektor, der nach Konstruktion orthonormal zu \vec{b}_1 steht.

Dieselbe Rechnung wie oben zeigt dann auch noch, daß $Q(\vec{b}_2) = \lambda_2$. Der Eigenwert ist also selbst das Maximum von Q auf der verkleinerten Sphäre.

3.7.5 Und so weiter

Wenn wir bereits $k < n - 1$ Eigenwerte $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_k$ mit den zugehörigen orthonormalen Eigenvektoren $\vec{b}_1, \vec{b}_2, \dots, \vec{b}_k$ haben, so suchen wir im $k + 1$ ten Schritt das Maximum der Form Q unter den $k + 1$ Nebenbedingungen

$$s(\vec{x}) = 0, \quad g_{\vec{b}_1}(\vec{x}) = 0, \quad g_{\vec{b}_2}(\vec{x}) = 0, \quad \dots, \quad g_{\vec{b}_k}(\vec{x}) = 0,$$

Diese definieren eine (wegen $k < n$ nichtleere) kompakte Menge, also gibt es darin einen Punkt \vec{b}_{k+1} , in dem Q maximal wird. Der zugehörige Ortsvektor \vec{b}_{k+1} ist automatisch (so sind die Nebenbedingungen gerade gewählt) orthonormal zu $\vec{b}_1, \dots, \vec{b}_k$. Die im Punkt \vec{b}_{k+1} ausgerechneten Gradienten der Nebenbedingungen sind

$$2\vec{b}_{k+1}, \quad \vec{b}_1, \quad \vec{b}_2, \quad \dots, \quad \vec{b}_k$$

also linear unabhängig (wegen der Orthogonalität), daher gibt es LAGRANGE-Multiplikatoren:

$$\vec{\text{grad}} Q(\vec{b}_{k+1}) = \lambda_{k+1} \vec{\text{grad}} s(\vec{b}_{k+1}) + \mu_1 \vec{\text{grad}} g_{\vec{b}_1}(\vec{b}_{k+1}) + \dots + \mu_k \vec{\text{grad}} g_{\vec{b}_k}(\vec{b}_{k+1})$$

also $2A\vec{b}_{k+1} = 2\lambda_{k+1}\vec{b}_{k+1} + \mu_1\vec{b}_1 + \dots + \mu_k\vec{b}_k$. Skalare Multiplikation mit $\vec{b}_1, \dots, \vec{b}_k$ zeigt, daß die μ_1, \dots, μ_k alle Null sind. Etwa für \vec{b}_i ergibt sich links

$$2A\vec{b}_{k+1} \bullet \vec{b}_i = 2\vec{b}_{k+1} \bullet A\vec{b}_i = 2\vec{b}_{k+1} \bullet \lambda_i \vec{b}_i = 2\lambda_i \vec{b}_{k+1} \bullet \vec{b}_i = 0$$

und rechts

$$2\lambda_{k+1} \vec{b}_{k+1} \bullet \vec{b}_i + \mu_1 \vec{b}_1 \bullet \vec{b}_i + \dots + \mu_i \vec{b}_i \bullet \vec{b}_i + \dots + \mu_k \vec{b}_k \bullet \vec{b}_i = 2\lambda_{k+1} \cdot 0 + \mu_1 \cdot 0 + \dots + \mu_i \cdot 1 + \dots + \mu_k \cdot 0 = \mu_i.$$

Es bleibt also $2A\vec{b}_{k+1} = 2\lambda_{k+1}\vec{b}_{k+1}$ bzw. $A\vec{b}_{k+1} = \lambda_{k+1}\vec{b}_{k+1}$ übrig. Da sich λ_{k+1} auch wieder als $Q(\vec{b}_{k+1})$ erweist, ist es nicht größer als das vorherige Maximum (das auf einer größeren Menge gesucht wurde).

3.7.6 Schluß

Die oben beschriebene Prozedur läßt sich solange fortsetzen, bis $n-1$ orthonormale Eigenvektoren $\vec{b}_1, \vec{b}_2, \dots, \vec{b}_{n-1}$ gefunden sind. Zu diesen gibt es nur noch zwei orthogonale (einander entgegengesetzte) Einheitsvektoren, einer davon sei \vec{b}_n . Ich zeige, daß dieser automatisch Eigenvektor von A ist. Da die \vec{b} s offensichtlich eine ON-Basis von \mathbb{R}^n bilden gilt

$$A\vec{b}_n = \alpha_1\vec{b}_1 + \alpha_2\vec{b}_2 + \dots + \alpha_{n-1}\vec{b}_{n-1} + \lambda_n\vec{b}_n$$

für geeignete Skalare. Wegen der ON-Eigenschaft kann man diese aber ausrechnen:

$$\alpha_i = A\vec{b}_n \bullet \vec{b}_i = \vec{b}_n \bullet A\vec{b}_i = \vec{b}_n \bullet \lambda_i\vec{b}_i = \lambda_i\vec{b}_n \bullet \vec{b}_i = 0.$$

Daher ist $A\vec{b}_n = \lambda_n\vec{b}_n$ und wir haben wirklich einen Eigenvektor vor uns.

Außerdem hatte \vec{b}_n bei allen vorherigen Runden auch die Chance gewählt zu werden. Da das nicht passiert ist, kann der Wert $Q(\vec{b}_n) = A\vec{b}_n \bullet \vec{b}_n = \lambda_n\vec{b}_n \bullet \vec{b}_n = \lambda_n$ nicht größer sein als eines der früher berücksichtigten λ s. Mithin ist λ_n der kleinste Eigenwert.

3.8 Höhere partielle Ableitungen

werden auf die naheliegende Weise definiert. Ich beschränke meine Diskussion auf ein Skalarfeld $f : D \rightarrow \mathbb{R}$ von dem wir annehmen wollen, daß in allen Punkten aus D die partielle Ableitung $\frac{\partial f}{\partial x_i}$ existiert. Dann definiert diese selbst wieder ein Skalarfeld, das man partiell ableiten kann. Wenn die partielle Ableitung von $\frac{\partial f}{\partial x_i}$ nach x_j in einem Punkt \bar{a} existiert, so bezeichnet man sie statt mit $\frac{\partial \frac{\partial f}{\partial x_i}}{\partial x_j}$ normalerweise mit $\frac{\partial^2 f}{\partial x_j \partial x_i}(\bar{a})$ und spricht von einer zweiten partiellen Ableitung von f . Auch die Schreibweisen f_{xy} oder f_{xx} sind gebräuchlich.

Was man zweimal gemacht hat, kann man versuchen öfter zu tun. Falls die entsprechenden Grenzwerte existieren, entstehen dritte, vierte, \dots , r -te partielle Ableitungen.

3.8.1 Der Satz von SCHWARZ

Die Reihenfolge in der die Ableitungen genommen werden ist im allgemeinen wichtig: $\frac{\partial^2 f}{\partial x \partial y}$ kann durchaus von $\frac{\partial^2 f}{\partial y \partial x}$ verschieden sein. In fast allen Beispielen, die im Leben vorkommen, kommt es aber nicht auf die Reihenfolge an. Der entsprechende Satz stammt von dem Berliner Mathematiker HERMANN AMANDUS SCHWARZ, den wir schon von der CAUCHY-SCHWARZschen Ungleichung kennen. In ausländischen Lehrbüchern heißt dieselbe Aussage oft auch **Satz von CLAIRAUT**.

Existieren die ersten partiellen Ableitungen $\frac{\partial f}{\partial x_i}$, $\frac{\partial f}{\partial x_j}$ und die gemischte Ableitung $\frac{\partial^2 f}{\partial x_i \partial x_j}$ in einer Umgebung eines Punktes \bar{a} und ist letztere im Punkt \bar{a} stetig, dann existiert in diesem Punkt auch die andere gemischte Ableitung $\frac{\partial^2 f}{\partial x_j \partial x_i}$ und ist der ersten gleich.

Wir sprechen von einer C^2 -Funktion, falls im (offen gedachten) Definitionsbereich alle zweiten partiellen Ableitungen existieren und stetig sind. Dann sind automatisch auch die ersten Ableitungen und die Funktion selbst stetig.

Der SCHWARZsche Satz sagt unter anderem, daß es bei C^2 -Funktionen nicht auf die Reihenfolge ankommt, in der die verschiedenen zweiten Ableitungen genommen werden.

Analog definiert man C^r -Funktionen für größere r und sieht (per Induktion), daß es nicht auf die Reihenfolge ankommt.

Für solche Funktionen muss man nur angeben, nach welcher Variablen wie oft abgeleitet wurde.

Man schreibt etwa $\frac{\partial^5 f}{\partial x^2 \partial z^3}(a, b, c)$ für jede der Ableitungen f_{xzxzz} , f_{zzxzx} , f_{xzzzz} , etc.

In 4.4.2 werden wir einen sehr durchsichtigen Beweis für eine etwas schwächere Variante des SCHWARZschen Satzes bekommen. Einen Beweis der Vollversion gibt es in der Literatur und in den Ergänzungen.

3.9 Die TAYLORSche Formel für Skalarfelder

wird aus der TAYLORSchen Formel für einstellige Funktionen hergeleitet. Da n jetzt üblicherweise die Dimension des Raumes bezeichnet, werden wir das r -te TAYLOR-Polynom und das entsprechende Restglied R_r untersuchen.

Wir betrachten wieder ein Skalarfeld $f : D \rightarrow \mathbb{R}$ mit offenem Definitionsbereich, dessen $r + 1$ -te partielle Ableitungen in D stetig sein sollen. \bar{a} sei ein fester Punkt von D . Wir wollen die Funktion $f(\bar{x})$ durch ein Polynom in $(x_i - a_i)$ approximieren, wobei sich die Koeffizienten mit Hilfe der partiellen Ableitungen von f im Punkt \bar{a} bestimmen lassen sollen. Die Formeln werden durchsichtiger, wenn wir \bar{x} als $\bar{a} + \vec{v}$ für einen passenden Vektor \vec{v} schreiben. Dann soll also ein Polynom $T_r(v_1, v_2, \dots, v_n)$ r -ten Grades gefunden werden, so daß $f(\bar{a} + \vec{v}) = T_r(\vec{v}) + R_r(\vec{v})$ und das Restglied $R_r(\vec{v})$ möglichst explizit bestimmt werden kann. ... und möglichst klein ist.

3.9.1 Der Trick

besteht darin den Vektor \vec{v} vorübergehend als fest anzusehen und die einstellige Funktion $\varphi(t) = f(\bar{a} + t\vec{v})$ zu betrachten. Das funktioniert, wenn die ganze Strecke von \bar{a} nach $\bar{a} + \vec{v}$ im Definitionsbereich von f enthalten ist.

Wenn f eine C^{r+1} -Funktion ist, dann ist auch φ $r + 1$ -mal stetig differenzierbar und es gilt (TAYLORSche Formel mit LAGRANGE-Restglied; die Potenzen 1^k weggelassen)

$$(*) \quad \varphi(1) = \varphi(0) + \varphi'(0) + \frac{1}{2}\varphi''(0) + \dots + \frac{1}{r!}\varphi^{(r)}(0) + \frac{1}{(r+1)!}\varphi^{(r+1)}(\xi)$$

für ein geeignetes ξ zwischen 0 und 1. Setzt man für φ wieder f ein, erhält man daraus $f(\bar{a} + \vec{v}) = f(\bar{a}) + \dots$. Um den Ausdruck der für \dots steht ansprechend und sachgerecht aufzuschreiben, müssen wir noch überlegen, wie sich die Ableitungen von φ durch die partiellen Ableitungen von f ausdrücken lassen. Die erste Ableitung von φ also $\frac{d}{dt}f(\bar{a} + tv_1, a_2 + tv_2, \dots, a_n + tv_n)$ ergibt sich nach der Kettenregel:

$$\begin{aligned} \frac{\partial f}{\partial x_1}(\bar{a} + t\vec{v}) \cdot \frac{d}{dt}(a_1 + tv_1) + \frac{\partial f}{\partial x_2}(\bar{a} + t\vec{v}) \cdot \frac{d}{dt}(a_2 + tv_2) + \dots + \frac{\partial f}{\partial x_n}(\bar{a} + t\vec{v}) \cdot \frac{d}{dt}(a_n + tv_n) \\ = \frac{\partial f}{\partial x_1}(\bar{a} + t\vec{v})v_1 + \frac{\partial f}{\partial x_2}(\bar{a} + t\vec{v})v_2 + \dots + \frac{\partial f}{\partial x_n}(\bar{a} + t\vec{v})v_n. \end{aligned}$$

Um die zweite Ableitung zu bestimmen, muß man jeden Summanden $\frac{\partial f}{\partial x_i}(\bar{a} + t\vec{v})v_i$ nach t ableiten. Das geht nach demselben Schema und liefert

$$\begin{aligned} \frac{d^2}{dt^2}f(\bar{a} + t\vec{v}) &= \frac{\partial}{\partial x_1} \frac{\partial f}{\partial x_1}(\bar{a} + t\vec{v})v_1v_1 + \frac{\partial}{\partial x_2} \frac{\partial f}{\partial x_1}(\bar{a} + t\vec{v})v_1v_2 + \dots + \frac{\partial}{\partial x_n} \frac{\partial f}{\partial x_1}(\bar{a} + t\vec{v})v_1v_n \\ &+ \frac{\partial}{\partial x_1} \frac{\partial f}{\partial x_2}(\bar{a} + t\vec{v})v_2v_1 + \frac{\partial}{\partial x_2} \frac{\partial f}{\partial x_2}(\bar{a} + t\vec{v})v_2v_2 + \dots + \frac{\partial}{\partial x_n} \frac{\partial f}{\partial x_2}(\bar{a} + t\vec{v})v_2v_n \\ &\vdots \\ &+ \dots \\ &\vdots \\ &+ \frac{\partial}{\partial x_1} \frac{\partial f}{\partial x_n}(\bar{a} + t\vec{v})v_nv_1 + \frac{\partial}{\partial x_2} \frac{\partial f}{\partial x_n}(\bar{a} + t\vec{v})v_nv_2 + \dots + \frac{\partial}{\partial x_n} \frac{\partial f}{\partial x_n}(\bar{a} + t\vec{v})v_nv_n \end{aligned}$$

Mit Summenzeichen wird es kurz und übersichtlich: $\sum_{i,j=1}^n \frac{\partial^2 f}{\partial x_i \partial x_j}(\bar{a} + t\vec{v})v_i v_j$.

Für die höheren Ableitungen verfährt man schrittweise genauso. Das allgemeine Resultat kann man so aufschreiben:

$$\varphi^{(k)}(t) = \frac{d^k}{dt^k}f(\bar{a} + t\vec{v}) = \sum_{i_1=1}^n \sum_{i_2=1}^n \dots \sum_{i_k=1}^n \frac{\partial^k f}{\partial x_{i_1} \partial x_{i_2} \dots \partial x_{i_k}}(\bar{a} + t\vec{v})v_{i_1}v_{i_2} \dots v_{i_k}.$$

Diese immer länger werdenden Summen kann man nun in (*) einsetzen und erhält die TAYLORSche Formel. Ich will das nicht allgemein machen, weil das Resultat ein Wust von Summenzeichen wird, aus dem nicht viel hervorgeht. In der Praxis kriegen Sie das bei Bedarf schon hin.

Wegen der C^{r+1} Voraussetzung dürfen die partiellen Ableitungen in der Reihenfolge vertauscht werden. Also kann man von den n^k Gliedern der oben aufgeschriebenen $\varphi^{(k)}$ - Summe noch viele zusammenfassen, nämlich jeweils alle, in denen gleich oft nach x_1, x_2 usw. abgeleitet wurde, wenn auch in verschiedenen Reihenfolgen. Auch das schreibe ich nicht allgemein auf. Die Koeffizienten sind dieselben, die entstehen, wenn die Summe $(x_1 + x_2 + \dots + x_n)^k$ ausmultipliziert wird. Ich beschränke mich auf

3.9.2 Die TAYLORSche Formel für Funktionen von zwei und drei Variablen

Für zweistellige Funktionen kann man alles einigermaßen übersichtlich sortieren. Zunächst ergibt sich bei den Ableitungen eine vollständige Analogie zum Binomischen Satz.

Für $\varphi(t) = f(a + tu, b + tv)$ gilt

$$\begin{aligned} \varphi(0) &= f(a, b) \\ \varphi'(0) &= \frac{\partial f}{\partial x}(a, b)u + \frac{\partial f}{\partial y}(a, b)v \\ \varphi''(0) &= \frac{\partial^2 f}{\partial x^2}(a, b)u^2 + 2 \frac{\partial^2 f}{\partial x \partial y}(a, b)uv + \frac{\partial^2 f}{\partial y^2}(a, b)v^2 \\ \varphi'''(0) &= \frac{\partial^3 f}{\partial x^3}(a, b)u^3 + 3 \frac{\partial^3 f}{\partial x^2 \partial y}(a, b)u^2v + 3 \frac{\partial^3 f}{\partial x \partial y^2}(a, b)uv^2 + \frac{\partial^3 f}{\partial y^3}(a, b)v^3 \\ &\vdots \\ \varphi^{(k)}(0) &= \frac{\partial^k f}{\partial x^k}u^k + k \frac{\partial^k f}{\partial x^{k-1} \partial y}u^{k-1}v + \binom{k}{2} \frac{\partial^k f}{\partial x^{k-2} \partial y^2}u^{k-2}v^2 + \dots + \frac{\partial^k f}{\partial y^k}v^k \\ &= \sum_{p+q=k} \frac{k!}{p!q!} \frac{\partial^k f}{\partial x^p \partial y^q} u^p v^q, \end{aligned}$$

wobei ∂x^0 bedeuten soll, daß nach x nicht abgeleitet wurde. Die Binomialkoeffizienten $\binom{k}{p}$ zählen ja gerade, wieviele k -Tupel aus I xen und Y psilon es gibt, bei denen p mal x und $q = k - p$ mal y auftaucht (bzw. umgekehrt).

Nun ist $\binom{k}{p} = \frac{k!}{p!q!}$ und in der TAYLORSchen Formel wird die k -te Ableitung durch $k!$ geteilt. Dann ergibt sich nach einsetzen in (*) folgendes:

$$f(a + u, b + v) = f(a, b) + \sum_{p+q \leq r} \frac{1}{p! \cdot q!} \frac{\partial^{p+q} f}{\partial x^p \partial y^q}(a, b) \cdot u^p v^q + R_r(u, v).$$

Wer nicht davor zurückschreckt, kann auch $f(a, b) = \frac{1}{0!0!} \frac{\partial^0 f}{\partial x^0 \partial y^0}(a, b) u^0 v^0$ in die Summe einbeziehen. Das Restglied ergibt sich als

$$R_r(u, v) = \sum_{p+q=r+1} \frac{1}{p! \cdot q!} \frac{\partial^{p+q} f}{\partial x^p \partial y^q}(a + \xi u, b + \xi v) \cdot u^p v^q.$$

Ganz analog ergibt sich das TAYLORpolynom (und das Restglied) in höheren Dimensionen. Ich schreibe noch den Fall von drei Veränderlichen auf (danach braucht man Indizes mit Indizes, wodurch die Sache nicht viel klarer wird):

$$T_r(u, v, w) = \sum_{p+q+s \leq r} \frac{1}{p! \cdot q! \cdot s!} \frac{\partial^{p+q+s} f}{\partial x^p \partial y^q \partial z^s}(a, b, c) \cdot u^p v^q w^s$$

Voraussetzung für die Gültigkeit ist C^{r+1} und daß die Funktion f auf der ganzen Strecke von \bar{a} nach $\bar{a} + \vec{v}$ definiert ist.

3.9.3 Die Restgliedabschätzung im allgemeinen Fall

Wir nehmen immer noch an, unser Skalarfeld ist C^{r+1} . Dann sind alle $r + 1$ -ten partiellen Ableitungen stetig und also in einer abgeschlossenen Kugel mit Mittelpunkt \bar{a} beschränkt, etwa durch die Konstante M . Das Restglied $R_r(\vec{v})$ bei der TAYLOR-Formel ist eine Summe aus $n^{\binom{r+1}{n}}$ dieser partiellen Ableitungen, geteilt durch Fakultäten und multipliziert mit jeweils $r + 1$ Faktoren v_i . Wegen $|v_i| \leq \|\vec{v}\|$, können wir jeden Summanden durch $M \cdot \|\vec{v}\|^{r+1}$ abschätzen. Da es nur endlich viele Summanden sind, ergibt sich folgendes.

Das Restglied R_r der TAYLOREntwicklung eines C^{r+1} -Skalarfeldes f in einem inneren Punkt \bar{a} des Definitionsbereiches kann für alle genügend kleinen Vektoren \vec{v} durch $C \cdot \|\vec{v}\|^{r+1}$ abgeschätzt werden, wo C von f und \bar{a} aber nicht von \vec{v} abhängt. Speziell ist $\lim_{\vec{v} \rightarrow \vec{0}} \frac{R_r(\vec{v})}{\|\vec{v}\|^r} = 0$.

Das ist, im Gegensatz zur oben angegebenen LAGRANGE-Form, die PEANO-Form des Restgliedes:

$$f(\bar{a} + \vec{v}) = T_r(\vec{v}) + \rho(\vec{v}) \cdot \|\vec{v}\|^r, \quad \text{wobei} \quad \lim_{\vec{v} \rightarrow \vec{0}} \rho(\vec{v}) = 0.$$

3.10 Hinreichende Bedingungen für Extremwerte

Es sei $D \subseteq \mathbb{R}^n$ $f : D \rightarrow \mathbb{R}$ eine C^2 -Funktion und \bar{a} ein innerer Punkt von D , in dem der Gradient von f verschwindet (sogenannter kritischer Punkt). Um zu untersuchen, ob in \bar{a} tatsächlich ein lokaler Extremwert angenommen wird, werden wir jetzt die zweiten Ableitungen von f heranziehen, genauer die aus ihnen gebildete quadratische und (wegen C^2 nach SCHWARZ) symmetrische sogenannte

3.10.1 HESSE-Matrix

$$H_f(\bar{a}) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(\bar{a}) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(\bar{a}) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(\bar{a}) \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(\bar{a}) & \frac{\partial^2 f}{\partial x_2^2}(\bar{a}) & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n}(\bar{a}) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(\bar{a}) & \frac{\partial^2 f}{\partial x_n \partial x_2}(\bar{a}) & \cdots & \frac{\partial^2 f}{\partial x_n^2}(\bar{a}) \end{pmatrix}.$$

Zu dieser Matrix gehört eine quadratische Form, die für den Vektor $\vec{v} \in \mathbb{R}^n$ nach der Vorschrift

$$(H_f(\bar{a})\vec{v}) \bullet \vec{v} \quad \text{bzw. ausführlich} \quad \sum_{i,j=1}^n \frac{\partial^2 f}{\partial x_i \partial x_j}(\bar{a}) v_i v_j$$

berechnet wird. Bis auf den Faktor $\frac{1}{2}$ ist die HESSE-Form der quadratische Teil des in \bar{a} entwickelten TAYLOR-Polynoms

$$f(\bar{a}) + \sum_{i=1}^n \frac{\partial f}{\partial x_i}(\bar{a}) v_i + \frac{1}{2} \sum_{i,j=1}^n \frac{\partial^2 f}{\partial x_i \partial x_j}(\bar{a}) v_i v_j + \dots$$

Ist \bar{a} kritisch, so wird der lineare Teil 0 und das Verhalten von f in der Nähe von \bar{a} läßt sich meist aus dem Verhalten des quadratischen Teils beurteilen. Dessen positive/negative (semi-)Definitheit ist ausschlaggebend für den Charakter des Punktes \bar{a} . Bevor ich das genauer formuliere eine kurze

3.10.2 Abschweifung: positiv definite quadratische Formen

Definition. Man nennt eine quadratische Form $Q(\vec{v}) = H\vec{v} \bullet \vec{v}$ *positiv definit*, wenn Sie außer für den Nullvektor nur strikt positive Werte annimmt. *Positiv semidefinit* bedeutet, daß $Q(\vec{v}) \geq 0$ für alle \vec{v} . Entsprechend wird *negativ (semi) definit* definiert. Quadratische Formen, die sowohl positive als auch negative Werte annehmen, nennt man *indefinit*.

Für jede symmetrische quadratische Matrix H mit zugehöriger quadratischer Form $Q(\vec{v}) = H\vec{v} \bullet \vec{v}$ sind folgende Aussagen äquivalent:

- (1) Q ist positiv definit.
- (2) Alle Eigenwerte von H sind positiv.

(3) Alle 'Eckdeterminanten' (vornehm 'Hauptminoren')

$$h_{11}, \quad \begin{vmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{vmatrix}, \quad \begin{vmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{vmatrix}, \quad \dots, \quad \det H$$

sind positiv.

Die Äquivalenz (1) \iff (3) nennt man **Kriterium von SYLVESTER**.

Für eine symmetrische 2×2 -Matrix $\begin{pmatrix} a & b \\ b & c \end{pmatrix}$ führen beide Kriterien auf

$$a > 0 \quad \text{und} \quad c > 0 \quad \text{und} \quad ac > b^2$$

(wobei $c > 0$ aus $a > 0$ und $ac > b^2$ folgt).

Wenn man der Matrix die negative Definitheit ansehen will, braucht man nur ihre entgegengesetzte auf positive Definitheit zu untersuchen. Da bei den Eckdeterminanten die -1 aus allen Spalten herausgezogen werden kann, ergibt sich je nach Spaltenzahl die Forderung eines positiven oder negativen Vorzeichens. Genauer: $H\vec{v} \bullet \vec{v}$ ist genau dann negativ definit, wenn alle Eigenwerte negativ sind, bzw. die Eckdeterminanten alternierende Vorzeichen haben (mit $-$ beginnend).

Beweis der Äquivalenzen.

Für die symmetrische Matrix H gibt es eine ON-Basis $\vec{b}_1, \vec{b}_2, \dots, \vec{b}_n$ von \mathbb{R}^n aus Eigenvektoren. Die Numerierung soll (schon im Hinblick auf die Anwendung unten) so vorgenommen sein, daß die zugehörigen Eigenwerte wachsen: $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$.

Sei jetzt Q positiv definit, dann ist $0 < Q(\vec{b}_i) = H\vec{b}_i \bullet \vec{b}_i = \lambda_i \vec{b}_i \bullet \vec{b}_i = \lambda_i$.

Sind umgekehrt alle λ s positiv und \vec{x} gegeben, so können wir $\vec{x} = \sum_{i=1}^n \beta_i \vec{b}_i$ schreiben und errechnen

$$Q(\vec{x}) = H\vec{x} \bullet \vec{x} = \left(\sum \beta_i H\vec{b}_i \right) \bullet \left(\sum \beta_i \vec{b}_i \right) = \sum \lambda_i \beta_i^2 > 0,$$

außer alle β_i sind null, d.h. für $\vec{x} = \vec{0}$. Damit ist (1) \iff (2) bewiesen.

Für (1) \Rightarrow (3) sei Q positiv definit. Da die Determinante von H gleich dem Produkt der Eigenwerte ist, folgt aus obiger Äquivalenz sofort $\det H > 0$. Mit Q ist auch die auf \mathbb{R}^k definierte Form $\vec{v} \mapsto Q(v_1, \dots, v_k, 0, \dots, 0)$ positiv definit. Die Determinante der zugehörigen Matrix, also der k -te Hauptminor von H ist dann auch positiv.

Es bleibt (3) \Rightarrow (1) zu sehen. Das geht induktiv nach n und der Fall $n = 1$ ist trivial. Nehmen wir an, die Aussage wäre für symmetrische $(n-1) \times (n-1)$ -Matrizen richtig. Sei H' das Resultat der Streichung von letzter Zeile und Spalte in der Matrix H . Dann sind alle Hauptminoren von H' auch Hauptminoren von H , daher positiv und die Induktionsvoraussetzung besagt, daß die zugehörige Form Q' positiv definit auf \mathbb{R}^{n-1} ist. Wenn wir \mathbb{R}^{n-1} mit dem Unterraum U von \mathbb{R}^n identifizieren, der aus allen Vektoren mit letzter Koordinate 0 besteht, dann ist Q auf diesem Unterraum mit Q' identisch: $Q(\vec{u}) = Q(u_1, \dots, u_{n-1}, 0) = Q'(u_1, \dots, u_{n-1}) > 0$ falls $U \ni \vec{u} \neq \vec{0}$. Sei weiterhin $V := \text{span}\{\vec{b}_1, \vec{b}_2\}$. Dann gilt für $\vec{v} = \beta_1 \vec{b}_1 + \beta_2 \vec{b}_2 \in V$ wie oben, daß

$$Q(\vec{v}) = \lambda_1 \beta_1^2 + \lambda_2 \beta_2^2 \leq \lambda_2 (\beta_1^2 + \beta_2^2) = \lambda_2 \|\vec{v}\|^2.$$

Der $n-1$ dimensionale Unterraum U und der zweidimensionale Unterraum V haben einen mindestens eindimensionalen Durchschnitt. Sei $\vec{w} \neq \vec{0}$ in diesem Durchschnitt. Dann gilt

$$0 < Q(\vec{w}) \quad \text{wegen} \quad \vec{w} \in U \quad \text{und} \quad Q(\vec{w}) \leq \lambda_2 \|\vec{w}\|^2 \quad \text{wegen} \quad \vec{w} \in V.$$

Das zeigt $\lambda_2 > 0$. Wegen der angenommenen Ordnung der Eigenwerte sind dann auch $\lambda_3, \dots, \lambda_n$ positiv.

Nach Voraussetzung ist aber auch das Produkt aller λ s positiv (nämlich $\det H$ als Hauptminor). Dann muß auch λ_1 und mithin alle Eigenwerte positiv sein.

3.10.3 Über den Charakter von kritischen Punkten

Mit den Bezeichnungen von oben:

- (1) Wenn f im kritischen Punkt \bar{a} ein lokales Minimum annimmt, so ist die zur HESSE-Matrix $H_f(\bar{a})$ gehörende quadratische Form positiv semidefinit.
- (2) Ist die zu $H_f(\bar{a})$ gehörende Form positiv definit, so liegt in \bar{a} ein striktes lokales Minimum vor.

Analoge Aussagen gelten für lokale Maxima und negative (semi-)Definitheit.

Beim **Beweis** beider Aussagen ist wieder folgende einstellige Hilfsfunktion φ nützlich, die nach Fixierung eines Vektors $\vec{v} \neq \vec{0}$ gebildet wird (eigentlich müsste man $\varphi_{\vec{v}}$ schreiben):

$$\varphi(t) := f(\bar{a} + t\vec{v}).$$

Da \bar{a} ein innerer Punkt des Definitionsbereiches von f ist, ist φ mindestens in einem kleinen Intervall $] -\delta, \delta[$ definiert, das 0 enthält. Da f als C^2 -Funktion vorausgesetzt wurde, ist auch φ zweimal stetig differenzierbar und die Ableitungen lassen sich nach der Kettenregel bestimmen (vgl. die Rechnungen im Zusammenhang mit der TAYLORSchen Formel):

$$\varphi'(t) = \sum_{i=1}^n \frac{\partial f}{\partial x_i}(\bar{a} + t\vec{v}) v_i = \vec{\text{grad}} f(\bar{a} + t\vec{v}) \bullet \vec{v}$$

und

$$\varphi''(t) = \sum_{j,i=1}^n \frac{\partial^2 f}{\partial x_j \partial x_i}(\bar{a} + t\vec{v}) v_j v_i = (H_f(\bar{a} + t\vec{v})\vec{v}) \bullet \vec{v}$$

Zu Aussage (1). Wenn f in \bar{a} ein lokales Minimum annimmt, dann nimmt φ in 0 ein lokales Minimum an. Nach unseren Erkenntnissen aus dem vorigen Semester ist dann $\varphi'(0) = 0$ und $\varphi''(0) \geq 0$ (letzteres weil sonst φ in 0 ein striktes lokales Maximum annehmen würde). Weil $\varphi''(0) = H_f(\bar{a})\vec{v} \bullet \vec{v} \geq 0$ unabhängig von also für alle $\vec{v} \neq \vec{0}$ gilt, muß die zu $H_f(\bar{a})$ gehörende Form positiv semidefinit sein.

Zu Aussage (2). Sei jetzt die zu $H_f(\bar{a})$ gehörende Form positiv definit. Um \bar{a} als striktes lokales Minimum nachzuweisen, suchen wir ein positives r so daß $f(\bar{a} + \vec{v}) > f(\bar{a})$ für alle \vec{v} mit $0 < \|\vec{v}\| < r$.

Die positive Definitheit von $H_f(\bar{a})$ macht sich an der Positivität der Eckdeterminanten der Matrix $H_f(\bar{x})$ im Punkt $\bar{x} = \bar{a}$ fest. Diese Determinanten sind Summen von \pm Produkten von zweiten Ableitungen von f und daher nach Voraussetzung stetig. Wegen der Robustheit des Vorzeichens stetiger Funktionen, bleiben die Eckdeterminanten von $H_f(\bar{x})$ auch in einer Umgebung von \bar{a} positiv. Es gibt also ein $r > 0$, so daß für alle $\bar{x} \in K_{\bar{a}}(r) \subseteq D$ und alle $\vec{v} \neq \vec{0}$ gilt $(H_f(\bar{x})\vec{v}) \bullet \vec{v} > 0$. Dieses r ist wie gewünscht. Sei nämlich $0 < \|\vec{v}\| < r$. Wir benutzen die TAYLORSche Formel für die (ausgehend von diesem \vec{v} gebildete) Hilfsfunktion φ :

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

wobei $0 < \xi < 1$. Oben hatten wir die Ableitungen von φ ausgerechnet. Setzt man das Ergebnis ein, ergibt sich

$$f(\bar{a} + \vec{v}) = f(\bar{a}) + \vec{\text{grad}} f(\bar{a}) \bullet \vec{v} + \frac{1}{2} (H_f(\bar{a} + \xi\vec{v})\vec{v}) \bullet \vec{v}$$

Nun ist aber $\vec{\text{grad}} f(\bar{a}) = \vec{0}$ und wegen $\|\xi\vec{v}\| = \xi\|\vec{v}\| \leq \|\vec{v}\| < r$ auch $(H_f(\bar{a} + \xi\vec{v})\vec{v}) \bullet \vec{v} > 0$. Daher $f(\bar{a} + \vec{v}) > f(\bar{a})$, wie gewünscht.

Bemerkungen. Wie gesehen, ist positive Semidefinitheit der HESSE-Form notwendig für das Vorliegen eines lokalen Minimums. Sie ist aber nicht hinreichend. Beispiel ist die Funktion $f(x, y) = y^2 - x^3$. Positive Definitheit ist umgekehrt auch nicht notwendig: $f(x, y) = x^4 + y^4$ hat in $(0, 0)$ ein absolutes Minimum, aber die HESSE-Form ist dort die Nullform.

3.10.4 Der Spezialfall zweier Veränderlicher

Gilt für die Funktion $f(x, y)$ im inneren Punkt (a, b) ihres Definitionsbereiches

$$\frac{\partial f}{\partial x}(a, b) = \frac{\partial f}{\partial y}(a, b) = 0 \quad \text{und} \quad \frac{\partial^2 f}{\partial x^2}(a, b) > 0 \quad \text{und} \quad \frac{\partial^2 f}{\partial x^2}(a, b) \cdot \frac{\partial^2 f}{\partial y^2}(a, b) > \left(\frac{\partial^2 f}{\partial x \partial y}(a, b) \right)^2,$$

so hat f in (a, b) ein striktes lokales Minimum. Die Bedingung

$$\frac{\partial f}{\partial x}(a, b) = \frac{\partial f}{\partial y}(a, b) = 0 \quad \text{und} \quad \frac{\partial^2 f}{\partial x^2}(a, b) < 0 \quad \text{und} \quad \frac{\partial^2 f}{\partial x^2}(a, b) \cdot \frac{\partial^2 f}{\partial y^2}(a, b) > \left(\frac{\partial^2 f}{\partial x \partial y}(a, b) \right)^2,$$

ist hinreichend für ein striktes lokales Maximum.

Wenn die Determinante der HESSE-Matrix negativ wird, ist die entsprechende Form indefinit. Als Übungsaufgabe können Sie das im Fall von zwei Variablen beweisen und anschließen daraus folgern, daß im Fall $\text{grad } f(a, b) = \vec{0}$ und $\det H_f(a, b) < 0$ ein Sattelpunkt vorliegt, d.h. es gibt dann zwei (aufeinander senkrecht stehende) Einheitsvektoren $\begin{pmatrix} u \\ v \end{pmatrix}$ und $\begin{pmatrix} p \\ q \end{pmatrix}$, so daß $f(a + tu, b + tv)$ für $t = 0$ ein striktes Minimum und $f(a + tp, b + tq)$ ein striktes Maximum annimmt.

3.10.5 Beispiel

Wir untersuchen die Funktion $x + y + \frac{1}{xy}$ auf Extremwerte im (offenen) Definitionsbereich $x, y > 0$. Nullsetzen der ersten Ableitungen liefert

$$\frac{\partial}{\partial x} : 1 - \frac{1}{x^2 y} = 0, \text{ also } x^2 y = 1 \quad \text{und} \quad \frac{\partial}{\partial y} : 1 - \frac{1}{x y^2} = 0, \text{ also } x y^2 = 1.$$

Dividiert man die Gleichungen durcheinander, so folgt $\frac{x}{y} = 1$, also $y = x$. Eingesetzt $x^3 = 1$ und daher $x = y = 1$. Das ist der einzige verdächtige Punkt. Ob wirklich ein Extremwert vorliegt und welcher Natur er ist, ergibt sich durch Untersuchung der HESSE-Matrix. Die zweiten partiellen Ableitungen sind

$$\frac{\partial^2}{\partial x^2} = \frac{2}{y x^3}, \quad \frac{\partial^2}{\partial y^2} = \frac{2}{y^3 x}, \quad \text{und} \quad \frac{\partial^2}{\partial x \partial y} = \frac{1}{x^2 y^2}.$$

Daher ist die HESSE-Matrix im Punkt $(1, 1)$ gleich $\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$. Nach dem oben genannten Kriterium ist sie positiv definit, also liegt ein striktes lokales Minimum vor.

4 Integrale

4.1 Darstellung von Funktionen durch Integrale

Gegeben sei eine zweistellige Funktion $f(x, t)$ die (mindestens) auf einem Rechteck $[a, b] \times [c, d]$ definiert sein soll. Wir nehmen an, daß für jedes fixierte $x \in [a, b]$ die einstellige Funktion

$$[c, d] \rightarrow \mathbb{R} \quad \text{mit} \quad t \mapsto f(x, t)$$

integrierbar ist. Unten wird diese Bedingung immer deshalb erfüllt sein, weil f auf dem Rechteck zumindest stetig ist. Das Integral $\int_c^d f(x, t) dt$ hängt von dem festgelassenen x ab; wir bezeichnen es mit $F(x)$. Man sagt, daß die neue Funktion $F : [a, b] \rightarrow \mathbb{R}$ durch das Integral $\int_c^d f(x, t) dt$ dargestellt wird. Wir wollen in diesem Abschnitt untersuchen, wie sich Eigenschaften von f auf F übertragen. Die eben beschriebenen Voraussetzungen und Bezeichnungen sind bis auf Widerruf fest und werden nicht dauernd wiederholt.

4.1.1 Vorbemerkungen

Der Hauptsatz der Differential- und Integralrechnung liefert folgende Formeln:

$$\int_c^d \frac{\partial f}{\partial t}(x, t) dt = f(x, d) - f(x, c) \quad \text{und} \quad \frac{\partial}{\partial y} \int_c^y f(x, t) dt = f(x, y).$$

Das sieht man sofort, wenn man bei festem x die Hilfsfunktion $\varphi(t) = f(x, t)$ einführt. Die zweite Formel gilt für stetiges f ; die erste setzt die Stetigkeit von $\frac{\partial f}{\partial t}$ voraus.

4.1.2 Stetigkeit

Satz. Ist f stetig auf dem Rechteck $[a, b] \times [c, d]$, so ist F stetig auf dem Intervall $[a, b]$.

Wir werden sogar die gleichmäßige Stetigkeit nachweisen und schätzen dazu ab:

$$\left| \int_c^d f(x_1, t) dt - \int_c^d f(x_2, t) dt \right| = \left| \int_c^d [f(x_1, t) - f(x_2, t)] dt \right| \leq \int_c^d |f(x_1, t) - f(x_2, t)| dt.$$

Wegen der Stetigkeit auf dem kompakten(!) Rechteck ist f sogar gleichmäßig stetig. Es gibt also für jedes vorgegebene positive ε ein $\delta > 0$, so daß für alle $a \leq x_1, x_2 \leq b$ und $c \leq t_1, t_2 \leq d$ aus $\sqrt{(x_1 - x_2)^2 + (t_1 - t_2)^2} < \delta$ folgt $|f(x_1, t_1) - f(x_2, t_2)| < \frac{\varepsilon}{2(d-c)}$. Gilt also $|x_1 - x_2| < \delta$, so ist der Integrand des letzten Integrals überall $< \frac{\varepsilon}{2(d-c)}$, daher

$$|F(x_1) - F(x_2)| \leq \int_c^d |f(x_1, t) - f(x_2, t)| dt \leq \int_c^d \frac{\varepsilon}{2(d-c)} dt = \frac{\varepsilon}{2} < \varepsilon,$$

wie verlangt.

Wenn man die Stetigkeit mit Hilfe von Grenzwerten ausdrückt, kann man die gerade bewiesene Aussage auch als

$$\lim_{x \rightarrow x_0} \int_a^b f(x, t) dt = \int_a^b f(x_0, t) dt$$

schreiben.

4.1.3 Verallgemeinerung auf variable Grenzen

Manchmal braucht man Integraldarstellungen etwas allgemeinerer Art:

$$F(x) = \int_{g(x)}^{h(x)} f(x, t) dt.$$

Dabei soll für die Funktionen $g, h : [a, b] \rightarrow \mathbb{R}$ stets $g(x) \leq h(x)$ gelten und $f(x, t)$ muß mindestens auf der Menge $\{(x, t) : a \leq x \leq b \text{ und } g(x) \leq t \leq h(x)\}$ definiert sein.

Satz. Wenn f, g und h stetig sind, so ist auch F stetig.

Das führt man am elegantesten durch eine Variablentransformation (Substitution) auf den Fall mit festen Grenzen zurück. Bei festem $x \in [a, b]$ führt man die neue Integrationsvariable u mit

$$t = g(x) + u(h(x) - g(x)) \quad \text{und} \quad dt = [h(x) - g(x)] du$$

ein. Dann variiert u zwischen 0 und 1 und das Integral wird zu

$$F(x) = \int_0^1 f(x, g(x) + u(h(x) - g(x))) \cdot [h(x) - g(x)] du$$

Jetzt steht rechts ein Integrationsintervall mit festen Grenzen und der Integrand hängt in $[a, b] \times [0, 1]$ stetig von x und u ab. Nach dem vorigen Satz ist $F(x)$ also stetig.

4.1.4 Differenzierbarkeit; LEIBNIZ-Formel

Wir betrachten jetzt wieder den Fall fester Grenzen.

Satz. Angenommen f ist auf einer das Rechteck umfassenden offenen Menge definiert und besitzt eine stetige partielle Ableitung nach x . Dann ist F (stetig) differenzierbar und es gilt die von LEIBNIZ stammende Formel

$$F'(x) = \frac{d}{dx} \int_c^d f(x, t) dt = \int_c^d \frac{\partial f}{\partial x}(x, t) dt.$$

Man sagt dann, daß Integration und Differentiation vertauschbar sind. Für C^1 -Funktionen f ist die Voraussetzung stets mehr als erfüllt. Falls $\frac{\partial f}{\partial x}$ allerdings unstetig ist, braucht die Formel nicht zu gelten.

Zum **Beweis** formen wir den Differenzenquotienten von F um, wobei der Mittelwertsatz der Differentialrechnung die ξ s liefert. Sie hängen jeweils von (h und) t ab, was durch den Index angedeutet werden soll. Jedes ξ_t liegt zwischen x und $x + h$.

$$\begin{aligned} \frac{F(x+h)-F(x)}{h} &= \frac{1}{h} \int_c^d [f(x+h, t) - f(x, t)] dt = \frac{1}{h} \int_c^d h \cdot \frac{\partial f}{\partial x}(\xi_t, t) dt \\ &= \int_c^d \frac{\partial f}{\partial x}(x, t) dt + \int_c^d \left(\frac{\partial f}{\partial x}(\xi_t, t) - \frac{\partial f}{\partial x}(x, t) \right) dt. \end{aligned}$$

Es bleibt zu sehen, daß der zweite Summand mit h gegen Null geht. Sein Betrag ist kleiner oder gleich

$$\int_c^d \left| \frac{\partial f}{\partial x}(\xi_t, t) - \frac{\partial f}{\partial x}(x, t) \right| dt.$$

Wegen der gleichmäßigen Stetigkeit von $\frac{\partial f}{\partial x}$ auf dem Rechteck wird der Integrand kleiner als $\frac{\varepsilon}{d-c}$ sobald $|\xi_t - x| < |h| < \delta$ ein passendem δ wird.

4.1.5 Anwendungsbeispiel

Wir wollen das Integral $\int_0^a t^2 \cos t dt$ berechnen. Die folgende Methode stellt eine Alternative zur partiellen Integration dar. Richtig gut funktioniert sie nur für spezielle Integrale. Ihre eigentliche Kraft entfaltet sie bei der Berechnung uneigentlicher Integrale. Die Idee besteht in der Einführung eines neuen Parameters. Wir berechnen $\int_0^a t^2 \cos(xt) dt$ und setzen hinterher $x = 1$.

$$\begin{aligned} \int_0^a t^2 \cos(xt) dt &= - \int_0^a \left(\frac{\partial^2}{\partial x^2} \cos(tx) \right) dt = - \frac{d^2}{dx^2} \int_0^a \cos(tx) dt = - \frac{d^2}{dx^2} \left[\frac{\sin(tx)}{x} \Big|_{t=0}^{t=a} \right] \\ &= - \frac{d^2}{dx^2} \left[\frac{\sin(ax)}{x} \right] = - \left[\frac{2 \sin(ax)}{x^3} - \frac{2a \cos(ax)}{x^2} - \frac{a^2 \sin(ax)}{x} \right]. \end{aligned}$$

Für $x = 1$ ergibt sich

$$\int_0^a t^2 \cos t dt = (a^2 - 2) \sin a + 2a \cos a.$$

4.1.6 Die Verallgemeinerung auf variable Grenzen

ist diesmal etwas länglicher zu begründen; ich deute das nur an. Ergebnis ist die Formel

$$\frac{d}{dx} \int_{g(x)}^{h(x)} f(x, t) dt = \int_{g(x)}^{h(x)} \frac{\partial f}{\partial x}(x, t) dt + f(x, h(x)) \cdot h'(x) - f(x, g(x)) \cdot g'(x),$$

wobei die Existenz und (sicherheitshalber) Stetigkeit aller Funktionen und Ableitungen auf der rechten Seite vorausgesetzt werden muß.

Man erhält sie, indem man die Transformation aus 4.1.3 vornimmt und dann 4.1.4 anwendet. Das Ergebnis hat zunächst keine Ähnlichkeit mit der angestrebten Formel, kann aber mit partieller Integration (rückwärts gelesen) auf diese gebracht werden. Das ist etwas unangenehm, aber machbar (Übungsaufgabe).

Ein anderer Weg zu dieser Formel, mit dessen Hilfe man sich das Resultat auch schnell wieder herleiten kann, besteht in der Betrachtung der dreistelligen Funktion $G(x, y, z) = \int_y^z f(x, t) dt$ und der Berechnung ihrer partiellen Ableitungen

$$\frac{\partial G}{\partial x}(x, y, z) = \int_y^z \frac{\partial f}{\partial x}(x, t) dt \quad \frac{\partial G}{\partial y}(x, y, z) = -f(x, y), \quad \frac{\partial G}{\partial z}(x, y, z) = f(x, z).$$

Bei der Ableitung von G nach x werden ja die Integrationsgrenzen y, z fest gelassen, also ist 4.1.4 anwendbar. Die beiden anderen Ableitungen kommen vom Hauptsatz der Differential- und Integralrechnung. Anschließend kann man $F(x) = G(x, g(x), h(x))$ nach der Kettenregel ableiten

$$F'(x) = \frac{\partial G}{\partial x}(x, g(x), h(x)) + \frac{\partial G}{\partial y}(x, g(x), h(x)) \cdot g'(x) + \frac{\partial G}{\partial z}(x, g(x), h(x)) \cdot h'(x),$$

und erhält durch Einsetzen die gewünschte Formel.

4.1.7 Integration; der Satz von FUBINI

Wir gehen aus von einem stetigen f . Dann hatten wir F als stetig erkannt, also existiert $\int_a^b F(x) dx$. Ausführlich sieht dieses sogenannte *iterierte Integral* so aus

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

Meist werden wir die inneren Klammern weglassen und nur $\int_a^b \int_c^d f(x, t) dt dx$ schreiben. Mit demselben Recht kann man aber x und t die Rollen tauschen lassen und bekommt dann ein iteriertes Integral

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

Der Satz von FUBINI sagt, daß bei stetigem f die beiden iterierten Integrale gleich sind, also nicht auf die Reihenfolge der Integration geachtet werden muß. Ist der Integrand unstetig, so kann das iterierte Integral sehr wohl von der Reihenfolge abhängen.

Wir **beweisen** formal allgemeiner, daß für alle $u \in [c, d]$ gilt

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

Nennen wir die linke Seite $g(u)$ und die rechte $h(u)$. Beide sind differenzierbare Funktionen auf $[c, d]$ mit den Ableitungen

$$g'(u) = \int_a^b f(x, u) dx \quad \text{nach Hauptsatz (und 4.1.2) und}$$

$$h'(u) = \int_a^b \left[\frac{\partial}{\partial u} \int_c^u f(x, t) dt \right] dx = \int_a^b f(x, u) dx \quad \text{nach LEIBNIZ und Hauptsatz.}$$

Die Ableitungen von g und h sind also gleich. Da beide Funktionen für $u = c$ übereinstimmen (nämlich beide Null ergeben), stimmen sie für alle u überein, was zu beweisen war.

4.1.8 Verallgemeinerung auf mehr Variable

Die bisherigen Integraldarstellungen lieferten eine einstellige Funktion ausgehend von einer zweistelligen. Mit demselben Ansatz kann man aber auch mehrstellige Funktionen darstellen. Die Formel liest sich dann

$$F(x_1, \dots, x_n) = \int_c^d f(x_1, \dots, x_n, t) dt.$$

Dabei soll f auf einem 'Zylinder' $D \times [c, d]$ definiert sein, wobei $D \subseteq \mathbb{R}^n$. F wird eine Funktion $D \rightarrow \mathbb{R}$. Die oben bewiesenen Resultate übertragen sich sinngemäß auf diesen Fall:

Ist f eine C^1 -Funktion so wird F stetig und C^1 , wobei

$$\frac{\partial F}{\partial x_j}(\bar{x}) = \frac{\partial}{\partial x_j} \int_c^d f(\bar{x}, t) dt = \int_c^d \frac{\partial f}{\partial x_j}(\bar{x}, t) dt.$$

Auch der Satz von FUBINI überträgt sich auf beliebige Quader:

Ist f stetig auf dem Quader $[\bar{a}, \bar{b}]$ so existiert das (n -fach) iterierte Integral

$$\int_{a_n}^{b_n} \int_{a_{n-1}}^{b_{n-1}} \dots \int_{a_1}^{b_1} f(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n$$

und ist nicht von der Reihenfolge der Integrationen abhängig.

Um die letzte Aussage als Formel aufzuschreiben, sei (i_1, i_2, \dots, i_n) ein Tupel, in dem jede der Zahlen $1, 2, \dots, n$ genau einmal vorkommt (eine sogenannte Permutation der Folge $1, 2, \dots, n$). Dann gilt

$$\begin{aligned} & \int_{a_n}^{b_n} \int_{a_{n-1}}^{b_{n-1}} \dots \int_{a_1}^{b_1} f(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n && \text{(Standardreihenfolge)} \\ = & \int_{a_{i_n}}^{b_{i_n}} \int_{a_{i_{n-1}}}^{b_{i_{n-1}}} \dots \int_{a_{i_1}}^{b_{i_1}} f(x_1, x_2, \dots, x_n) dx_{i_1} dx_{i_2} \dots dx_{i_n} && \text{(permutierte Reihenfolge)} \end{aligned}$$

4.2 Integrale über Rechtecke

Gegeben sei eine Funktion (Skalarfeld) $f : D \rightarrow \mathbb{R}$ und eine in $D \subseteq \mathbb{R}^n$ enthaltene kompakte Menge B , auf der f beschränkt sein soll. Wir wollen das RIEMANNSCHE Integral $\int_B f$ oder ausführlich $\int_B f(\bar{x}) d\bar{x}$ definieren. Ein gangbarer Weg beginnt mit der Integration über Quader und führt die allgemeineren Integrationsbereiche dann auf Quader zurück. Die erste Hälfte dieses Programms ist weitgehend parallel zum eindimensionalen Fall (Unterteilungen, RIEMANNSCHE Summen, etc.). Deshalb werde ich mich relativ kurz fassen. Da das Wesentliche des Verfahrens schon bei zweistelligen Funktionen klar wird und dieser Fall den Vorteil hat sich sehr durchsichtig hinschreiben zu lassen, z.B. weil man keine Doppelindizes braucht, werde ich zunächst relativ ausführlich beschreiben, wie zweistellige Funktionen $f(x, y)$ erst auf Rechtecken und später auch auf allgemeineren Mengen integriert werden.

Ein Rechteck $R = [a, b] \times [c, d] \subseteq \mathbb{R}^2$ ist für den Rest dieses Abschnittes fixiert.

4.2.1 Zerlegungen des Rechtecks

werden bei uns immer dadurch entstehen, daß man beide Intervalle durch jeweils endlich viele Punkte unterteilt: $a = x_0 < x_1 < x_2 < \dots < x_p = b$ und $c = y_0 < y_1 < \dots < y_q = d$. Dadurch entstehen $p \cdot q$ Teilrechtecke (auch 'Kästchen' genannt) $R_{ij} = [x_{i-1}, x_i] \times [y_{j-1}, y_j]$ für $i = 1, \dots, p$; $j = 1, \dots, q$. Anschaulich stelle man sich einen Blechkuchen vor, der durch endlich viele achsenparallele Schnitte in Kuchenstücke zerschnitten ist. Die Stücke müssen durchaus nicht gleich groß sein. Unten brauchen wir Namen für Zerlegungen. Wie im eindimensionalen Fall benutze ich Skriptbuchstaben, meist \mathcal{Z} . Formal ist das das Paar, daß aus den beiden Folgen $(x_i)_{i=0}^p$ und $(y_j)_{j=0}^q$ besteht. Die Vereinigung aller Kästchen ist R . Die Flächeninhalte $|R_{ij}| = (x_i - x_{i-1}) \cdot (y_j - y_{j-1})$ der Kästchen addieren sich zum Flächeninhalt $|R| = (b - a) \cdot (d - c)$ des gesamten Rechtecks.

4.2.2 Unter- und Obersummen

Ist außerdem eine *beschränkte* Funktion (Skalarfeld) f gegeben, die mindestens auf R definiert ist, so wollen wir setzen

$$\begin{aligned} M &:= \sup\{f(x, y) : (x, y) \in R\} & m &:= \inf\{f(x, y) : (x, y) \in R\} \\ M_{ij} &:= \sup\{f(x, y) : (x, y) \in R_{ij}\} & m_{ij} &:= \inf\{f(x, y) : (x, y) \in R_{ij}\} \\ OS(f, \mathcal{Z}) &:= \sum_{i,j} M_{ij} \cdot |R_{ij}| & US(f, \mathcal{Z}) &:= \sum_{i,j} m_{ij} \cdot |R_{ij}|, \end{aligned}$$

wobei jeweils über $i = 1, \dots, p$ und $j = 1, \dots, q$, d.h. über alle Kästchen summiert wird. Die letzten beiden Zahlen heißen die zu f und \mathcal{Z} gehörende Ober- bzw. Untersumme.

Da nach Definition stets $m \leq m_{ij} \leq M_{ij} \leq M$ gilt, folgt nach Multiplikation dieser Ungleichungen mit $|R_{ij}|$ und anschließender Addition

$$\sum_{i,j} m \cdot |R_{ij}| \leq \sum_{i,j} m_{ij} \cdot |R_{ij}| \leq \sum_{i,j} M_{ij} \cdot |R_{ij}| \leq \sum_{i,j} M \cdot |R_{ij}|,$$

also

$$(*) \quad m \cdot |R| \leq US(f, \mathcal{Z}) \leq OS(f, \mathcal{Z}) \leq M \cdot |R|.$$

4.2.3 Unteres, oberes und RIEMANNsches Integral

Die Ungleichung (*) zeigt, daß alle Untersummen bei allen denkbaren Zerlegungen \mathcal{Z} eine durch $M |R|$ nach oben beschränkte Menge bilden. Folglich gibt es ein endliches Supremum, das als *unteres Integral* der Funktion f definiert wird. Ich¹⁸ bezeichne es mit

$$u \int_R f := \sup\{US(f, \mathcal{Z}) : \mathcal{Z} \text{ Zerlegung von } R\}.$$

Analog ist die Menge aller Obersummen nach unten beschränkt. Ihr Infimum heißt *oberes Integral*:

$$o \int_R f := \inf\{OS(f, \mathcal{Z}) : \mathcal{Z} \text{ Zerlegung von } R\}.$$

Wenn unteres und oberes Integral zusammenfallen, so nennt man die Funktion f auf dem Rechteck $R = [a, b] \times [c, d]$ *RIEMANN-integrierbar* und definiert ihr *RIEMANNsches Integral* durch

$$\int_R f := u \int_R f = o \int_R f.$$

Bemerkung zur Schreibweise. Besonders wenn man konkrete Integrale ausrechnet und die Funktion irgendwie beschreiben muß, ist es praktisch statt $\int_R f$ ausführlicher $\iint_R f(x, y) d(x, y)$ zu schreiben. Die beiden Integralzeichen sind Geschmackssache; ich benutze sie gern, um anzudeuten, dass es sich um ein Flächenintegral handelt. Im dreidimensionalen Raum schreibe ich dann \iiint etc.

4.2.4 Nicht alle Funktionen sind integrierbar

Die Beschränktheit der Funktion f ist für die Definition des Integrals notwendig, weil sonst Obersummen oder Untersummen gar nicht gebildet werden könnten. Aber nicht jede beschränkte Funktion ist auch integrierbar. Das **Beispiel** der DIRICHLET-Funktion aus dem letzten Semester überträgt sich sofort auf Rechtecke. Ist

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

¹⁸Das ist nicht allgemein üblich.

so sind alle Untersummen 0 und alle Obersummen gleich $|R|$. Also unterscheiden sich unteres und oberes Integral.

Unser nächstes Ziel ist ein Integrierbarkeitstest. Er kann wörtlich aus dem ersten Semester abgeschrieben werden, nur daß sich die Zerlegungen damals auf das Integrationsintervall bezogen und jetzt auf das Rechteck. Auch der Beweis läuft letztlich wie im eindimensionalen ab. Ich skizziere nur und beginne mit

4.2.5 Verfeinerungen

Alle Zerlegungen sind solche von R .

Die Zerlegung \mathcal{U} heißt *Verfeinerung* der Zerlegung \mathcal{Z} , wenn sie durch die Hinzunahme weiterer Punkte auf der x - oder y - Achse entsteht. Offensichtlich gelangt man von \mathcal{Z} zu jeder Verfeinerung, indem man schrittweise jeweils einen weiteren Unterteilungspunkt hinzunimmt.

Was passiert mit Ober- und Untersummen, wenn die Zerlegung verfeinert wird? Angenommen, es kommt ein Punkt u auf der x -Achse hinzu (machen Sie ein Bild!):

$$\begin{array}{ll} \mathcal{Z}: & a = x_0 < x_1 < \dots < x_{r-1} < x_r < \dots < x_p = b & c = y_0 < y_1 < \dots < y_q = d \\ \mathcal{U}: & a = x_0 < x_1 < \dots < x_{r-1} < u < x_r < \dots < x_n = b & c = y_0 < y_1 < \dots < y_q = d. \end{array}$$

Das \mathcal{Z} -Kästchen $R_{rj} = [x_{r-1}, x_r] \times [y_{j-1}, y_j]$ zerfällt in die Vereinigung der beiden \mathcal{U} -Kästchen $R'_{rj} := [x_{r-1}, u] \times [y_{j-1}, y_j]$ und $R''_{rj} := [u, x_r] \times [y_{j-1}, y_j]$. Dementsprechend wird der Summand $m_{rj} \cdot |R_{rj}|$ von $US(f, \mathcal{Z})$ beim Übergang zu $US(f, \mathcal{U})$ durch die Summe

$$m'_{rj} \cdot |R'_{rj}| + m''_{rj} \cdot |R''_{rj}|$$

ersetzt, wobei mit

$$m'_{rj} := \inf\{f(x, y) : (x, y) \in R'_{rj}\} \quad \text{und} \quad m''_{rj} := \inf\{f(x, y) : (x, y) \in R''_{rj}\}$$

gemeint sind. Da auf der umfassenderen Menge R_{rj} jeweils mehr (also potentiell kleinere) Funktionswerte angenommen werden als auf R'_{rj} bzw. R''_{rj} , gilt $m'_{rj} \geq m_{rj}$ bzw. $m''_{rj} \geq m_{rj}$, daher

$$m'_{rj} \cdot |R'_{rj}| + m''_{rj} \cdot |R''_{rj}| \geq m_{rj} \cdot |R'_{rj}| + m_{rj} \cdot |R''_{rj}| = m_{rj} \cdot (|R'_{rj}| + |R''_{rj}|) = m_{rj} \cdot |R_{rj}|.$$

D.h. die gegenüber der \mathcal{Z} -Untersumme veränderten Summanden der \mathcal{U} -Untersumme liefern jeweils einen mindestens genauso großen Beitrag. Also wird durch Hinzunahme eines Unterteilungspunktes auf der x -Achse die Untersumme höchstens größer. Derselbe Effekt tritt ein, wenn die Verfeinerung durch Hinzunahme eines Punktes auf der y -Achse entsteht. Nimmt man mehrere Punkte hinzu, kann die Untersumme in jedem Schritt nur größer werden.

Fazit. *Ist \mathcal{U} Verfeinerung von \mathcal{Z} , so ist $US(f, \mathcal{Z}) \leq US(f, \mathcal{U})$. Dual dazu wird die Obersumme bei Verfeinerungen höchstens kleiner.*

Einfach aber wichtig ist folgendes **Lemma**. *Zu je zwei Unterteilungen gibt es eine gemeinsame Verfeinerung.* Man nimmt einfach alle an wenigstens einer der beiden Unterteilungen beteiligten Punkte zusammen und ordnet sie in wachsender Reihenfolge neu.

4.2.6 Jede Untersumme ist kleiner als jede Obersumme.

Sind nämlich \mathcal{Z}_1 und \mathcal{Z}_2 beliebige Zerlegungen, so erlaubt das Lemma, eine gemeinsame Verfeinerung \mathcal{U} zu wählen. Mit ihrer Hilfe schätzen wir ab:

$$US(f, \mathcal{Z}_1) \leq US(f, \mathcal{U}) \leq OS(f, \mathcal{U}) \leq OS(f, \mathcal{Z}_2).$$

Die beiden äußeren Ungleichungen bestehen, weil \mathcal{U} jeweils Verfeinerung ist; die innere nach (*) aus 4.2.2.

Geht man in der Ungleichung $US(f, \mathcal{Z}_1) \leq OS(f, \mathcal{Z}_2)$ erst links zum Supremum und dann rechts zum Infimum über, erhält man die wichtige Ungleichung

$$\int_R f \leq \int_R f.$$

4.2.7 Integrierbarkeitstest

Die beschränkte Funktion f ist genau dann integrierbar, wenn für jedes $\varepsilon > 0$ eine Zerlegung \mathcal{Z} gefunden werden kann, so daß $OS(f, \mathcal{Z}) - US(f, \mathcal{Z}) < \varepsilon$.

In der Formel

$$OS(f, \mathcal{Z}) - US(f, \mathcal{Z}) = \sum_{i,j} (M_{ij} - m_{ij}) \cdot |R_{ij}|$$

sind die jeweiligen Differenzen $M_{ij} - m_{ij}$ und $|R_{ij}|$ und daher alle Summanden stets positiv. Das ist eine für das Hantieren ganz nützliche Hintergrundinformation, die demnächst stillschweigend benutzt werden wird. Ebenso die oben schon benutzte offensichtliche Beziehung $\sum_{i,j} |R_{ij}| = |R|$.

Beweis des Kriteriums. Nehmen wir zunächst an, daß f integrierbar ist. Wegen $\sup\{US\} = \int_R f = \inf\{OS\}$ gibt es \mathcal{Z}_1 und \mathcal{Z}_2 mit $\int_R f - \frac{\varepsilon}{2} < US(f, \mathcal{Z}_1) \leq \int_R f \leq OS(f, \mathcal{Z}_2) < \int_R f + \frac{\varepsilon}{2}$. Ist dann \mathcal{Z} eine gemeinsame Verfeinerung von \mathcal{Z}_1 und \mathcal{Z}_2 so gelten diese Ungleichungen erst recht für \mathcal{Z} und also haben wir $OS(f, \mathcal{Z}) - US(f, \mathcal{Z}) < \varepsilon$.

Wenn umgekehrt die Bedingung erfüllt ist, so muß f integrierbar sein, denn ansonsten wäre

$$\varepsilon := o \int f - u \int f > 0$$

(nach 4.2.6) und dieses ε könnte wegen

$$US(f, \mathcal{Z}) \leq u \int f \leq o \int f \leq OS(f, \mathcal{Z})$$

von keiner Differenz $OS - US$ unterboten werden.

Auch der

4.2.8 Rest der Integrationstheorie

zweistelliger Funktionen auf Rechtecken ist dem eindimensionalen Fall sehr analog; die Beweise aus dem vorigen Semester übertragen sich mit offensichtlichen Modifikationen (Übungsaufgaben!). Ich liste nur noch einige Ergebnisse auf¹⁹

Die Menge der auf R integrierbaren Funktionen ist ein Vektorraum, auf dem das Integral eine monotone und lineare Abbildung ist. Mit f ist auch $|f|$ integrierbar und es gilt $|\int_R f| \leq \int_R |f|$. Produkte von integrierbaren Funktionen sind integrierbar.

4.2.9 Integrierbarkeit vererbt sich auf Teilrechtecke

Sind $R_1 \subseteq R \subseteq \mathbb{R}^2$ Rechtecke und ist die Funktion f auf R integrierbar, so ist sie auch auf R_1 integrierbar. Der Beweis ist eine einfache Übungsaufgabe.

4.2.10 Das Integral ist als Funktion des Rechtecks additiv.

Damit ist folgendes gemeint. Sind R_1 und R_2 zwei in einem großen Rechteck R , enthaltene Rechtecke, die keine inneren Punkte gemeinsam haben aber eventuell eine Seite oder einen Eckpunkt) und ist die Funktion $f : R \rightarrow \mathbb{R}$ außerhalb $R_1 \cup R_2$ gleich Null, so gilt

$$\int_R f = \int_{R_1} f + \int_{R_2} f$$

in dem Sinne, daß die Existenz jeweils einer Seite die Existenz der jeweils andern und die Gleichheit nach sich zieht. Für den Fall, daß R_1 und R_2 eine gemeinsame Seite haben, ist $R_1 \cup R_2$ selbst Rechteck und es gilt

$$\int_{R_1 \cup R_2} f = \int_{R_1} f + \int_{R_2} f.$$

Auch den Nachweis dieser (trotzdem sehr wichtigen) Eigenschaft überlasse ich Ihnen.

¹⁹Sie sind trotzdem ganz wichtig, die Linearität wird gleich angewendet.

4.2.11 Dünne Mengen sind Integralen egal

Etwas wissenschaftlicher ausgedrückt: *Unterscheiden sich zwei beschränkten Funktionen $R \rightarrow \mathbb{R}$ nur auf einer dünnen Menge, so sind sie gleichzeitig integrierbar oder nicht integrierbar und haben dieselben Integrale.* Dabei soll eine Teilmenge $A \subseteq \mathbb{R}^2$ *dünn*²⁰ heißen, wenn sie sich durch endlich viele achsenparallele Rechtecke überdecken läßt, deren summarischer Flächeninhalt beliebig klein gemacht werden kann. Genauer: für jedes $\varepsilon > 0$ muß es endlich viele Rechtecke A_k derart geben, daß $A \subseteq \bigcup_k A_k$ und $\sum_k |A_k| < \varepsilon$.

Beispiele für dünne Mengen sind achsenparallele Strecken (offensichtlich), Graphen integrierbarer Funktionen $[a, b] \rightarrow \mathbb{R}$ (das ist gerade das Integrierbarkeitskriterium aus dem letzten Semester) und Bahnen regulärer Kurven (etwas knifflig).

Beweis. Wir betrachten zunächst einen Spezialfall. $f : R \rightarrow \mathbb{R}$ sei beschränkt und überhaupt nur auf der dünnen Menge A von Null verschieden (die zweite Funktion kommt nicht explizit vor; sie ist konstant Null). Wir zeigen, daß f integrierbar ist und $\int_R f = 0$. Sei dazu S eine feste Schranke für $|f|$ auf R . Dann ist bei jeder Zerlegung

$$(*) \quad -S \leq m_{ij} \leq M_{ij} \leq S \text{ für alle } ij \text{ mit } R_{ij} \cap A \neq \emptyset \text{ und } m_{ij} = M_{ij} = 0 \text{ falls } R_{ij} \cap A = \emptyset.$$

Sei $\varepsilon > 0$ vorgegeben. Weil A dünn ist, gibt es endlich viele Rechtecke $A_k = [a_k, b_k] \times [c_k, d_k]$ mit $k = 1, \dots, N$, die A überdecken und $\sum_{k=1}^N |A_k| < \frac{\varepsilon}{2S}$ erfüllen. Diese A_k wollen wir ein wenig aufblähen. Für $\delta \geq 0$ setzen wir $A_k^\delta := [a_k - \delta, b_k + \delta] \times [c_k - \delta, d_k + \delta]$.

Der summarische Flächeninhalt aller A_k^δ ist $\sum_{k=1}^N |A_k^\delta| = \sum_{k=1}^N (b_k - a_k + 2\delta) \cdot (d_k - c_k + 2\delta)$ hängt quadratisch von δ ab. Für $\delta = 0$ stimmt er mit $\sum |A_k|$ überein und ist deshalb kleiner als $\frac{\varepsilon}{2S}$.

Wegen der stetigen Abhängigkeit von δ , bleibt $\sum_{k=1}^N |A_k^\delta|$ für alle hinreichend kleinen positiven δ kleiner als $\frac{\varepsilon}{2S}$. Ein derartiges $\delta > 0$ nehmen wir uns her und wählen eine Zerlegung \mathcal{Z} von R , bei der alle Kästchen R_{ij} Kantenlängen $\leq \delta$ haben.

Jedes \mathcal{Z} -Kästchen R_{ij} , das A schneidet, schneidet (erst recht) ein A_k und ist deshalb in A_k^δ enthalten (Bild!). Es ist also

$$(**) \quad \text{die summarische Fläche der } \mathcal{Z}\text{-Kästchen } R_{ij}, \text{ die } A \text{ schneiden, höchstens so groß, wie die summarische Fläche der } A_k^\delta, \text{ also } < \frac{\varepsilon}{2S}.$$

Die Obersumme läßt sich in zwei Teile aufspalten:

$$OS(f, \mathcal{Z}) = \sum_{ij} M_{ij} \cdot |R_{ij}| = \sum'_{ij} M_{ij} \cdot |R_{ij}| + \sum''_{ij} M_{ij} \cdot |R_{ij}|,$$

wobei in \sum' diejenigen R_{ij} berücksichtigt werden, die A schneiden, in \sum'' der Rest. Dann folgt mit (*) und (**):

$$OS \leq \sum' S \cdot |R_{ij}| + \sum'' 0 \cdot |R_{ij}| = S \cdot \sum' |R_{ij}| \leq S \cdot \sum_k |A_k^\delta| < S \cdot \frac{\varepsilon}{2S} = \frac{\varepsilon}{2}.$$

Analog sieht man $US(f, \mathcal{Z}) \geq -\frac{\varepsilon}{2}$. Dann ist aber $OS - US < \varepsilon$, womit die Integrierbarkeit bewiesen ist.

Wegen $-\frac{\varepsilon}{2} < US \leq \int_R f \leq OS < \frac{\varepsilon}{2}$, folgt nun auch $\int_R f = 0$, indem man ε gegen Null gehen läßt. Damit ist der Spezialfall bewiesen. Der allgemeine Fall folgt daraus wegen der Additivität von Integrierbarkeit und Integral (4.2.8). Ist f_1 integrierbar und unterscheidet sich f_2 von f_1 nur auf einer dünnen Menge, so ist $f_2 - f_1$ nur auf dieser dünnen Menge von Null verschieden, also integrierbar und mit Integral Null. Daher ist $f_2 = f_1 + (f_2 - f_1)$ als Summe integrierbarer Funktionen integrierbar und $\int_R f_2 = \int_R f_1 + \int_R (f_2 - f_1) = \int_R f_1$.

4.2.12 Jede auf R stetige Funktion ist integrierbar.

Im nächsten Abschnitt werden wir ein allgemeineres Resultat beweisen; die folgenden Betrachtungen haben eher propädeutischen Wert. Ich werde auf eine Weise argumentieren, die einen anderen Aspekt des RIEMANNschen Integrals aufblitzen läßt. Dazu brauchen wir noch einen neuen Begriff.

²⁰Das Wort ist nicht allgemein üblich.

Als *Feinheit* oder *Maschenweite* einer Unterteilung \mathcal{Z} bezeichnet man den größten Durchmesser eines Kästchens, d.h. die Zahl

$$\mu(\mathcal{Z}) := \max \left\{ \sqrt{(x_i - x_{i-1})^2 + (y_j - y_{j-1})^2} : i = 1, \dots, p; j = 1, \dots, q \right\}.$$

Satz. *Ist f eine auf R stetige Funktion, so gibt es zu jedem $\varepsilon > 0$ ein $\delta > 0$ mit*

$$OS(f, \mathcal{Z}) - US(f, \mathcal{Z}) < \varepsilon \quad \text{sobald} \quad \mu(\mathcal{Z}) < \delta.$$

Daraus folgt die Integrierbarkeit der stetigen Funktionen dann sofort. Da das Rechteck $R = [a, b] \times [c, d]$ kompakt ist, muß jede stetige Funktion beschränkt sein. Wenn man $OS - US < \varepsilon$ machen möchte, läßt man sich vom Satz ein passendes δ geben und wählt dann eine Zerlegung mit Maschenweite $< \delta$ (z.B. *äquidistant*: $x_i := a + \frac{i}{N}(b-a)$; $y_j := c + \frac{j}{N}(d-c)$; $i, j = 0, 1, \dots, N$ für ein genügend großes N).

Der **Beweis** des Satzes beruht auf der gleichmäßigen Stetigkeit der auf dem kompakten(!) Rechteck stetigen Funktion f . Sie liefert zu dem gegebenen $\varepsilon > 0$ (und der im Hinterkopf gebildeten ebenfalls positiven Zahl $\frac{\varepsilon}{2|R|}$) ein $\delta > 0$, so daß $|f(u, v) - f(u', v')| < \frac{\varepsilon}{2|R|}$, falls $\sqrt{(u - u')^2 + (v - v')^2} < \delta$. Hat nun \mathcal{Z} eine Maschenweite $< \delta$, so haben alle Kästchen R_{ij} einen Durchmesser kleiner δ , also so kann die Funktion f auf den Kästchen R_{ij} um höchstens $\frac{\varepsilon}{2|R|}$ schwanken. Daher $M_{ij} - m_{ij} \leq \frac{\varepsilon}{2|R|}$ und die Differenz von Ober- und Untersumme wird

$$\sum_{ij} (M_{ij} - m_{ij}) |R_{ij}| \leq \frac{\varepsilon}{2|R|} \sum_{ij} |R_{ij}| = \frac{\varepsilon}{2|R|} \cdot |R| = \frac{\varepsilon}{2} < \varepsilon,$$

wie verlangt. Hier gleich der versprochene neue Aspekt.

4.2.13 Integrale stetiger Funktionen als Grenzwerte RIEMANNscher Summen

Wir bekommen eine etwas einfachere Möglichkeit der Integralberechnung. Dabei werden die lästigen Suprema und Infima durch Grenzwerte ersetzt. Eine zur Funktion f und Zerlegung \mathcal{Z} gehörende RIEMANNsche Summe ist eine Summe der Art

$$\sum_{ij} f(u_{ij}, v_{ij}) \cdot |R_{ij}|,$$

wobei die Punkte (u_{ij}, v_{ij}) jeweils zu R_{ij} gehören. Da wir nicht ernsthaft damit arbeiten werden, erlaube ich mir die Abkürzung $RS(f, \mathcal{Z})$, die unvollständig ist, weil sie die markierten Punkte (u_{ij}, v_{ij}) nicht erwähnt.

Bei stetigem f sind sowohl OS als auch US spezielle RIEMANNsche Summen, denn M_{ij} und m_{ij} werden auf R_{ij} als Funktionswerte angenommen. Allgemein gilt

$$m_{ij} \leq f(u_{ij}, v_{ij}) \leq M_{ij} \quad \text{und daher} \quad US(f, \mathcal{Z}) \leq RS(f, \mathcal{Z}) \leq OS(f, \mathcal{Z}).$$

Satz. *Für eine beliebige Folge $(\mathcal{Z}_N)_{N=0}^\infty$ von Unterteilungen des Intervalls $[a, b]$, deren Feinheit gegen Null geht, und alle stetigen Funktionen f gilt*

$$\int_R f = \lim_{N \rightarrow \infty} RS(f, \mathcal{Z}_N).$$

Man schreibt den Satz auch in der Form

$$\lim_{\mu(\mathcal{Z}) \rightarrow 0} RS(f, \mathcal{Z}) = \iint_R f(x, y) d(x, y).$$

Beweis. Sei $\varepsilon > 0$ vorgegeben und dazu ein δ wie im vorigen Satz bestimmt. Die hinreichend späten Glieder der Folge (\mathcal{Z}_N) haben nach Voraussetzung Maschenweiten $< \delta$. Also gilt für sie $OS - US < \varepsilon$. Wegen $US \leq \int_R \leq OS$ und $US \leq RS \leq OS$ folgt daraus dann

$$\left| \int_R f - RS(f, \mathcal{Z}_N) \right| \leq OS(f, \mathcal{Z}_N) - US(f, \mathcal{Z}_N) < \varepsilon.$$

In 4.3.9 kommen wir noch einmal auf die Idee der RIEMANNschen Summen zurück, werden dann aber etwas liberaler sein und nicht nur Rechtecke in den Zerlegungen zulassen.

4.2.14 Rückführung auf iterierte Integrale

Bisher sind wir ausgehend von der Definition kaum in der Lage, irgendwelche Integrale über Rechtecke auszurechnen. Das ändert sich jetzt schlagartig.

Wenn f auf R integrierbar ist und für jedes feste $x \in [a, b]$ das Integral $\int_c^d f(x, y) dy$ existiert und die entstehende Funktion $x \mapsto \int_c^d f(x, y) dy$ auf $[a, b]$ integrierbar ist, so gilt

$$\iint_R f(x, y) d(x, y) = \int_a^b \int_c^d f(x, y) dy dx.$$

Natürlich gilt auch die symmetrische Aussage:

$$\iint_R f(x, y) d(x, y) = \int_c^d \int_a^b f(x, y) dx dy,$$

sofern beide Seiten existieren. Speziell ist das für stetige Funktionen erfüllt. Für diese sind also die in beiden Reihenfolgen berechneten iterierten Integrale beide gleich dem Flächenintegral und daher untereinander gleich. Damit haben wir einen neuen, natürlicheren Beweis des Satzes von FUBINI.

Beweis. Es genügt für beliebige Zerlegungen \mathcal{Z} von R nachzuweisen, daß

$$US(f, \mathcal{Z}) \leq \int_a^b \int_c^d f(x, y) dy dx \leq OS(f, \mathcal{Z}).$$

Dann folgt nämlich

$$\iint_R f(x, y) d(x, y) = \sup\{US\} \leq \int_a^b \int_c^d f(x, y) dy dx \leq \inf\{OS\} = \iint_R f(x, y) d(x, y)$$

und also die Gleichheit des Flächenintegrals mit dem iterierten.

Beide Ungleichungen gehen völlig analog; ich beschränke mich auf die rechte. Sei \mathcal{Z} gegeben. Für jedes feste $x \in [a, b]$ zerfällt das Integral

$$\int_c^d f(x, y) dy \quad \text{in die Summe} \quad \sum_{j=1}^q \int_{y_{j-1}}^{y_j} f(x, y) dy.$$

Wenn wir die entsprechenden Funktionen aus psychologischen Gründen mit $F(x)$ bzw $F_j(x)$ bezeichnen, haben wir also $F(x) = \sum_{j=1}^q F_j(x)$ und daher $\int_a^b F(x) dx = \sum_{j=1}^q \int_a^b F_j(x) dx$. Alle \int_a^b -Integrale zerfallen ihrerseits in die Summen $\sum_{i=1}^p \int_{x_{i-1}}^{x_i}$. Dann ist also

$$\int_a^b F(x) dx = \sum_{j=1}^q \int_a^b F_j(x) dx = \sum_{j=1}^q \sum_{i=1}^p \int_{x_{i-1}}^{x_i} F_j(x) dx$$

oder, wenn man die ursprüngliche Bedeutung resubstituiert,

$$(*) \quad \int_a^b \left(\int_c^d f(x, y) dy \right) dx = \sum_{j=1}^q \sum_{i=1}^p \int_{x_{i-1}}^{x_i} \left(\int_{y_{j-1}}^{y_j} f(x, y) dy \right) dx.$$

Jetzt halten wir ein Indexpaar (i, j) fest. Wie oben soll M_{ij} das Supremum von f auf R_{ij} sein. Dann ist für alle $x \in [x_{i-1}, x_i]$ und $y \in [y_{j-1}, y_j]$ die Ungleichung $f(x, y) \leq M_{ij}$ erfüllt. Diese Ungleichung kann man nun zuerst für festes $x \in [x_{i-1}, x_i]$ nach y integrieren:

$$\int_{y_{j-1}}^{y_j} f(x, y) dy \leq \int_{y_{j-1}}^{y_j} M_{ij} dy = M_{ij}(y_j - y_{j-1})$$

und dann das Resultat (die rechte Seite hängt nicht von x ab) nach x :

$$\int_{x_{i-1}}^{x_i} \left(\int_{y_{j-1}}^{y_j} f(x, y) dy \right) dx \leq \int_{x_{i-1}}^{x_i} M_{ij}(y_j - y_{j-1}) dx = M_{ij}(y_j - y_{j-1})(x_i - x_{i-1}) = M_{ij} |R_{ij}|.$$

Das gilt für alle (i, j) . Aufsummiert über alle i, j , ergibt sich für die rechte Seite von $(*)$, daß

$$\int_a^b \left(\int_c^d f(x, y) dy \right) dx = \sum_{j=1}^q \sum_{i=1}^p \int_{x_{i-1}}^{x_i} \left(\int_{y_{j-1}}^{y_j} f(x, y) dy \right) dx \leq \sum_{j=1}^q \sum_{i=1}^p M_{ij} |R_{ij}| = OS(f, \mathcal{Z}),$$

wie versprochen.

4.3 Integration zweistelliger Funktionen über krummlinig begrenzte Gebiete

Im Gegensatz zu einstelligen Funktionen, bei denen Intervalle die einzigen natürlichen Integrationsbereiche waren, möchte man in der Ebene sicher auch über andere Bereiche als Rechtecke integrieren. Den 'uneigentlichen' Fall lassen wir vorerst außen vor und beschränken uns auf beschränkte Mengen, auf denen auch die zu integrierenden Funktionen beschränkt sein sollen. Die Idee besteht darin, eine solche Menge, etwa B , in ein großes Rechteck R einzuschließen, die Funktion f so abzuändern, daß sie außerhalb B Null wird und dann die modifizierte Funktion über R zu integrieren, wie das im vorigen Abschnitt besprochen wurde.

$$\int_B f \text{ wird definiert als } \int_R f_B$$

wobei $R \supset B$ ein Rechteck ist und $f_B(x, y) = \begin{cases} f(x, y), & (x, y) \in B \\ 0, & \text{sonst} \end{cases}$ gesetzt ist. Die Abhängigkeit von R ist nur scheinbar. Es gibt immer ein minimales Rechteck, in das G hineinpaßt. Wird über ein größeres R integriert, macht das aber auch nichts, weil f_B in dem überflüssigen Teil konstant Null ist (vgl. 4.2.10).

Diese Idee funktioniert für einfache B und ziemlich stetige f . Wir werden die Grenzen der Methode nicht exakt ausloten, sondern den wichtigsten Fall beschreiben und etwas näher untersuchen. Vorher jedoch noch einmal offiziell die

4.3.1 Integraldefinition

$B \subseteq \mathbb{R}^2$ sei beschränkt. Eine mindestens auf B definierte und dort beschränkte Funktion f heißt *auf/über B RIEMANN-integrierbar*, wenn die Funktion f_B auf einem/jedem B umfassenden Rechteck im Sinne des vorigen Abschnittes RIEMANN-integrierbar ist. Die von $R \supset B$ unabhängige Zahl $\int_R f_B$ wird dann mit $\int_B f$ oder $\iint_B f(x, y)d(x, y)$ bezeichnet und RIEMANNSches Integral von f über B genannt.

4.3.2 Einfache Eigenschaften

Aus den Ergebnissen des vorigen Abschnittes bekommt man ohne Mühe die folgenden Verallgemeinerungen:

Die Menge der auf B integrierbaren Funktionen ist ein Vektorraum, auf dem das Integral eine monotone und lineare Abbildung ist. Mit f ist auch $|f|$ integrierbar und es gilt $|\int_B f| \leq \int_B |f|$. Produkte von integrierbaren Funktionen sind integrierbar.

Änderungen auf dünnen Mengen haben keinen Einfluß auf Integrierbarkeit und Integrale.

Das Integral ist als Funktion des Integrationsbereiches additiv: Sind A und B zwei disjunkte beschränkte Mengen auf denen die Funktion f jeweils integrierbar ist, so ist f auch auf $A \cup B$ integrierbar und es gilt

$$(*) \int_{A \cup B} f = \int_A f + \int_B f.$$

Das liegt daran, daß $f_{A \cup B} = f_A + f_B$ und man ein Rechteck wählen darf, das von vornherein beide Mengen umfaßt. Tatsächlich dürften sich A und B sogar auf einer dünnen Menge überschneiden.

Warnung: Nach 4.2.9 ist jede auf einem Rechteck integrierbare Funktion auch auf jedem Teilrechteck integrierbar. Für beliebige Bereiche gilt das nicht! Man kann etwa ein Rechteck so in zwei (wilde) disjunkte Teile zerlegen, daß selbst die konstante Funktion 1 auf keinem von ihnen integrierbar ist. Im Gegensatz zu Rechtecken ist die Additivität im allgemeinen Fall also eine Einbahnstraße: aus der Existenz der linken Seite von (*) folgt nicht die der rechten. Ist etwa $B = \{(x, y) : 0 \leq x, y \leq 1 \text{ und } x \text{ rational}\}$ so wird 1_B die in 4.2.4 betrachtete DIRICHLET-Funktion. Das Beispiel zeigt, daß, auch im Unterschied zu Rechtecken, keinesfalls jede auf (irgendeiner wilden Menge) B stetige und beschränkte Funktion integrierbar sein muß.

4.3.3 Gutartige Bereiche

Die Sprechweise ist nicht allgemein gebräuchlich. Jeder Autor beschreibt und benennt die ihm zur Integration besonders geeignet erscheinenden Mengen auf seine Weise. Das Wort ‘Normalbereich’ ist relativ verbreitet (aber auch ziemlich nichtssagend).

Wir nennen eine Teilmenge von \mathbb{R}^2 *einfach gutartig*, wenn sie durch ein abgeschlossenes Intervall auf einer der beiden Achsen und zwei stetige(!) Funktionen beschrieben werden kann. Je nachdem auf welcher Achse das Intervall liegt gibt es zwei Typen (Skizze!):

$$G = \{(x, y) : a \leq x \leq b; g(x) \leq y \leq h(x)\} \quad \text{und} \quad H = \{(x, y) : c \leq y \leq d; g(y) \leq x \leq h(y)\}.$$

Eine Menge ist gutartig wenn sie durch endlich viele achsenparallele Schnitte in einfach gutartige zerlegt werden kann. Formaler: Eine Menge B soll *gutartig* heißen, wenn es endlich viele einfach gutartige Mengen B_1, B_2, \dots, B_k gibt, die paarweise höchstens eine achsenparallele Strecke gemeinsam haben, und deren Vereinigung B ergibt.

Man bemerke, daß einfach gutartige Mengen abgeschlossen und beschränkt also kompakt sind. Da sich diese Eigenschaft auf endliche Vereinigungen überträgt, gilt das auch für gutartige.

Das Hauptresultat dieses Abschnittes besagt, daß die oben erläuterte Idee für stetige Funktionen auf gutartigen Mengen funktioniert. Ausserdem lernen wir, das Integral $\int_B f$ auf eine Summe von iterierten Integralen zurückzuführen.

4.3.4 Integration stetiger Funktionen über einfach gutartige Mengen

Wir beginnen mit einer einfach gutartigen Menge $G = \{(x, y) : a \leq x \leq b; g(x) \leq y \leq h(x)\}$ (stillschweigend, wird überall $g < h$ vorausgesetzt) und einem sie umfassenden Rechteck $R = [u, v] \times [c, d]$ (Skizze!). Weiter sei eine Funktion $f : R \rightarrow \mathbb{R}$ gegeben, die auf G stetig sein soll und außerhalb G konstant Null (Prototyp ist natürlich f_G für ein auf G stetiges f).

Eine derartige Funktion f ist auf R integrierbar und ihr Integral kann nach der Formel

$$\iint_R f(x, y) d(x, y) = \int_a^b \left(\int_{g(x)}^{h(x)} f(x, y) dy \right) dx.$$

berechnet werden.

Beweis. Falls $u < a$ und $b < v$ so zerfällt das Rechteck R in drei:

$$R = [u, a] \times [c, d] \cup [a, b] \times [c, d] \cup [b, v] \times [c, d].$$

Auf den beiden äußeren ist f (außer auf einer Seite, die aber für Integrierbarkeit und Integral keine Rolle spielt) konstant gleich Null. Daher ist f auf den beiden äußeren Rechtecken integrierbar mit dem Integral 0. Also brauchen wir uns nur um das mittlere Rechteck $[a, b] \times [c, d]$ zu kümmern. Wir werden von nun an annehmen, daß es mit R übereinstimmt, d.h. $u = a$ und $v = b$ setzen. Das Hauptproblem ist die Integrierbarkeit von f auf R .

Wir lassen uns $\varepsilon > 0$ vorgeben und müssen eine Zerlegung finden, für die $OS - US < \varepsilon$ wird (tatsächlich wird nur eine Konstante mal ε herauskommen, aber das ist nicht schlimm). Indem wir die gleichmäßige Stetigkeit von f (auf G) und g, h (auf $[a, b]$) benutzen, können wir derart feine Unterteilungen $a = x_0 < x_1 < \dots < x_p = b$ und $c = y_0 < y_1 < \dots < y_p = d$ wählen, daß

- (1) g und h auf jedem Teilintervall $[x_{i-1}, x_i]$ um höchstens ε schwanken,
- (2) f auf jeder Menge $R_{ij} \cap G$ um höchstens ε schwankt und
- (3) die Höhen $|y_j - y_{j-1}|$ aller Kästchen R_{ij} kleiner als ε sind.

Die Funktion f ist auf der kompakten Menge G beschränkt, weil stetig. Da außerhalb der Wert 0 angenommen wird, ist f auf ganz R beschränkt und wir finden K so daß $|f| \leq K$ auf R . Aus technischen Gründen, nehmen wir $K > \varepsilon$ an, was sicher kein Problem ist.

Wir werden zeigen, daß die gewählte Zerlegung \mathcal{Z}

$$OS(f, \mathcal{Z}) - US(f, \mathcal{Z}) \leq (|R| + 6K(b-a)) \cdot \varepsilon$$

erfüllt. Dann sind wir fertig, da diese Zahl mit ε beliebig klein wird.

Die Summe

$$OS - US = \sum_{ij} (M_{ij} - m_{ij}) \cdot |R_{ij}|$$

zerfällt in einen unproblematischen und einen problematischen Teil. Unproblematisch sind alle R_{ij} , die entweder G gar nicht schneiden, für die ist $M_{ij} = m_{ij} = 0$ und der entsprechende Summand 0, oder die ganz im Inneren von G liegen, dann ist $M_{ij} - m_{ij} \leq \varepsilon$ nach (2). Die unproblematischen Summanden lassen sich alle durch $\varepsilon \cdot |R_{ij}|$ abschätzen, ergeben daher aufsummiert höchstens $\varepsilon \cdot |R|$.

Problematisch sind die R_{ij} , die G schneiden aber nicht in G enthalten sind. Sie enthalten Punkte in denen f Null wird und (normalerweise) solche, wo f nicht verschwindet. Die Differenz $M_{ij} - m_{ij}$ kann hier maximal K werden (warum nicht $2K$?), und das haben wir nicht unter Kontrolle. Wir werden aber sehen, daß die problematischen Summanden keinen großen Anteil an der Gesamtsumme liefern.

Fixieren wir dazu zunächst ein $i \leq p$. Dann haben sowohl g als auch h auf dem abgeschlossenen Intervall $[x_{i-1}, x_i]$ ein Minimum und ein Maximum, etwa g_- und g_+ bzw. h_- und h_+ . Nach Bedingung (1) sind die beiden Differenzen $g_+ - g_-$ und $h_+ - h_-$ höchstens ε . Problematisch können nur Kästchen R_{ij} sein, bei denen die vertikale Seite $[y_{j-1}, y_j]$ eine der beiden Strecken $[g_-, g_+]$ oder $[h_-, h_+]$ schneidet. Von der ersten Sorte gibt es eines, das g_- und eines (eventuell dasselbe) das g_+ enthält. Wegen (3) haben beide eine Höhe $< \varepsilon$. Bei eventuell vorhandenen weiteren problematischen Kästchen muß $[y_{j-1}, y_j]$ dann ein Teilintervall von $[g_-, g_+]$ sein. Ihre Gesamthöhe kann daher $g_+ - g_-$ nicht übersteigen. Die Gesamthöhe aller bei fixiertem i wegen

g' problematischen Kästchen R_{ij} ist also höchstens $\varepsilon + \varepsilon + \varepsilon = 3\varepsilon$. Dasselbe gilt für die 'wegen h' problematischen. Also ist die Gesamthöhe der bei festem i problematischen R_{ij} höchstens 6ε . Der Beitrag dieser problematischen Kästchen an der Summe $\sum_{j=1}^q (M_{ij} - m_{ij}) |R_{ij}|$ ist also höchstens $K \cdot 6\varepsilon \cdot (x_i - x_{i-1})$. Summiert man daher den Beitrag aller problematischen Kästchen über alle i (d.h. die ganze Breite von $[a, b]$) so kann höchstens $6K\varepsilon \cdot (b - a)$ herauskommen. Damit haben wir als Summe der unproblematischen und problematischen Summanden die versprochene Abschätzung

$$OS - US \leq (|R| + 6K(b - a)) \cdot \varepsilon,$$

eingesehen und sind fertig; f ist also integrierbar.

Es bleibt die Formel

$$\iint_R f(x, y) d(x, y) = \int_a^b \left(\int_{g(x)}^{h(x)} f(x, y) dy \right) dx$$

einzusehen. Die Existenz des iterierten Integrals auf der rechten Seite ist unproblematisch, denn das innere Integral ist von uns früher als stetige Funktion von x nachgewiesen worden. Bei festem x ist $f(x, y)$ in den Intervallen $[c, g(x)[$ und $]h(x), d]$ konstant Null. Diese Teile leisten also keinen Beitrag zum (eindimensionalen, ganz gewöhnlichen) Integral $\int_c^d f(x, y) dy = \int_c^{g(x)} + \int_{g(x)}^{h(x)} + \int_{h(x)}^d$. Daher können wir die angestrebte Formel auch in der Form

$$\iint_R f(x, y) d(x, y) = \int_a^b \int_c^d f(x, y) dy dx$$

aufschreiben. Da beide Seiten existieren, ist die Gleichheit aber bereits in 4.2.14 etabliert worden.

Ein ganz analoges Resultat gilt für Funktionen auf einfach gutartigen Mengen des anderen Typs: Ist $f : R = [a, b] \times [c, d] \rightarrow \mathbb{R}$ stetig auf $H = \{(x, y) : c \leq y \leq d \text{ und } g(y) \leq x \leq h(y)\} \subseteq R$ und Null außerhalb von H , so ist f auf R integrierbar und es gilt

$$\int_R f = \int_c^d \left(\int_{g(y)}^{h(y)} f(x, y) dx \right) dy.$$

4.3.5 Integrale stetiger Funktionen über gutartige Mengen

Ist f stetig auf der gutartigen Menge $B = B_1 \cup B_2 \cup \dots \cup B_k$ mit einfach gutartigen B_i , so ist die Modifikation f_B auf jedem B umfassenden Rechteck integrierbar. Das Integral $\int_R f_B$ hängt dann nicht von R ab und ist gleich der Summe der Integrale $\sum_{i=1}^k \int_R f_{B_i} = \sum_{i=1}^k \int_{B_i} f$. Letztere lassen sich als iterierte Integrale ausrechnen.

Der **Beweis** besteht in einer einfachen Rückführung auf das vorige Resultat. Bis auf die endlich vielen achsenparallelen Schnittstrecken, in denen sich zwei oder mehrere der B_i vielleicht überlappen ist f_B die Summe der f_{B_i} . Wir hatten aber gerade gesehen, daß alle Summanden f_{B_i} über R integrierbar sind und daß ihr Integral nicht vom umfassenden Rechteck abhängt. Dasselbe gilt dann für die Summe.

Bemerkung. Eine gutartige Menge kann normalerweise auf verschiedene Weisen in einfach gutartige zerlegt werden (probieren Sie etwa einen Kreisring). Unsere Art das Integral auszurechnen hängt von der Zerlegung ab: Die Zerlegungen $B = B_1 \cup B_2 \cup \dots \cup B_k = A_1 \cup A_2 \cup \dots \cup A_h$ führen auf die Summen $\sum_{i=1}^k \int_{B_i} f$ und $\sum_{j=1}^h \int_{A_j} f$. Beide Summen müssen aber gleich sein, denn sie sind, wie gerade bewiesen, gleich dem Integral $\int_B f$.

4.3.6 Ein echter Schönheitsfehler

der Methode besteht darin, daß wir sie auf Unterteilungen in (viele kleine) achsenparallele Rechtecke gründen. Wenn die Menge B und die Funktion f in einem anderen Koordinatensystem beschrieben werden, sind ganz andere Rechtecke achsenparallel. Hoffentlich kommt trotzdem dasselbe Integral heraus. Rechtecke werden ja überhaupt nur aus Bequemlichkeit benutzt, weil sich ihre Flächeninhalte so einfach hinschreiben lassen und die Rückführung auf iterierte Integrale über Intervalle so durchsichtig wird. Eine Zerlegung in kleine Dreiecke hätte aber zum Beispiel den Vorteil, daß man dann über jede geradlinig begrenzte Figur (Polygon) sofort integrieren könnte und nicht auf das größere Rechteck fortsetzen müßte. Manchmal scheint auch eine Zerlegung in viele kleine Segmente von Kreisringen passend, z.B. wenn B rotationssymmetrisch ist und f sehr durchsichtig von r, φ abhängt.

Die gute Nachricht zuerst: alle Methoden, den Integrationsbereich in (halbwegs reguläre) kleine Stücke zu zerlegen, entsprechende Ober- und Untersummen $\sum_{i=1}^n \sup f(\Delta_i) |\Delta_i|$ bzw. $\sum_{i=1}^n \inf f(\Delta_i) |\Delta_i|$ zu betrachten und deren Infimum = Supremum als Integral zu definieren, sind (jedenfalls für anständige Funktionen) gleichwertig²¹ zu unserer. Wir werden das später als heuristisches Prinzip benutzen.

Die schlechte Nachricht: Die zugehörige ‘seriöse’ Integrationstheorie ist eher schwer und liefert dem Physiker nicht unbedingt neue Einsichten, so daß wir darauf verzichten (bis auf 4.3.9).

4.3.7 Der Flächeninhalt einer ebenen Figur,

den wir durch Betragsstriche bezeichnen, kann als Integral $\iint_B 1 \, d(x, y)$ definiert werden. Für einfach gutartige Mengen läßt er sich nach der Formel

$$|B| = \int_a^b \int_{g(x)}^{h(x)} 1 \, dy \, dx = \int_a^b [h(x) - g(x)] \, dx$$

ausrechnen. Das kann uns allerdings nicht überraschen, denn diese Formel war ja im vorigen Semester quasi die Motivation für die ganze Integralrechnung: ‘Integral = Fläche unter dem Funktionsgraphen’. Bei gutartigen Mengen wird zerschnitten, integriert und dann addiert.

Nicht jede ebene Menge hat einen Flächeninhalt, wie man schon an der Menge unter dem Graphen der DIRICHLET-Funktion sieht. Sie besteht aus abzählbar vielen Strecken, die ganz dicht nebeneinander liegen.

4.3.8 Mittelwertsatz der Integralrechnung

$B \subseteq \mathbb{R}^2$ sei kompakt und zusammenhängend, $f : B \rightarrow \mathbb{R}$ stetig. Wenn B einen Flächeninhalt hat und $\int_B f$ existiert, so gibt es einen Punkt $\bar{b} \in B$ mit

$$\int_B f = f(\bar{b}) \cdot |B|.$$

Der **Beweis** geht genauso wie im letzten Semester: Als stetige Funktion hat f Maximum und Minimum: $f(\bar{b}_1) \leq f(x, y) \leq f(\bar{b}_2)$ für passende \bar{b}_1 und \bar{b}_2 und alle $(x, y) \in B$. Diese Ungleichungen darf man integrieren und den konstanten Faktor herausziehen:

$$f(\bar{b}_1) \int_B 1 \, d(x, y) = \int_B f(\bar{b}_1) \, d(x, y) \leq \int_B f(x, y) \, d(x, y) \leq \int_B f(\bar{b}_2) \, d(x, y) = f(\bar{b}_2) \int_B 1 \, d(x, y).$$

Ist $|B| = 0$, so folgt $\int_B f = 0$ und jedes $\bar{b} \in B$ leistet das Verlangte. Ist $|B| \neq 0$, so darf man in den Ungleichungen dividieren und erhält

$$f(\bar{b}_1) \leq \frac{1}{|B|} \int_B f \leq f(\bar{b}_2).$$

²¹Hier sind natürlich haufenweise Probleme versteckt: was bedeutet ‘halbwegs regulär’ und was ist $|\Delta|$ wenn Δ etwas komplizierter ist und noch kein Integralbegriff vorhanden ist, mit dem man die Fläche ausrechnen kann.

Wegen des Zusammenhangs liefert der Zwischenwertsatz ein $\bar{b} \in B$, so daß $f(\bar{b}) = \frac{1}{|B|} \int_B f$. Dieser Punkt ist wie verlangt.

Dieses Argument war sehr allgemein: es benutzt nur die Monotonie des Integrals und den Zwischenwertsatz für stetige Funktionen auf zusammenhängenden Mengen. Daher gilt der Mittelwertsatz auch für alle anderen Typen von Integralen, die noch kommen.

4.3.9 Integrale als Grenzwerte RIEMANNscher Summen

$B \subseteq \mathbb{R}^2$ sei kompakt und gutartig $f : B \rightarrow \mathbb{R}$ sei stetig. Dann wissen wir schon, daß B einen Flächeninhalt hat und $\int_B f$ existiert.

Wir denken uns B zerlegt in viele kleine Teilflächen B_i , die kompakt und zusammenhängend sein sollen und nur unwesentlich überlappen (in dem Sinne, daß der Teil den ein B_i mit den anderen Teilflächen gemeinsam hat, den Flächeninhalt Null besitzt: $|B_i \cap \bigcup_{j \neq i} B_j| = 0$ für alle i .)

Wählt man nun in jedem B_i einen Punkt \bar{c}_i so nennt man $\sum_i f(\bar{c}_i) |B_i|$ eine RIEMANNsche Summe für das Integral $\int_B f$. Man hofft, daß diese RIEMANNschen Summen das Integral immer besser approximieren, je feiner die Zerlegung ist.

Um einzusehen, daß das so ist, benutzen wir die Additivität²² des Integrals und den Mittelwertsatz

$$\int_B f(x, y) d(x, y) = \sum_i \int_{B_i} f(x, y) d(x, y) = \sum_i f(\bar{b}_i) |B_i|,$$

wobei die $\bar{b}_i \in B_i$ vom Mittelwertsatz gelieferte Punkte sind. Das Integral ist also sogar genau gleich einer RIEMANNschen Summe, die zur Zerlegung gehört.

Ist nun die Zerlegung hinreichend fein, d.h. haben alle B_i einen hinreichend kleinen Durchmesser, so unterscheiden sich die zugehörigen RIEMANNschen Summen beliebig wenig. Da eine von ihnen, wie gesehen mit dem Integral übereinstimmt, unterscheiden sich die anderen auch nur beliebig wenig vom Integral.

Genauer: Ist $\varepsilon > 0$ vorgegeben, so gibt es wegen der gleichmäßigen Stetigkeit von f auf B eine positive Zahl δ , so daß $|f(\bar{b}) - f(\bar{c})| < \varepsilon$ sobald $d(\bar{b}, \bar{c}) < \delta$. Haben dann alle B_i einen Durchmesser $< \delta$, so gilt

$$\left| \int_B f(x, y) d(x, y) - \sum_i f(\bar{c}_i) |B_i| \right| = \left| \sum_i f(\bar{b}_i) |B_i| - \sum_i f(\bar{c}_i) |B_i| \right| \leq \sum_i |f(\bar{b}_i) - f(\bar{c}_i)| |B_i| < \varepsilon |B|.$$

4.4 Zwei kleine Zusätze

4.4.1 Bemerkung,

von der ich nicht weiß, wo ich sie sonst unterbringen soll. Sind $g : [a, b] \rightarrow \mathbb{R}$ und $h : [c, d] \rightarrow \mathbb{R}$ einstellige integrierbare Funktionen, so ist $f(x, y) = g(x) \cdot h(y)$ auf $R = [a, b] \times [c, d]$ integrierbar und es gilt

$$\iint_R g(x) \cdot h(y) d(x, y) = \int_a^b g(x) dx \cdot \int_c^d h(y) dy.$$

Setzt man beide Funktionen als stetig voraus, so ist das nach dem vorangehenden trivial. Sind aber g und h nur integrierbar (im Sinne des vorigen Semesters), dann wird es eine nette kleine Übung.

4.4.2 Schwacher SCHWARZ

Aus der Vertauschbarkeit der Reihenfolge der Integrationen bei iterierten Integralen kann man auf die Vertauschbarkeit der Reihenfolge der partiellen Differentiationen schließen.

²²Hier wird ein wenig geschummelt; wir haben sie nur für nicht überlappende Zerlegungen bewiesen. Die B_i dürfen (müssen sogar) überlappen, aber nur am Rand.

Nööö, wissen wir noch nicht. Oben war das noch angenommen worden.

Genauer leiten wir jetzt aus FUBINI die folgende (zugegeben etwas schwache) Version von SCHWARZ her.

Angenommen $f : D \rightarrow \mathbb{R}$ ist auf einer offenen Menge $D \subseteq \mathbb{R}^2$ definiert und die beiden gemischten Ableitungen $\frac{\partial^2 f}{\partial x \partial y}$ und $\frac{\partial^2 f}{\partial y \partial x}$ sind stetig auf D . Dann stimmen sie überein.

Wir führen die gegenteilige Annahme zum Widerspruch. Sollten sich die beiden Ableitungen irgendwo unterscheiden, so gäbe es einen Punkt, in dem oBdA

$$\frac{\partial^2 f}{\partial x \partial y}(x_0, y_0) > \frac{\partial^2 f}{\partial y \partial x}(x_0, y_0)$$

gilt. Zwischen die beiden Zahlen passen zwei weitere, etwa $u > v$. Wegen der Stetigkeit beider Ableitungen gelten die Ungleichungen nicht nur in dem einen Punkt, sondern setzen sich auf eine Umgebung fort. Wir finden also ein (vielleicht nur kleines aber das macht nichts!) Rechteck $R = [a, b] \times [c, d]$ so daß für alle $(x, y) \in R$ gilt

$$\frac{\partial^2 f}{\partial x \partial y}(x, y) > u > v > \frac{\partial^2 f}{\partial y \partial x}(x, y).$$

Dann ist aber auch

$$\iint_R \frac{\partial^2 f}{\partial x \partial y} d(x, y) \geq v |R| > u |R| \geq \iint_R \frac{\partial^2 f}{\partial y \partial x} d(x, y).$$

Andererseits kann man die Integrale über R auf iterierte Integrale zurückführen und nach Hauptsatz ausrechnen:

$$\iint_R \frac{\partial^2 f}{\partial x \partial y} d(xy) = \int_c^d \left(\int_a^b \frac{\partial}{\partial x} \frac{\partial f}{\partial y} dx \right) dy = \int_c^d \left[\frac{\partial f}{\partial y}(b, y) - \frac{\partial f}{\partial y}(a, y) \right] dy = f(b, d) - f(b, c) - f(a, d) + f(a, c)$$

und

$$\iint_R \frac{\partial^2 f}{\partial y \partial x} d(xy) = \int_a^b \left(\int_c^d \frac{\partial}{\partial y} \frac{\partial f}{\partial x} dy \right) dx = \int_a^b \left[\frac{\partial f}{\partial x}(x, d) - \frac{\partial f}{\partial x}(x, c) \right] dx = f(b, d) - f(a, d) - f(b, c) + f(a, c).$$

Beide Male kommt derselbe Wert heraus, Widerspruch.

4.5 Integration im \mathbb{R}^n

Für $n > 2$ geht man im Prinzip genauso vor, wie im zweidimensionalen Fall, nur daß aus den Rechtecken Quader werden. Auch die Beweise sind praktisch dieselben nur schreibt sich alles viel unangenehmer auf²³.

4.5.1 Integration über Quader

Zunächst lernt man, beschränkte Funktionen, über Quader zu integrieren. Das macht man mit Hilfe von Zerlegungen in viele kleine Teilquader und die zugehörigen Ober- und Untersummen. Die Obersumme ist etwa definiert als

$$\sum \text{Supremum von } f \text{ auf dem Teilquader} \cdot \text{mal} \cdot \text{dessen Volumen,}$$

wobei über alle Teilquader der Zerlegung summiert wird. Dabei ist das Volumen eines Quaders definiert als Produkt seiner Seitenlängen also

$$\text{für } Q = [a_1, b_1] \times [a_2, b_2] \times \dots \times [a_n, b_n] \text{ als } |Q| := (b_1 - a_1) \cdot (b_2 - a_2) \cdot \dots \cdot (b_n - a_n).$$

²³Es sei denn, man führt clevere Bezeichnungen ein, die aber vorher auch erst gelernt und verdaut werden müssen.

Eine auf Q definierte beschränkte Funktion f heißt RIEMANN-integrierbar, wenn das Supremum der Menge aller Untersummen (bei allen möglichen Zerlegungen in Teilquader) gleich dem Infimum der Menge aller Obersummen ist, wobei die Zahl $\sup = \inf$ dann Integral genannt und mit $\int_Q f$ oder $\int_Q f(\bar{x}) d\bar{x}$ bezeichnet wird. Um die Dimension anzudeuten, schreibt man manchmal n Integralzeichen hintereinander: $\iint \dots \int_Q f(\bar{x}) d\bar{x}$. In \mathbb{R}^3 werde ich wahlweise $d(x, y, z)$ oder dV schreiben (und dann, wie es die Physiker gern tun vom ‘Volumenelement’ sprechen). Der Integrierbarkeitstest bleibt wörtlich erhalten und mit seiner Hilfe zeigt man, daß alle stetigen Funktionen auf allen Quadern integrierbar sind. Die entsprechenden Integrale lassen sich als (n fach) iterierte Integrale ausrechnen.

$$\iint \dots \int_Q f = \int_{a_1}^{b_1} \int_{a_2}^{b_2} \dots \int_{a_n}^{b_n} f(x_1, x_2, \dots, x_n) dx_n \dots dx_2 dx_1.$$

Hier ist ein Induktionsbeweis fällig, die Idee kann man 4.3.4 entnehmen.

4.5.2 Integration über allgemeinere Mengen

Nachdem man eine Integrationstheorie über Quader hat, macht man sich an allgemeinere beschränkte Gebiete. Integrierbarkeit und Integral werden mit demselben Trick wie in der Ebene definiert: Das Gebiet B wird zu einem großen Quader Q erweitert und nachgesehen, ob f_B auf Q im bisherigen Sinne integrierbar ist. Wenn ja, so nennt man f auf B integrierbar und definiert $\int_B f$ als $\int_Q f_B$.

Die Frage welche Funktionen auf welchen Bereichen integrierbar sind ist wieder heikel aber man beweist, mit denselben Schritten wie in der Ebene, daß *stetige Funktionen auf gutartigen Mengen integrierbar sind*.

Gutartig bedeutet wieder, daß man die Menge durch endlich viele Schnitte in einfach gutartige zerlegen kann. Die **Definition** von *einfach gutartig* wird rekursiv gegeben.

Die eindimensionalen einfach gutartigen Mengen sind die abgeschlossenen Intervalle.

Eine n -dimensionale einfach gutartige Menge G entsteht aus einer $n-1$ dimensionalen einfach gutartigen Menge H , zwei auf H stetigen Funktionen g, h und der Auswahl einer Koordinatenachse nach folgendem Schema:

$$G = \{\bar{x} = (\bar{u}, y, \bar{v}) \in \mathbb{R}^n : (\bar{u}, \bar{v}) \in H \text{ und } g(\bar{u}, \bar{v}) \leq y \leq h(\bar{u}, \bar{v})\}$$

4.5.3 Rückführung auf iterierte Integrale (von innen)

Jede stetige Funktion f ist auf jeder einfach gutartigen Menge G integrierbar und das Integral läßt sich auf das iterierte Integral

$$\int \dots \iint_G f(\bar{x}) d\bar{x} = \int \dots \int_H \left(\int_{g(\bar{u}, \bar{v})}^{h(\bar{u}, \bar{v})} f(\bar{u}, y, \bar{v}) dy \right) d(\bar{u}, \bar{v})$$

zurückführen. Das äußere Integral \int_H läßt sich nun seinerseits wieder als iteriertes Integral schreiben. In n Schritten kommt man zu ganz normalen Integralen über Intervalle herunter, die man mit Glück und BRONSTEINS Hilfe ausrechnen kann. Das wird vielleicht durchsichtiger an einem

Beispiel. Im dreidimensionalen könnte eine einfach gutartige Menge so aussehen:

$$G = \{(x, y, z) : (x, y) \in H \text{ und } g(x, y) \leq z \leq h(x, y)\},$$

wobei $H = \{(x, y) : a \leq y \leq b \text{ und } p(y) \leq x \leq q(y)\}$. Ausführlich aufgeschrieben ist

$$G = \{(x, y, z) : a \leq y \leq b \text{ und } p(y) \leq x \leq q(y) \text{ und } g(x, y) \leq z \leq h(x, y)\}$$

und das Integral $\iiint_G f(x, y, z) d(x, y, z)$ wird zu

$$\iint_H \left(\int_{g(x, y)}^{h(x, y)} f(x, y, z) dz \right) d(x, y) = \int_a^b \left(\int_{p(y)}^{q(y)} \left(\int_{g(x, y)}^{h(x, y)} f(x, y, z) dz \right) dx \right) dy.$$

4.5.4 Rückführung auf iterierte Integrale (von außen); Prinzip des CAVALIERI

Stellen wir uns diese Rückführung bis zum Ende ausgeführt vor. Dann steht ganz außen ein Integral über ein Intervall mit festen Grenzen. Im letzten Beispiel $\int_a^b \dots dy$. Wenn wir einen y -Wert zwischen a und b fixieren, so verwandelt sich \dots in ein $(n-1)$ -fach iteriertes Integral, im Beispiel $\int_{p(y)}^{q(y)} \left(\int_{g(x,y)}^{h(x,y)} f(x,y,z) dz \right) dx$, das der Integration über einen $n-1$ -dimensionalen einfach gutartigen Raumbereich entspricht, nämlich dem der aus G entsteht, wenn die y -Koordinate den fixierten konstanten Wert bekommt.

Das ist der Hintergrund des folgenden Verfahrens, das mit dem Namen CAVALIERI verbunden ist. Wir beschreiben es unabhängig von den bisherigen Betrachtungen noch einmal von vorn.

Wir wollen das n -dimensionale Integral $\int_B f d\bar{x}$ berechnen (Existenz vorausgesetzt). Seien a und b der größte und der kleinste Wert, den die Koordinate x_n in Punkten der Menge B annehmen kann. Für $t \in [a, b]$ bezeichne B^t den $n-1$ dimensional Schnitt von B in Höhe t . Mit selbsterklärenden Bezeichnungen

$$B^t = \{\bar{y} \in \mathbb{R}^{n-1} : (\bar{y}, t) \in B\}.$$

Weiter sei $f^t : B^t \rightarrow \mathbb{R}$ durch $f^t(\bar{y}) = f(\bar{y}, t)$ definiert. War f stetig, so ist auch jedes f^t stetig. War B gutartig, so auch jedes B^t . Also existiert für jedes t das $n-1$ -dimensionale Integral $\int_{B^t} f^t(\bar{y}) d\bar{y}$ und stellt eine stetige Funktion von t dar (bei Quadern ist das ganz klar; auf einen allgemeinen Beweis verzichten wir). Für diese gilt schließlich:

$$\iint \dots \int_B f(\bar{x}) d\bar{x} = \iint \dots \int_B f(\bar{y}, t) d(\bar{y}, t) = \int_a^b \left(\int \dots \int_{B^t} f^t(\bar{y}) d\bar{y} \right) dt.$$

In Anwendungen muß man nicht unbedingt die letzte Koordinate herausgenommen werden, das schreibt sich nur leichter auf.

4.5.5 Beispiel: das Volumen der n -dimensionalen Kugel

Ist $B \subseteq \mathbb{R}^n$ so nennen wir die Zahl $\int_B 1 d\bar{x}$ sofern sie existiert, das *Volumen*²⁴ von B und schreiben dafür wieder mal $|B|$. Hier soll das Volumen der n -dimensionalen Kugel

$$K_0^n(r) = \{\bar{x} \in \mathbb{R}^n : \sum x_i^2 \leq r^2\}$$

berechnet werden. Wir bezeichnen es mit $k_n(r)$ und schreiben, solange eventuell mehrere Dimensionen im Spiel sind, kurzzeitig $K_0^n(r)$.

Die Kugel vom Radius r entsteht aus der Einheitskugel durch Aufblähen oder Schrumpfen. Dabei wird das Volumen mit r^n multipliziert: $k_n(r) = r^n k_n(1)$. Das ist geometrisch ziemlich plausibel und folgt im Übrigen aus der Transformationsformel des nächsten Abschnitts, angewendet auf die Aufblähungsabbildung $(x_1, \dots, x_n) \mapsto (rx_1, \dots, rx_n)$. Ihre JACOBI-Matrix ist das r -fache der Einheitsmatrix, hat also die Determinante r^n .

Deswegen genügt es, $k_n(1)$ zu berechnen, das Volumen der Einheitskugel. Kugeln sind sicher so gutartig, wie es nur geht. Wir haben also zwei Möglichkeiten das Problem auf ein iteriertes Integral zu reduzieren. Versuchen wir es zunächst ‘von innen’:

$$K_0^n(1) = \left\{ (\bar{y}, t) \in \mathbb{R}^n : \bar{y} \in K_0^{n-1}(1) \text{ und } -\sqrt{1 - y_1^2 - \dots - y_{n-1}^2} \leq t \leq \sqrt{1 - y_1^2 - \dots - y_{n-1}^2} \right\}$$

liefert die Darstellung

$$k_n(1) = \int_{K_0^{n-1}(1)} \int_{-\sqrt{\dots}}^{\sqrt{\dots}} 1 dt d\bar{y} = \int_{K_0^{n-1}(1)} 2\sqrt{1 - y_1^2 - \dots - y_{n-1}^2} d\bar{y}.$$

Dieses Integral kann nun weiter aufgedröselt werden. Eine geeignete Rekursionsformel ist aber erstmal nicht in Sicht.

²⁴Notfalls das n -dimensionale Volumen. 1-dimensionales Volumen ist Länge, 2-dimensionales Flächeninhalt.

Das klappt besser beim Versuch ‘von außen’. Alle Koordinaten variieren zwischen -1 und 1 , der Schnitt von $K_0^n(1)$ in Höhe t ist $\{\bar{y} \in \mathbb{R}^{n-1} : y_1^2 + \dots + y_{n-1}^2 + t^2 \leq 1\} = K_0^{n-1}(\sqrt{1-t^2})$. Also

$$k_n(1) = \int_{-1}^1 \left(\int \dots \int_{K_0^{n-1}(\sqrt{1-t^2})} 1 \, d\bar{y} \right) dt = \int_{-1}^1 k_{n-1}(\sqrt{1-t^2}) \, dt.$$

Wegen $k_{n-1}(\sqrt{1-t^2}) = (\sqrt{1-t^2})^{n-1} k_{n-1}(1)$ und mit der Substitution $t = \sin s$, $dt = \cos s \, ds$ folgt für das Kugelvolumen weiter

$$k_n(1) = k_{n-1}(1) \cdot \int_{-1}^1 (\sqrt{1-t^2})^{n-1} dt = k_{n-1}(1) \cdot \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cos^n s \, ds.$$

Wenn wir das Integral von \cos^n mit c_n abkürzen wird das zu: $k_n(1) = c_n k_{n-1}(1)$. Im Abschnitt über partielle Integration hatten wir im letzten Semester rekursiv die Stammfunktion von $\cos^n s$ bestimmt:

$$\int \cos^n s \, ds = \frac{1}{n} \left(\cos^{n-1} s \sin s + (n-1) \int \cos^{n-2} s \, ds \right).$$

Daraus ergibt sich für c_n die Rekursionsgleichung $c_n = \frac{n-1}{n} c_{n-2}$. Die ersten beiden c 's rechnen wir direkt aus:

$$c_1 = \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cos s \, ds = 2 \quad c_2 = \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cos^2 s \, ds = \frac{\pi}{2}.$$

Danach läuft die Rekursion:

$$c_3 = \frac{2}{3} c_1 = \frac{4}{3}, \quad c_4 = \frac{3}{4} c_2 = \frac{3\pi}{4 \cdot 2}, \quad c_5 = \frac{4}{5} c_3 = \frac{4}{5} \cdot \frac{4}{3}, \quad \dots$$

Diese Formeln führen auf die Vermutung $c_n \cdot c_{n-1} = \frac{2\pi}{n}$, die man mit der obigen Rekursionsgleichung leicht induktiv bestätigt. Das bringt dann

$$k_n(1) = c_n \cdot k_{n-1}(1) = c_n \cdot c_{n-1} \cdot k_{n-2}(1) = \frac{2\pi}{n} k_{n-2}(1).$$

Die Volumina für $n = 1, 2$ kennen wir aus der Schule: die eindimensionale Kugel ist einfach das Intervall $[-1, 1]$ und hat das eindimensionale Volumen (Länge) $k_1(1) = 2$. Die zweidimensionale Kugel ist der Einheitskreis mit der Fläche $k_2(1) = \pi$. Danach kommen rekursiv

$$\left. \begin{array}{l} k_3(1) = \frac{2\pi}{3} k_1(1) = \frac{4}{3}\pi \\ k_5(1) = \frac{2\pi}{5} k_3(1) = \frac{8}{15}\pi^2 \\ \vdots \\ k_{2n+1}(1) = \frac{2^{n+1}\pi^n}{1 \cdot 3 \cdot 5 \cdot \dots \cdot (2n+1)} \end{array} \right\| \left. \begin{array}{l} k_4(1) = \frac{2\pi}{4} k_2(1) = \frac{\pi^2}{2}, \\ k_6(1) = \frac{2\pi}{6} k_4(1) = \frac{1}{6}\pi^3, \\ \vdots \\ k_{2n}(1) = \frac{\pi^n}{n!} \end{array} \right.$$

Insgesamt folgt die unerwartete Tatsache, daß $\lim_{n \rightarrow \infty} k_n(1) = 0$.

Solange $\frac{2\pi}{n} > 1$ steigen die Volumina (gerader Dimension) und ungerader Dimension). Ab $n = 7$ beginnen sie zu fallen. Also sind $k_5 = \frac{8}{15}\pi^2$ und $k_6 = \frac{1}{6}\pi^3$ die beiden Kandidaten für das maximale Kugelvolumen. Der Taschenrechner zeigt, daß k_5 etwas größer ist. Unter allen Einheitskugeln hat also die fünfdimensionale das größte Volumen.

4.6 Die Transformationsformel

Eine wichtige Methode, eindimensionale Integrale auszurechnen besteht in der Variablentransformation, die auf der Substitutionsformel

$$(*) \quad \int_a^b f(g(u)) \cdot g'(u) \, du = \int_{g(a)}^{g(b)} f(x) \, dx$$

beruht. In diesem Abschnitt wollen wir das höherdimensionale Analogon dieser Methode kennenlernen. Ein Integral $\int_C f(\bar{x}) \, d\bar{x}$ soll auf ein Integral $\int_B \dots \, d\bar{u}$ einer umgerechneten Funktion über einen anderen n -dimensionalen Bereich zurückgeführt werden. Um auf die richtige Spur zu kommen, beschäftigen wir uns noch ein wenig mit (*).

4.6.1 Interpretation der Substitutionsformel

Wir stellen uns zwei Geraden vor, auf denen die Koordinaten u bzw. x etabliert sind²⁵. Eine Transformation g eines genügend großen Stückes der einen Achse auf ein Stück der anderen Achse rechnet u -Koordinaten in x -Koordinaten um: $x = g(u)$. Das Ganze soll bijektiv passieren und in beiden Richtungen differenzierbar. Dann ist g entweder streng monoton wachsend (die Orientierung kleiner/größer der beiden Geraden bleibt erhalten) oder streng monoton fallend (die Orientierung dreht sich um). Ein abgeschlossenes Intervall $[a, b]$ auf der u -Achse wird von g in ein abgeschlossenes Intervall $[c, d]$ auf der x -Achse überführt. Ist $f(x)$ auf $[c, d]$ gegeben so entsteht durch Umrechnung die Funktion $\tilde{f}(u) := f(g(u))$ auf $[a, b]$.

Die Substitutionsformel sagt uns nun, wie das Integral $\int_c^d f(x) \, dx$ umgerechnet werden muß. Für wachsendes g ist $c = g(a)$ und $d = g(b)$ daher

$$\int_c^d f(x) \, dx = \int_{g(a)}^{g(b)} f(x) \, dx = \int_a^b \tilde{f}(u) \cdot g'(u) \, du.$$

Für fallendes g ist $c = g(b)$ und $d = g(a)$, daher

$$\int_c^d f(x) \, dx = \int_{g(b)}^{g(a)} f(x) \, dx = \int_b^a \tilde{f}(u) \cdot g'(u) \, du = - \int_a^b \tilde{f}(u) \cdot g'(u) \, du = \int_a^b \tilde{f}(u) \cdot |g'(u)| \, du,$$

weil g' jetzt negativ ist und daher $-g' = |g'|$. Im wachsenden Fall ist g' durchgängig positiv, daher schadet das Einfügen des Betrages nichts und die Formel

$$\int_c^d f(x) \, dx = \int_a^b \tilde{f}(u) \cdot |g'(u)| \, du$$

ist immer richtig. Neben dem f -spezifischen Teil \tilde{f} kommt der für alle f gültige zur Koordinatentransformation gehörende Umrechnungsfaktor $|g'|$ ins Spiel.

4.6.2 Heuristik für den mehrdimensionalen Fall

Jetzt betrachten wir eine (Koordinaten)transformation $\bar{g} : U \rightarrow V$, die eine offene Teilmenge $U \subseteq \mathbb{R}^n$ bijektiv und in beiden Richtungen stetig differenzierbar auf eine offene Teilmenge $V \subseteq \mathbb{R}^n$ abbildet. Am besten stellt man sich verschiedene Kopien von \mathbb{R}^n vor. In der ersten mögen die Koordinaten \bar{u} heißen, in der zweiten \bar{x} . Angenommen wir sind an einem stetigen Skalarfeld $f : C \rightarrow \mathbb{R}$ interessiert, das auf einer gutartigen Teilmenge von V definiert ist. Seine Beschreibung in \bar{u} -Koordinaten ist ein Skalarfeld $\tilde{f} : B \rightarrow \mathbb{R}$, wobei $B \subseteq U$ und $\tilde{f}(\bar{u}) = f(\bar{g}(\bar{u}))$. Ausgerechnet werden soll $\int_C f(\bar{x}) \, d\bar{x}$ durch Transformation in ein Integral von \tilde{f} über B . Der eindimensionale Fall lehrt, daß eine Formel der Art

$$\int_C f(\bar{x}) \, d\bar{x} = \int_B \tilde{f}(\bar{u}) \dots \, d\bar{u}$$

zu erwarten ist. Wir wollen herausbekommen, was anstelle der \dots zu stehen hat.

²⁵Die beiden Geraden können auch dieselbe sein; dann haben wir zwei Koordinatensysteme vor uns. Die Vorstellung mit zwei Geraden ist aber anschaulicher.

Dazu stellen wir uns eine Zerlegung von B in viele kleine höchstens am Rande überlappende gutartige Teile vor: $B = \bigcup_{i=1}^N B_i$. Diese wird von \bar{g} in eine analoge Zerlegung von C überführt. Mit $C_i := \bar{g}(B_i)$ haben wir $C = \bigcup_{i=1}^N C_i$. Das Integral über C zerfällt dann in die Summe der Integrale über die C_i und jedes von diesen kann nach dem Mittelwertsatz umgeschrieben werden:

$$\int_C f(\bar{x}) d\bar{x} = \sum_{i=1}^N f(\bar{c}_i) \cdot |C_i|$$

für einen jeweils geeigneten Punkt $\bar{c}_i \in C_i$. Klar, kommen diese Punkte via \bar{g} von Punkten $\bar{b}_i \in B_i$: $\bar{c}_i = \bar{g}(\bar{b}_i)$. Damit wird das Integral zu

$$\int_C f(\bar{x}) d\bar{x} = \sum_{i=1}^N f(\bar{g}(\bar{b}_i)) \cdot |C_i| = \sum_{i=1}^N \tilde{f}(\bar{b}_i) \cdot \frac{|C_i|}{|B_i|} \cdot |B_i|$$

Bis auf den noch zu bestimmenden Faktor, ist das eine RIEMANNSche Summe für ein Integral über B .

Betrachten wir ein festes i und stellen uns für den Moment vor, \bar{g} wäre auf B_i linear²⁶:

$$\bar{g}(\bar{u}) = \bar{c}_i + A_i \cdot (\bar{u} - \bar{b}_i)$$

für eine geeignete Matrix A_i . Für derartige Abbildungen hatten wir im ersten Semester gelernt, wie sie auf das Volumen wirken: es wird mit der Determinante multipliziert. Da wir hier am absoluten, nicht vorzeichenbehafteten Volumen interessiert sind, muß der Betrag der Determinante benutzt werden. Wir hätten also $|C_i| = |\det A_i| \cdot |B_i|$. In Wahrheit ist nun \bar{g} nicht linear. Wegen der Differenzierbarkeit ist die Transformation aber in einer Umgebung jedes Punktes gut durch eine lineare Transformation approximierbar:

$$\bar{g}(\bar{u}) \approx \bar{c}_i + J_{\bar{g}}(\bar{b}_i) \cdot (\bar{u} - \bar{b}_i) \quad \text{also} \quad |C_i| \approx |\det J_{\bar{g}}(\bar{b}_i)| \cdot |B_i|$$

umso genauer, je kleiner die Menge B_i ist. Für das Integral heißt das

$$\int_C f(\bar{x}) d\bar{x} \approx \sum_{i=1}^N \tilde{f}(\bar{b}_i) \cdot |\det J_{\bar{g}}(\bar{b}_i)| \cdot |B_i|,$$

je genauer, je feiner die Zerlegung von B ist. Im Grenzwert wird daraus eine Gleichheit und die RIEMANNSche Summe rechts verwandelt sich in das angestrebte Integral über B .

Diese Betrachtungen machen das folgende Resultat plausibel. Ein wasserdichter Beweis ist schwierig. Ich empfehle, in das Buch von JÄNICH zu schauen. Dort sind dieselben Plausibilitätsbetrachtungen mit vielen netten Bildern garniert.

4.6.3 Das exakte Resultat liest sich mathematisch folgendermaßen:

$\bar{g} : U \rightarrow V$ sei eine bijektive, in beiden Richtungen stetig differenzierbare (Koordinaten-) Transformation der offenen Menge $U \subseteq \mathbb{R}^n$ auf die offene Menge $V \subseteq \mathbb{R}^n$. Ist $B \subseteq U$ kompakt und f eine auf der (automatisch ebenfalls) kompakten Teilmenge $C = \bar{g}[B] \subseteq V$ integrierbare Funktion, so ist auch die auf B definierte Funktion $f \circ \bar{g} \cdot |\det J_{\bar{g}}|$ integrierbar und es gilt

$$\int_C f(\bar{x}) d\bar{x} = \int_B f(\bar{g}(\bar{u})) \cdot |\det J_{\bar{g}}(\bar{u})| d\bar{u}.$$

Das sieht sehr abstrakt und vielleicht verwirrend aus. Daher schreiben wir die

²⁶Was die Analytiker so linear nennen; eigentlich sollte man *affin* sagen.

4.6.4 zweidimensionale Variante mit Koordinaten

x, y bzw. u, v auf und geben auch die Transformation koordinatenweise an, etwa als $(u, v) \mapsto (g(u, v), h(u, v))$ bzw. klassisch $x = g(u, v)$ $y = h(u, v)$. Dann steht da

$$\int_C f(x, y) d(x, y) = \int_B f(g(u, v), h(u, v)) \cdot \left| \det \begin{pmatrix} \frac{\partial g}{\partial u}(u, v) & \frac{\partial g}{\partial v}(u, v) \\ \frac{\partial h}{\partial u}(u, v) & \frac{\partial h}{\partial v}(u, v) \end{pmatrix} \right| d(u, v)$$

Ganz klassisch schreibt man die Koordinatentransformation einfach als $x = x(u, v)$
 $y = y(u, v)$

und bezeichnet die sog. *Funktionaldeterminante* $\left| \begin{matrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{matrix} \right|$ mit $\frac{\partial(x, y)}{\partial(u, v)}$.

Setzt man noch $\tilde{f}(u, v) = f(x(u, v), y(u, v))$, so wird unsere Transformationsformel zu

$$\iint_C f(x, y) d(x, y) = \iint_B \tilde{f}(u, v) \left| \frac{\partial(x, y)}{\partial(u, v)} \right| d(u, v).$$

Wir gehen jetzt die drei häufigsten Koordinatentransformationen durch und rechnen jeweils die Funktionaldeterminante aus.

4.6.5 Polarkoordinaten

Hier spielen r und φ die Rolle von u und v . $x = r \cos \varphi$
 $y = r \sin \varphi$

hat die Funktionaldeterminante $\frac{\partial(x, y)}{\partial(r, \varphi)} = \left| \begin{matrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \varphi} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \varphi} \end{matrix} \right| = \left| \begin{matrix} \cos \varphi & -r \sin \varphi \\ \sin \varphi & r \cos \varphi \end{matrix} \right| = r$.

Damit wird die Transformationsformel zu

$$\iint_C f(x, y) d(x, y) = \iint_B \tilde{f}(r, \varphi) \cdot r d(r, \varphi),$$

wobei B das Gebiet der (r, φ) -Ebene ist, das dem (x, y) -Gebiet C entspricht. Man muß aufpassen, daß B in einem Streifen enthalten ist, auf dem die Transformation eindeutig ist; die Punkte von C dürfen nicht doppelt überdeckt werden²⁷!

Beispiel. Die Funktion $f(x, y) = \frac{1}{\sqrt{x^2 + y^2}^3}$ soll über den Viertelkreisring $C: a^2 \leq x^2 + y^2 \leq b^2$

$x, y \geq 0$ integriert werden.

Das wird in x, y -Koordinaten eher unangenehm, ist aber maßgeschneidert für die Umrechnung: $\tilde{f}(r, \varphi) = \frac{1}{r^3}$ und der (x, y) -Kreissausschnittssektor wird durch das (r, φ) -Rechteck $B: a \leq r \leq b; 0 \leq \varphi \leq \frac{\pi}{2}$ beschrieben.

$$\iint_C f(x, y) d(x, y) = \iint_B \frac{1}{r^3} \cdot r d(r, \varphi) = \int_0^{\frac{\pi}{2}} \int_a^b \frac{dr}{r^2} d\varphi = \frac{\pi}{2} \cdot \left(\frac{1}{a} - \frac{1}{b} \right).$$

²⁷So will es jedenfalls die reine Lehre: Bei der Berechnung von Integralen machen 'dünne' Störungen nichts aus. Will man etwa über einen Vollkreis C integrieren, so darf man getrost das Rechteck $B = [0, R] \times [0, 2\pi]$ benutzen. Auch wenn die ganze Unterkante $\{0\} \times [0, 2\pi]$ auf den Kreismittelpunkt abgebildet werden und auch die Punkte $(r, 0)$ und $(r, 2\pi)$ jeweils auf demselben Punkt des Kreises liegen. Dagegen darf man kein größeres Rechteck, etwa $[0, r] \times [0, 3\pi]$ als Integrationsgebiet benutzen. Hier werden zu viele Punkte des Kreises doppelt überdeckt und das fällt schon ins Gewicht. Ich will keine Regel aufschreiben; in der Praxis wird das selten falsch gemacht.

4.6.6 Anwendung: die LEIBNIZsche Sektorformel

$r = f(\varphi)$ sei die Gleichung einer Kurve in Polarkoordinaten und $\varphi_1 < \varphi_2$ zwei Winkel, die höchstens 2π auseinanderliegen. Dann beschreiben diese Daten ein krummliniges Dreieck C (vornehm: einen Sektor), dessen krumme Seite der Kurvenbogen von φ_1 bis φ_2 ist. Der dritte Eckpunkt ist der Koordinatenursprung. Gesucht ist der Flächeninhalt dieses Sektors.

C ergibt sich als Bild von $B = \{(r, \varphi) : \varphi_1 \leq \varphi \leq \varphi_2; 0 \leq r \leq f(\varphi)\}$ bei der oben schon betrachteten Abbildung $(r, \varphi) \mapsto (r \cos \varphi, r \sin \varphi)$ mit der Funktionaldeterminante r . Also

$$|C| = \int_C 1 d(x, y) = \int_B r d(r, \varphi) = \int_{\varphi_1}^{\varphi_2} \int_0^{f(\varphi)} r dr d\varphi = \int_{\varphi_1}^{\varphi_2} \frac{r^2}{2} \Big|_0^{f(\varphi)} d\varphi = \frac{1}{2} \int_{\varphi_1}^{\varphi_2} f(\varphi)^2 d\varphi.$$

Das ist die sogenannte *LEIBNIZsche Sektorformel*, die klassisch als $|C| = \frac{1}{2} \int_{\varphi_1}^{\varphi_2} r^2 d\varphi$ geschrieben wird, ohne die Funktion f explizit zu erwähnen. In Anwendungen geht es meist um zu geschlossenen Kurven entartete Sektoren; entweder $f(\varphi_1) = f(\varphi_2) = 0$ oder $f(\varphi_1) = f(\varphi_2)$ und $\varphi_2 = \varphi_1 + 2\pi$.

Als **konkretes Beispiel** betrachten wir die Herzkurve $r = a(1 - \sin \varphi)$. Nach Obigem umschließt sie eine Fläche von

$$\frac{1}{2} \int_0^{2\pi} a^2 (1 - \sin \varphi)^2 d\varphi = \frac{a^2}{2} \left[\int_0^{2\pi} 1 d\varphi - 2 \int_0^{2\pi} \sin \varphi d\varphi + \int_0^{2\pi} \sin^2 \varphi d\varphi \right] = \frac{a^2}{2} [2\pi + 0 + \pi] = \frac{3}{2} a^2 \pi.$$

4.6.7 Zylinderkoordinaten

sind dreidimensional und durch die Formeln $x = r \cos \varphi$, $y = r \sin \varphi$, $z = z$ gegeben. Die Funktionaldeterminante errechnet sich zu

$$\begin{vmatrix} \cos \varphi & -r \sin \varphi & 0 \\ \sin \varphi & r \cos \varphi & 0 \\ 0 & 0 & 1 \end{vmatrix} = r.$$

Die Formel für das Integral sieht bis auf die Dreistelligkeit aus wie bei Polarkoordinaten:

$$\iiint_C f(x, y, z) d(x, y, z) = \iiint_B f(r \cos \varphi, r \sin \varphi, z) \cdot r d(r, \varphi, z).$$

4.6.8 Kugelkoordinaten

Hier wird die Lage des Punktes P bestimmt durch seinen Abstand r vom Koordinatenursprung und seiner Lage auf der Späre $S_0(r)$, die wiederum durch zwei Winkel beschrieben wird. Aus der Geographie kennt man Längen- und Breitengrad, die könnte man benutzen. Traditionsgemäß nehmen die Mathematiker meist zwei andere Winkel. φ ist die Polarkoordinate der Projektion in die x, y -Ebene und θ der Winkel zwischen dem Ortsvektor des fraglichen Punktes und der positiven z -Richtung.

Es sind also (standardmäßig) $r \geq 0$, $0 \leq \varphi \leq 2\pi$, $0 \leq \theta \leq \pi$ und die Umrechnung in kartesische Koordinaten erfolgt nach der Vorschrift:

$$x = r \cos \varphi \sin \theta \quad y = r \sin \varphi \sin \theta \quad z = r \cos \theta$$

Daraus ergibt sich die Funktionaldeterminante

$$\begin{aligned} \frac{\partial(x, y, z)}{\partial(r, \varphi, \theta)} &= \begin{vmatrix} \cos \varphi \sin \theta & -r \sin \varphi \sin \theta & r \cos \varphi \cos \theta \\ \sin \varphi \sin \theta & r \cos \varphi \sin \theta & r \sin \varphi \cos \theta \\ \cos \theta & 0 & -r \sin \theta \end{vmatrix} \\ &= \cos \theta (-r^2 \sin \theta \cos \theta) - r \sin \theta (r \sin^2 \theta) = -r^2 \sin \theta \end{aligned}$$

4.7 Beispiel: Integration über einen Torus

Wir wollen jetzt eine kurze Pause einlegen und die bisherigen Integrationstechniken auf eine konkrete dreidimensionale Menge beziehen, nämlich einen Torus. Dieses Wort ist, ähnlich wie Kreis und Kugel, zweideutig; es bezeichnet sowohl einen Körper als auch dessen Oberfläche. Im Moment meine ich den Körper, den man der Deutlichkeit halber manchmal auch ‘Volltorus’ nennt. Anschaulich gesprochen handelt es sich um einen Autoreifen (Vollgummi) oder Donut. Mathematischer ist es ein Rotationskörper, der entsteht, wenn ein Vollkreis um eine außerhalb gelegene Achse rotiert. Er wird durch zwei Radien beschrieben: r = Radius des rotierenden Kreises und R = Radius des Kreises den der Kreismittelpunkt bei der Rotation beschreibt. Weil die Achse den Kreis nicht schneidet, ist stets $R > r$. Wir fixieren jetzt diese Daten und nennen den entstehenden Torus T . Spätestens ab hier müssen Sie Bilder malen; da ich dazu nicht fähig bin, habe ich am Rand Platz gelassen.

4.7.1 Darstellung in kartesischen Koordinaten; T als gutartiger Körper

Wir benutzen die z -Achse des Koordinatensystems als Rotationsachse und legen den rotierenden Kreis in die (x, z) -Ebene mit Mittelpunkt $(R, 0)$.

Als Projektion von T auf die (x, y) -Ebene ergibt sich ein Kreisring mit den beiden Radien $R \pm r$, d.h. die Menge

$$H := \{(x, y) \in \mathbb{R}^2 : R - r \leq \sqrt{x^2 + y^2} \leq R + r\}$$

Für $(x, y) \in H$ gehört (x, y, z) zu T , falls

$$z^2 + \left(\sqrt{x^2 + y^2} - R\right)^2 \leq r^2.$$

T ist also der zwischen den Funktionsgraphen von $\pm\sqrt{r^2 - \left(\sqrt{x^2 + y^2} - R\right)^2}$ auf H eingeschlossene gutartige Bereich. Als Formel

$$T = \left\{ (x, y, z) \in \mathbb{R}^3 : (x, y) \in H \text{ und } |z| \leq \sqrt{r^2 - \left(\sqrt{x^2 + y^2} - R\right)^2} \right\}.$$

Diese Darstellung, erlaubt, Integrale über T als iterierte Integrale darzustellen:

$$\iiint_T f(x, y, z) d(x, y, z) = \iint_H \left(\int_{-g(x, y)}^{g(x, y)} f(x, y, z) dz \right) d(x, y),$$

wobei ich $\sqrt{r^2 - \left(\sqrt{x^2 + y^2} - R\right)^2}$ mit $g(x, y)$ abgekürzt habe. Will man speziell das Volumen von T bestimmen, so wird 1 integriert und man erhält

$$|T| = \iint_H \int_{-g}^g 1 dz d(x, y) = 2 \iint_H \sqrt{r^2 - \left(\sqrt{x^2 + y^2} - R\right)^2} d(x, y),$$

ein Integral über den Kreisring, das man mit Polarkoordinaten berechnen könnte. Ich überspringe das.

4.7.2 Integration über T 'nach CAVALIERI'

Der größte und kleinste Wert von z auf T sind $\pm r$. Schneidet man T in Höhe z , so entsteht ein Kreisring, über dessen Radien man sich am schnellsten mit einer kleinen Skizze klar wird.

$$T^z = \left\{ (x, y) : R - \sqrt{r^2 - z^2} \leq \sqrt{x^2 + y^2} \leq R + \sqrt{r^2 - z^2} \right\}$$

Ein Integral über T läßt sich also auch als

$$\iiint_T f(x, y, z) d(x, y, z) = \int_{-r}^r \left(\iint_{T^z} f(x, y, z) d(x, y) \right) dz$$

berechnen. Speziell für die konstante Funktion 1 ergibt sich innen der Flächeninhalt des Kreisrings $|T^z|$, den man auch ohne zu integrieren hinschreiben kann (Differenz der beiden Kreisflächen). Das liefert für das Torusvolumen die Formel

$$\int_{-r}^r \pi \left(\left(R + \sqrt{r^2 - z^2} \right)^2 - \left(R - \sqrt{r^2 - z^2} \right)^2 \right) dz = \int_{-r}^r \pi 4R \sqrt{r^2 - z^2} dz$$

Das Integral von $\sqrt{r^2 - z^2}$ erkennen wir als halbe Fläche des Kreises vom Radius r also $\frac{\pi r^2}{2}$. Daher wird das Torusvolumen

$$|T| = 4\pi R \frac{\pi r^2}{2} = 2\pi^2 R r^2.$$

4.7.3 Beschreibung von T durch Zylinderkoordinaten

Da der Buchstabe r für den kleinen Radius vergeben ist, benutze ich jetzt t für die Zylinderkoordinaten: (t, φ, z) . Da t den Abstand des dargestellten Punktes von der z -Achse mißt, variiert dieser Wert für Punkte aus T zwischen $R-r$ und $R+r$. Die z -Koordinate des zugehörigen Punktes hängt nur von t ab (das ist immer so bei Rotationskörpern) und zwar gilt, wie oben, $z^2 + (t - R)^2 \leq r^2$. Damit ergibt sich die Darstellung

$$T = \left\{ (t \cos \varphi, t \sin \varphi, z) : \varphi \in [0, 2\pi], \quad R - r \leq t \leq R + r, \quad |z| \leq \sqrt{r^2 - (t - R)^2} \right\}.$$

Bei der Integration über T mittels dieser Darstellung kommt der für Zylinderkoordinaten gültige Korrekturfaktor t (statt r !) ins Spiel und das Integral wird zu

$$\int_0^{2\pi} \int_{R-r}^{R+r} \int_{-\sqrt{\dots}}^{\sqrt{\dots}} \tilde{f}(t, \varphi, z) \cdot t dz dt d\varphi.$$

Speziell für das Torusvolumen erhalten wir

$$\int_0^{2\pi} \int_{R-r}^{R+r} \int_{-\sqrt{\dots}}^{\sqrt{\dots}} 1 \cdot t dz dt d\varphi = \int_0^{2\pi} \int_{R-r}^{R+r} t 2\sqrt{r^2 - (t - R)^2} dt d\varphi = 2\pi \int_{R-r}^{R+r} t 2\sqrt{r^2 - (t - R)^2} dt.$$

Mit der Substitution $s = t - R$ wird daraus

$$4\pi \int_{-r}^r (s + R) \sqrt{r^2 - s^2} ds = 4\pi R \int_{-r}^r \sqrt{r^2 - s^2} ds,$$

wie oben (der erste Summand ist eine ungerade Funktion, die, auf einem symmetrischen Intervall integriert, 0 ergibt).

4.7.4 Toruskoordinaten

Schließlich gibt es noch eine spezifische Möglichkeit den Torus darzustellen (zu parametrisieren, Koordinaten einzuführen), die dem ‘geographischen’ Vorgehen der Kugelkoordinaten analog ist. Man benutzt zwei Winkel φ, ψ und einen Abstand t . Um deren Bedeutung zu verstehen, stellt man sich am besten noch einmal den rotierenden kleinen Kreis vor. Auf ihm läßt sich die Lage eines Punktes durch Polarkoordinaten beschreiben (t, φ) . ψ gibt dann an, um welchen Winkel der kleine Kreis gedreht wurde um den zu beschreibenden Punkt zu erreichen. Das liefert dann den Zusammenhang:

$$x = (R + t \cos \varphi) \cos \psi, \quad y = (R + t \cos \varphi) \sin \psi, \quad z = t \sin \varphi,$$

wobei $t \in [0, r]$ $\varphi, \psi \in [0, 2\pi]$. Im (t, φ, ψ) Raum entspricht dem Torus also ein Quader, über den es sich besonders angenehm integrieren läßt. Dafür kommt jetzt allerdings ein Korrekturfaktor ins Spiel, den wir neu ausrechnen müssen. Es ist der Betrag der Funktionaldeteminante

$$\frac{\partial(x, y, z)}{\partial(t, \varphi, \psi)} = \begin{vmatrix} \cos \varphi \cos \psi & -t \sin \varphi \cos \psi & -(R + t \cos \varphi) \sin \psi \\ \cos \varphi \sin \psi & -t \sin \varphi \sin \psi & (R + t \cos \varphi) \cos \psi \\ \sin \varphi & t \cos \varphi & 0 \end{vmatrix} = -t(R + t \cos \varphi).$$

Wegen $t \leq r < R$ ist $R + t \cos \varphi$ immer positiv. Daher wird der Korrekturfaktor $t(R + t \cos \varphi)$ und Integrale in Toruskoordinaten berechnen sich so:

$$\int_0^{2\pi} \int_0^{2\pi} \int_0^r \tilde{f}(t, \varphi, \psi) \cdot t(R + t \cos \varphi) dt d\psi d\varphi.$$

Benutzt man diese Koordinaten zur Volumenberechnung, so entsteht das Integral

$$\begin{aligned} \int_0^{2\pi} \int_0^{2\pi} \int_0^r t(R + t \cos \varphi) dt d\psi d\varphi &= \int_0^{2\pi} \int_0^{2\pi} \left(\frac{r^2}{2} R + \frac{r^3}{3} \cos \varphi \right) d\psi d\varphi = \\ &= \int_0^{2\pi} 2\pi \left(\frac{r^2}{2} R + \frac{r^3}{3} \cos \varphi \right) d\varphi = 4\pi^2 \frac{r^2 R}{2} = 2\pi^2 R r^2. \end{aligned}$$

4.8 Zwischenspiel: Flächen in \mathbb{R}^3 ; Tangentialebene und Normalenvektor

Als Höhepunkt unserer Integrationstheorie wollen wir über Oberflächen integrieren. Die Mathematiker behandeln ganz allgemein k -dimensionale Gebilde, die irgendwie schief und ausgebeult im n -dimensionalen Raum liegen. Wir haben schon $k = 1$ (Kurven) für beliebige n betrachtet. Ansonsten schränken wir uns aber auf $k = 2$ und $n = 3$ ein: normale Flächen im normalen Raum. Das macht die ganze Theorie viel anschaulicher und ist der für die Physik wichtigste Fall. Dieser Abschnitt dient dazu, uns erst einmal über Flächen in \mathbb{R}^3 zu verständigen.

Wir werden nicht versuchen zu erklären, welche Punkt Mengen in \mathbb{R}^3 man sinnvollerweise als Flächen bezeichnen kann. Wir gehen sofort davon aus, daß die Flächen, mit denen wir zu tun haben wollen, durch differenzierbare Funktionen dargestellt werden können. Mit einer Art der Darstellung haben wir es bei den Nebenbedingungen schon zu tun gehabt: *Niveauflächen* von Abbildungen

$$\Gamma = \{(x, y, z) \in \mathbb{R}^3 : G(x, y, z) = c\}$$

für eine Funktion G und einen Wert c . Der zweite Prototyp sind die Funktionsgraphen

$$\Gamma = \{(x, y, z) : (x, y) \in B \text{ und } z = g(x, y)\} = \{(x, y, g(x, y)) : (x, y) \in B\},$$

wobei g die Funktion ist, um deren Graphen es sich handelt und $B \subseteq \mathbb{R}^2$ ihr Definitionsbereich (oder ein Teil davon). Der Satz über implizite Funktionen sagt, daß man eine Niveaufläche lokal als Funktionsgraph schreiben kann, allerdings nicht unbedingt mit x, y als unabhängigen Variablen, sondern eventuell x, z oder y, z , je nachdem, welche partielle Ableitung von G gerade $\neq 0$ ist.

4.8.1 Beschreibung

Das allgemeinere Konzept eines *parametrisierten Flächenstücks* geht aus von einem gutartigen (platten) Bereich B einer Parameterebene (deren Koordinaten zur Unterscheidung u, v heißen) und einer Funktion²⁸ Φ die dieses B gekrümmt und verzogen aber nicht verklebt (also injektiv) in \mathbb{R}^3 legt. Die parametrisierte Fläche ist $\Gamma = \Phi(B)$. Von der Funktion Φ setzen wir voraus, daß sie auf einer B umfassenden offenen Menge definiert und stetig differenzierbar ist. Ihre JACOBI-Matrix $J_\Phi(\vec{b})$ soll in jedem Punkt vollen Rang (d.h. Rang 2) haben. Das wird garantieren, daß in jedem Punkt von Γ eine Tangentialebene existiert (s. unten). Später, wenn wir über Flächenstücke integrieren wollen, werden wir auch noch voraussetzen, daß der ebene Bereiche B einfach ist; vorerst spielt das keine Rolle. Wir werden nur solche *Flächen* betrachten, die sich als endliche Vereinigungen von (eventuell überlappenden) parametrisierten Flächenstücken schreiben lassen. Die Gleichung $(x, y, z) = \Phi(u, v)$, welche die Parametrisierung beschreibt, wird koordinatenweise als $x = \Phi_1(u, v), y = \Phi_2(u, v), z = \Phi_3(u, v)$ geschrieben. In der klassischen Mathematik hatte die Funktion Φ oft auch gar keinen Namen, man schrieb $x = x(u, v), y = y(u, v), z = z(u, v)$. Viele Praktiker und Physiker tun das heute noch.

Die Funktionsgraphen von C^1 -Funktionen fallen natürlich unter das Schema: dann ist $\Phi(u, v) = (u, v, g(u, v))$. Der Satz über implizite Funktionen sagt, daß (sofern $\text{grad } G \neq \vec{0}$) auch Niveauflächen unter das Schema fallen, wobei allerdings eventuell gestückelt werden muß.

Ein anderer Aspekt desselben Sachverhalts: durch Φ werden auf der Fläche Γ Koordinaten eingeführt; jeder Punkt $\Phi(u, v)$ von Γ ist durch das Zahlenpaar (u, v) eindeutig bestimmt. Ohne präziser zu werden, sollte ich hier die Ausdrücke (lokale) Karte und Atlas erwähnen. Die Paare (B, Φ) bilden die Karten; sie beschreiben Teilbereiche (Stücke) der Fläche. Ihre Gesamtheit ist ein Atlas.

4.8.2 Beispiel Sphäre

Als Niveaufläche wird sie durch

$$\Gamma = \{(x, y, z) : x^2 + y^2 + z^2 = R^2\}$$

dargestellt. Sie zerfällt in die nördliche ($\Gamma_+; z \geq 0$) und südliche ($\Gamma_-; z \leq 0$) Hemisphäre, die sich am Äquator ($z = 0$) überschneiden. Beide sind Funktionsgraphen. Für beide kann als Parameterbereich der Vollkreis $B = \{(x, y) : x^2 + y^2 \leq R^2\}$ in der x, y -Ebene gewählt werden. Die beiden Parametrisierungen sind

$$\Phi_+(x, y) = (x, y, \sqrt{R^2 - x^2 - y^2}) \quad \text{bzw.} \quad \Phi_-(x, y) = (x, y, -\sqrt{R^2 - x^2 - y^2}).$$

Wenn man genau hinguckt, sieht man aber, daß nicht alle oben gemachten Voraussetzungen erfüllt sind. Die Funktionen Φ_\pm sind nicht in einer B enthaltenden offenen Menge definiert und auch nicht auf ganz B differenzierbar. Die gleich zu besprechenden Tangentialebenen und Normalenvektoren können mit dieser Parametrisierung nur für innere Punkte von B also Sphärenpunkte außerhalb des Äquators bestimmt werden.

Der Ausweg besteht darin, stärker zu stückeln, etwa die westliche und östliche Hemisphäre (und zwei weitere) einzubeziehen. Dann fällt jeder Punkt in 'die Mitte' eines Flächenstücks, wo die entsprechenden Φ 's stetig differenzierbar sind.

Ein anderer Versuch, die Späre zu parametrisieren, könnte (wie in der Geographie) von einem Rechteck $B = [-\pi, \pi] \times [-\frac{\pi}{2}, \frac{\pi}{2}]$ der φ, ψ -Ebene ausgehen und

$$\Phi(\varphi, \psi) = (R \cos \varphi \cos \psi, R \sin \varphi \cos \psi, R \sin \psi)$$

benutzen. Hierbei ist Φ sogar wie verlangt auf einer größeren offenen Menge differenzierbar. Aber auch diese Parametrisierung ist nicht ganz im Rahmen der oben skizzierten Theorie, weil zuviele Punkte des Rechtecks identifiziert werden: Nordpol, Südpol und die Punkte der Datumsgrenze haben jeweils mehr als ein Urbild auf dem Rechteck.

²⁸Eigentlich sollten wir $\vec{\Phi}$ schreiben, aber da Φ nur in dieser Bedeutung vorkommt, werde ich mir das sparen.

4.8.3 Tangentialebene und Tangentialraum

Sei jetzt ein durch die Funktion $\Phi : B \rightarrow \mathbb{R}^3$ parametrisiertes Flächenstück Γ gegeben, $\bar{b} = (u_0, v_0)$ ein (sicherheitshalber) innerer Punkt von B und $\bar{p} = \Phi(\bar{b}) = (x_0, y_0, z_0)$ der entsprechende Punkt auf der Fläche. Wir wollen die *Tangentialebene* an die Fläche im Punkt definieren und beschreiben. Wenn der Punkt \bar{x} diese Tangentialebene durchläuft, so durchläuft der Vektor $\vec{v} = \bar{x} - \bar{p}$ einen gewissen zweidimensionalen Unterraum von \mathbb{R}^3 , den man *Tangentialraum* nennt und mit $T_{\bar{p}}\Gamma$ bezeichnet. Umgekehrt ergeben sich alle Punkte der Tangentialebene als $\bar{p} + \vec{v}$, wobei $\vec{v} \in T_{\bar{p}}\Gamma$. Tangentialebene und -raum kodieren sich also gegenseitig, aber letzterer erweist sich als leichter zugänglich. Die Idee besteht wieder darin, Kurven vorzuschicken. Verläuft eine Kurve ganz in der Fläche Γ und geht sie durch den Punkt \bar{p} , so liegt die Kurventangente in der Tangentialebene an die Fläche, ihre Ableitung (als Richtungsvektor dieser Tangente) also im Tangentialraum. Alle $\dot{\gamma}(0)$ für ganz in Γ verlaufende Kurven mit $\bar{\gamma}(0) = \bar{p}$, sind also sicher in $T_{\bar{p}}\Gamma$. Wenn wir uns noch davon überzeugen, daß diese Vektoren einen zweidimensionalen Unterraum von \mathbb{R}^3 bilden, wird die Definition

$$T_{\bar{p}}\Gamma = \{\dot{\gamma}(0) : \bar{\gamma} :]-\varepsilon, \varepsilon[\rightarrow \Gamma \text{ und } \bar{\gamma}(0) = \bar{p}\}$$

gerechtfertigt sein. Bei dem folgenden kleinen Argument werden wir gleich noch eine Basis finden, die uns nachträglich von der Betrachtung der Kurven wieder befreit. Man bemerke, daß die Parametrisierung von Γ in diese Überlegungen gar nicht eingegangen ist. Verschiedene Parametrisierungen führen deshalb automatisch zum selben Tangentialraum. Allerdings hängen die gleich zu bestimmenden Basen von $T_{\bar{p}}\Gamma$ sehr wohl von der Parametrisierung ab.

Für beliebige $c_1, c_2 \in \mathbb{R}$ definiert die Festsetzung $\bar{\gamma}(t) = \Phi(\bar{b} + t \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}) = \Phi(u_0 + c_1 t, v_0 + c_2 t)$ für genügend kleine $|t|$ eine differenzierbare Kurve die gezwungenermaßen innerhalb $\Gamma = \Phi(B)$ verläuft und zum Zeitpunkt $t = 0$ durch den Punkt $\bar{\gamma}(0) = \Phi(\bar{b}) = \bar{p}$ geht. Ihre Ableitung ergibt sich nach der Kettenregel als

$$\dot{\gamma}(0) = c_1 \frac{\partial \Phi}{\partial u}(\bar{b}) + c_2 \frac{\partial \Phi}{\partial v}(\bar{b}).$$

Das zeigt, daß jedenfalls alle Linearkombinationen der beiden Vektoren $\frac{\partial \Phi}{\partial u}(\bar{b})$ und $\frac{\partial \Phi}{\partial v}(\bar{b})$ von der Form $\dot{\gamma}(0)$ sind.

Umgekehrt läßt sich jede zulässige Kurve $\bar{\gamma}$ in der Form $\bar{\gamma}(t) = \Phi(\bar{\alpha}(t))$ schreiben, wo $\bar{\alpha} = (\alpha_1, \alpha_2)$ eine Kurve in B ist, die zum Zeitpunkt $t = 0$ durch \bar{b} geht²⁹. Die Kettenregel zeigt,

$$\dot{\gamma}(0) = \dot{\alpha}_1(0) \frac{\partial \Phi}{\partial u}(\bar{\alpha}(0)) + \dot{\alpha}_2(0) \frac{\partial \Phi}{\partial v}(\bar{\alpha}(0)) = \dot{\alpha}_1(0) \frac{\partial \Phi}{\partial u}(\bar{b}) + \dot{\alpha}_2(0) \frac{\partial \Phi}{\partial v}(\bar{b}).$$

Tatsächlich sind also alle Tangentenvektoren von Kurven durch \bar{p} Linearkombinationen der beiden partiellen Ableitungen. Damit ist die oben angegebene Definition gerechtfertigt und wir haben sogar eine Basis des Tangentialraumes gefunden. Die beiden partiellen Ableitungen $\frac{\partial \Phi}{\partial u}$ und $\frac{\partial \Phi}{\partial v}$ sind nämlich gerade die Spalten der JACOBI-Matrix $J_{\Phi}(\bar{b})$ und die hatte nach unserer Voraussetzung vollen Rang. Also

$$T_{\bar{p}}\Gamma = \text{span} \left\{ \frac{\partial \Phi}{\partial u}(\bar{b}), \frac{\partial \Phi}{\partial v}(\bar{b}) \right\}.$$

Die Frage was mit Punkten ist, an denen die JACOBI-Matrix keinen vollen Rang hat (die Mathematiker nennen sie *singulär*, im Gegensatz zu den bisher betrachteten *regulären* Punkten), müssen wir aus Zeitgründen unbeantwortet lassen. Für unser eigentliches Ziel, die Integration über Flächen, spielen sie normalerweise keine Rolle, weil es so wenige sind, daß die Integrale gar nichts davon merken.

4.8.4 Normalen

Wir betrachten dasselbe Flächenstück wie eben und den Punkt $\bar{p} = \Phi(\bar{b})$. Jeden $\neq \vec{0}$ Vektor, der auf $T_{\bar{p}}\Gamma$ senkrecht steht, nennt man *normal* im Punkt \bar{p} . Trägt man die normalen Vektoren an den

²⁹Das ist jedenfalls sehr plausibel; beweisen will ich es nicht.

Punkt \bar{p} an, so füllen sie zusammen mit \bar{p} eine Gerade aus, die auf der Tangentialebene senkrecht steht. Das ist spezifisch für zweidimensionale Flächen im dreidimensionalen Raum³⁰. Ebenso spezifisch dreidimensional ist die Möglichkeit, einen Normalenvektor sofort hinzuschreiben. Da er senkrecht auf den beiden Basisvektoren von $T_{\bar{p}}\Gamma$ stehen soll, nimmt man einfach das Vektorprodukt³¹:

$$\frac{\partial\Phi}{\partial u}(\bar{b}) \times \frac{\partial\Phi}{\partial v}(\bar{b}).$$

Alle anderen Normalenvektoren sind skalare Vielfache von diesem. Speziell gibt es zwei Normalenvektoren der Länge 1, die sogenannten *Normaleneinheitsvektoren*, die sich im Vorzeichen unterscheiden und auf die eine oder andere Seite des Flächenstücks zeigen.

Definition. Eine Fläche heißt *orientierbar* oder *zweiseitig*, wenn man auf stetige Weise in jedem Punkt durch Angabe eines der beiden Normaleneinheitsvektoren eine Seite auszeichnen kann. Anders gesagt, wenn es ein auf Γ definiertes stetiges Vektorfeld \vec{n} derart gibt, daß in jedem Punkt $\bar{p} \in \Gamma$

$$\|\vec{n}(\bar{p})\| = 1 \quad \text{und} \quad \vec{n}(\bar{p}) \perp T_{\bar{p}}\Gamma.$$

Die üblichen Flächen, die einem so einfallen, sind orientierbar. Sie haben ein natürliches oben/unten bzw. innen/außen. Das bekannte MÖBIUSband ist das populärste Beispiel einer nicht orientierbaren Fläche.

4.8.5 Gleichung der Tangentialebene

Kennt man einen Normalenvektor \vec{N} im Punkt \bar{p} so kann man die Gleichung der Tangentialebene sofort hinschreiben. Sie besteht aus allen Punkten \bar{x} , für die $\bar{x} - \bar{p}$ senkrecht zu \vec{N} steht:

$$\vec{N} \bullet (\bar{x} - \bar{p}) = 0 \quad \text{ausführlich} \quad N_1 \cdot (x - p_1) + N_2 \cdot (y - p_2) + N_3 \cdot (z - p_3) = 0.$$

4.8.6 Spezialfall Funktionsgraphen

Ist die betrachtete Fläche Graph der Funktion $g(x, y)$ über der Menge B . Dann benutzt man natürlich auch x, y als Parameter. Seien $\bar{b} = (x_0, y_0) \in B$ und $\bar{p} = (x_0, y_0, g(x_0, y_0)) \in \Gamma$ gegeben. Die Abbildung $\Phi(x, y)$ ist definiert als $(x, y, g(x, y))$ und die obigen Formeln werden zu folgenden (alle partiellen Ableitungen im Punkt $\bar{b} = (x_0, y_0)$ ausgerechnet):

Basis des Tangentialraumes $T_{\bar{p}}\Gamma$.

$$\begin{pmatrix} 1 \\ 0 \\ \frac{\partial g}{\partial x} \end{pmatrix} \quad \text{und} \quad \begin{pmatrix} 0 \\ 1 \\ \frac{\partial g}{\partial y} \end{pmatrix}$$

Normalenvektor:

$$\begin{pmatrix} 1 \\ 0 \\ \frac{\partial g}{\partial x} \end{pmatrix} \times \begin{pmatrix} 0 \\ 1 \\ \frac{\partial g}{\partial y} \end{pmatrix} = \begin{pmatrix} -\frac{\partial g}{\partial x} \\ -\frac{\partial g}{\partial y} \\ 1 \end{pmatrix}$$

Die **Gleichung der Tangentialebene im Punkt \bar{p}** lautet, wenn man oben formal einsetzt

$$-\frac{\partial g}{\partial x} \cdot (x - x_0) - \frac{\partial g}{\partial y} \cdot (y - y_0) + 1 \cdot (z - g(x_0, y_0)) = 0.$$

Das kann man nach z umstellen und findet, gar nicht verwunderlich, auf der rechten Seite die lineare Approximation von g in der Nähe von (x_0, y_0) wieder:

$$z = g(x_0, y_0) + \frac{\partial g}{\partial x}(x_0, y_0) \cdot (x - x_0) + \frac{\partial g}{\partial y}(x_0, y_0) \cdot (y - y_0)$$

Funktionsgraphen sind immer orientierbar, es gibt eine sinnvolle Unterscheidung von oben und unten, vermittelt durch die z -Achse.

³⁰Die bisherigen Betrachtungen über Tangentialräume würden dagegen ganz analog auch für andere Kombinationen von k und n durchgehen.

³¹Die wichtigsten Eigenschaften stehen noch einmal in 4.8.9.

4.8.7 Spezialfall Niveauflächen

Sei $\bar{p} = (x_0, y_0, z_0)$ ein Punkt auf der Niveaufläche $\Gamma = \{(x, y, z) : G(x, y, z) = c\}$. Bei dieser Darstellung hat der Tangentialraum keine kanonische Basis (aber natürlich, wie jeder Vektorraum viele Basen, s. auch unten). Dafür werden wir im Normalenvektor einen alten Bekannten wiedertreffen.

Stellen wir uns zunächst einen Tangentialvektor $\dot{\bar{\gamma}}(0)$ vor, wobei $\bar{\gamma}$ eine in Γ verlaufende Kurve ist, die zum Zeitpunkt $t = 0$ durch unseren Punkt \bar{p} geht. Die Funktion $G(\bar{\gamma}(t))$ ist dann konstant gleich c also verschwindet ihre Ableitung:

$$0 = \frac{d}{dt}G(\bar{\gamma})(0) = \vec{\text{grad}} G(\bar{p}) \bullet \dot{\bar{\gamma}}(0).$$

Mithin steht der Gradient von G senkrecht auf $\dot{\bar{\gamma}}(0)$. Da der Tangentialvektor beliebig war, ist $\vec{\text{grad}} G(\bar{p})$ als Normalenvektor der Niveaufläche identifiziert, jedenfalls solange er $\neq \vec{0}$ ist, was ich stillschweigend voraussetze. Die Gleichung der Tangentialebene ergibt sich dann als $\vec{\text{grad}} G(\bar{p}) \bullet (\bar{x} - \bar{p}) = 0$ oder koordinatenweise

$$\frac{\partial G}{\partial x}(x_0, y_0, z_0)(x - x_0) + \frac{\partial G}{\partial y}(x_0, y_0, z_0)(y - y_0) + \frac{\partial G}{\partial z}(x_0, y_0, z_0)(z - z_0) = 0.$$

Niveauflächen sind immer orientierbar; es gibt Innen $\{G < c\}$ und Außen $\{G > c\}$.

Wenn man eine Basis des Tangentialraumes braucht, sucht man sich beliebige zwei Vektoren, die linear unabhängig sind und auf dem Gradienten senkrecht stehen. Ist etwa $\vec{\text{grad}} G(\bar{p}) = \begin{pmatrix} a \\ b \\ c \end{pmatrix}$ mit $a, b \neq 0$, so kann man z.B.

$$\begin{pmatrix} -b \\ a \\ 0 \end{pmatrix} \quad \text{und} \quad \begin{pmatrix} a \\ b \\ c \end{pmatrix} \times \begin{pmatrix} -b \\ a \\ 0 \end{pmatrix} = \begin{pmatrix} -ac \\ -bc \\ a^2 + b^2 \end{pmatrix}$$

benutzen.

4.8.8 Abschweifung: LAGRANGE-Multiplikatoren aus geometrischer Sicht

Wir wollen die relativen Extremwerte der Zielfunktion $f(x, y, z)$ auf der Niveaufläche $\Gamma := \{(x, y, z) \in \mathbb{R}^3 : G(x, y, z) = 0\}$ finden. Beide Funktionen werden als C^1 vorausgesetzt.

Hintergrund ist die Beobachtung, daß in Extrempunkten alle Richtungsableitungen in Tangentialrichtungen verschwinden. Um das genauer auszuführen sei $\bar{p} \in \Gamma$ der Extrempunkt von f und $\vec{\text{grad}} G(\bar{p}) \neq \vec{0}$ (Regularitätsbedingung). Jeder Einheitsvektor $\vec{e} \in T_{\bar{p}}\Gamma$ läßt sich als $\dot{\bar{\gamma}}(0)$ schreiben, wobei $\bar{\gamma}$ eine in Γ verlaufende Kurve mit $\bar{\gamma}(0) = \bar{p}$ ist. Die Funktion $f(\bar{\gamma}(t))$ nimmt dann für $t = 0$ einen relativen Extremwert an, also verschwindet ihre Ableitung nach t :

$$0 = \left. \frac{d}{dt}f(\bar{\gamma}(t)) \right|_{t=0} = \vec{\text{grad}} f(\bar{\gamma}(0)) \bullet \dot{\bar{\gamma}}(0) = \vec{\text{grad}} f(\bar{p}) \bullet \vec{e} = \frac{\partial f}{\partial \vec{e}}(\bar{p}).$$

Da nun \vec{e} ein beliebiger Einheitsvektor aus $T_{\bar{p}}\Gamma$ war, folgt aus $\vec{\text{grad}} f(\bar{p}) \bullet \vec{e} = 0$, daß $\vec{\text{grad}} f(\bar{p})$ auf $T_{\bar{p}}\Gamma$ senkrecht steht. Das tun aber nur die Vielfachen des Gradienten von G , daher $\vec{\text{grad}} f(\bar{p}) = \lambda \vec{\text{grad}} G(\bar{p})$, für ein passendes λ .

4.8.9 Anhang: Vektorprodukt und GRAMSche Determinante

Im Laufe dieses und des nächsten Abschnittes werden das Vektorprodukt und die GRAMSche Determinante eine Rolle spielen. Zu Ihrer Bequemlichkeit habe ich das Wichtigste noch einmal kurz zusammengestellt.

Jede Meteorologin ist seit dem ersten Semester mit Vektorprodukten vertraut:

$$\begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} \times \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} = \begin{pmatrix} a_2 b_3 - a_3 b_2 \\ a_3 b_1 - a_1 b_3 \\ a_1 b_2 - a_2 b_1 \end{pmatrix}$$

Diese Formel läßt sich leicht merken, wenn man die rechte Seite formal als Determinante schreibt, in deren erster Spalte die kanonischen Basisvektoren stehen:

$$\begin{vmatrix} \vec{e}_1 & a_1 & b_1 \\ \vec{e}_2 & a_2 & b_2 \\ \vec{e}_3 & a_3 & b_3 \end{vmatrix}$$

Die wichtigsten **Eigenschaften** sind Ihnen sicher auch vertraut:

- (1) $(\alpha \vec{a} + \alpha' \vec{a}') \times \vec{b} = \alpha(\vec{a} \times \vec{b}) + \alpha'(\vec{a}' \times \vec{b})$ und $\vec{a} \times \vec{b} = -\vec{b} \times \vec{a}$
- (2) $\vec{a} \times \vec{b} = \vec{0} \iff \vec{a}, \vec{b}$ linear abhängig
- (3) Das Vektorprodukt steht senkrecht auf den Faktoren.
- (4) Seine Länge ist gleich dem Flächeninhalt des von den Faktoren aufgespannten Parallelogramms : $\|\vec{a} \times \vec{b}\| = \|\vec{a}\| \cdot \|\vec{b}\| \cdot \sin \angle(\vec{a}, \vec{b})$
- (5) $(\vec{a} \times \vec{b}) \times \vec{c} = -(\vec{b} \bullet \vec{c}) \vec{a} + (\vec{a} \bullet \vec{c}) \vec{b}$.
- (6) $(\vec{a} \times \vec{b}) \bullet \vec{c} = \det(\vec{a}, \vec{b}, \vec{c})$ (Spatprodukt)

Die GRAMSCHE IDENTITÄT. Wir brauchen noch eine andere Formel für die Norm des Vektorproduktes bzw. den Flächeninhalt des von den beiden Vektoren \vec{a} und \vec{b} im Raum(!) aufgespannten Parallelogramms.

$$\|\vec{a} \times \vec{b}\|^2 = \|\vec{a}\|^2 \cdot \|\vec{b}\|^2 - (\vec{a} \bullet \vec{b})^2 = \begin{vmatrix} \vec{a} \bullet \vec{a} & \vec{a} \bullet \vec{b} \\ \vec{b} \bullet \vec{a} & \vec{b} \bullet \vec{b} \end{vmatrix}.$$

Um das nachzurechnen, nennen wir den von \vec{a} und \vec{b} eingeschlossenen Winkel α . Dann folgt alles aus der obigen Formel (4) und einem Additionstheorem:

$$\|\vec{a} \times \vec{b}\|^2 = \|\vec{a}\|^2 \|\vec{b}\|^2 \sin^2 \alpha = \|\vec{a}\|^2 \|\vec{b}\|^2 (1 - \cos^2 \alpha) = \|\vec{a}\|^2 \|\vec{b}\|^2 - \underbrace{\|\vec{a}\|^2 \|\vec{b}\|^2 \cos^2 \alpha}_{(\vec{a} \bullet \vec{b})^2}.$$

4.9 Der Inhalt einer gekrümmten Oberfläche im Raum

Wir betrachten eine Oberfläche Γ im Raum und wollen ihr, wenn möglich, eine nicht-negative reelle Zahl zuordnen, die man mit einiger Plausibilität als Flächeninhalt ansehen kann. Speziell sollen Polyeder, d.h. aus lauter platten Polygonen (Dreiecke genügen) zusammengeklebte Oberflächen, die üblichen elementargeometrischen Flächeninhalte zugeteilt bekommen.

Bei gekrümmten Flächen besteht die Idee darin, diese durch Polyeder zu approximieren und die Polyederfläche als Approximation an die gesuchte Fläche zu betrachten. Diese ergibt sich dann durch einen Grenzübergang. Das Ausführen dieser Idee macht im Vergleich zum Verfahren bei Kurven, wo durch Polygonzüge approximiert wurde, erhebliche Schwierigkeiten. Ich will das nur auf heuristischem Niveau diskutieren.

4.9.1 Der Flächeninhalt als Integral

Es wird bequem sein, eine Bezeichnung für Dreiecke zu haben: $\Delta(\bar{a}, \bar{b}, \bar{c})$ ist das Dreieck mit den Eckpunkten $\bar{a}, \bar{b}, \bar{c}$ entweder in der u, v -Parameterebene oder im x, y, z -Raum in dem die Fläche Γ liegt. Wie schon früher ist $|\Delta(\bar{a}, \bar{b}, \bar{c})|$ der Flächeninhalt des Dreiecks, den wir ohne weiteres verstehen und berechnen können, auch wenn das Dreieck im Raum liegt.

Wir nehmen jetzt an, daß unsere Oberfläche eine stetig differenzierbare Parametrisierung Φ besitzt, die auf dem Parameterbereich B der u, v -Ebene definiert ist. Um nebensächlichen Problemen von vornherein aus dem Weg zu gehen, setzen wir B als Polygon voraus (etwa ein Rechteck). Dann läßt sich B bequem in Dreiecke zerlegen, die paarweise jeweils höchstens eine Seite gemeinsam haben: $B = \bigcup_{i=1}^N \Delta_i$ (man spricht von einer *Triangulierung*). Mittels Φ werden die Eckpunkte der Δ_i auf die Fläche transportiert. Sie bestimmen ein (räumliches) Dreieck, das ich Δ_i^Φ nennen werde. Zusammen bilden die ‘hochgehobenen’ Dreiecke eine Polyederfläche $\bigcup_{i=1}^N \Delta_i^\Phi$, die wir als Approximation an Γ auffassen. Ihr Flächeninhalt ist $\sum_{i=1}^N |\Delta_i^\Phi|$ und wir wollen überlegen, was passiert, wenn $N \rightarrow \infty$ geht und gleichzeitig die beteiligten Dreiecke Δ_i immer kleiner werden.

Die Idee besteht in der zunächst trivialen Umformung $\sum_{i=1}^N |\Delta_i^\Phi| = \sum_{i=1}^N \frac{|\Delta_i^\Phi|}{|\Delta_i|} \cdot |\Delta_i|$ und in der Interpretation der rechten Seite als RIEMANNSCHE Summe $\sum_{i=1}^N \rho(\bar{s}_i) |\Delta_i|$ für eine (noch zu bestimmende) Funktion $\rho: B \rightarrow \mathbb{R}$ und jeweils passend gewählte Punkte $\bar{s}_i \in \Delta_i$.

Bei dem oben beschriebenen Grenzübergang gehen die RIEMANNSCHE Summen dann in das Integral $\iint_B \rho(u, v) d(u, v)$ über. Die Funktion ρ ist eine Art Dichte. Ich werde sie als den zu Φ gehörenden *Verzerrungsfaktor* bezeichnen, was insofern nicht ganz korrekt ist, als daß dieser nicht konstant ist, sondern vom Punkt abhängt.

4.9.2 Die Bestimmung des Verzerrungsfaktors

wäre nun vorrangige Aufgabe. Ich will aber weder die Vorlesung noch das Skript damit belasten, sondern verweise auf die Literatur, bzw. die Ergänzungen und Zusätze im Netz. Dort habe ich sogar drei verschiedene Herleitungen skizziert. Ergebnis ist die Formel

$$|\Gamma| = \iint_B \left\| \frac{\partial \Phi}{\partial u}(u, v) \times \frac{\partial \Phi}{\partial v}(u, v) \right\| d(u, v).$$

Der Verzerrungsfaktor ist bemerkenswerterweise der Flächeninhalt des von den kanonischen Basisvektoren von $T_{\Phi(u,v)}\Gamma$ aufgespannten Parallelogramms, das sich (dreidimensionaler Zufall) als die Norm des oben ausgerechneten Standardnormalenvektors berechnen läßt.

4.9.3 Spezialfall Funktionsgraphen

Der **Verzerrungsfaktor** ergibt sich als Norm des in 4.8.6 ausgerechneten Normalenvektors

$$\sqrt{1 + \left(\frac{\partial g}{\partial x}\right)^2 + \left(\frac{\partial g}{\partial y}\right)^2}.$$

4.9.4 Alternative Form des Verzerrungsfaktors

Die Norm eines Vektorproduktes hatten wir im ersten Semester schon einmal durch eine andere Formel ausgedrückt (GRAMSCHE Identität), die im gegebenen Zusammenhang manchmal bequemer ist:

$$\left\| \frac{\partial \Phi}{\partial u} \times \frac{\partial \Phi}{\partial v} \right\| = \sqrt{\left(\frac{\partial \Phi}{\partial u} \bullet \frac{\partial \Phi}{\partial u}\right) \cdot \left(\frac{\partial \Phi}{\partial v} \bullet \frac{\partial \Phi}{\partial v}\right) - \left(\frac{\partial \Phi}{\partial u} \bullet \frac{\partial \Phi}{\partial v}\right)^2}$$

Die traditionelle Schreibweise hierfür ist $\sqrt{EG - F^2}$, wobei

$$E = \frac{\partial \Phi}{\partial u} \bullet \frac{\partial \Phi}{\partial u} = \left(\frac{\partial \Phi_1}{\partial u}\right)^2 + \left(\frac{\partial \Phi_2}{\partial u}\right)^2 + \left(\frac{\partial \Phi_3}{\partial u}\right)^2 \quad G = \frac{\partial \Phi}{\partial v} \bullet \frac{\partial \Phi}{\partial v} = \left(\frac{\partial \Phi_1}{\partial v}\right)^2 + \left(\frac{\partial \Phi_2}{\partial v}\right)^2 + \left(\frac{\partial \Phi_3}{\partial v}\right)^2$$

$$\text{und } F = \frac{\partial \Phi}{\partial u} \bullet \frac{\partial \Phi}{\partial v} = \frac{\partial \Phi_1}{\partial u} \cdot \frac{\partial \Phi_1}{\partial v} + \frac{\partial \Phi_2}{\partial u} \cdot \frac{\partial \Phi_2}{\partial v} + \frac{\partial \Phi_3}{\partial u} \cdot \frac{\partial \Phi_3}{\partial v}$$

gesetzt ist. Die zweite Variante hat gegenüber dem Vektorprodukt den Vorteil auch für zwei-dimensionale Flächen zu gelten, die in höherdimensionale Räume eingebettet sind: denn das von zwei Vektoren \vec{a} und \vec{b} aus \mathbb{R}^n aufgespannte Parallelogramm hat immer den Flächeninhalt $\sqrt{(\vec{a} \bullet \vec{a}) \cdot (\vec{b} \bullet \vec{b}) - (\vec{a} \bullet \vec{b})^2}$ (Wurzel aus der GRAMschen Determinante).

4.9.5 Der höhere Standpunkt

Die aus den beiden Spalten $\frac{\partial \Phi}{\partial u}$ und $\frac{\partial \Phi}{\partial v}$ der JACOBI-Matrix von Φ gebildete GRAMsche Determinante kann man auch als Determinante des Matrizenproduktes $(J_\Phi)^T \cdot J_\Phi$ schreiben. Dann wird der Verzerrungsfaktor zu $\sqrt{\det((J_\Phi)^T \cdot J_\Phi)}$. Das mag im Moment abgehoben erscheinen, ist aber von einem allgemeineren Standpunkt aus gesehen die richtige Form. Sie hat den Vorteil, auch zur ‘Volumen’berechnung von k -dimensionalen Gebilden im n -dimensionalen Raum zu taugen. Sei $B \subseteq \mathbb{R}^k$ der Parameterbereich und $\Phi : B \rightarrow \Gamma \subseteq \mathbb{R}^n$ die Parametrisierung eines derartigen Gebildes³², so wird für die ‘Volumen’berechnung von Γ gerade das Integral

$$\int_B \sqrt{\det(J_\Phi(\vec{u})^T \cdot J_\Phi(\vec{u}))} d\vec{u}$$

benutzt. Für $k = 2, n = 3$ haben wir das gerade diskutiert. Für $n = k$ ist die Matrix J_Φ quadratisch aus $\sqrt{\det(J_\Phi^T \cdot J_\Phi)}$ wird $\sqrt{\det J_\Phi^T \cdot \det J_\Phi} = \sqrt{(\det J_\Phi)^2} = |\det J_\Phi|$, der Korrekturfaktor aus der Transformationsformel. Für $k = 1$ und beliebiges n handelt es sich um eine Kurve $\Phi = \vec{\gamma}$, deren JACOBI-Matrix der Geschwindigkeitsvektor $\dot{\vec{\gamma}}$ ist. Aus $(J_\Phi)^T \cdot J_\Phi$ wird die 1×1 -Matrix mit dem einzigen Eintrag $\dot{\vec{\gamma}} \bullet \dot{\vec{\gamma}}$ und der Verzerrungsfaktor wird zu $\sqrt{\dot{\vec{\gamma}} \bullet \dot{\vec{\gamma}}} = \|\dot{\vec{\gamma}}\|$, wie bei der Berechnung der Kurvenlänge gelernt.

4.9.6 Beispiele

Wir wollen die Plausibilität der gefundenen Formel nachprüfen, indem wir einige Flächeninhalte ausrechnen. Zunächst für die **Sphäre vom Radius R** . Wir benutzen die Zerlegung in die nördliche und südliche Hemisphäre, die beide denselben Flächeninhalt haben. Es genügt also einen von ihnen zu berechnen: S_+ ist Graph der Funktion $g(x, y) = \sqrt{R^2 - x^2 - y^2}$ auf dem Definitionsbereich $B = \{(x, y) : x^2 + y^2 \leq R^2\}$. Die partiellen Ableitungen dieser Funktion sind

$$\frac{\partial g}{\partial x} = \frac{-x}{\sqrt{R^2 - x^2 - y^2}} \quad \text{und} \quad \frac{\partial g}{\partial y} = \frac{-y}{\sqrt{R^2 - x^2 - y^2}} \quad \text{und}$$

Daher der Flächeninhalt der oberen Halbkugel

$$\iint_B \sqrt{1 + \frac{x^2}{R^2 - x^2 - y^2} + \frac{y^2}{R^2 - x^2 - y^2}} d(x, y) = \iint_B R \frac{d(x, y)}{\sqrt{R^2 - x^2 - y^2}} = \dots = 2\pi R^2.$$

Die ... füllen Sie vielleicht selbst aus, durch Umwandlung in ein iteriertes Integral oder Einführung von Polarkoordinaten. (Haben Sie bemerkt, daß es sich um ein uneigentliches Integral handelt? Die Funktion g ist am Rand nicht ganz koscher.)

Nochmal die Späre vom Radius R : Diesmal wählen wir die Parametrisierung, die auch den Kugelkoordinaten zugrunde liegt (die ist zwar auf den Randpunkten des Parameterrechtecks nicht injektiv, aber die spielen für das Integral keine Rolle):

$$B = [-\pi, \pi] \times [0, \pi] \quad \text{und} \quad \Phi(\varphi, \theta) = (R \cos \varphi \sin \theta, R \sin \varphi \sin \theta, R \cos \theta).$$

³²Auf Feinheiten kommt es hier nicht an.

Als Basisvektoren des Tangentialraumes ergeben sich

$$\frac{\partial \Phi}{\partial \varphi} = \begin{pmatrix} -R \sin \varphi \sin \theta \\ R \cos \varphi \sin \theta \\ 0 \end{pmatrix}, \quad \frac{\partial \Phi}{\partial \theta} = \begin{pmatrix} R \cos \varphi \cos \theta \\ R \sin \varphi \cos \theta \\ -R \sin \theta \end{pmatrix},$$

daraus $E = R^2 \sin^2 \theta$, $G = R^2$, $F = 0$ und der Verzerrungsfaktor

$$\sqrt{EG - F^2} = \sqrt{R^4 \sin^2 \theta} = R^2 \sin \theta,$$

denn der Sinus ist für die zugelassenen Thetas immer positiv Also ist der Flächeninhalt der ganzen Sphäre jetzt

$$\int_0^{2\pi} \int_0^\pi R^2 \sin \theta \, d\theta \, d\varphi = 2\pi R^2 \int_0^\pi \sin \theta \, d\theta = 4\pi R^2,$$

wie es sich gehört.

Ein **Torus** ist anschaulich ein aufgeblasener Autoreifen. Seine Größe wird bestimmt durch zwei Radien $r < R$. Dabei soll r der Radius des Schlauches sein und R der Radius des in der Mitte des Schlauches gedachten Kreises (der sogenannten Seele des Torus). Aus dieser Beschreibung ergibt sich leicht die folgende Parametrisierung

$$x = (R + r \cos \varphi) \cos \psi, \quad y = (R + r \cos \varphi) \sin \psi, \quad z = r \sin \varphi$$

wobei $(\varphi, \psi) \in [0, 2\pi] \times [0, 2\pi]$. Wieder ist das nicht ganz injektiv, denn die Ränder des Parameterrechtecks werden miteinander verklebt. Aber das ist für das Integral unerheblich. Als Tangentialvektoren ergeben sich

$$\frac{\partial}{\partial \varphi} = \begin{pmatrix} -r \sin \varphi \cos \psi \\ -r \sin \varphi \sin \psi \\ r \cos \varphi \end{pmatrix}, \quad \frac{\partial}{\partial \psi} = \begin{pmatrix} -(R + r \cos \varphi) \sin \psi \\ (R + r \cos \varphi) \cos \psi \\ 0 \end{pmatrix},$$

daher

$$E = \frac{\partial}{\partial \varphi} \bullet \frac{\partial}{\partial \varphi} = r^2, \quad G = \frac{\partial}{\partial \psi} \bullet \frac{\partial}{\partial \psi} = (R + r \cos \varphi)^2, \quad F = \frac{\partial}{\partial \varphi} \bullet \frac{\partial}{\partial \psi} = 0.$$

Der Flächeninhalt des Torus ist gleich dem Integral $\int_0^{2\pi} \int_0^{2\pi} \sqrt{EG - F^2} \, d\varphi \, d\psi$

$$\int_0^{2\pi} \int_0^{2\pi} |r(R + r \cos \varphi)| \, d\varphi \, d\psi = 2\pi r \int_0^{2\pi} (R + r \cos \varphi) \, d\varphi = 2\pi r 2\pi R = 4\pi^2 r R$$

(wegen $R > r$ ist $R + r \cos \varphi$ immer positiv).

4.10 Oberflächenintegrale

gibt es zwei Typen je nachdem ob Skalar- oder Vektorfelder integriert werden sollen. Es besteht eine Analogie zu den zwei Typen von Kurvenintegralen $\int_\gamma f \, ds$ bzw. $\int_\gamma \vec{f} \bullet d\vec{s}$. Wir beginnen mit der Integration von Skalarfeldern.

4.10.1 Wie man zum Integral kommt

Γ sei eine Oberfläche in \mathbb{R}^3 und $f : \Gamma \rightarrow \mathbb{R}$ eine darauf definierte skalare Funktion, die integriert werden soll. Man macht das Übliche. Die Fläche Γ wird in viele kleine paarweise höchstens am Rand überlappende Teilflächen zerlegt. Dann bildet man RIEMANNsche Summen, indem jeweils ein Funktionswert $f(\bar{p})$ der auf dem Teilstück angenommen wird mit dem Flächeninhalt des Teilstücks multipliziert wird und dann über alle Teilstücke aufsummiert. Wenn diese Summen bei immer feiner werdenden Zerlegungen der Fläche Γ gegen einen Grenzwert streben und dieser nicht von der Art und Weise der Zerlegung abhängt, dann nennt man ihn Integral der Funktion f über die Oberfläche Γ und schreibt $\int_\Gamma f$. Zur Betonung wird manchmal auch etwas wie $\iint_\Gamma f(x, y, z) \, dF$ geschrieben.

4.10.2 Anforderungen an die Fläche

Wenn schon bei der Integration über platte zweidimensionale Gebiete Anforderungen an die Fläche (hauptsächlich ihren Rand) gestellt werden müssen, so ist das bei Oberflächen umso mehr der Fall. Wie bei anderen Arten von Integralen werden wir erlauben zu stückeln. Die Integrale sollen sich stückweise ausrechnen lassen und werden dann addiert. Also kommt es darauf an, sich mit der Definition des Integrals auf einem geeigneten Flächenstück zu befassen.

Die Flächenstücke sollen so sein, wie im vorigen Abschnitt besprochen, also eine Parametrisierung $\Phi : B \rightarrow \Gamma$ haben mit einem stetig differenzierbaren Φ , dessen JACOBI-Matrix in jedem Punkt von B Rang 2 hat. Der ebene Parameterbereich B soll jetzt gutartig sein, so daß die normale Integration stetiger Funktionen über B jedenfalls keine Schwierigkeiten macht.

Inoffizielle Zwischenbemerkung. Hinsichtlich der im vorigen Abschnitt geforderten Injektivität der Parametrisierung Φ sind wir nur in der Theorie streng. In der Praxis ist es nicht so schlimm, wenn manche Punkte aus Γ viele Urbilder haben. Wichtig ist nur, daß die Menge der verklebten Punkte in B dünn ist (etwa eine Vereinigung von endlich vielen Strecken) und beim Integrieren über B nicht ins Gewicht fällt. Ähnlich großzügig können wir sein, was den Rang der JACOBI-Matrix angeht. Offiziell werden wir darauf bestehen, daß er überall zwei ist. Wenn der Rang aber auf einer dünnen Menge von Punkten aus B kleiner als 2 ist, so merken die Integrale das gar nicht.

Diese Schummelei ist in der Praxis bequem, weil sie erlaubt mit einem oder doch wenigen Flächenstücken auszukommen. Um die reine Lehre zu retten, kann man die Flächen meist in mehrere (viele) Stücke zerschneiden, für die alle Bedingungen erfüllt sind.

4.10.3 Definition

Nun stellen wir uns vor, daß die in 4.10.1 beschriebene Zerlegung von Γ dadurch entsteht, daß eine Zerlegung von B mittels Φ auf die Fläche hochgehoben wird. Wenn die dazu benötigten Teile von B nur klein genug sind, so unterscheidet sich der hochgehobene Flächeninhalt von dem seines Urbilds in der Ebene um einen Verzerrungsfaktor, den wir im vorigen Abschnitt ausführlich diskutiert hatten.

Setzt man diesen Faktor ein, so verwandelt sich die Integralsumme der Funktion f über Γ in eine Integralsumme einer korrigierten Funktion über B . Das entsprechende Integral wird als Oberflächenintegral von f definiert.

Formale Definition.

$$\iint_{\Gamma} f(x, y, z) dF = \iint_B f(\Phi(u, v)) \cdot \left\| \frac{\partial \Phi}{\partial u} \times \frac{\partial \Phi}{\partial v} \right\| d(u, v) = \iint_B f(\Phi(u, v)) \cdot \sqrt{EG - F^2} d(u, v),$$

wobei

$$E = \frac{\partial \Phi}{\partial u} \bullet \frac{\partial \Phi}{\partial u}, \quad G = \frac{\partial \Phi}{\partial v} \bullet \frac{\partial \Phi}{\partial v} \quad \text{und} \quad F = \frac{\partial \Phi}{\partial u} \bullet \frac{\partial \Phi}{\partial v}.$$

Dieser Zugang hat den Nachteil, daß das definierte Integral zunächst von der Parametrisierung abhängt. Man muß nachweisen, daß in Wahrheit jede Parametrisierung denselben Wert liefert. Wir werden das nicht tun.

Den formalen Ausdruck $\left\| \frac{\partial \Phi}{\partial u} \times \frac{\partial \Phi}{\partial v} \right\| d(u, v)$ bzw. $\sqrt{EG - F^2} d(u, v)$ werden wir *Oberflächenelement* nennen und als dF notieren (so wie ds Standardnotation für das Bogenelement war).

4.10.4 Wenn Γ ein Funktionsgraph ist

wird die Formel für das Oberflächenintegral einfacher. Angenommen

$$\Gamma = \{(x, y, g(x, y)) : (x, y) \in B\}.$$

Dann haben wir den Verzerrungsfaktor in 4.9.3 ausgerechnet und setzten ihn nun ein:

$$\iint_{\Gamma} f(x, y, z) dF = \iint_B f(x, y, g(x, y)) \cdot \sqrt{1 + \left(\frac{\partial g}{\partial x}\right)^2 + \left(\frac{\partial g}{\partial y}\right)^2} d(x, y).$$

Klassisch schreibt man oft nur $z = z(x, y)$. Dann liest sich die Formel

$$\iint_{\Gamma} f(x, y, z) dF = \iint_B f(x, y, z) \cdot \sqrt{1 + \left(\frac{\partial z}{\partial x}\right)^2 + \left(\frac{\partial z}{\partial y}\right)^2} d(x, y).$$

4.10.5 Skalare Oberflächenintegrale von Vektorfeldern

werden auch **Flussintegrale** genannt (auch das Wort ‘Durchsatz’ an Stelle von Fluß habe ich gelesen). Als motivierendes Beispiel denken wir an das Geschwindigkeitsfeld $\vec{v}(\vec{x})$ einer stationär strömenden Flüssigkeit. In die Strömung hängen wir irgendwie eine volldurchlässige Membran (was immer das ist; eher ein Sieb) mit bestimmter Oberfläche Γ . Die Frage ist, wieviel Flüssigkeit pro Zeiteinheit durch diese Fläche strömt.

Wäre die Geschwindigkeit konstant und die Fläche senkrecht zur Strömungsrichtung aufgehängt, so wäre die Antwort: $\|\vec{v}\| \cdot \text{Flächeninhalt}$. Ist die Geschwindigkeit konstant, die Fläche platt aber geneigt zur Strömung, so muß man statt \vec{v} die Komponente senkrecht zur Fläche benutzen. Schreiben wir \vec{n} für den Normaleneinheitsvektor der Fläche (nach der Strömungsseite), so ist die Flüssigkeitsmenge

$$\|\vec{v}\| \cdot \cos \angle(\vec{v}, \vec{n}) \cdot \text{Flächeninhalt} = \vec{v} \cdot \vec{n} \cdot \text{Flächeninhalt}.$$

Bei nicht konstantem \vec{v} und gekrümmtem Γ zerlegt man die Fläche in viele kleine Plättchen, am Ende kommt ein Oberflächenintegral $\iint_{\Gamma} \vec{v} \cdot \vec{n} dF$ heraus.

Im Gegensatz zum skalaren Kurvenintegral wo das Vektorfeld auf die Tangentenrichtung projiziert wurde, nimmt man hier die Normalenrichtung. So wie man bei der Kurve einen Umlaufsinn festlegen mußte, von dem das Vorzeichen des Integrals abhing, so muß man auch auf der Oberfläche eine Orientierung festlegen. Technisch geschieht das durch Auswahl eines stetigen Feldes aus Normaleneinheitsvektoren. Flußintegrale lassen sich daher nur für zweiseitige Oberflächen berechnen.

Da das bei den Physikern sehr beliebt ist, werde auch ich ein ‘vektorielles Oberflächenelement’ $d\vec{F}$ einführen (eigentlich wäre die Bezeichnung $d\vec{F}$ angemessener); es ist definiert als $\vec{n} dF$ hat also den Betrag des skalaren Oberflächenelementes dF und die Richtung des ausgewählten Normalenfeldes. Mit dieser Bezeichnung sieht das Flußintegral dann so aus:

$$\iint_{\Gamma} \vec{f} \cdot d\vec{F}.$$

In der klassischen Schreibweise hat das (dreidimensionale) Vektorfeld, dessen Flußintegral berechnet wird die Koordinaten P, Q, R . Die Koordinaten des ausgewählten Normaleneinheitsvektors \vec{n} , die ich etwas nichtssagend n_1, n_2, n_3 schreiben werde, heißen in vielen älteren Lehrbüchern ‘Richtungskosinus’ $\cos \angle(n, x), \cos \angle(n, y), \cos \angle(n, z)$. Gemeint sind die Winkel zwischen der ausgewählten Normalenrichtung und der positiven x -, y -, z -Richtung.

Das entsprechende Integral könnte also in klassischer Schreibweise so aussehen:

$$\iint_{\Gamma} (P \cos(n, x) + Q \cos(n, y) + R \cos(n, z)) dF.$$

Will man das Flussintegral $\iint_{\gamma} \vec{f} \cdot \vec{n} dF$ mit Hilfe einer Parametrisierung $\Phi : B \rightarrow \mathbb{R}^3$ ausrechnen, so ergeben sich gewisse Vereinfachungen. Der Normaleneinheitsvektor \vec{n} entsteht als $\pm \frac{\vec{N}}{\|\vec{N}\|}$, wobei $\vec{N} = \frac{\partial \Phi}{\partial u} \times \frac{\partial \Phi}{\partial v}$ der ‘kanonische’ Normalenvektor ist. Die Norm von \vec{N} taucht gleichzeitig als Verzerrungsfaktor auf, läßt sich also herauskürzen:

$$\iint_{\Gamma} \vec{f} \cdot d\vec{F} = \iint_B \vec{f}(\Phi(u, v)) \cdot \vec{n}(u, v) \|\vec{N}(u, v)\| d(u, v) = \pm \iint_B \vec{f}(\Phi(u, v)) \cdot \vec{N}(u, v) d(u, v).$$

Das Vorzeichen richtet sich danach, ob der kanonisch gegebene Normalenvektor \vec{N} in dieselbe Richtung zeigt, wie die für den Fluß ausgesuchte Einheitsnormale \vec{n} .

Beachtet man noch, daß \vec{N} ein Vektorprodukt ist, so steht unter dem letzten Integralzeichen ein Spatprodukt, das als Determinante geschrieben werden kann:

$$\iint_{\Gamma} \vec{f} \bullet d\vec{F} = \pm \iint_B \det \left(\vec{f}(\Phi(u, v)), \frac{\partial \Phi}{\partial u}(u, v), \frac{\partial \Phi}{\partial v}(u, v) \right) d(u, v)$$

Das ist manchmal nützlich manchmal auch schwerfälliger.

4.10.6 Beispiel

Wir wollen den Fluß $\iint_E \vec{f} \bullet d\vec{F}$ des Vektorfeldes $\vec{f}(x, y, z) = \begin{pmatrix} 1/x \\ 1/y \\ 1/z \end{pmatrix}$ berechnen, wobei \vec{n} die

äußere Einheitsnormale und E das Ellipsoid mit der Gleichung $\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1$ ist. Es entsteht aus der Späre, indem in den drei Koordinatenrichtungen jeweils gestreckt bzw gestaucht wird. Also entsteht die Parametrisierung aus der oben besprochenen Parametrisierung der Spähre, indem in x, y und z jeweils mit den passenden Faktoren multipliziert werden:

$$B = [-\pi, \pi] \times [0, \pi] \quad \text{und} \quad \Phi(\varphi, \theta) = (a \cos \varphi \sin \theta, b \sin \varphi \sin \theta, c \cos \theta).$$

Als Basisvektoren des Tangentialraumes ergeben sich

$$\frac{\partial \Phi}{\partial \varphi} = \begin{pmatrix} -a \sin \varphi \sin \theta \\ b \cos \varphi \sin \theta \\ 0 \end{pmatrix}, \quad \frac{\partial \Phi}{\partial \theta} = \begin{pmatrix} a \cos \varphi \cos \theta \\ b \sin \varphi \cos \theta \\ -c \sin \theta \end{pmatrix},$$

Weiter ist

$$\frac{\partial \Phi}{\partial \varphi} \times \frac{\partial \Phi}{\partial \theta} = \begin{pmatrix} -bc \cos \varphi \sin^2 \theta \\ -ac \sin \varphi \sin^2 \theta \\ -ab [\cos \theta \sin \theta \sin^2 \varphi + \cos \theta \sin \theta \cos^2 \varphi] \end{pmatrix} = - \begin{pmatrix} bc \cos \varphi \sin^2 \theta \\ ac \sin \varphi \sin^2 \theta \\ ab \cos \theta \sin \theta \end{pmatrix}.$$

Um einzusehen, ob es sich um die innere oder äußere Normale handelt (die andere würde sich nur im Vorzeichen unterscheiden), betrachten wir den Punkt mit den Parametern $\varphi = 0, \theta = \pi/2$.

Das ist der Punkt $(a, 0, 0)$, also der Extrempunkt auf der x -Achse. Dort ist $\frac{\partial \Phi}{\partial \varphi} \times \frac{\partial \Phi}{\partial \theta} = - \begin{pmatrix} bc \\ 0 \\ 0 \end{pmatrix}$,

zeigt also zum Nullpunkt hin, d.h. nach innen. Um die äußere Normale zu bekommen, wird das obige Minuszeichen weggelassen.

Dann wird der Fluß gleich $\iint_B \vec{f}(x(\varphi, \theta), y(\varphi, \theta), z(\varphi, \theta)) \bullet \vec{N}(\varphi, \theta) d(\varphi, \theta)$

$$\begin{aligned} &= \int_{-\pi}^{\pi} \int_0^{\pi} \frac{bc \cos \varphi \sin^2 \theta}{a \cos \varphi \sin \theta} + \frac{ac \sin \varphi \sin^2 \theta}{b \sin \varphi \sin \theta} + \frac{ab \cos \theta \sin \theta}{c \cos \theta} d\theta d\varphi \\ &= 2\pi \left(\frac{bc}{a} + \frac{ac}{b} + \frac{ab}{c} \right) \int_0^{\pi} \sin \theta d\theta = 4\pi \left(\frac{bc}{a} + \frac{ac}{b} + \frac{ab}{c} \right). \end{aligned}$$

Allerdings ist das Vektorfeld \vec{f} gar nicht auf der ganzen Oberfläche definiert, was die obige Rechnung etwas anrühlich macht. Wir haben das deshalb nicht bemerkt, weil $\vec{f} \cdot \vec{n}$ eine stetige Fortsetzung auf die ganze Fläche hat (wegkürzen der Nenner unter dem Integral).

5 Integralsätze

setzen das Integral von etwas über einem höherdimensionalen Gebilde mit dem Integral von etwas verwandtem über dem niedrigerdimensionalen Rand dieses Gebildes in Beziehung. Vorbild, wenn auch leicht entartet, ist der Hauptsatz der Differential- und Integralrechnung in der Form

$$\int_a^b f'(t) dt = f(t) \Big|_a^b.$$

Hier ist die Strecke $[a, b]$ das eindimensionale Gebilde und die beiden Punkte a und b bilden den nulldimensionalen Rand. Die beiden verwandten ‘Etwasse’, die ‘integriert’ werden, sind f und f' .

Die moderne Mathematik behandelt die Integralsätze dieses Abschnittes in einem anderen Rahmen, als wir das tun werden. Es werden Differentialformen über Mannigfaltigkeiten bzw. deren Ränder integriert. Am eingängigsten können Sie das bei JÄNICH nachlesen. Wissenschaftlicher sieht es bei KERNER/VON WAHL aus. Dieses Herangehen ist gleichzeitig mathematisch exakter und allgemeiner als die folgenden Betrachtungen.

Die exakte Theorie bräuchte eine ganze Latte neuer ‘Vokabeln und Grammatik’, um überhaupt zu sagen, was man will. Noch mehr Aufwand ist für die Beweise der entsprechenden Sätze nötig. Das wollte ich Ihnen und mir ersparen.

Ich will aber noch eine schicke Bezeichnung erwähnen, die oft auch dann benutzt wird, wenn die zugehörige Theorie gar nicht erklärt wird. Ist A eine ebene Fläche, eine Oberfläche (Haube) im Raum oder ein dreidimensionaler Körper (exakt: eine Mannigfaltigkeit), so wird ihr Rand mit ∂A bezeichnet. Stillschweigend wird vorausgesetzt, daß A mit einer Parametrisierung (exakt: Atlas) kommt, aus der sich kanonisch eine Parametrisierung des Randes ergibt. Die Integralsätze liefern dann Beziehungen zwischen Integralen über A bzw. ∂A und könnten als Formeln so aussehen:

$$\int_A \operatorname{rot} \vec{f} \bullet d\vec{F} = \int_{\partial A} \vec{f} \bullet d\vec{s} \quad \text{oder} \quad \int_A \operatorname{div} \vec{f} \, dV = \int_{\partial A} \vec{f} \bullet d\vec{F}.$$

Meine Darstellung (derselben Sätze!) ist etwas hausbacken und benutzt den Randoperator ∂ nicht, sondern gibt den Randkurven bzw. -flächen eigene Namen.

5.1 Die GREENSche Formel (in der Ebene)

Die³³ GREENSche Formel ist der nach dem eindimensionalen Hauptsatz nächst einfache Fall. Damit wir uns um die Existenz der Integrale keine Sorgen machen müssen, setzen wir Stetigkeit und Glattheit auch dort voraus, wo man eventuell mit weniger auskäme.

5.1.1 Die Situation

Gegeben sei eine ebene, stückweise glatte, geschlossene JORDAN-Kurve (d.h. bis auf Anfangs- und Endpunkt doppeltpunktfrei) $\bar{\gamma}$. Sie berandet einen kompakten Teil G der Ebene (und einen unbeschränkten Teil, für den wir uns im Moment nicht interessieren). Diese auf den ersten Blick völlig einleuchtende (weil man viel zu einfache Kurven vor Augen hat) Aussage ist in Wahrheit schwierig zu beweisen und heißt JORDANScher Kurvensatz (Glattheit wird dafür nicht gebraucht). Wir gehen einfach davon aus, daß das so ist. Wegen der Glätte von $\bar{\gamma}$ ist G gutartig. Hinreichend stetige Funktionen lassen sich also über G integrieren und stetige Vektorfelder auch entlang $\bar{\gamma}$.

5.1.2 Die klassische Formulierung der GREENSchen Formel

geht von zwei C^1 -Funktionen P und Q aus, die auf einer G umfassenden offenen Menge definiert sein sollen. Für diese gilt

$$\iint_G \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) d(x, y) = \oint_{\bar{\gamma}} P \, dx + Q \, dy.$$

Dabei ist der Umlaufsinn von $\bar{\gamma}$ so festgelegt, daß der Bereich G immer auf der Fahrerseite (links der Kurve) bleibt.

³³Der bestimmte Artikel ist mit Vorsicht zu genießen: vom Herrn GREEN gibt es noch mehr Formeln (s. später in diesem Kapitel). Auch hat die Formel dieses Abschnittes bei fast jedem Lehrbuchautor einen anderen Namen. Zum Beispiel habe ich auch die Bezeichnungen RIEMANNSche und GAUSSSche Formel gefunden. Andere Autoren sprechen von der STOKESSchen Formel in der Ebene.

Tatsächlich gelten zwei asymmetrische Formeln, deren Differenz die oben hingeschriebene ergibt. Weil es dann übersichtlicher wird, werden wir es auch so beweisen. Sie lassen sich aus der symmetrischen Variante zurückgewinnen, indem einmal P und dann Q Null gesetzt werden:

$$\iint_G \frac{\partial Q}{\partial x} d(x, y) = \oint_{\bar{\gamma}} Q dy \quad \text{und} \quad \iint_G \frac{\partial P}{\partial y} d(x, y) = - \oint_{\bar{\gamma}} P dx.$$

Vielleicht stutzen Sie wegen des Minuszeichens. x, y sowie P, Q sind doch gleichberechtigt wieso kann man sie dann nicht einfach vertauschen? Das liegt am festgelegten Umlaufsinn: **von** positiver x -Richtung **zu** positiver y Richtung. Damit ist die Symmetrie zwischen x und y zerstört. Wird die Kurve andersherum durchlaufen, so steht das Minuszeichen in der anderen Gleichung.

5.1.3 Beweis für einfach gutartiges G

Wir setzen zunächst voraus, daß G eine Beschreibung der Art

$$G = \{(x, y) : a \leq x \leq b \text{ und } g(x) \leq y \leq h(x)\}$$

hat. Dabei sollen g, h stetig differenzierbar (stückweise genügt) sein und können zur stückweisen Parametrisierung der Randkurve benutzt werden (Skizze!). Die Randkurve $\bar{\gamma}$ zerfällt in vier Teile mit den Parametrisierungen

$$\begin{array}{ll} \bar{\gamma}_1 & x \mapsto (x, g(x)) \quad x \in [a, b] \\ \bar{\gamma}_2 & y \mapsto (b, y) \quad y \in [g(b), h(b)] \\ \bar{\gamma}_3 & x \mapsto (x, h(x)) \quad x \in [a, b] \\ \bar{\gamma}_4 & y \mapsto (a, y) \quad y \in [g(a), h(a)] \end{array}$$

Dabei müssen $\bar{\gamma}_3$ und $\bar{\gamma}_4$ rückwärts durchlaufen werden: $\bar{\gamma} = \bar{\gamma}_1 \bar{\gamma}_2 \bar{\gamma}_3^- \bar{\gamma}_4^-$. Benutzt man diese Parametrisierungen, so wird die Identität $\iint_G \frac{\partial P}{\partial y} d(x, y) = - \oint_{\bar{\gamma}} P dx$ zu

$$\begin{aligned} & \int_a^b \left(\int_{g(x)}^{h(x)} \frac{\partial P}{\partial y} dy \right) dx \\ &= - \left[\int_a^b P(x, g(x)) \cdot 1 dx + \int_{g(b)}^{h(b)} P(b, y) \cdot 0 dy - \int_a^b P(x, h(x)) \cdot 1 dx - \int_{g(a)}^{h(a)} P(a, y) \cdot 0 dy \right] \end{aligned}$$

Auf beiden Seiten (links nach Hauptsatz) bleibt $\int_a^b [P(x, h(x)) - P(x, g(x))] dx$, also dasselbe. Mit der zweiten Identität wird es schwieriger, weil die Parametrisierungen dem Problem schlechter angepaßt sind. Rechnet man erst einmal analog zum gerade erfolgreichen Vorgehen, so wird aus der angestrebten Beziehung $\iint_G \frac{\partial Q}{\partial x} d(x, y) = \oint_{\bar{\gamma}} Q dy$ die Gleichung

$$\begin{aligned} & \int_a^b \left(\int_{g(x)}^{h(x)} \frac{\partial Q}{\partial x} dy \right) dx \\ &= \int_a^b Q(x, g(x)) \cdot g'(x) dx + \int_{g(b)}^{h(b)} Q(b, y) \cdot 1 dy - \int_a^b Q(x, h(x)) \cdot h'(x) dx - \int_{g(a)}^{h(a)} Q(a, y) \cdot 1 dy. \end{aligned}$$

Auf den ersten Blick ist keine Vereinfachung abzusehen, mit deren Hilfe wir die Gleichheit feststellen könnten. Der Trick besteht im Heranziehen der Formel 4.1.6:

$$\frac{d}{dx} \left(\int_{g(x)}^{h(x)} Q(x, y) dy \right) = \int_{g(x)}^{h(x)} \frac{\partial Q}{\partial x}(x, y) dy + Q(x, h(x)) \cdot h'(x) - Q(x, g(x)) \cdot g'(x).$$

Integriert man beide Seiten $\int_a^b \dots dx$, so entsteht links (nach Hauptsatz)

$$\int_a^b \frac{d}{dx} \left(\int_{g(x)}^{h(x)} Q(x, y) dy \right) dx = \int_{g(b)}^{h(b)} Q(b, y) dy - \int_{g(a)}^{h(a)} Q(a, y) dy.$$

Rechts ergibt sich

$$\int_a^b \int_{g(x)}^{h(x)} \frac{\partial Q}{\partial x}(x, y) dy dx + \int_a^b Q(x, h(x)) \cdot h'(x) dx - \int_a^b Q(x, g(x)) \cdot g'(x) dx.$$

Nach Gleichsetzen und Umstellen, ist das die verlangte Gleichung. Damit ist die GREENSche Formel für einfach gutartige Mengen bewiesen, denn der symmetrische Fall

$$G = \{(x, y) : c \leq y \leq d \text{ und } u(y) \leq x \leq v(y)\}.$$

geht ganz analog.

5.1.4 Die Rückführung komplizierterer Bereiche auf einfache

passiert durch Zerschneiden. Ist G nicht so einfach wie oben beschrieben, so läßt sich die Menge vielleicht durch Schnitte in endlich viele einfache Teile zerlegen $G_1 \cup G_2 \cup \dots \cup G_k$, auf die dann einzeln die obige Formel anwendbar ist $\iint_{G_i} = \oint_{\bar{\gamma}_i}$. Summiert man alle diese Gleichungen so entsteht links das Integral über die Gesamtfläche und rechts die Summe über die Kurvenintegrale der einzelnen Randkurven

$$\iint_G = \sum_{i=1}^k \oint_{\bar{\gamma}_i}.$$

Die Randkurven $\bar{\gamma}_i$ der Teilmengen G_i sind selbst zusammengestückt aus Teilen, die zum Rand von G gehören (und sich insgesamt zu $\bar{\gamma}$ zusammenstücken) und Strecken, die beim Zerschneiden entstanden sind. Letztere kommen jeweils zweimal vor als gemeinsamer Teil des Randes von G_i und G_j . Da der Rand beider Teilflächen jeweils im positiven Sinn durchlaufen wird, muß die gemeinsame Randstrecke dabei einmal hin und einmal zurück durchlaufen werden. Die diesen Durchläufen entsprechenden Kurvenintegrale haben entgegengesetzte Vorzeichen, heben sich also in der Summe $\sum_{i=1}^k \oint_{\bar{\gamma}_i}$ auf. Dann bleibt aber von der großen Summe nach Wegheben und Zusammenfassen wirklich nur $\oint_{\bar{\gamma}}$ übrig.

Machen Sie sich das ganze Argument an einer Skizze klar! Legen Sie zunächst nur einen Schnitt!

5.1.5 Gebiete mit Löchern

sind in der obigen Situationsbeschreibung nicht erfaßt, da wir von einer einfach geschlossenen Randkurve $\bar{\gamma}$ ausgegangen waren. Die GREENSche Formel gilt aber auch für Gebieten mit endlich vielen Löchern (denken Sie an eine Scheibe Schweizer Käse), die wir am besten als $G = A \setminus (L_1 \cup L_2 \cup \dots \cup L_p)$ beschreiben, wobei A und die L_i solche Gebiete wie bisher sind, jeweils mit linksherum orientierten einfach geschlossenen Randkurven $\bar{\alpha}$ bzw. $\bar{\lambda}_i$. Alle Löcher L_i sollen dabei im Inneren von A enthalten sein und sich gegenseitig nicht berühren.

Der Rand von G zerfällt jetzt in den Rand von A , der durch $\bar{\alpha}$ richtig parametrisiert wird, und die Ränder der Löcher. Da $\bar{\lambda}_i$ das Loch L_i so umrundet, daß es links bleibt, muß man umdrehen, wenn man den Käse auf der Fahrerseite haben will. Als Randstück von G muß also $\bar{\lambda}_i^-$ benutzt werden. Im ursprünglichen Sinne der Definition, handelt es sich beim Rand von G gar nicht um eine Kurve. Das Kurvenintegral über derartige 'unzusammenhängende' Kurven kann aber trotzdem sinnvoll definiert werden: es ist natürlich als Summe der Kurvenintegrale der zusammenhängenden Teilstücke zu verstehen.

Wenn wir davon ausgehen, daß die Funktionen P und Q stetig differenzierbar auf ganz A definiert sind oder sich wenigstens fortsetzen lassen (eigentlich brauchen sie nur in einer offenen Menge erklärt sein, die G enthält), dann erhalten wir die GREENSche Formel für G durch einfache Subtraktion:

$$\iint_G = \iint_A - \sum_{i=1}^p \iint_{L_i} = \oint_{\bar{\alpha}} - \sum_{i=1}^p \oint_{\bar{\lambda}_i} = \oint_{\bar{\alpha}} + \sum_{i=1}^p \oint_{\bar{\lambda}_i^-} = \oint_{\bar{\gamma}}.$$

Wenn Sie lieber nicht an die Fortsetzbarkeit von C^1 -Funktionen glauben wollen, müssen Sie sich überlegen, daß man das im vorigen Abschnitt benutzte Zerschneiden so organisieren kann, daß die Löcher dabei aufgeschnitten werden (Käsescheibe vorstellen!). Die Teilstücke von G sind dann alle einfach berandet ohne Löcher und dasselbe Argument wie oben greift.

5.1.6 Ein Spezialfall: Flächeninhalt als Kurvenintegral

Setzt man in der GREENSchen Formel

$$\iint_G \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) d(x, y) = \oint_{\bar{\gamma}} P dx + Q dy.$$

$Q = x$ und $P = -y$ so ergibt sich

$$\iint_G (1 + 1) d(x, y) = \oint_{\bar{\gamma}} -y dx + x dy.$$

Links steht der doppelte Flächeninhalt von G , also

$$|G| = \frac{1}{2} \oint_{\bar{\gamma}} -y dx + x dy.$$

falls G von der stückweise glatten Kurve $\bar{\gamma}$ berandet wird. Alternativ kann man auch $Q = x$ und $P = 0$ bzw. $Q = 0$ und $P = -y$ setzen. Dann entstehen asymmetrische Flächenformeln

$$|G| = \oint_{\bar{\gamma}} x dy = - \oint_{\bar{\gamma}} y dx.$$

Beispiel. Die Gleichung einer zentral gelegenen Ellipse mit den Halbachsen a und b lautet $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$. Die von dieser Ellipse berandete Fläche, nennen wir sie E , ist einfach gutartig, denn

$$E = \left\{ (x, y) : -a \leq x \leq a; -\frac{b}{a} \sqrt{a^2 - x^2} \leq y \leq \frac{b}{a} \sqrt{a^2 - x^2} \right\}.$$

Der Flächeninhalt ist das Integral

$$|E| = \iint_E 1 d(x, y) = \int_{-a}^a \frac{2b}{a} \sqrt{a^2 - x^2} dx,$$

das man mit BRONSTEINS Hilfe wohl ausrechnen kann. Die Randkurve wird von $\bar{\varepsilon}(t) = (a \cos t, b \sin t); t \in [0, 2\pi]$ glatt parametrisiert. Also berechnet sich das Kurvenintegral

$$\begin{aligned} \oint_{\bar{\varepsilon}} -y dx + x dy &= \int_0^{2\pi} [-y\dot{x} + x\dot{y}] dt \\ &= \int_0^{2\pi} [-b \sin t \cdot (-a \sin' t) + a \cos t \cdot (b \sin' t)] dt = \int_0^{2\pi} ab[\sin^2 t + \cos^2 t] dt = 2\pi ab. \end{aligned}$$

Damit wird der Flächeninhalt der Ellipse zu πab , was sich für den Kreis zu πr^2 spezialisiert, also sehr plausibel ist.

Nebenbei bemerkt: Der Umfang der Ellipse führt auf ein Kurvenintegral, das sich nicht in geschlossener Form berechnen läßt. Da man den Ellipsenumfang aber gerne kennen möchte (Planetenbahnen), haben die Mathematiker im 19. Jahrhundert eine umfangreiche Theorie der sogenannten *elliptischen Integrale* entwickelt. Leider haben wir dafür überhaupt keine Zeit.

5.2 Der STOKESSche Integralsatz

handelt von einer kompakten zweiseitigen Oberfläche Γ im dreidimensionalen Raum und ihrer geschlossenen Randkurve $\bar{\gamma}$. Das Wort ‘Haube’ ist fürs erste ganz suggestiv. Die Fläche Γ soll im Definitionsbereich eines stetig differenzierbaren Vektorfeldes \vec{f} liegen, dessen Komponenten in der klassischen Formulierung P, Q, R heißen. In Beziehung gesetzt wird die Zirkulation, d.h. das skalare Kurvenintegral dieses Vektorfeldes entlang $\bar{\gamma}$, zu dem ‘Flußintegral’ eines verwandten Vektorfeldes durch die Oberfläche Γ . Das verwandte Vektorfeld hat die auf den ersten Blick nicht sonderlich naheliegenden Komponenten $\frac{\partial R}{\partial y} - \frac{\partial Q}{\partial z}, \frac{\partial P}{\partial z} - \frac{\partial R}{\partial x}, \frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}$. Es wird *Rotor* oder *Rotationsfeld* von \vec{f} genannt und mit $\vec{rot} \vec{f}$ bezeichnet³⁴. Wir kommen später darauf zurück.

³⁴Der Vektorpfeil über dem rot ist meine kleine Extavaganz und nicht allgemein üblich. Im Englischen heißt die Operation übrigens *curl* also ‘Locke’.

5.2.1 Die Formel

Bezeichnet $\vec{n} = \begin{pmatrix} n_1 \\ n_2 \\ n_3 \end{pmatrix}$ ein stetiges Normaleneinheitsfeld von Γ , durch das ein vektorielles Flächenelement $d\vec{F}$ festgelegt wird, so besagt die STOKESSche Formel, daß

$$\iint_{\Gamma} \text{rot } \vec{f} \bullet d\vec{F} = \oint_{\bar{\gamma}} \vec{f} \bullet d\vec{s}.$$

In klassischer Formulierung:

$$\iint_{\Gamma} \left\{ \left(\frac{\partial R}{\partial y} - \frac{\partial Q}{\partial z} \right) n_1 + \left(\frac{\partial P}{\partial z} - \frac{\partial R}{\partial x} \right) n_2 + \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) n_3 \right\} dF = \oint_{\bar{\gamma}} P dx + Q dy + R dz.$$

An Γ und $\bar{\gamma}$ werden dabei die uns schon bekannten Forderungen gestellt, die die Existenz der Integrale garantieren (glatte Parametrisierbarkeit, etc.).

Es gibt aber noch ein **Problem**. Auf einer zweiseitigen zusammenhängenden Fläche gibt es immer zwei Normaleneinheitsvektorfelder, die in entgegengesetzte Richtungen zeigen. Da sie in $d\vec{F}$ stecken, ändert das Flußintegral sein Vorzeichen, wenn die andere Normale ausgewählt wird. Das Kurvenintegral merkt davon aber gar nichts. Dieses wiederum ändert sein Vorzeichen, wenn die Kurve in der entgegengesetzten Richtung durchlaufen wird. Damit die STOKESSche Formel exakt und nicht nur bis auf das Vorzeichen gilt, müssen beide Orientierungen aufeinander abgestimmt werden. Das geschieht so, daß eine ‘Rechtsschraubung’ entsteht. Zeigt der Daumen der rechten Hand in Richtung \vec{n} , so zeigen die gekrümmten Finger der rechten Hand in die Umlaufrichtung von $\bar{\gamma}$.

5.2.2 Bemerkungen

Wie die GREENSche so gilt auch die STOKESSche Formel, für Hauben mit Löchern. Dann zerfällt das Kurvenintegral in mehrere Teile. Die Integrale über die Ränder der Löcher sind mit dem passenden Vorzeichen zu versehen.

Die GREENSche Formel selbst ist flacher Spezialfall der STOKESSchen. Die Oberfläche B liegt in der x, y -Ebene und das Vektorfeld hat keine Komponente in z -Richtung $R = 0$. Der Normaleneinheitsvektor zeigt dann konstant in positive z -Richtung, also sind $n_1 = n_2 = 0$ und $n_3 = 1$. Die Formel wird zu

$$\iint_B \left\{ \left(\frac{\partial Q}{\partial y} - \frac{\partial P}{\partial x} \right) \cdot 0 + \left(\frac{\partial R}{\partial z} - \frac{\partial P}{\partial x} \right) \cdot 0 + \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) \cdot 1 \right\} dF = \oint_{\bar{\gamma}} P dx + Q dy + 0 dz,$$

also

$$\iint_B \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) d(x, y) = \oint_{\bar{\gamma}} P dx + Q dy.$$

5.2.3 Beweisvorbereitungen

Unser Beweis wird allerdings gerade andersherum verlaufen; die ebene Formel wird ‘in den Raum gehoben.’ Dazu fixieren wir eine mindestens zweimal stetig differenzierbare Parametrisierung $\Phi = (\Phi_1, \Phi_2, \Phi_3) : B \rightarrow \Gamma$. Die Randkurve des u, v - Bereiches B sei durch $\bar{\beta} = (\beta_1, \beta_2) : [a, b] \rightarrow \mathbb{R}^2$ glatt parametrisiert. Dann wird die Randkurve von $\Gamma = \Phi(B)$ durch $\bar{\gamma} = \Phi \circ \bar{\beta}$ parametrisiert³⁵. Der übliche Normalenvektor auf Γ im Punkt $\Phi(u, v)$ ist gegeben durch

$$\vec{N}(u, v) = \frac{\partial \Phi}{\partial u}(u, v) \times \frac{\partial \Phi}{\partial v}(u, v) = \begin{pmatrix} \frac{\partial \Phi_1}{\partial u} \\ \frac{\partial \Phi_2}{\partial u} \\ \frac{\partial \Phi_3}{\partial u} \end{pmatrix} \times \begin{pmatrix} \frac{\partial \Phi_1}{\partial v} \\ \frac{\partial \Phi_2}{\partial v} \\ \frac{\partial \Phi_3}{\partial v} \end{pmatrix} = \begin{pmatrix} \frac{\partial \Phi_2}{\partial u} \frac{\partial \Phi_3}{\partial v} - \frac{\partial \Phi_3}{\partial u} \frac{\partial \Phi_2}{\partial v} \\ \frac{\partial \Phi_3}{\partial u} \frac{\partial \Phi_1}{\partial v} - \frac{\partial \Phi_1}{\partial u} \frac{\partial \Phi_3}{\partial v} \\ \frac{\partial \Phi_1}{\partial u} \frac{\partial \Phi_2}{\partial v} - \frac{\partial \Phi_2}{\partial u} \frac{\partial \Phi_1}{\partial v} \end{pmatrix}.$$

³⁵Wenn die Haube nicht in einem Stück parametrisiert werden kann, muß man zerschneiden, die Teile einzeln behandeln und die entstehenden Formeln addieren.

Wenn $\bar{\beta}$ den Bereich B so umrundet, daß B immer links bleibt, so sind $\bar{\gamma}$ und \bar{N} automatisch rechtsschraubig orientiert. Wir erinnern uns weiter, daß $\|\bar{N}\|$ gerade der Verzerrungsfaktor ist, der beim Oberflächenintegral benutzt wird: $dF = \|\bar{N}\|d(u, v)$. Der Normaleneinheitsvektor, den wir für die STOKESSche Formel brauchen, entsteht durch Normieren: $\bar{n} = \frac{\bar{N}}{\|\bar{N}\|}$.

Jetzt haben wir alles bereit, die Formel nachzurechnen.

5.2.4 Die Rechnung

könnte nun einfach durchgezogen werden. Die dabei entstehenden Formeln werden ziemlich lang und unübersichtlich. Nach dem Motto 'Teile und Herrsche' kann man die Rechnung in drei etwas übersichtlichere aufspalten, indem man jeweils zwei der Komponenten P, Q, R Null setzt. Anschließend werden die erhaltenen Ergebnisse addiert. Ich führe nur den Fall $Q = R = 0$ vor. Für diesen ist nachzurechnen, daß

$$\iint_{\Gamma} \left(\frac{\partial P}{\partial z} n_2 - \frac{\partial P}{\partial y} n_3 \right) dF = \oint_{\bar{\gamma}} P dx.$$

Mit Hilfe der oben fixierten Parametrisierungen, werden Kurven- und Oberflächenintegral folgendermaßen ausgerechnet:

$$\iint_B \left[\frac{\partial P}{\partial z} (\Phi(u, v)) \cdot n_2(u, v) - \frac{\partial P}{\partial y} (\Phi(u, v)) \cdot n_3(u, v) \right] \|\bar{N}\| d(u, v) \quad \text{und} \quad \int_a^b P(\bar{\gamma}(t)) \cdot \dot{\gamma}_1(t) dt$$

Wir werden mit dem Kurvenintegral $\oint_{\bar{\gamma}} P dx$ starten und es schrittweise in das Gebietsintegral umrechnen. Zunächst mittels Parametrisierung

$$\oint_{\bar{\gamma}} P dx = \int_a^b P(\bar{\gamma}(t)) \cdot \dot{\gamma}_1(t) dt.$$

Wegen $\gamma_1(t) = \Phi_1(\beta_1(t), \beta_2(t))$ ist $\dot{\gamma}_1(t) = \frac{\partial \Phi_1}{\partial u} \dot{\beta}_1(t) + \frac{\partial \Phi_1}{\partial v} \dot{\beta}_2(t)$ also

$$= \int_a^b P(\Phi(\bar{\beta}(t))) \cdot \left[\frac{\partial \Phi_1}{\partial u} \dot{\beta}_1(t) + \frac{\partial \Phi_1}{\partial v} \dot{\beta}_2(t) \right] dt = \oint_{\bar{\beta}} P(\Phi(u, v)) \cdot \frac{\partial \Phi_1}{\partial u} du + P(\Phi(u, v)) \cdot \frac{\partial \Phi_1}{\partial v} dv.$$

Jetzt benutzen wir die GREENSche Formel in der u, v -Ebene, um das Randintegral in ein Flächenintegral umzuwandeln:

$$= \iint_B \left[\frac{\partial}{\partial u} \left(P(\Phi(u, v)) \cdot \frac{\partial \Phi_1}{\partial v} \right) - \frac{\partial}{\partial v} \left(P(\Phi(u, v)) \cdot \frac{\partial \Phi_1}{\partial u} \right) \right] d(u, v).$$

Die beiden partiellen Ableitungen werden nach der Produktregel ausgerechnet. wegen $\frac{\partial^2 \Phi_1}{\partial u \partial v} = \frac{\partial^2 \Phi_1}{\partial v \partial u}$ heben sich zwei der entstehenden Summanden auf und es bleibt

$$= \iint_B \left[\frac{\partial P(\Phi(u, v))}{\partial u} \cdot \frac{\partial \Phi_1}{\partial v} - \frac{\partial P(\Phi(u, v))}{\partial v} \cdot \frac{\partial \Phi_1}{\partial u} \right] d(u, v).$$

Als nächstes differenzieren wir $P(\Phi(u, v)) = P(\Phi_1(u, v), \Phi_2(u, v), \Phi_3(u, v))$ mit Hilfe der Kettenregel nach u und v . Statt $\frac{\partial P}{\partial x}(\Phi(u, v))$ etc. schreiben wir dabei $\frac{\partial \widetilde{P}}{\partial x}$ etc.

$$\begin{aligned} \frac{\partial P(\Phi(u, v))}{\partial u} &= \frac{\partial \widetilde{P}}{\partial x} \frac{\partial \Phi_1}{\partial u} + \frac{\partial \widetilde{P}}{\partial y} \frac{\partial \Phi_2}{\partial u} + \frac{\partial \widetilde{P}}{\partial z} \frac{\partial \Phi_3}{\partial u} \\ \frac{\partial P(\Phi(u, v))}{\partial v} &= \frac{\partial \widetilde{P}}{\partial x} \frac{\partial \Phi_1}{\partial v} + \frac{\partial \widetilde{P}}{\partial y} \frac{\partial \Phi_2}{\partial v} + \frac{\partial \widetilde{P}}{\partial z} \frac{\partial \Phi_3}{\partial v} \end{aligned}$$

Setzt man das Ergebnis in die vorige Formel ein, ergibt sich

$$= \iint_B \left[\left(\frac{\partial \widetilde{P}}{\partial x} \frac{\partial \Phi_1}{\partial u} + \frac{\partial \widetilde{P}}{\partial y} \frac{\partial \Phi_2}{\partial u} + \frac{\partial \widetilde{P}}{\partial z} \frac{\partial \Phi_3}{\partial u} \right) \cdot \frac{\partial \Phi_1}{\partial v} - \left(\frac{\partial \widetilde{P}}{\partial x} \frac{\partial \Phi_1}{\partial v} + \frac{\partial \widetilde{P}}{\partial y} \frac{\partial \Phi_2}{\partial v} + \frac{\partial \widetilde{P}}{\partial z} \frac{\partial \Phi_3}{\partial v} \right) \cdot \frac{\partial \Phi_1}{\partial u} \right] d(u, v)$$

Hier hebt sich einiges auf und der Rest läßt sich sortieren.

$$= \iint_B \left[\frac{\partial \widetilde{P}}{\partial z} \cdot \left(\frac{\partial \Phi_3}{\partial u} \frac{\partial \Phi_1}{\partial v} - \frac{\partial \Phi_1}{\partial u} \frac{\partial \Phi_3}{\partial v} \right) - \frac{\partial \widetilde{P}}{\partial y} \cdot \left(\frac{\partial \Phi_1}{\partial u} \frac{\partial \Phi_2}{\partial v} - \frac{\partial \Phi_2}{\partial u} \frac{\partial \Phi_1}{\partial v} \right) \right] d(u, v)$$

Das ist aber dasselbe wie

$$= \iint_B \left[\frac{\partial \widetilde{P}}{\partial z} \cdot N_2(u, v) - \frac{\partial \widetilde{P}}{\partial y} \cdot N_3(u, v) \right] d(u, v).$$

Indem man die Norm des Normalenvektors \vec{N} ausklammert, entstehen innen die Komponenten des Normaleneinheitsvektors und außen das Oberflächendifferential:

$$= \iint_B \left[\frac{\partial \widetilde{P}}{\partial z} \cdot n_2(u, v) - \frac{\partial \widetilde{P}}{\partial y} \cdot n_3(u, v) \right] \|\vec{N}(u, v)\| d(u, v).$$

Das ist aber das mit Hilfe der Parametrisierung Φ ausgerechnete Oberflächenintegral

$$\iint_{\Gamma} \left(\frac{\partial P}{\partial z} n_2 - \frac{\partial P}{\partial y} n_3 \right) dF,$$

bei dem wir ankommen wollten.

5.2.5 Eine vektorielle Variante

der STOKESSchen Formel, die eventuell in der Elektrodynamik vorkommt. Die Haube Γ und ihre Randkurve $\bar{\gamma}$ seien wie bisher und P ein C^1 -Skalarfeld, das auf einer offenen Menge definiert ist,

die Γ umfaßt. Dann kann man auf die drei Vektorfelder $\begin{pmatrix} P \\ 0 \\ 0 \end{pmatrix}$, $\begin{pmatrix} 0 \\ P \\ 0 \end{pmatrix}$ und $\begin{pmatrix} 0 \\ 0 \\ P \end{pmatrix}$ jeweils

den STOKESSchen Satz anwenden. Er liefert die drei Formeln $\oint P dx = \iint (\frac{\partial P}{\partial z} n_2 - \frac{\partial P}{\partial y} n_3) dF$, $\oint P dy = \iint (-\frac{\partial P}{\partial z} n_1 + \frac{\partial P}{\partial x} n_3) dF$ und $\oint P dz = \iint (\frac{\partial P}{\partial y} n_1 - \frac{\partial P}{\partial x} n_2) dF$. Indem man die rechten und die linken Seiten jeweils untereinander schreibt, entsteht eine Vektorgleichung. Die soll jetzt noch etwas effektvoller aufgeschrieben werden, wozu wir Bezeichnungen benutzen, die sich dem Physiker (hoffentlich) selbst erklären. Der links entstehende Vektor

$$\begin{pmatrix} \oint P dx \\ \oint P dy \\ \oint P dz \end{pmatrix} \text{ wird geschrieben als } \oint P d \begin{pmatrix} x \\ y \\ z \end{pmatrix} \text{ oder } \oint P d\vec{r}.$$

Rechts entsteht, gleich als Vektorintegral geschrieben,

$$\iint \begin{pmatrix} \frac{\partial P}{\partial z} n_2 - \frac{\partial P}{\partial y} n_3 \\ -\frac{\partial P}{\partial z} n_1 + \frac{\partial P}{\partial x} n_3 \\ \frac{\partial P}{\partial y} n_1 - \frac{\partial P}{\partial x} n_2 \end{pmatrix} dF = \iint (\vec{n} \times \text{grad } P) dF = - \iint \text{grad } P \times d\vec{F}.$$

Zusammen liest sich das so

$$\oint_{\bar{\gamma}} P d\vec{r} = - \iint_{\Gamma} \text{grad } P \times d\vec{F},$$

wenn man denn diese Bezeichnungen benutzen will.

5.3 Der Integralsatz von GAUSS

Es sei H ein kompakter räumlicher Bereich (Körper, Volumen), der von einer geschlossenen Fläche Γ berandet wird. Von beiden Mengen wird wieder vorausgesetzt daß sie gutartig sind, so daß man darüber integrieren kann. Auf einer H umfassenden offenen Menge sei ein C^1 -Vektorfeld \vec{f} definiert, dessen Komponenten in der klassischen Version wieder P, Q, R heißen.

Ausgehend von diesem Vektorfeld wird ein Skalarfeld gebildet, das *Divergenz* von \vec{f} heißt, mit $\operatorname{div} \vec{f}$ bezeichnet wird und sich als Summe der Ableitungen der Komponenten ergibt:

$$\operatorname{div} \vec{f} = \operatorname{div} \begin{pmatrix} P \\ Q \\ R \end{pmatrix} = \frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} + \frac{\partial R}{\partial z}.$$

In Vektorschreibweise besagt der GAUSSsche Satz, daß dann

$$\iiint_H \operatorname{div} \vec{f} \, dV = \iint_{\Gamma} \vec{f} \cdot d\vec{F}$$

(hier habe ich dV statt $d(x, y, z)$ geschrieben, wie das die Physiker gern tun). Klassisch aufgeschrieben, lautet die Formel

$$\iiint_H \left(\frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} + \frac{\partial R}{\partial z} \right) d(x, y, z) = \iint_{\Gamma} (Pn_1 + Qn_2 + Rn_3) \, dF.$$

Der Normaleneinheitsvektor \vec{n} auf der Oberfläche Γ von H ist dabei so zu wählen, daß er nach außen zeigt (er steckt auch in $d\vec{F}$!).

5.3.1 Der Beweis für Quader

Sei $Q = [a_1, a_2] \times [b_1, b_2] \times [c_1, c_2]$ ein Quader. Ich werde nachrechnen, daß

$$\iiint_Q \frac{\partial P}{\partial x} \, d(x, y, z) = \iint_{\Gamma} Pn_1 \, dF$$

gilt. Für Q und R fällt die Rechnung ganz analog aus. Die drei Gleichungen ergeben dann aufaddiert den GAUSSschen Satz für Quader.

Das Volumenintegral wird in ein iteriertes verwandelt und dann nach Hauptsatz ausgerechnet:

$$\int_{b_1}^{b_2} \int_{c_1}^{c_2} \left(\int_{a_1}^{a_2} \frac{\partial P}{\partial x} \, dx \right) \, dz \, dy = \int_{b_1}^{b_2} \int_{c_1}^{c_2} (P(a_2, y, z) - P(a_1, y, z)) \, dz \, dy.$$

Das Oberflächenintegral zerfällt in eine Summe, entsprechend den sechs Rechtecken, die den Rand von Q ausmachen: R_r, R_l, R_o, R_u, R_v und R_h (die Indizes deuten an, daß das entsprechende Rechteck rechts, links, oben, unten, vorn oder hinten liegt). Die äußere Einheitsnormale \vec{n} zeigt auf jedem der sechs Rechtecke in Richtung einer Koordinatenachse. In zwei Fällen, nämlich R_r und R_l ist $\vec{n} = \pm \vec{e}_1$. Auf den anderen vier Rechtecken ist $n_1 = 0$. Also geben auch nur R_r/l einen Beitrag zum Integral $\iint_{\Gamma} Pn_1 \, dF$. Mithin ist

$$\iint_{\Gamma} Pn_1 \, dF = \iint_{R_r} P \cdot 1 \, dF + \iint_{R_l} P \cdot (-1) \, dF$$

Mit der offensichtlichen Parametrisierung (da die Seitenflächen platt sind gibt es keinen Verzerrungsfaktor) wird das

$$\int_{b_1}^{b_2} \int_{c_1}^{c_2} P(a_2, y, z) \, dz \, dy - \int_{b_1}^{b_2} \int_{c_1}^{c_2} P(a_1, y, z) \, dz \, dy$$

gleich dem oben bestimmten Volumenintegral. Das war der ganz Beweis.

5.3.2 Warum der GAUSSsche Satz wahr ist

Jetzt stellen wir uns vor, der Körper H wäre die Vereinigung (ganz vieler ganz kleiner) gleichgroßer Würfel, die sich entweder nicht schneiden oder eine Seitenfläche gemeinsam haben: $H = \bigcup_{i=1}^N W_i$.

Die Oberfläche von W_i soll Ω_i heißen. Jeder dieser Würfel ist ein Quader, also gilt der GAUSSsche Satz:

$$\iiint_{W_i} \operatorname{div} \vec{f} \, d(x, y, z) = \iint_{\Omega_i} \vec{f} \cdot d\vec{F}.$$

Alle diese Gleichungen kann man aufsummieren. Dann ergibt sich links

$$\sum_{i=1}^N \iiint_{W_i} \operatorname{div} \vec{f} \, d(x, y, z) = \iiint_H \operatorname{div} \vec{f} \, d(x, y, z)$$

und rechts

$$\sum_{i=1}^N \iint_{\Omega_i} \vec{f} \cdot d\vec{F}.$$

Von den $6 \cdot N$ Seitenflächen der kleinen Würfel kommen viele doppelt vor, nämlich als gemeinsame Seitenfläche zweier miteinander verklebter Würfel, etwa W_i und W_j . Die äußere Normale des einen Würfels zeigt dabei in den anderen hinein und umgekehrt. Die der gemeinsamen Fläche entsprechenden Anteile am Fluß $\iint_{\Omega_i} \vec{f} \cdot d\vec{F}$ bzw. $\iint_{\Omega_j} \vec{f} \cdot d\vec{F}$ heben sich also auf. Es bleiben also in der großen Summe nur Oberflächenintegrale über solche Seiten der beteiligten Würfel übrig, die nicht mit einer Seite eines anderen kleinen Würfels verklebt sind. Diese Seitenflächen bilden die Oberfläche von H . Also

$$\sum_{i=1}^N \iint_{\Omega_i} \vec{f} \cdot d\vec{F} = \iint_{\Gamma} \vec{f} \cdot d\vec{F}.$$

Damit haben wir nachgewiesen, daß das Divergenzintegral über H gleich dem Flußintegral durch die Oberfläche von H ist, wie vom GAUSSschen Satz behauptet.

Wenn man die Bausteine nur genügend klein macht, so läßt sich jeder Körper beliebig gut durch eine Vereinigung von Würfeln approximieren (im Grenzfall genau). Deshalb überträgt sich der Satz auf beliebige Körper. Auch solche mit Löchern.

Mit diesem Argument sind die meisten Naturwissenschaftler zufrieden. Bei näherem Hinsehen ist es allerdings nicht ganz unproblematisch (warum und in welchem Sinn wird auch die Oberfläche gut approximiert?). Deshalb finden Kenner und Liebhaber in den Ergänzungen einen exakten Beweis für gutartige Körper H

5.3.3 Volumen als Oberflächenintegral

Das Vektorfeld $\begin{pmatrix} x \\ y \\ z \end{pmatrix}$ wird standardmäßig mit \vec{r} bezeichnet. Man spricht vom Radiusvektor oder Ortsvektor des Punktes (x, y, z) . Setzt man es in die GAUSSsche Formel ein, erhält man

$$\iiint_H 3 \, d(x, y, z) = \iint_{\Gamma} \vec{r} \cdot d\vec{F}$$

Links steht das dreifache Volumen des Körpers H also ergibt sich das Volumen von H als

$$\frac{1}{3} \iint_{\Gamma} \vec{r} \cdot d\vec{F}.$$

Alternativ kann man jedes der drei Vektorfelder $\begin{pmatrix} x \\ 0 \\ 0 \end{pmatrix}$, $\begin{pmatrix} 0 \\ y \\ 0 \end{pmatrix}$ oder $\begin{pmatrix} 0 \\ 0 \\ z \end{pmatrix}$ benutzen. Ihre Divergenzen sind jeweils 1 also ergibt ihr Flußintegral durch Γ gerade das Volumen von G .

Beispiel Vollkugel vom Radius R : $K = \{(x, y, z) : x^2 + y^2 + z^2 \leq R^2\}$. Für jeden Punkt der Kugeloberfläche zeigt der Radiusvektor senkrecht nach außen. Es ist also

$$\vec{n} = \frac{\vec{r}}{|\vec{r}|} \quad \text{daher} \quad \vec{r} \cdot \vec{n} = |\vec{r}| = R.$$

konstant auf der Kugeloberfläche. Daher ist

$$\frac{1}{3} \iint_{\Gamma} \vec{r} \bullet d\vec{F} = \frac{1}{3} \iint_{\Gamma} R dF.$$

also ist das Volumen der Kugel gleich (wegen der Oberfläche vgl. 4.9.6)

$$\frac{1}{3} \cdot R \cdot \text{Oberfläche} = \frac{1}{3} R 4\pi R^2 = \frac{4}{3} \pi R^3,$$

wie wir es in der Schule gelernt haben.

5.3.4 Beispiel

Gegeben sei das außer im Nullpunkt definierte Vektorfeld $\vec{f} = \frac{\vec{r}}{r^3}$ (das z.B. im COULOMBSchen Gesetz vorkommt). Wir wollen seinen Fluß durch eine geschlossenen Oberfläche Γ bestimmen und nehmen zunächst an, daß diese den Nullpunkt nicht im Inneren enthält. Dann sagt der GAUSSsche Satz

$$\iint_{\Gamma} \vec{f} \bullet d\vec{F} = \iiint_H \operatorname{div} \vec{f} dV.$$

Nun rechnet man schnell nach (6.2.8), daß $\operatorname{div} \frac{\vec{r}}{r^3} = 0$, also ist das gesuchte Flußintegral auch Null.

Offenbar können wir nicht so schließen, wenn der Nullpunkt in H liegt. Dann ist jedenfalls der GAUSSsche Satz nicht anwendbar, obwohl der Fluß problemlos definiert ist, sofern der Nullpunkt nicht auf Γ selbst liegt.

Für die Sphäre $S_{\bar{0}}(R)$ ist $r = R$, die äußere Normale gleich $\frac{\vec{r}}{R}$ und daher:

$$\iint \vec{f} \bullet d\vec{F} = \iint \frac{\vec{r}}{R^3} \bullet \frac{\vec{r}}{R} dF = \iint \frac{1}{R^2} dF = \frac{1}{R^2} 4\pi R^2 = 4\pi,$$

unabhängig von R .

Mit einem Trick, erweitern wir das letzte Ergebnis auf alle geschlossenen Flächen Γ , die einen Körper H beranden, der den Nullpunkt im Inneren enthält. Dazu wird aus H eine kleine Vollkugel $K_{\bar{0}}(\varepsilon)$ herausgeschnitten und auf $H_{\varepsilon} = H \setminus K_{\bar{0}}(\varepsilon)$ der GAUSSsche Integralsatz angewendet. Das geht, denn \vec{f} ist in der Menge $\mathbb{R}^3 \setminus \{\bar{0}\} \supseteq H_{\varepsilon}$ definiert und stetig. Zwar hat H_{ε} jetzt ein Loch, aber das war erlaubt. Der Rand von H_{ε} besteht aus Γ und der Sphäre $S_{\bar{0}}(\varepsilon)$, die jetzt aber im Vergleich zu eben entgegengesetzt orientiert ist, weil die äußere Normale von H_{ε} natürlich in die herausgeschnittene Kugel hineinzeigt. Daher das Minuszeichen in

$$0 = \iiint_{H_{\varepsilon}} \operatorname{div} \vec{f} dV = \iint_{\Gamma} \vec{f} \bullet d\vec{F} - \iint_{S_{\bar{0}}(\varepsilon)} \vec{f} \bullet d\vec{F} = \iint_{\Gamma} \vec{f} \bullet d\vec{F} - 4\pi.$$

Liegt der Nullpunkt auf Γ selbst, so ist das Flußintegral uneigentlich und daher in der Definition problematisch. Der gängige Trick schneidet (immer kleinere) Halbkugeln heraus und bestimmt das Flußintegral als 2π ; darauf gehe ich nicht näher ein.

Ohne schlafende Hunde wecken zu wollen, doch eine kurze **Bemerkung**. Es ist in manchen Zusammenhängen zweckmäßig, sowohl dem Vektorfeld $\frac{\vec{r}}{r^3}$ als auch seiner Divergenz im Nullpunkt unendliche Werte zuzuordnen und den GAUSSschen Satz formal doch anzuwenden. Dann haben wir, nach obigen Rechnung

$$\iiint_H \operatorname{div} \frac{\vec{r}}{r^3} dV = \begin{cases} 0, & \text{falls } \bar{0} \notin H \\ 4\pi, & \text{falls } \bar{0} \in H \end{cases}.$$

Dieses Ergebnis wird als $\operatorname{div} \frac{\vec{r}}{r^3} = 4\pi\delta$ interpretiert, wo δ die berühmte im Nullpunkt konzentrierte dreidimensionale DIRAC-Funktion ist. Wegen $\frac{\vec{r}}{r^3} = -\operatorname{grad} \frac{1}{r}$ schreibt man diese Formel auch als $\Delta \frac{1}{r} = 4\pi\delta$.

5.3.5 Die ebene Variante des GAUSSschen Satzes

ergibt sich nur durch eine Vorzeichenänderung aus der GREENESchen Formel. Wir betrachten ein ebenes C^1 -Vektorfeld $\vec{f} = \begin{pmatrix} P \\ Q \end{pmatrix}$, das auf einer die Fläche G umfassenden offenen Menge definiert sein soll, und die G im positiven Sinn umrandende Kurve $\bar{\gamma} = (\gamma_1, \gamma_2)$. Dreht man den Tangentenvektor $\dot{\gamma} = (\dot{\gamma}_1, \dot{\gamma}_2)$ um 90° nach rechts, so entsteht ein äußerer Normalenvektor $\vec{N}(t) = \begin{pmatrix} \dot{\gamma}_2 \\ -\dot{\gamma}_1 \end{pmatrix}$, der dieselbe Norm hat, wie der Tangentenvektor. Also ist $\vec{n}(t) = \frac{\vec{N}(t)}{\|\dot{\gamma}(t)\|}$ der äußere Normaleneinheitsvektor. Nun rechnen wir mit Hilfe der GREENESchen Formel aus:

$$\begin{aligned} \iint_G \operatorname{div} \vec{f} \, d(x, y) &= \iint_G \left(\frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} \right) d(x, y) = \oint_{\bar{\gamma}} (-Q) dx + P dy \\ &= \int_a^b [-Q(\bar{\gamma}(t)) \cdot \dot{\gamma}_1(t) + P(\bar{\gamma}(t)) \cdot \dot{\gamma}_2(t)] dt \\ &= \int_a^b \vec{f}(t) \bullet \vec{N}(t) dt = \int_a^b \vec{f} \bullet \vec{n} \cdot \|\dot{\gamma}(t)\| dt = \oint_{\bar{\gamma}} \vec{f} \bullet \vec{n} ds. \end{aligned}$$

Wenn wir die Zwischenschritte weglassen, ergibt sich die GAUSSsche Formel

$$\iint_G \operatorname{div} \vec{f} \, d(x, y) = \oint_{\bar{\gamma}} \vec{f} \bullet \vec{n} ds,$$

wobei $\vec{n}(t)$ die im Randpunkt $\bar{\gamma}(t)$ errichtete äußere Einheitsnormale an G ist.

5.3.6 Eine vektorielle Variante

des GAUSSschen Satzes, die in Analogie zu 5.2.5 entsteht. H und Γ behalten ihre Bedeutung, P ist ein C^1 -Skalarfeld. Indem man GAUSS auf die drei Vektorfelder $\begin{pmatrix} P \\ 0 \\ 0 \end{pmatrix}$, $\begin{pmatrix} 0 \\ P \\ 0 \end{pmatrix}$ und $\begin{pmatrix} 0 \\ 0 \\ P \end{pmatrix}$ anwendet, erhält man

$$\iiint \frac{\partial P}{\partial x} d(x, y, z) = \iint P n_1 dF \quad \iiint \frac{\partial P}{\partial y} d(x, y, z) = \iint P n_2 dF \quad \text{und} \quad \iiint \frac{\partial P}{\partial z} d(x, y, z) = \iint P n_3 dF.$$

Die angestrebte Formel entsteht, indem diese drei Gleichungen zu einer Vektorgleichung zusammengefaßt werden. Mit effektvollen Bezeichnungen liest sich das dann netter. Links wird aus

$$\begin{pmatrix} \iiint \frac{\partial P}{\partial x} d(x, y, z) \\ \iiint \frac{\partial P}{\partial y} d(x, y, z) \\ \iiint \frac{\partial P}{\partial z} d(x, y, z) \end{pmatrix} \quad \text{erst} \quad \iiint \begin{pmatrix} \frac{\partial P}{\partial x} \\ \frac{\partial P}{\partial y} \\ \frac{\partial P}{\partial z} \end{pmatrix} d(x, y, z) \quad \text{und dann} \quad \iiint \vec{\operatorname{grad}} P d(x, y, z).$$

Rechts schreibt man

$$\begin{pmatrix} \iint P n_1 dF \\ \iint P n_2 dF \\ \iint P n_3 dF \end{pmatrix} \quad \text{als} \quad \iint P \begin{pmatrix} n_1 \\ n_2 \\ n_3 \end{pmatrix} dF = \iint P \vec{n} dF \quad \text{oder sogar} \quad \iint P d\vec{F}.$$

Dann entsteht also die Formel

$$\iiint_H \vec{\operatorname{grad}} P d(x, y, z) = \iint_{\Gamma} P d\vec{F},$$

wie sie z.B. in der Elektrodynamik benutzt wird.

Man hätte unter Benutzung der schicken Schreibweisen auch eleganter argumentieren können. Das mache ich bei der nächsten Variante vor.

5.3.7 Noch eine vektorielle Variante

H und Γ bleiben wie gehabt, \vec{f} ist jetzt ein C^1 -Vektorfeld. Die angestrebte Formel besagt

$$\iiint_H \operatorname{rot} \vec{f} \, d(x, y, z) = - \iint_{\Gamma} (\vec{f} \times \vec{n}) \, dF = - \iint_{\Gamma} \vec{f} \times d\vec{F}$$

(auf beiden Seiten stehen Vektoren!). Um das zu beweisen, halten wir einen beliebigen Vektor \vec{a} fest und wenden den GAUSSSchen Satz auf das Vektorfeld $\vec{f} \times \vec{a}$ an:

$$\iiint_H \operatorname{div} (\vec{f} \times \vec{a}) \, d(x, y, z) = \iint_{\Gamma} (\vec{f} \times \vec{a}) \cdot d\vec{F}.$$

Eine kleine Rechnung (Übungsaufgabe, vgl. auch 6.2.5) zeigt $\operatorname{div} (\vec{f} \times \vec{a}) = \operatorname{rot} \vec{f} \cdot \vec{a}$. Auf der rechten Seite ist $(\vec{f} \times \vec{a}) \cdot \vec{n} = \det(\vec{f}, \vec{a}, \vec{n}) = -\det(\vec{f}, \vec{n}, \vec{a}) = -(\vec{f} \times \vec{n}) \cdot \vec{a}$. Setzt man das ein, ergibt sich

$$\iiint_H \operatorname{rot} \vec{f} \cdot \vec{a} \, d(x, y, z) = - \iint_{\Gamma} (\vec{f} \times \vec{n}) \cdot \vec{a} \, dF$$

Da \vec{a} konstant ist, kann man das Skalarprodukt herausziehen (jetzt greift die vektorielle Notation), und bekommt

$$\vec{a} \cdot \iiint_H \operatorname{rot} \vec{f} \, d(x, y, z) = \vec{a} \cdot - \iint_{\Gamma} \vec{f} \times \vec{n} \, dF.$$

Da das für alle Vektoren \vec{a} gilt, kann man diese herauskürzen; es entsteht die versprochene Formel.

5.4 Die GREENSchen Formeln im Raum

5.4.1 Der LAPLACE-Operator

Ist f ein C^2 -Skalarfeld, so kann man daraus zunächst das Vektorfeld $\vec{\operatorname{grad}} f$ machen und dieses hinterher wieder zu einem Skalarfeld: $\operatorname{div} \vec{\operatorname{grad}} f$. Das kommt häufiger vor, deshalb hat man ein kurzes Symbol eingeführt, das LAPLACE-Operator genannt wird:

$$\Delta f = \operatorname{div} \vec{\operatorname{grad}} f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} + \frac{\partial^2 f}{\partial z^2},$$

wie man durch Nachrechnen sofort bestätigt. Genau wie bei Divergenz und Gradient ist diese Definition nicht spezifisch dreidimensional. Ist das Skalarfeld f auf einer Teilmenge von \mathbb{R}^n definiert, wird Δf zu $\sum_{i=1}^n \frac{\partial^2 f}{\partial x_i^2}$.

5.4.2 Herleitung der asymmetrischen (ersten) GREENSchen Formel

Gegeben ist ein räumlicher Bereich (Körper, Volumen) H und seine Randfläche Γ , wie im GAUSSSchen Integralsatz, den wir gleich anwenden wollen. Außerdem spielen zwei auf einer H umfassenden offenen Menge definierte C^2 -Skalarfelder f und g eine Rolle.

Wendet man auf das Vektorfeld $f \vec{\operatorname{grad}} g$ den GAUSSSchen Satz an, so ergibt sich

$$\iiint_H \operatorname{div} (f \vec{\operatorname{grad}} g) \, dV = \iint_{\Gamma} f \vec{\operatorname{grad}} g \cdot d\vec{F} = \iint_{\Gamma} f \vec{\operatorname{grad}} g \cdot \vec{n} \, dF$$

Beiden Seiten wollen wir noch eine etwas andere Form geben. Eine kleine Nebenrechnung zeigt

$$\operatorname{div} (f \vec{\operatorname{grad}} g) = \frac{\partial}{\partial x} \left(f \frac{\partial g}{\partial x} \right) + \dots = \frac{\partial f}{\partial x} \frac{\partial g}{\partial x} + f \frac{\partial^2 g}{\partial x^2} + \dots = \vec{\operatorname{grad}} f \cdot \vec{\operatorname{grad}} g + f \Delta g.$$

Auf der rechten Seite können wir f aus dem Skalarprodukt herausziehen und erinnern uns dann an die Richtungsableitung:

$$f \vec{\operatorname{grad}} g \cdot \vec{n} = f \vec{\operatorname{grad}} g \cdot \vec{n} = f \frac{\partial g}{\partial \vec{n}}.$$

Setzt man das ein, entsteht die angestrebte GREENSche Formel:

$$\iiint_H (\vec{\operatorname{grad}} f \cdot \vec{\operatorname{grad}} g + f \Delta g) \, dV = \iint_{\Gamma} f \frac{\partial g}{\partial \vec{n}} \, dF.$$

5.4.3 Varianten

Für $f = g = u$ liest sich die GREENSche Formel so:

$$\iiint_H \left(\|\vec{\text{grad}} u\|^2 + u\Delta u \right) d(x, y, z) = \iint_{\Gamma} u \frac{\partial u}{\partial \vec{n}} dF.$$

Auch die folgende Formel wird oft benutzt:

$$\iiint_H \Delta u d(x, y, z) = \iint_{\Gamma} \frac{\partial u}{\partial \vec{n}} dF.$$

Sie entsteht, wenn $f = 1$ und $g = u$ gesetzt wird.

5.4.4 Die symmetrische (zweite) GREENSche Formel

ergibt sich, wenn man erst f und g die Rollen tauschen läßt und dann die zweite Formel von der ersten abzieht. Dabei hebt sich das Volumenintegral von $\text{grad } f \bullet \text{grad } g$ auf und es entsteht

$$\iiint_H (f\Delta g - g\Delta f) dV = \iint_{\Gamma} \left(f \frac{\partial g}{\partial \vec{n}} - g \frac{\partial f}{\partial \vec{n}} \right) dF$$

5.4.5 Ebene Varianten

Dieselbe Herleitung unter Verwendung der ebenen GAUSSSchen Formel 5.3.5 liefert für zweistellige C^2 -Funktionen z.B.

$$\iint_G (f\Delta g - g\Delta f) d(x, y) = \oint_{\vec{\gamma}} \left(f \frac{\partial g}{\partial \vec{n}} - g \frac{\partial f}{\partial \vec{n}} \right) ds.$$

Daraus folgen, die den oben angegebenen analogen Varianten.

6 Vektoranalysis

wird von uns vorzugsweise im dreidimensionalen Raum betrieben und handelt von den oben schon eingeführten Differentialoperatoren grad , div und rot . Bevor wir mathematisch einwandfrei loslegen, werde ich auf etwas laxer Weise eine

6.1 Anschaulich-physikalische Deutung von Divergenz und Rotation

geben. Die analoge Diskussion des Gradienten haben wir schon hinter uns (3.3.7, 3.3.8).

6.1.1 Divergenz als Ergiebigkeitsdichte

Wir betrachten ein Vektorfeld \vec{f} , das wir uns als Geschwindigkeitsfeld einer stationär strömenden Flüssigkeit vorstellen wollen. Ist dann H ein von der geschlossenen Fläche Γ begrenztes Gebiet durch das die Flüssigkeit strömt, so ist $\iint_{\Gamma} \vec{f} \bullet d\vec{F}$ die Bilanz zwischen der pro Zeiteinheit aus H ausströmenden und der in H einströmenden Flüssigkeitsmenge. Wir wollen anschaulich von der *Ergiebigkeit* von H sprechen. Setzt man die Flüssigkeit als inkompressibel voraus, so kann eine von Null verschiedene Ergiebigkeit nur zustande kommen, wenn es innerhalb von H Quellen gibt (aus denen Flüssigkeit kommt) oder Senken (in die Flüssigkeit verschwindet).

Jetzt fixieren wir einen Punkt \bar{p} in der Strömung und schauen uns kleine Volumina H an, die \bar{p} enthalten (etwa Würfel oder Kugeln). Wenn der Quotient aus

$$\frac{\text{Ergiebigkeit von } H}{\text{Volumen von } H}$$

gegen einen Grenzwert strebt, wenn man H immer mehr auf \bar{p} zusammenzieht, so wird man diesen als *Ergiebigkeitsdichte* des Feldes \vec{f} im Punkt \bar{p} interpretieren. Schreiben wir dafür provisorisch $\text{erg } \vec{f}(\bar{p})$.

Angenommen $\text{erg } \vec{f}$ existiert in jedem Punkt der Strömung. Ist S ein von Σ berandeter Teil des Feldes, so wird die Ergiebigkeit von S gleich der über S integrierten Ergiebigkeitsdichte sein:

$$\iint_{\Sigma} \vec{f} \bullet d\vec{F} = \iiint_S \text{erg } \vec{f}(\bar{x}) d\bar{x}.$$

Nun ist aber andererseits nach GAUSS

$$\iint_{\Sigma} \vec{f} \bullet d\vec{F} = \iiint_S \text{div } \vec{f}(\bar{x}) dV.$$

Da das für alle S gilt, folgt $\text{erg } \vec{f} = \text{div } \vec{f}$. Die Divergenz mißt die Ergiebigkeitsdichte des Feldes; und wir können die Bezeichnung erg wieder vergessen.

6.1.2 Alternative Herleitung

Man kann die Divergenz eines C^1 -Vektorfeldes auch ohne Rückgriff auf den GAUSSschen Integralatz als Ergiebigkeitsdichte identifizieren. Sei $\bar{p} = (x, y, z)$ der entsprechende Punkt und W_h der Würfel mit Mittelpunkt \bar{p} und Kantenlänge $2h$:

$$W_h = [x - h, x + h] \times [y - h, y + h] \times [z - h, z + h].$$

Wir berechnen die Ergiebigkeit von W_h , teilen durch das Volumen $(2h)^3$ und lassen $h \rightarrow 0$ gehen. Als Grenzwert sollte $\text{div } \vec{f}(\bar{p}) = \frac{\partial f_1}{\partial x}(\bar{p}) + \frac{\partial f_2}{\partial y}(\bar{p}) + \frac{\partial f_3}{\partial z}(\bar{p})$ herauskommen.

Für die Ergiebigkeit müssen wir $\vec{f} \bullet \vec{n}$ über die Oberfläche des Würfels integrieren. Diese besteht aus 6 Quadraten, auf denen die äußere Normale jeweils \pm einer der kanonischen Basisvektoren \vec{e}_i ist. Ich berechne ausführlich nur den Anteil der beiden Würfel­flächen, die man aus positiver bzw negativer x -Richtung sieht. Der Normaleneinheitsvektor ist $\pm \vec{e}_1$, deshalb bleibt vom Skalarprodukt $\vec{f} \bullet \vec{n}$ nur $\pm f_1$ übrig. Die Flächen werden parametrisiert durch $(u, v) \mapsto (x \pm h, u, v)$,

wobei (u, v) das Quadrat $Q_h = [y - h, y + h] \times [z - h, z + h]$ durchläuft. Offenbar gibt es keine Verzerrung, daher liefern die beiden Seiten des Würfels den Anteil

$$\iint_{Q_h} f_1(x+h, u, v) d(u, v) + \iint_{Q_h} -f_1(x-h, u, v) d(u, v) = \iint_{Q_h} [f_1(x+h, u, v) - f_1(x-h, u, v)] d(u, v)$$

an der Ergiebigkeit. Nun wenden wir den Mittelwertsatz der Integralrechnung an. Er liefert von h abhängige $(u_h, v_h) \in Q_h$, so daß

$$\iint_{Q_h} [f_1(x+h, u, v) - f_1(x-h, u, v)] d(u, v) = [f_1(x+h, u_h, v_h) - f_1(x-h, u_h, v_h)] \cdot |Q_h|.$$

Der Mittelwertsatz der Differentialrechnung liefert ein $w_h \in [x-h, x+h]$, so daß

$$f_1(x+h, u_h, v_h) - f_1(x-h, u_h, v_h) = \frac{\partial f_1}{\partial x}(w_h, u_h, v_h) \cdot 2h$$

Setzt man das oben ein und berücksichtigt $|Q_h| = 4h^2$, so bleibt für den Anteil der beiden Flächen an der Ergiebigkeit $\frac{\partial f_1}{\partial x}(w_h, u_h, v_h) \cdot 8h^3$. Division des Flußanteils der beiden Quadrate durch das Würfelvolumen $(2h)^3$ liefert $\frac{\partial f_1}{\partial x}(w_h, u_h, v_h)$. Für $h \rightarrow 0$ bleibt den Punkten (w_h, u_h, v_h) nichts übrig, als gegen den Würfelmittelpunkt (x, y, z) zu streben. Also ergibt sich im Grenzwert $\frac{\partial f_1}{\partial x}(\bar{p})$. Analog ergeben die beiden Seitenflächen, die man von der $\pm y$ -Achse sieht, den Anteil $\frac{\partial f_2}{\partial y}(\bar{p})$. Deckel und Boden des Würfels liefern $\frac{\partial f_3}{\partial z}(\bar{p})$. Insgesamt kommt also tatsächlich die Divergenz heraus.

Ein Physiker würde wahrscheinlich die beiden Hinweise auf die Mittelwertsätze nicht machen, sondern die TAYLORSche Formel benutzen und ‘Glieder höherer Ordnung unter den Tisch fallen lassen’. Diese etwas direktere Art zu argumentieren werde ich bei der Rotation demonstrieren.

6.1.3 Die Rotation als Wirbeldichte

zu interpretieren wird dadurch kompliziert, daß nicht der Vektor $\text{rot } \vec{f}$ selbst, sondern nur seine Komponente in eine (allerdings beliebige) Richtung beschrieben wird. So ähnlich wie der Gradient als derjenige Vektor beschrieben werden kann, dessen Komponente in \vec{e} -Richtung die Richtungsableitung ist. Seien also ein Vektorfeld \vec{f} , ein innerer Punkt \bar{p} des Definitionsbereiches und ein Einheitsvektor \vec{e} vorgegeben. Wir wollen erklären, was die Zahl $\text{rot } \vec{f}(\bar{p}) \cdot \vec{e}$ mißt.

In der durch \bar{p} gehenden senkrecht auf \vec{e} stehenden Ebene werden kleine geschlossenen Kurven (etwa Kreise oder Quadrate) $\bar{\gamma}$ betrachtet, die \bar{p} umrunden, wobei der Umlaufsinn mit \vec{e} abgestimmt ist (Rechtsschraube). Die umrandete Fläche sei Γ . Sie ist platt und hat \vec{e} als Einheitsnormale in jedem Punkt. Nach STOKES gilt also

$$\iint_{\Gamma} \text{rot } \vec{f} \cdot \vec{e} \, dF = \oint_{\bar{\gamma}} \vec{f} \cdot d\vec{s}.$$

Man betrachte
Zirkulation * delta t

Der Mittelwertsatz liefert ein $\bar{a} \in \Gamma$, so daß

$$\iint_{\Gamma} \text{rot } \vec{f} \cdot \vec{e} \, dF = \text{rot } \vec{f}(\bar{a}) \cdot \vec{e} \cdot |\Gamma|$$

Zieht man jetzt die Kurve $\bar{\gamma}$ immer enger um \bar{p} herum, so bleibt dem Punkt \bar{a} nichts übrig, als sich immer weiter \bar{p} zu nähern. Damit wird $\text{rot } \vec{f}(\bar{p}) \cdot \vec{e}$ zum Grenzwert des Quotienten aus Zirkulation des Feldes durch umrandete Fläche:

$$\text{rot } \vec{f}(\bar{p}) \cdot \vec{e} = \lim \frac{1}{|\Gamma|} \oint_{\bar{\gamma}} \vec{f} \cdot d\vec{s}.$$

Wie beim Gradienten schließen wir, daß $\text{rot } \vec{f}(\bar{p})$ die Richtung der Achse durch \bar{p} angibt, senkrecht zu der die ‘Verwirbelung’ maximal ist. Der Betrag von rot ist ein Maß für die Stärke dieser Verwirbelung.

6.1.4 Alternative Herleitung

Wieder kann man auch ohne den Satz von STOKES auskommen, um umgekehrt die Zirkulationsdichte im eben besprochenen Sinn als $\vec{\text{rot}}$ zu identifizieren. Sei dazu $\bar{p} = (x, y, z)$ der fragliche Punkt. Das Vektorfeld habe die Koordinatenfunktionen P, Q, R . Ich berechne die x -Koordinate der Zirkulationsdichte.

Dazu wähle ich Kreise $\bar{\gamma}(t) = (x, y + r \cos t, z + r \sin t)$, die den Punkt \bar{p} senkrecht zu x -Achse umrunden. Der eingeschlossenen Kreis hat den Flächeninhalt πr^2 , daher ist nachzuweisen, daß

$$\lim_{r \rightarrow 0} \frac{1}{\pi r^2} \oint_{\bar{\gamma}} P dx + Q dy + R dz = \frac{\partial R}{\partial y}(\bar{p}) - \frac{\partial Q}{\partial z}(\bar{p}).$$

Da die Kreisbewegung senkrecht zur x -Achse erfolgt, ist $dx = 0$; der erste Summand im Kurvenintegral braucht nicht berücksichtigt zu werden. Für den Rest ergibt sich nach Definition

$$\oint_{\bar{\gamma}} Q dy + R dz = \int_0^{2\pi} [Q(x, y + r \cos t, z + r \sin t)(-r \sin t) + R(x, y + r \cos t, z + r \sin t)(r \cos t)] dt$$

Nun wird Q nach der TAYLORSchen Formel entwickelt:

$$Q(x, y + r \cos t, z + r \sin t) = Q(x, y, z) + \frac{\partial Q}{\partial y}(x, y, z)r \cos t + \frac{\partial Q}{\partial z}(x, y, z)r \sin t + \dots$$

In den verbleibenden Gliedern kommt immer r^2 vor. Mit dem r aus $dy = -r \sin t$ ergibt sich der Faktor r^3 . Da das Kurvenintegral durch πr^2 geteilt wird und danach der Grenzwert für $r \rightarrow 0$ genommen werden soll, liefern diese Glieder (einschließlich des Restgliedes) keinen Beitrag zum Grenzwert. Daher brauchen wir nicht konkret zu wissen, wie sie aussehen. Dasselbe gilt für R . Setzt man das Ergebnis in das Integral, ein so folgt

$$\begin{aligned} \oint_{\bar{\gamma}} Q dy + R dz &= \int_0^{2\pi} \left[\left(Q(\bar{p}) + \frac{\partial Q}{\partial y}(\bar{p})r \cos t + \frac{\partial Q}{\partial z}(\bar{p})r \sin t + r^2 \dots \right) \cdot (-r \sin t) \right. \\ &\quad \left. + \left(R(\bar{p}) + \frac{\partial R}{\partial y}(\bar{p})r \cos t + \frac{\partial R}{\partial z}(\bar{p})r \sin t + r^2 \dots \right) \cdot (r \cos t) \right] dt \end{aligned}$$

Da die Integrale $\int_0^{2\pi}$ über $\sin t$, $\cos t$ und $\sin t \cos t$ alle Null sind, bleibt nur übrig

$$\oint_{\bar{\gamma}} Q dy + R dz = -r^2 \frac{\partial Q}{\partial z}(\bar{p}) \int_0^{2\pi} \sin^2 t dt + r^2 \frac{\partial R}{\partial y}(\bar{p}) \int_0^{2\pi} \cos^2 t dt + r^3 \dots$$

Da die Integrale von \sin^2 und \cos^2 gleich π sind, ergibt sich nach Division durch πr^2 und Grenzübergang $r \rightarrow 0$ tatsächlich $\frac{\partial R}{\partial y} - \frac{\partial Q}{\partial z}$, wie behauptet.

Nach diesen etwas heuristischen Einlassungen werden wir wieder streng.

6.2 Formeln

Generalvoraussetzung. Wir vereinbaren, daß alle in diesem Kapitel vorkommenden Vektor- und Skalarfelder auf einer offenen Menge D definiert sind und (um damit keinen Stress zu haben) alle partiellen Ableitungen aller Ordnungen besitzen, die sogar stetig sein sollen (C^∞ -Felder). Wenn nichts anderes gesagt ist, liegt D in \mathbb{R}^3 und die Vektorfelder nehmen ihre Werte in $\vec{\mathbb{R}}^3$ an. Die C^∞ -Skalarfelder bilden dann selbst einen (natürlich unendlichdimensionalen) Vektorraum, den ich kurzzeitig \mathcal{S} nenne. Genauso bilden die (dreidimensionalen) C^∞ -Vektorfelder einen Vektorraum \mathcal{V} . Die drei Operatoren lassen sich auffassen als Abbildungen

$$\vec{\text{grad}} : \mathcal{S} \rightarrow \mathcal{V}, \quad \vec{\text{rot}} : \mathcal{V} \rightarrow \mathcal{V} \quad \text{und} \quad \text{div} : \mathcal{V} \rightarrow \mathcal{S}.$$

Es erweist sich, daß diese Abbildungen linear sind (die entsprechenden Formeln stehen unten). Wir werden uns um ihre Kerne und Bilder kümmern.

6.2.1 Nabla

In manchen Zusammenhängen ist es bequem, das von HAMILTON eingeführte Nabla-Symbol³⁶ zu benutzen. Es ist der formale Vektor

$$\nabla = \begin{pmatrix} \frac{\partial}{\partial x} \\ \frac{\partial}{\partial y} \\ \frac{\partial}{\partial z} \end{pmatrix}$$

der folgendermaßen formal auf Skalar und Vektorfelder angewendet wird:

$$\vec{\text{grad}} f = \nabla f = \begin{pmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \\ \frac{\partial f}{\partial z} \end{pmatrix}, \quad \text{div } \vec{f} = \nabla \bullet \vec{f} = \frac{\partial}{\partial x} f_1 + \frac{\partial}{\partial y} f_2 + \frac{\partial}{\partial z} f_3$$

und

$$\vec{\text{rot}} \vec{f} = \nabla \times \vec{f} = \begin{vmatrix} \vec{e}_1 & \frac{\partial}{\partial x} & f_1 \\ \vec{e}_2 & \frac{\partial}{\partial y} & f_2 \\ \vec{e}_3 & \frac{\partial}{\partial z} & f_3 \end{vmatrix} = \begin{pmatrix} \frac{\partial}{\partial y} f_3 - \frac{\partial}{\partial z} f_2 \\ \frac{\partial}{\partial z} f_1 - \frac{\partial}{\partial x} f_3 \\ \frac{\partial}{\partial x} f_2 - \frac{\partial}{\partial y} f_1 \end{pmatrix}.$$

Darin wollen wir keinen tieferen Sinn suchen. Immerhin hilft diese Darstellung dabei, sich einige wichtige Formeln zu merken. Unter anderem die Definitionen von div und $\vec{\text{rot}}$.

Bei Rechnungen wie

$$\text{div}(\vec{\text{rot}} \vec{f}) = \nabla \bullet (\nabla \times \vec{f}) = \det(\nabla, \nabla, \vec{f}) = 0$$

(Eigenschaft des Spatprodukts) oder (Anwendung des dreifachen Vektorprodukts)

$$\vec{\text{rot}}(\vec{f} \times \vec{g}) = \nabla \times (\vec{f} \times \vec{g}) = (\nabla \bullet \vec{g}) \vec{f} - (\nabla \bullet \vec{f}) \vec{g} = (\text{div } \vec{g}) \vec{f} - (\text{div } \vec{f}) \vec{g},$$

ist dagegen Vorsicht geboten (die zweite ist falsch!). Es stimmt auch nicht, daß $\vec{\text{rot}} \vec{f}$ immer senkrecht auf \vec{f} steht, wie $\nabla \times \vec{f}$ vermuten ließe. Ich rate, sich zwar von den Nablas inspirieren zu lassen, aber am Ende immer mit den ordentlichen Definitionen zu argumentieren.

Die folgenden Formeln sollte man sich nicht unbedingt merken, sondern bei Bedarf immer wieder herleiten. Das gilt auch für viele weitere Formeln, die man in der Literatur finden kann.

In der kleinen Liste sind a, b reelle Zahlen, f, g Skalarfelder (d.h. Elemente von \mathcal{S}) und \vec{f}, \vec{g} Vektorfelder aus \mathcal{V} .

6.2.2 Eigenschaften des Gradienten

- (1) $\vec{\text{grad}}(af + bg) = a \vec{\text{grad}} f + b \vec{\text{grad}} g$ (Linearität)
- (2) $\vec{\text{grad}}(f \cdot g) = f \cdot \vec{\text{grad}} g + g \cdot \vec{\text{grad}} f$ (Produktregel)

6.2.3 Eigenschaften der Divergenz

- (1) $\text{div}(a\vec{f} + b\vec{g}) = a \text{div } \vec{f} + b \text{div } \vec{g}$ (Linearität)
- (2) $\text{div}(f \cdot \vec{g}) = \vec{\text{grad}} f \bullet \vec{g} + f \text{div } \vec{g}$ (Produktregel)

6.2.4 Eigenschaften der Rotation

- (1) $\vec{\text{rot}}(a\vec{f} + b\vec{g}) = a \vec{\text{rot}} \vec{f} + b \vec{\text{rot}} \vec{g}$ (Linearität)
- (2) $\vec{\text{rot}}(f \cdot \vec{g}) = (\vec{\text{grad}} f) \times \vec{g} + f \cdot \vec{\text{rot}} \vec{g}$ (Produktregel)

³⁶Nabla ist ein altjüdisches harfenähnliches Musikinstrument.

6.2.5 Weitere Produktregeln

Da mit \vec{f} und \vec{g} auch $\vec{f} \times \vec{g}$ ein Vektorfeld ist, liegt es nahe zu fragen, wie sich röt und div dazu verhalten. Für beides gibt es Formeln. Die für $\text{röt}(\vec{f} \times \vec{g})$ läßt sich allerdings nicht ohne Einführung neuer Bezeichnungen aufschreiben, weshalb ich darauf verzichte. Gleiches gilt übrigens für $\text{grad}(\vec{f} \bullet \vec{g})$. Dagegen ist

Wieso denn das?

$$\text{div}(\vec{f} \times \vec{g}) = \vec{g} \bullet \text{röt} \vec{f} - \vec{f} \bullet \text{röt} \vec{g},$$

wie sie selbst schnell nachrechnen.

6.2.6 Hintereinanderausführung

$$(1) \quad \text{röt}(\text{grad} \vec{f}) = \vec{0} \quad (2) \quad \text{div}(\text{röt} \vec{f}) = 0 \quad (3) \quad \text{div}(\text{grad} \vec{f}) = \Delta f.$$

Die letzte dieser Formeln ist nur eine Erinnerung an den im Zusammenhang mit der GREENSchen Formel schon aufgetauchten LAPLACE-Operator Δ der durch $\Delta f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} + \frac{\partial^2 f}{\partial z^2}$ (bzw. das höherdimensionale Analogon) definiert ist.

Die ersten beiden Identitäten rechnen wir nach:

$$\text{röt}(\text{grad} \vec{f}) = \text{röt} \begin{pmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \\ \frac{\partial f}{\partial z} \end{pmatrix} = \begin{pmatrix} \frac{\partial \frac{\partial f}{\partial x}}{\partial y} - \frac{\partial \frac{\partial f}{\partial y}}{\partial x} \\ \frac{\partial \frac{\partial f}{\partial x}}{\partial z} - \frac{\partial \frac{\partial f}{\partial z}}{\partial x} \\ \frac{\partial \frac{\partial f}{\partial y}}{\partial z} - \frac{\partial \frac{\partial f}{\partial z}}{\partial y} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix},$$

wegen der zweimaligen stetigen Differenzierbarkeit aller beteiligten Funktionen.

$$\begin{aligned} \text{div}(\text{röt} \vec{f}) &= \frac{\partial}{\partial x} \left(\frac{\partial f_3}{\partial y} - \frac{\partial f_2}{\partial z} \right) + \frac{\partial}{\partial y} \left(\frac{\partial f_1}{\partial z} - \frac{\partial f_3}{\partial x} \right) + \frac{\partial}{\partial z} \left(\frac{\partial f_2}{\partial x} - \frac{\partial f_1}{\partial y} \right) \\ &= \frac{\partial^2 f_3}{\partial x \partial y} - \frac{\partial^2 f_2}{\partial x \partial z} + \frac{\partial^2 f_1}{\partial y \partial z} - \frac{\partial^2 f_3}{\partial y \partial x} + \frac{\partial^2 f_2}{\partial z \partial x} - \frac{\partial^2 f_1}{\partial z \partial y} = 0, \end{aligned}$$

wieder nach dem SCHWARZschen Satz.

Eher ungewöhnlich aber manchmal abkürzend, wendet man den LAPLACE-Operator auch auf Vektorfelder an; natürlich koordinatenweise:

$$\Delta \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_n \end{pmatrix} := \begin{pmatrix} \Delta f_1 \\ \Delta f_2 \\ \vdots \\ \Delta f_n \end{pmatrix}.$$

Mit dieser Bezeichnung ergibt sich

$$(4) \quad \text{röt}(\text{röt} \vec{f}) = \text{grad}(\text{div} \vec{f}) - \Delta \vec{f}.$$

Hier muß man nur nach Definition ausrechnen und beide Seiten vergleichen; das machen Sie mal bei Regenwetter.

6.2.7 Vornehme Interpretation

Betrachtet man grad , div und röt als lineare Abbildungen der entsprechenden Vektorräume \mathcal{S} bzw. \mathcal{V} , so sagen die Aussagen (1) und (2) gerade

$$\text{Im}(\text{grad}) \subseteq \text{Ker}(\text{röt}) \quad \text{und} \quad \text{Im}(\text{röt}) \subseteq \text{Ker}(\text{div}).$$

Im Rest dieses Kapitels werden wir hauptsächlich mit dem Versuch beschäftigt sein, die Gleichheit der Kerne und Bilder nachzuweisen. Wir werden das Problem unten in einer weniger abgehobene Sprache formulieren und dann erkennen Sie sofort, warum das wichtig ist. Es erweist sich, daß die Antwort von der 'Geometrie' des bisher bescheiden im Hintergrund gebliebenen Definitionsbereiches D abhängt.

6.2.8 Kleine Rechnungen

Hier stellen wir kurz einige Formeln zusammen, die in späteren Beispielen benutzt werden. Sie betreffen den (dreidimensionalen!) Ortsvektor \vec{r} und seine Norm r . Zunächst ist (überall wo der Nenner definiert ist)

$$\frac{\partial r}{\partial x} = \frac{\partial}{\partial x} \sqrt{x^2 + y^2 + z^2} = \frac{2x}{2\sqrt{x^2 + y^2 + z^2}} = \frac{x}{r} \quad \text{analog} \quad \frac{\partial r}{\partial y} = \frac{y}{r} \quad \text{und} \quad \frac{\partial r}{\partial z} = \frac{z}{r}.$$

Weiter gilt

$$\frac{\partial \frac{1}{r}}{\partial x} = -\frac{1}{r^2} \cdot \frac{\partial r}{\partial x} = -\frac{x}{r^3} \quad \text{analog} \quad \frac{\partial \frac{1}{r}}{\partial y} = -\frac{y}{r^3} \quad \text{und} \quad \frac{\partial \frac{1}{r}}{\partial z} = -\frac{z}{r^3}$$

und

$$\frac{\partial^2 \frac{1}{r}}{\partial x^2} = -\frac{\partial}{\partial x} \left(\frac{x}{r^3} \right) = -\frac{r^3 - x3r^2 \frac{\partial r}{\partial x}}{r^6} = \frac{r^3 - x3r^2 \frac{x}{r}}{r^6} = -\frac{r^3 - 3x^2 r}{r^6}$$

analog

$$\frac{\partial^2 \frac{1}{r}}{\partial y^2} = \frac{r^3 - 3y^2 r}{r^6} \quad \text{und} \quad \frac{\partial^2 \frac{1}{r}}{\partial z^2} = \frac{r^3 - 3z^2 r}{r^6}.$$

Als Vektorgleichungen

$$\operatorname{grad} \frac{1}{r} = -\frac{\vec{r}}{r^3} \quad \text{und} \quad \Delta \frac{1}{r} = -\operatorname{div} \frac{\vec{r}}{r^3} = \frac{r^3 - 3x^2 r}{r^6} + \frac{r^3 - 3y^2 r}{r^6} + \frac{r^3 - 3z^2 r}{r^6} = \frac{3r^3 - 3r^3}{r^6} = 0.$$

6.3 Potentialfelder

Fast alles, was in diesem Abschnitt gesagt wird, gilt auch in \mathbb{R}^n . Da es sich dort kompakter aufschreiben läßt, werde ich n -dimensional argumentieren (bis auf weiteres ist also $D \subseteq \mathbb{R}^n$ offen und $\vec{f} : D \rightarrow \mathbb{R}^n$ ein C^1 -Vektorfeld, oft genügt sogar die Stetigkeit von \vec{f}).

Definition. Ein Vektorfeld \vec{f} heißt *Potentialfeld*, falls es ein Skalarfeld p gibt, das dann *Potential* genannt wird, so daß $\vec{f} = -\operatorname{grad} p$.

Das Minuszeichen könnte auch in p hineingenommen werden; es hat keine mathematische Bedeutung, sorgt aber dafür, daß Potentialdifferenzen als $p(\text{Anfangspunkt}) - p(\text{Endpunkt})$ geschrieben werden und nicht umgekehrt (s. unten).

6.3.1 Bemerkung zur Eindeutigkeit.

Da Gradienten von Konstanten Null sind, kann das Potential nur bis auf eine additive Konstante eindeutig bestimmt sein. Aus 3.3.3 folgt umgekehrt, daß sich je zwei Potentiale desselben Vektorfeldes auf einer zusammenhängenden offenen Menge nur um eine Konstante unterscheiden können.

6.3.2 Wegeunabhängigkeit von Kurvenintegralen durch Potentialfelder

Ist $\bar{\gamma}$ eine durch den Definitionsbereich des Potentialfeldes \vec{f} verlaufende stückweise glatte Kurve mit Anfangspunkt \bar{a} und Endpunkt \bar{b} , so gilt

$$\int_{\bar{\gamma}} \vec{f} \bullet d\vec{s} = p(\bar{a}) - p(\bar{b}).$$

Kurvenintegrale von Potentialfeldern entlang geschlossener (stückweise glatter) Kurven verschwinden: $\oint_{\bar{\gamma}} \vec{f} \bullet d\vec{s} = 0$.

Wir **beweisen** die erste Aussage zunächst für glatte Kurven, etwa $\bar{\gamma} : [u, v] \rightarrow D$. Nach Definition des Kurvenintegrals und wegen $\vec{f} = -\text{grad } p$

$$\int_{\bar{\gamma}} \vec{f} \bullet d\vec{s} = \int_u^v \vec{f}(\bar{\gamma}(t)) \bullet \dot{\bar{\gamma}}(t) dt = \int_u^v -\text{grad } p(\bar{\gamma}(t)) \bullet \dot{\bar{\gamma}}(t) dt$$

Minuszeichen vorgeholt, Integrationsgrenzen vertauscht und Kettenregel rückwärts gelesen:

$$= \int_v^u \text{grad } p(\bar{\gamma}(t)) \bullet \dot{\bar{\gamma}}(t) dt = \int_v^u \frac{d}{dt} p(\bar{\gamma}(t)) dt = p(\bar{\gamma}(u)) - p(\bar{\gamma}(v)) = p(\bar{a}) - p(\bar{b}).$$

Ist die Kurve nur stückweise glatt, so kann man die gewonnene Formel auf jedes der endlich vielen Stücke anwenden. Wenn man dann addiert, heben sich die Zwischenpunkte wieder auf. Führt etwa $\bar{\alpha}$ glatt von \bar{a} nach \bar{b} und $\bar{\beta}$ glatt von \bar{b} nach \bar{c} , so folgt:

$$\int_{\bar{\alpha}\bar{\beta}} \vec{f} \bullet d\vec{s} = \int_{\bar{\alpha}} \vec{f} \bullet d\vec{s} + \int_{\bar{\beta}} \vec{f} \bullet d\vec{s} = (p(\bar{a}) - p(\bar{b})) + (p(\bar{b}) - p(\bar{c})) = p(\bar{a}) - p(\bar{c}).$$

Bei geschlossenen Kurven stimmen Anfangs- und Endpunkt überein; die Potentialdifferenz wird Null.

6.3.3 Zusatz: Eine geometrisch-physikalische Folgerung

betrifft *Feldlinien* von Potentialfeldern. Darunter versteht man solche regulären Kurven durch den Definitionsbereich eines Vektorfeldes, deren Tangentenvektoren parallel zu den Vektoren des Feldes sind: $\dot{\bar{\gamma}}(t) = \lambda_t \vec{f}(\bar{\gamma}(t))$ für alle t , wobei $\lambda_t > 0$.

Beobachtung. Für Potentialfelder gibt es keine geschlossenen Feldlinien.

Wäre nämlich $\bar{\gamma}$ eine solche, so müßte $\oint \vec{f} \bullet d\vec{s} = 0$ sein. Andererseits wäre der Integrand in $\int_a^b \vec{f}(\bar{\gamma}(t)) \bullet \dot{\bar{\gamma}}(t) dt$ überall strikt positiv. Das geht nicht.

6.3.4 Die Umkehrung

Es erweist sich, daß die gerade bewiesene Wegeunabhängigkeit der Kurvenintegrale charakteristisch für Potentialfelder ist. Ist $\oint_{\bar{\gamma}} \vec{f} \bullet d\vec{s} = 0$ für alle durch D führenden stückweise glatten geschlossenen Kurven, so muß \vec{f} ein Potentialfeld sein. Der Beweis wird zeigen, daß die Wegeunabhängigkeit nur für Polygonzüge gefordert werden muß (achten Sie darauf).

Beweis. Wenn Kurvenintegrale entlang geschlossener Kurven verschwinden, so liefern Kurvenintegrale über Kurven mit gemeinsamem Anfangs- und Endpunkt identische Ergebnisse. Seien nämlich $\bar{\alpha}$ und $\bar{\beta}$ zwei derartige Kurven, so ist $\bar{\alpha}\bar{\beta}^-$ geschlossen, daher

$$0 = \oint_{\bar{\alpha}\bar{\beta}^-} \vec{f} \bullet d\vec{s} = \int_{\bar{\alpha}} \vec{f} \bullet d\vec{s} + \int_{\bar{\beta}^-} \vec{f} \bullet d\vec{s} = \int_{\bar{\alpha}} \vec{f} \bullet d\vec{s} - \int_{\bar{\beta}} \vec{f} \bullet d\vec{s} \Rightarrow \int_{\bar{\alpha}} \vec{f} \bullet d\vec{s} = \int_{\bar{\beta}} \vec{f} \bullet d\vec{s}.$$

Wir können also unzweideutig $\int_{\bar{a}}^{\bar{b}} \vec{f} \bullet d\vec{s}$ schreiben, solange \bar{a} und \bar{b} durch eine in D verlaufende stückweise glatte Kurve verbunden werden können.

Sei \bar{a} ein fester Anfangspunkt aus D und \bar{x} ein 'laufender' Punkt. Gäbe es ein Potential p , so wäre $\int_{\bar{a}}^{\bar{x}} \vec{f} \bullet d\vec{s} = p(\bar{a}) - p(\bar{x})$. Lesen wir diese Gleichung rückwärts, so ergibt sich $p(\bar{x}) = p(\bar{a}) - \int_{\bar{a}}^{\bar{x}} \vec{f} \bullet d\vec{s}$. Da p sowieso nur bis auf eine additive Konstante bestimmt ist (die beim grad wegfällt), können wir den Wert $p(\bar{a}) = 0$ annehmen. Dann kennen wir aber alle Daten auf der rechten Seite der letzten Gleichung und p wäre gleich dem negativen Kurvenintegral. Wenn es also ein p gibt, so ist klar wie es aussehen muß. Prüfen wir, daß dieser Ansatz funktioniert. Bei gegebenem \vec{f} und \bar{a} setzen wir

$$q(\bar{x}) = \int_{\bar{a}}^{\bar{x}} \vec{f} \bullet d\vec{s}$$

und prüfen grad $q = \vec{f}$ also $\frac{\partial q}{\partial x_i} = f_i$ für alle i . Dann wird $p = -q$ gesetzt, um dem konventionellen Minuszeichen Rechnung zu tragen. Die partielle Ableitung wird nach Definition ausgerechnet:

$$\frac{\partial q}{\partial x_i}(\bar{x}) = \lim_{h \rightarrow 0} \frac{q(\bar{x} + h\vec{e}_i) - q(\bar{x})}{h}.$$

Um den Differenzenquotienten zu bestimmen benutzen wir zur Berechnung von $q(\bar{x} + h\vec{e}_i)$ eine Kurve die zunächst von \bar{a} nach \bar{x} führt und danach geradlinig gleichförmig von \bar{x} nach $\bar{x} + h\vec{e}_i$. Der erste Teil des Kurvenintegrals liefert gerade $q(\bar{x})$ und fliegt daher beim Differenzenquotienten wieder heraus. Das Kurvenintegral über die verbleibende Strecke können wir sofort hinschreiben (der Geschwindigkeitsvektor auf der Strecke ist \vec{e}_i):

$$\frac{q(\bar{x} + h\vec{e}_i) - q(\bar{x})}{h} = \frac{1}{h} \int_0^h \vec{f}(\bar{x} + t\vec{e}_i) \bullet \vec{e}_i dt = \frac{1}{h} \int_0^h f_i(\bar{x} + t\vec{e}_i) dt = f_i(\bar{x} + \tau\vec{e}_i),$$

für einen geeigneten Wert τ zwischen 0 und h , der vom Mittelwertsatz der Integralrechnung geliefert wird. Wenn $h \rightarrow 0$, so muß der Zwischenpunkt gegen \bar{x} konvergieren und wir erhalten $\frac{\partial q}{\partial x_i}(\bar{x}) = f_i(\bar{x})$, wie verlangt.

Die Sache hat einen kleinen Haken, weil D vielleicht nicht zusammenhängend ist. Wir haben $q(\bar{x})$ erfolgreich für alle \bar{x} definiert, die mit \bar{a} durch eine Kurve in D verbunden werden können. Diese bilden eine zusammenhängende offene (weil D offen ist) Teilmenge von D , eine sogenannte Zusammenhangskomponente. Wenn D nicht zusammenhängend ist, gibt es mehrere derartige Komponenten. Die gesuchte Funktion wird dann auf jeder Zusammenhangskomponente von D definiert, indem man jeweils einen Startpunkt wählt und wie oben verfährt. Das klappt, denn $\frac{\partial q}{\partial x_i} = f_i$ ist eine lokale Eigenschaft. Die braucht jedesmal nur zu wissen, wie sich \vec{f} und q in einer kleinen Umgebung des gerade betrachteten Punktes verhalten und diese Umgebungen sind jeweils in einer Komponente enthalten.

6.3.5 Die Integrabilitätsbedingungen

Ist das stetig differenzierbare Vektorfeld \vec{f} Gradient eines Skalarfeldes q , so erhalten wir nach SCHWARZ

$$\frac{\partial f_i}{\partial x_j} = \frac{\partial}{\partial x_j} \frac{\partial q}{\partial x_i} = \frac{\partial}{\partial x_i} \frac{\partial q}{\partial x_j} = \frac{\partial f_j}{\partial x_i}.$$

diese sogenannte *Integrabilitätsbedingung* ist also notwendig für Potentialfelder.

6.3.6 Für sternförmige Mengen sind die Integrabilitätsbedingungen auch hinreichend.

Eine Teilmenge M von \mathbb{R}^n wird *sternförmig* genannt, wenn es einen Punkt $\bar{a} \in M$ gibt, dessen Verbindungsstrecke zu jedem anderen Punkt aus M ganz in M liegt. Anschaulich: von \bar{a} aus kann man alle anderen Punkte aus M sehen.

$D \subseteq \mathbb{R}^n$ sei offen und sternförmig und $\vec{f} : D \rightarrow \mathbb{R}^n$ erfülle die Integrabilitätsbedingung $\frac{\partial f_i}{\partial x_j} = \frac{\partial f_j}{\partial x_i}$ für alle i, j . Dann ist \vec{f} ein Potentialfeld.

Wir nehmen der Einfachheit halber an, daß alle Punkte aus D vom Nullpunkt aus gesehen werden können (sonst verschiebt man). Wir können also folgendermaßen ansetzen:

$$q(\bar{x}) = \int_0^1 \sum_{i=1}^n f_i(tx_1, tx_2, \dots, tx_n) x_i dt.$$

Das ist durchaus naheliegend, weil dasselbe wie $\int_{\vec{\gamma}} \vec{f} \bullet d\vec{s}$ für die spezielle Kurve (Strecke von $\vec{0}$ nach \bar{x}) $\vec{\gamma}(t) = (tx_1, \dots, tx_n)$, die nach Voraussetzung ganz in D verläuft.

Nun muß 'nur' noch $\frac{\partial q}{\partial x_i}(\bar{x}) = f_i(\bar{x})$ nachgewiesen werden. Ich mache das für $i = 1$:

Differenzieren unter dem Integralzeichen liefert:

$$\begin{aligned}
 \frac{\partial q}{\partial x_1} &= \frac{\partial}{\partial x_1} \int_0^1 \sum_{i=1}^n f_i(tx_1, tx_2, \dots, tx_n) x_i dt = \int_0^1 \sum_{i=1}^n \frac{\partial}{\partial x_1} [f_i(tx_1, tx_2, \dots, tx_n) x_i] dt \\
 &= \int_0^1 \left\{ \frac{\partial}{\partial x_1} [f_1(tx_1, tx_2, \dots, tx_n) x_1] + \sum_{i=2}^n \frac{\partial}{\partial x_1} [f_i(tx_1, tx_2, \dots, tx_n) x_i] \right\} dt \\
 &\quad \text{erster Summand nach Produktregel} \\
 &= \int_0^1 \left\{ tx_1 \frac{\partial f_1}{\partial x_1}(tx_1, \dots, tx_n) + f_1(tx_1, tx_2, \dots, tx_n) + \sum_{i=2}^n tx_i \frac{\partial f_i}{\partial x_1}(tx_1, tx_2, \dots, tx_n) \right\} dt \\
 &\quad \text{umsortiert und Integrabilitätsbedingung } \frac{\partial f_i}{\partial x_1} = \frac{\partial f_1}{\partial x_i} \\
 &= \int_0^1 f_1(tx_1, tx_2, \dots, tx_n) dt + \int_0^1 \left\{ tx_1 \frac{\partial f_1}{\partial x_1}(tx_1, tx_2, \dots, tx_n) + \sum_{i=2}^n tx_i \frac{\partial f_1}{\partial x_i}(tx_1, tx_2, \dots, tx_n) \right\} dt \\
 &= \int_0^1 f_1(tx_1, tx_2, \dots, tx_n) dt + \int_0^1 t \sum_{i=1}^n x_i \frac{\partial f_1}{\partial x_i}(tx_1, tx_2, \dots, tx_n) dt \\
 &\quad \text{im zweiten Integral Kettenregel rückwärts} \\
 &= \int_0^1 f_1(tx_1, tx_2, \dots, tx_n) dt + \int_0^1 t \frac{d}{dt} f_1(tx_1, tx_2, \dots, tx_n) dt.
 \end{aligned}$$

Schreibt man das als ein Integral, so sieht man weiter:

$$\begin{aligned}
 \int_0^1 \left[f_1(tx_1, tx_2, \dots, tx_n) + t \frac{d}{dt} f_1(tx_1, tx_2, \dots, tx_n) \right] dt &= \int_0^1 \frac{d}{dt} [t f_1(tx_1, tx_2, \dots, tx_n)] dt \\
 &= 1 \cdot f_1(1x_1, 1x_2, \dots, 1x_n) - 0 \cdot f_1(0x_1, 0x_2, \dots, 0x_n) = f_1(x_1, x_2, \dots, x_n),
 \end{aligned}$$

wie behauptet.

Im dreidimensionalen Raum läßt sich das Ergebnis insofern verbessern, als daß von D weniger verlangt werden muß.

6.3.7 Beweis eines noch nicht formulierten Satzes

Wir betrachten im Rest des Abschnittes wieder dreidimensionale Vektorfelder die auf dreidimensionalen offenen Mengen definiert sind. Offensichtlich kann man die Integrabilitätsbedingung auch als $\frac{\partial f_i}{\partial x_j} - \frac{\partial f_j}{\partial x_i} = 0$ schreiben. Im dreidimensionalen Fall sind diese Differenzen aber gerade die Koordinaten von $\vec{\text{rot}} \vec{f}$. Daher liest sich die dreidimensionale Integrabilitätsbedingung:

*Potentialfelder erfüllen $\vec{\text{rot}} = \vec{0}$. Derartige Felder nennt man *wirbelfrei*.*

(Das ist natürlich gerade die Formel $\vec{\text{rot}} \vec{\text{grad}} = \vec{0}$, die wir oben schon nachgerechnet hatten.)

Wir wollen jetzt beweisen, daß ein wirbelfreies Vektorfeld $\vec{f} : D \rightarrow \mathbb{R}^3$ Potentialfeld sein muß. Dazu muß $\oint_{\vec{\gamma}} \vec{f} \bullet d\vec{s} = 0$ für jede in D verlaufende (stückweise glatte) geschlossenen Kurve nachgewiesen werden. Sei $\vec{\gamma}$ eine solche Kurve. Sie berandet ein Flächenstück Γ und wir können den STOKESSchen Satz anwenden:

$$\iint_{\Gamma} \vec{\text{rot}} \vec{f} \bullet d\vec{F} = \oint_{\vec{\gamma}} \vec{f} \bullet d\vec{s}.$$

Da $\vec{\text{rot}} = \vec{0}$, ist aber das Oberflächenintegral Null, daher auch das Kurvenintegral und wir sind fertig.

Der Haken. Es stimmt zwar, daß jede stückweise glatte geschlossenen Kurve in \mathbb{R}^3 eine stückweise glatte Fläche berandet³⁷. Um STOKES anzuwenden, brauchen wir aber eine Fläche, die *in D enthalten ist*. Die gibt es nicht immer.

³⁷Das ist bei näherem Hinsehen nicht wirklich klar. Aber beim Beweis von 6.3.4 hatte ich angemerkt, daß die Wegeunabhängigkeit für Polygonzüge ausreicht. Geschlossene Polygonzüge in \mathbb{R}^3 beranden aber immer eine Fläche, nämlich ein aus lauter Dreiecken zusammengesetztes Polyeder (Induktion nach der Anzahl der Strecken des Polygons).

6.3.8 Ein Beispiel

das sofort einleuchtet, liefert die Einheitskreislinie in der x, y -Ebene von \mathbb{R}^3 . Klar berandet sie den Vollkreis dieser Ebene, die obere und untere Hemisphäre in \mathbb{R}^3 und noch viele andere 'Hauben'. Aber jede dieser Oberflächen wird irgendwo von der z -Achse durchstoßen. Für $D = \mathbb{R}^3 \setminus z$ -Achse berandet die Kreislinie keine in D enthaltene Fläche mehr.

Tatsächlich gibt es wirbelfreie Vektorfelder auf diesem D , die keine Potentialfelder sind. Ein **Beispiel** ist

$$\vec{f}(x, y, z) = \begin{pmatrix} \frac{-y}{x^2+y^2} \\ \frac{x}{x^2+y^2} \\ 0 \end{pmatrix}.$$

Die Integrabilitätsbedingung reduziert sich hier auf

$$\frac{\partial}{\partial y} \frac{-y}{x^2+y^2} = \frac{\partial}{\partial x} \frac{x}{x^2+y^2}$$

und ist leicht verifiziert. Der Einheitskreis in der x, y -Ebene hat die auf $[0, 2\pi]$ definierte Parametrisierung $t \mapsto (\cos t, \sin t, 0)$. Daher ist das Kurvenintegral um diesen Kreis

$$\oint = \int_0^{2\pi} \left(\frac{-\sin t}{\cos^2 t + \sin^2 t} \cos' t + \frac{\cos t}{\cos^2 t + \sin^2 t} \sin' t \right) dt = \int_0^{2\pi} 1 dt = 2\pi \neq 0.$$

Also haben wir kein Potentialfeld vor uns. Ursache für die Diskrepanz ist der große Wirbel um die z -Achse, den \vec{rot} als lokaler Operator nicht sieht, der aber geschlossene Feldlinien (Kreise) liefert. Sobald man D derart verkleinert, daß keine geschlossenen Kurven in D mehr die z -Achse umrunden können, wird \vec{f} zum Potentialfeld. Z.B. ist $\arctan \frac{y}{x}$ Potential, aber nicht auf ganz D sondern nur für $x \neq 0$.

6.3.9 Einfach zusammenhängende Mengen

Um das oben gegebene Argument zu retten, werden wir das Hindernis durch eine zusätzliche Voraussetzung aus dem Weg räumen.

D soll so beschaffen sein, daß jede in D verlaufende geschlossene Kurve eine in D liegende zweiseitige Fläche berandet. Mathematisch exakter kann man fordern, daß sich jede stetige Abbildung $\tilde{\gamma} : S_0(1) \rightarrow D$ der Einheitskreislinie in D zu einer stetigen Abbildung Φ des Vollkreises $K_0(1) \rightarrow D$ fortsetzen läßt. $\tilde{\gamma}$ repräsentiert die gegebene geschlossene Kurve; Φ parametrisiert die von $\tilde{\gamma}$ berandete Fläche.

Derartige Mengen nennt man *einfach zusammenhängend*³⁸. Anschaulich charakterisiert man sie durch die Forderung, daß jede in D verlaufende geschlossene Kurve sich innerhalb von D stetig auf einen Punkt zusammenziehen läßt. Die bei diesem Prozeß von der Kurve überstrichene Punktmenge ist die berandete Fläche.

Beispiele für einfach zusammenhängende Mengen sind ganz \mathbb{R}^3 , $\mathbb{R}^3 \setminus$ endlich viele Punkte, Kugeln und Sphären, konvexe Mengen und etwas allgemeiner sternförmige Mengen.

Nicht einfach zusammenhängend sind $\mathbb{R}^3 \setminus$ Gerade, $\mathbb{R}^3 \setminus$ geschlossene Kurve, Torus, geschlossene Kurve.

6.3.10 Der oben bereits bewiesene Satz lautet

Jedes auf einer offenen einfach zusammenhängenden Menge in \mathbb{R}^3 definierte wirbelfreie C^1 -Feld ist ein Potentialfeld.

³⁸Die Mathematiker fordern normalerweise zusätzlich, daß die Menge zusammenhängend ist; für unsere Zwecke ist das unerheblich.

6.3.11 Ergänzung für Praktiker

In vielen Fällen ist das Vektorfeld, für das ein Potential gesucht wird, im ganzen Raum gegeben. Dann kann man alternativ zum Verfahren aus 6.3.6 einfacher rechnen. Ich mache das im Dreidimensionalen vor.

Für das Vektorfeld $\begin{pmatrix} P \\ Q \\ R \end{pmatrix}$ suchen wir ein Potential U auf ganz \mathbb{R}^3 . Damit wir überhaupt Chancen haben, müssen wir Wirbelfreiheit voraussetzen:

$$(wf) \quad \frac{\partial P}{\partial y} = \frac{\partial Q}{\partial x}, \quad \frac{\partial P}{\partial z} = \frac{\partial R}{\partial x} \quad \text{und} \quad \frac{\partial Q}{\partial z} = \frac{\partial R}{\partial y}.$$

Nachdem ein Punkt $(x_0, y_0, z_0) \in D$ fixiert ist (meist nimmt man $(0, 0, 0)$), läßt sich ein Potential hinschreiben (das ist auch ein Kurvenintegral $(x_0, y_0, z_0) \rightsquigarrow (x, y, z)$, aber diesmal nicht geradlinig gleichförmig sondern auf achsenparallelem Umweg):

$$U(x, y, z) = \int_{x_0}^x P(t, y, z) dt + \int_{y_0}^y Q(x_0, t, z) dt + \int_{z_0}^z R(x_0, y_0, t) dt.$$

Beim Nachrechnen benutzen wir den Hauptsatz der Differential- und Integralrechnung, das Heranziehen der Ableitungen und, an den mit $=^{wf}$ den gekennzeichneten Stellen, die Integrierbarkeitsbedingung.

$$\frac{\partial U}{\partial x} = \frac{\partial}{\partial x} \int_{x_0}^x P(t, y, z) dt + \frac{\partial}{\partial x} \int_{y_0}^y Q(x_0, t, z) dt + \frac{\partial}{\partial x} \int_{z_0}^z R(x_0, y_0, t) dt = P(x, y, z) + 0 + 0,$$

denn der zweite und der dritte Summand von U hängen nicht von x ab.

$$\begin{aligned} \frac{\partial U}{\partial y} &= \frac{\partial}{\partial y} \int_{x_0}^x P(t, y, z) dt + \frac{\partial}{\partial y} \int_{y_0}^y Q(x_0, t, z) dt + \frac{\partial}{\partial y} \int_{z_0}^z R(x_0, y_0, t) dt \\ &= \int_{x_0}^x \frac{\partial P}{\partial y}(t, y, z) dt + Q(x_0, y, z) + 0 =^{wf} \int_{x_0}^x \frac{\partial Q}{\partial x}(t, y, z) dt + Q(x_0, y, z) \\ &= \left(Q(x, y, z) - Q(x_0, y, z) \right) + Q(x_0, y, z) = Q(x, y, z). \end{aligned}$$

Schließlich

$$\begin{aligned} \frac{\partial U}{\partial z} &= \frac{\partial}{\partial z} \int_{x_0}^x P(t, y, z) dt + \frac{\partial}{\partial z} \int_{y_0}^y Q(x_0, t, z) dt + \frac{\partial}{\partial z} \int_{z_0}^z R(x_0, y_0, t) dt \\ &= \int_{x_0}^x \frac{\partial P}{\partial z}(t, y, z) dt + \int_{y_0}^y \frac{\partial Q}{\partial z}(x_0, t, z) dt + R(x_0, y_0, z) \\ &=^{wf} \int_{x_0}^x \frac{\partial R}{\partial x}(t, y, z) dt + \int_{y_0}^y \frac{\partial R}{\partial y}(x_0, t, z) dt + R(x_0, y_0, z) \\ &= \left(R(x, y, z) - R(x_0, y, z) \right) + \left(R(x_0, y, z) - R(x_0, y_0, z) \right) + R(x_0, y_0, z) = R(x, y, z). \end{aligned}$$

In Wahrheit rechnet der Praktiker das Potential mit unbestimmten Integralen aus. Zuerst wird $a(x, y, z)$ derart gefunden, daß $\frac{\partial a}{\partial x} = P$. Dazu werden y und z wie Konstanten behandelt und a als x -Stammfunktion von P gesucht. Danach sucht man auf ähnliche Weise $b(y, z)$, so daß $\frac{\partial b}{\partial y} = Q - \frac{\partial a}{\partial y}$ (wegen (wf) hängt die rechte Seite nicht von x ab). Schließlich wird $c(z)$ als Stammfunktion von $R(x, y, z) - \frac{\partial a}{\partial z} - \frac{\partial b}{\partial z}$ bestimmt (der Ausdruck hängt nicht von x, y ab). Das gesuchte Potential ist $U = a + b + c$.

6.4 Quellenfreiheit und Vektorpotential

Vektorfelder, deren Divergenz Null ist nennt man *solenoidal*³⁹. Ich benutze lieber das traditionelle deutsche Wort *quellenfrei*. Ein solches Feld hat weder Quellen noch Senken. Ist das Vektorfeld \vec{f} von der Form $\text{rot } \vec{g}$, so nennt man \vec{g} ein *Vektorpotential* von \vec{f} .

Wir hatten oben nachgerechnet, daß $\text{div rot } \vec{g} = 0$ für alle Vektorfelder \vec{g} gilt. Also sind Vektorfelder mit Vektorpotential quellenfrei. Ob die Umkehrung gilt, hängt von der Geometrie des Definitionsbereiches ab. Bevor wir darauf eingehen, eine

6.4.1 Bemerkung zur Eindeutigkeit des Vektorpotentials

Wegen $\text{rot grad } = \vec{0}$ ist jedes Vektorpotential höchstens bis auf einen Summanden $\text{grad } p$, d.h. bis auf ein Potentialfeld eindeutig bestimmt. Die Ergebnisse des vorigen Abschnittes zeigen, daß die Differenz zweier Vektorpotentiale des gleichen Feldes auf einer einfach zusammenhängenden offenen Menge tatsächlich ein Potentialfeld sein muß.

6.4.2 Eine notwendige Bedingung

So wie Kurvenintegrale von Potentialfeldern längs geschlossener Kurven verschwinden, so verschwinden Flußintegrale von Feldern mit Vektorpotential über geschlossenen Oberflächen. Um das einzusehen, sei $\vec{f} = \text{rot } \vec{g}$ und Γ eine geschlossene Fläche in \mathbb{R}^3 . Indem wir eine geschlossene Kurve $\vec{\gamma}$ innerhalb Γ betrachten, zerlegen wir die Fläche in zwei Hauben Γ_+ und Γ_- (so wie der Äquator die Sphäre in die beiden Hemisphären zerlegt). Auf beide ist der STOKESSche Satz anwendbar. Da die Normale in beiden Fällen nach außen zeigen soll, wird aber die Randkurve $\vec{\gamma}$ für Γ_+ und Γ_- in entgegengesetzten Richtungen durchlaufen. Addiert man also die beiden STOKESSchen Formeln

$$\iint_{\Gamma_+} \text{rot } \vec{g} \bullet d\vec{F} = \oint_{\vec{\gamma}} \vec{g} \bullet d\vec{s} \quad \text{und} \quad \iint_{\Gamma_-} \text{rot } \vec{g} \bullet d\vec{F} = \oint_{\vec{\gamma}^-} \vec{g} \bullet d\vec{s}$$

so heben sich die Kurvenintegrale auf und das Oberflächenintegral über die gesamte Fläche wird Null.

Bemerkung. Man hätte auch mit dem GAUSSschen Satz argumentieren können und $\text{div rot } = 0$ anwenden:

$$0 = \iiint_H \text{div}(\text{rot } \vec{g}) dV = \iint_{\Gamma} \text{rot } \vec{g} \bullet d\vec{F}.$$

Das setzt aber voraus, daß $\text{rot } \vec{g}$ auf dem ganzen Körper definiert ist, dessen Rand Γ darstellt, was nicht immer gegeben ist.

6.4.3 Beispiel

Wir betrachten das auf $\mathbb{R}^3 \setminus \{\vec{0}\}$ definierte Vektorfeld

$$\frac{\vec{r}(\vec{x})}{r^3(\vec{x})} = \frac{1}{(x^2 + y^2 + z^2)^{3/2}} \begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

Nach 6.2.8 ist $\text{div} \left(\frac{\vec{r}}{r^3} \right) = 0$, das Feld also quellenfrei. Die große Quelle im Nullpunkt $\notin D$ kann div nicht sehen. Wird aber über die Einheitskugel $S = \{\vec{x} : r(\vec{x}) = 1\}$ integriert, so ist die äußere Normale gleich \vec{r} daher

$$\iint_S \frac{\vec{r}}{r^3} \bullet \vec{r} dF = \iint_S \frac{1}{r} dF = \iint_S 1 dF = \text{Oberfläche von } S = 4\pi \neq 0.$$

Also gibt es kein Vektorpotential. Die Menge $\mathbb{R}^3 \setminus \{\vec{0}\}$ ist wohlgermerkt einfach zusammenhängend, jedoch nicht sternförmig: kein Punkt sieht sein Spiegelbild am Nullpunkt.

³⁹Englisch nur so; das Wort kommt von griechisch Röhre.

6.4.4 Sternförmige Definitionsbereiche sind gut

Jedes auf einem offenen sternförmigen Definitionsbereich quellenfreie Vektorfeld besitzt ein Vektorpotential.

Sei $\vec{f} : D \rightarrow \mathbb{R}^3$ das gegebene Vektorfeld mit $\operatorname{div} \vec{f} = 0$. OBdA gehen wir wieder davon aus, daß jeder Punkt $(x, y, z) \in D$ vom Nullpunkt aus gesehen wird. Dann ist $\vec{f}(tx, ty, tz)$ für alle $t \in [0, 1]$ und alle $(x, y, z) \in D$ definiert. Wir definieren zunächst das Hilfsfeld

$$\vec{c}(\bar{x}) = \int_0^1 t \vec{f}(tx, ty, tz) dt = \begin{pmatrix} \int_0^1 t f_1(tx, ty, tz) dt \\ \int_0^1 t f_2(tx, ty, tz) dt \\ \int_0^1 t f_3(tx, ty, tz) dt \end{pmatrix}$$

und rechnen nach, daß es mit \vec{f} quellenfrei ist: $\operatorname{div} \vec{c} = \frac{\partial c_1}{\partial x} + \frac{\partial c_2}{\partial y} + \frac{\partial c_3}{\partial z}$

$$= \int_0^1 \left\{ \frac{\partial}{\partial x} [t f_1(tx, ty, tz)] + \frac{\partial}{\partial y} [t f_2(tx, ty, tz)] + \frac{\partial}{\partial z} [t f_3(tx, ty, tz)] \right\} dt = \int_0^1 t^2 \operatorname{div} \vec{f}(tx, ty, tz) dt = 0.$$

Das gesuchte Vektorpotential ergibt sich als $\vec{c}(\bar{x}) \times \vec{r}(\bar{x})$, wobei $\vec{r}(\bar{x}) = \bar{x} - \bar{0}$ der Radiusvektor oder Ortsvektor des Punktes \bar{x} ist. Wir müssen also nachrechnen, daß

$$\operatorname{rot}(\vec{c} \times \vec{r}) = \operatorname{rot} \begin{pmatrix} c_2 z - c_3 y \\ c_3 x - c_1 z \\ c_1 y - c_2 x \end{pmatrix} = \vec{f}$$

Ich beschränke mich auf die erste Koordinate. Für diese ist nachzurechnen

$$\frac{\partial}{\partial y} (c_1 y - c_2 x) - \frac{\partial}{\partial z} (c_3 x - c_1 z) = f_1.$$

Die linke Seite wird ehrlich differenziert, wobei bei zwei Summanden die Produktregel zur Anwendung kommt. Dann ergibt sich

$$\left(\frac{\partial c_1}{\partial y} y + c_1 - \frac{\partial c_2}{\partial y} x \right) - \left(\frac{\partial c_3}{\partial z} x - \frac{\partial c_1}{\partial z} z - c_1 \right) = 2c_1 - \frac{\partial c_2}{\partial y} x - \frac{\partial c_3}{\partial z} x + \frac{\partial c_1}{\partial y} y + \frac{\partial c_1}{\partial z} z.$$

Wegen der Quellenfreiheit von \vec{c} können wir $-\frac{\partial c_2}{\partial y} x - \frac{\partial c_3}{\partial z} x$ durch $\frac{\partial c_1}{\partial x} x$ ersetzen und haben dann für die erste Koordinate von $\operatorname{rot}(\vec{c} \times \vec{r})$ den Ausdruck

$$2c_1 + \left(\frac{\partial c_1}{\partial x} x + \frac{\partial c_1}{\partial y} y + \frac{\partial c_1}{\partial z} z \right).$$

Setzt man für c_1 das Integral $\int_0^1 t f_1(tx, ty, tz) dt$ ein und differenziert unter dem Integralzeichen, so folgt

$$\frac{\partial c_1}{\partial x} x = x \frac{\partial}{\partial x} \int_0^1 t f_1(tx, ty, tz) dt = x \int_0^1 t \frac{\partial}{\partial x} f_1(tx, ty, tz) dt = \int_0^1 t^2 \frac{\partial f_1}{\partial x}(tx, ty, tz) \cdot x dt.$$

Analoge Formeln gelten für die beiden anderen partiellen Ableitungen. Addiert man sie alle drei, so ergibt sich

$$\frac{\partial c_1}{\partial x} x + \frac{\partial c_1}{\partial y} y + \frac{\partial c_1}{\partial z} z = \int_0^1 t^2 \left(\frac{\partial f_1}{\partial x}(tx, ty, tz) \cdot x + \frac{\partial f_1}{\partial y}(tx, ty, tz) \cdot y + \frac{\partial f_1}{\partial z}(tx, ty, tz) \cdot z \right) dt.$$

Nach rückwärts gelesener Kettenregel

$$= \int_0^1 t^2 \frac{d}{dt} f_1(tx, ty, tz) dt.$$

Partielle Integration liefert weiter

$$= t^2 f_1(tx, ty, tz) \Big|_{t=0}^{t=1} - 2 \int_0^1 t f_1(tx, ty, tz) dt = f_1(x, y, z) - 2c_1.$$

Und damit

$$2c_1 + \left(\frac{\partial c_1}{\partial x} x + \frac{\partial c_1}{\partial y} y + \frac{\partial c_1}{\partial z} z \right) = 2c_1 + (f_1(x, y, z) - 2c_1) = f_1,$$

wie behauptet.

6.4.5 Praktische Berechnung von Vektorpotentialen auf \mathbb{R}^3

Die Formel eben war relativ kompliziert. Ist der Definitionsbereich des gegebenen quellenfreien Feldes $\vec{f} = \begin{pmatrix} P \\ Q \\ R \end{pmatrix}$ ganz \mathbb{R}^3 , so kann man ein Vektorpotential in der Form $\vec{p} = \begin{pmatrix} a(x, y, z) \\ b(x, y, z) \\ 0 \end{pmatrix}$ bekommen. Die Gleichung $\text{rot } \vec{p} = \vec{f}$ zerfällt bei diesem Ansatz in die drei skalaren Gleichungen:

$$-\frac{\partial b}{\partial z} = P, \quad \frac{\partial a}{\partial z} = Q \quad \text{und} \quad \frac{\partial b}{\partial x} - \frac{\partial a}{\partial y} = R.$$

Die ersten beiden dieser Gleichungen sind leicht zu erfüllen, durch z -Stammfunktionen von $-P$ bzw. Q . Um konkret zu sein:

$$a(x, y, z) := \int_{z_0}^z Q(x, y, t) dt \quad b(x, y, z) := - \int_{z_0}^z P(x, y, t) dt,$$

für ein irgendwie gewähltes z_0 . Die letzte Gleichung wird dann allerdings normalerweise nicht erfüllt sein. Die erreicht man, indem zu a eine von z unabhängige Korrekturfunktion $\alpha(x, y)$ addiert wird. Dadurch bleibt $\frac{\partial a + \alpha}{\partial z} = \frac{\partial a}{\partial z} + 0 = Q$ unberührt, aber α kann so gewählt werden, daß

$$\frac{\partial b}{\partial x} - \frac{\partial(a + \alpha)}{\partial y} = R, \quad \text{also} \quad \frac{\partial \alpha}{\partial y} = \frac{\partial b}{\partial x} - \frac{\partial a}{\partial y} - R.$$

Das klappt, denn die rechte Seite hängt (wegen der Quellenfreiheit) nicht von z ab. Mit a und b wie oben wird nämlich $\frac{\partial b}{\partial x} - \frac{\partial a}{\partial y} - R$ zu

$$- \int_{z_0}^z \frac{\partial P}{\partial x}(x, y, t) dt - \int_{z_0}^z \frac{\partial Q}{\partial y}(x, y, t) dt - R(x, y, z) = \int_{z_0}^z \frac{\partial R}{\partial z}(x, y, t) dt - R(x, y, z) = -R(x, y, z_0)$$

Mit anderen Worten: $\alpha(x, y) := - \int_{y_0}^y R(x, t, z_0) dt$ tut, was wir wollen.

Wenn Ihnen das lieber ist, können Sie auch unabhängig von allen gegebenen Argumenten nachrechnen, daß

$$\text{rot} \begin{pmatrix} \int_{z_0}^z Q(x, y, t) dt - \int_{y_0}^y R(x, t, z_0) dt \\ - \int_{z_0}^z P(x, y, t) dt \\ 0 \end{pmatrix} = \begin{pmatrix} P \\ Q \\ R \end{pmatrix}.$$

Was muß man eigentlich vom Definitionsbereich von \vec{f} wirklich voraussetzen, damit die angegebene Konstruktion klappt?

6.5 Der Satz von HELMHOLTZ

heißt auch **Fundamentalsatz der Vektoranalysis**. Unter diesem Namen sind bei Physikern mehrere Aussagen in Umlauf, um deren genauen Beweis, sie sich meist drücken. Ich werde jedenfalls einen Teil beweisen, allerdings unter stärkeren Voraussetzungen, als man sie wirklich

braucht. Es geht um die Frage, ob und wie ein Vektorfeld eindeutig aus seiner Rotation und Divergenz rekonstruiert werden kann. Die meisten Physiker antworten spontan mit ja. Das Vektorfeld $\vec{\text{grad}}(x^2 - y^2) = \begin{pmatrix} 2x \\ -2y \\ 0 \end{pmatrix}$ zeigt aber, daß das falsch ist; seine Rotation und Divergenz verschwinden, ohne daß es darum konstant wäre. Die eindeutige Rekonstruierbarkeit ist also nur unter zusätzlichen Bedingungen gegeben. Bei einem beschränkten einfach zusammenhängenden Definitionsbereich muß eine zusätzliche Randbedingung erfüllt sein und Felder, die im ganzen Raum definiert sind, müssen für $r \rightarrow \infty$ schnell genug abklingen.

6.5.1 Eindeutigkeit im beschränkten Bereich

H und Γ seien wie immer aber H einfach zusammenhängend. \vec{f} und \vec{g} seien auf einer H umfassenden offenen Menge definierte C^1 -Felder, so daß

$$\text{auf } H: \quad \text{rot } \vec{f} = \text{rot } \vec{g} \text{ und } \text{div } \vec{f} = \text{div } \vec{g} \quad \text{sowie auf } \Gamma: \quad \vec{f} \bullet \vec{n} = \vec{g} \bullet \vec{n}.$$

Dann stimmen \vec{f} und \vec{g} auf H überein.

Mit anderen Worten: Es kann höchstens ein Vektorfeld mit vorgegebener Divergenz und Rotation geben, das auf dem Rand des betrachteten Bereiches eine vorgegebene Normalenkomponente hat.

Zum **Beweis** betrachten wir $\vec{h} := \vec{f} - \vec{g}$. Dann verschwinden Divergenz und Rotation von \vec{h} und die Normalenkomponente ist überall auf dem Rand Null. Wir zeigen, daß konstant $\vec{h} = \vec{0}$ sein muß.

Aus $\text{rot } \vec{h} = \vec{0}$ und dem einfachen Zusammenhang von H folgt $\vec{h} = \vec{\text{grad}} q$ für ein passendes Skalarfeld. Auf q wenden wir die GREENSche Formel 5.4.3 an:

$$\iiint_H \left(\|\vec{\text{grad}} q\|^2 + q \Delta q \right) d(x, y, z) = \iint_{\Gamma} q \frac{\partial q}{\partial \vec{n}} dF.$$

Mit $\vec{h} = \vec{\text{grad}} q$ und $\Delta q = \text{div}(\vec{\text{grad}} q) = \text{div } \vec{h}$, sowie $\frac{\partial q}{\partial \vec{n}} = \vec{\text{grad}} q \bullet \vec{n} = \vec{h} \bullet \vec{n}$ wird daraus

$$\iiint_H \left(\|\vec{h}\|^2 + \underbrace{q \text{div } \vec{h}}_{=0} \right) dV = \iint_{\Gamma} q \cdot \underbrace{\vec{h} \bullet \vec{n}}_{=0} dF.$$

Also ist $\iiint_H \|\vec{h}\|^2 = 0$ und daher $\vec{h} = \vec{0}$ (eine irgendwo positive stetige Funktion hätte auch ein positives Integral), wie verlangt.

6.5.2 Kontrastierendes Beispiel

Es sei H der Hohlzylinder $H := \{(x, y, z) : 0 \leq z \leq 1 \text{ und } 1 \leq x^2 + y^2 \leq 4\}$ und \vec{f} das darauf definierte Vektorfeld

$$\vec{f}(x, y, z) := \begin{pmatrix} \frac{-y}{x^2+y^2} \\ \frac{x}{x^2+y^2} \\ 0 \end{pmatrix}.$$

In 6.3.8 hatten wir für dieses \vec{f} bereits die Integrabilitätsbedingung geprüft, d.h. $\text{rot } \vec{f} = \vec{0}$. Weiter ist

$$\text{div } \vec{f} = \frac{\partial}{\partial x} \left(\frac{-y}{x^2+y^2} \right) + \frac{\partial}{\partial y} \left(\frac{x}{x^2+y^2} \right) + \frac{\partial 0}{\partial z} = \frac{y2x}{(x^2+y^2)^2} + \frac{-x2y}{(x^2+y^2)^2} = 0.$$

Auf dem Deckel bzw. dem Boden von H ist $\vec{n} = \pm \vec{e}_3$, daher $\vec{n} \bullet \vec{f} = 0$. Auf den beiden Mantelflächen ist $\vec{n} = -\vec{r}$ (innen) und $\vec{n} = \frac{1}{2}\vec{r}$ (außen). Daher

$$\vec{n} \bullet \vec{f} = \begin{cases} - \left(x \frac{-y}{x^2+y^2} + y \cdot \frac{x}{x^2+y^2} + z \cdot 0 \right) = 0 & \text{innen} \\ \frac{1}{2} \left(x \frac{-y}{x^2+y^2} + y \cdot \frac{x}{x^2+y^2} + z \cdot 0 \right) = 0 & \text{außen} \end{cases}$$

Trotzdem ist \vec{f} nicht konstant $\vec{0}$. Dieses Beispiel widerspricht nicht dem vorigen Satz, da H nicht einfach zusammenhängend ist.

6.5.3 Eindeutigkeit für den ganzen Raum

Ich formuliere etwas verwaschen:

Sind \vec{f} und \vec{g} auf ganz \mathbb{R}^3 definierte C^1 -Vektorfelder mit gleicher Rotation und gleicher Divergenz, die zudem schnell genug abklingen, so stimmen sie überein.

Im **Beweis** werde ich voraussetzen⁴⁰, daß sich die Normen von \vec{f} und \vec{g} durch $\frac{C}{r^{2+\varepsilon}}$ beschränken lassen, wobei C eine passende Konstante und $\varepsilon > 0$ ist. Dann folgt, daß auch die Differenz $\vec{h} := \vec{f} - \vec{g}$ so schnell abklingt: $\|\vec{h}(\vec{x})\| \leq \frac{C}{r^{2+\varepsilon}}$.

Wegen $\text{rot } \vec{h} = 0$ können wir wieder $\vec{h} = \text{grad } q$ für ein C^1 -Skalarfeld annehmen. Wir zeigen zunächst, daß q beschränkt sein muß. Wegen der Stetigkeit gibt es jedenfalls eine Konstante A , mit $|q(\vec{x})| \leq A$ für alle \vec{x} aus der abgeschlossenen Einheitskugel $K_{\vec{0}}(1)$. Ist dann $\vec{x} = (x, y, z)$ nicht in der Einheitskugel, also der Abstand r von \vec{x} zum Nullpunkt größer 1, so können wir schreiben

$$q(x, y, z) = q\left(\frac{x}{r}, \frac{y}{r}, \frac{z}{r}\right) + \int_1^r \frac{d}{dt} q\left(t \frac{x}{r}, t \frac{y}{r}, t \frac{z}{r}\right) dt = q\left(\frac{x}{r}, \frac{y}{r}, \frac{z}{r}\right) + \int_1^r \underbrace{\text{grad } q\left(t \frac{x}{r}, t \frac{y}{r}, t \frac{z}{r}\right)}_{=\vec{h}} \cdot \vec{r}_0 dt,$$

wobei $\vec{r}_0 = \frac{\vec{r}}{r}$ der normierte Ortsvektor von \vec{x} ist.

Der erste Summand ist betragsmäßig $\leq A$. Das Integral läßt sich durch

$$\int_1^r \left| \vec{h}\left(t \frac{x}{r}, t \frac{y}{r}, t \frac{z}{r}\right) \cdot \vec{r}_0 \right| dt \leq \int_1^r \left\| \vec{h}\left(t \frac{x}{r}, t \frac{y}{r}, t \frac{z}{r}\right) \right\| \cdot 1 dt \leq \int_1^r \frac{C}{t^{2+\varepsilon}} dt \leq C \int_1^\infty \frac{dt}{t^{2+\varepsilon}}$$

abschätzen. Damit haben wir eine von \vec{x} unabhängige obere Schranke für $|q|$ gefunden, die unten B heißen wird.

Jetzt geht das Argument richtig los. Wäre $\vec{h} \neq \vec{0}$ an irgend einer Stelle, so gäbe es eine Kugel $K_{\vec{0}}(T)$ für die $\iiint_{K_{\vec{0}}(T)} \|\vec{h}\|^2 > 0$. Andererseits gilt für alle $R > T$

$$\begin{aligned} \iiint_{K_{\vec{0}}(T)} \|\vec{h}\|^2 dV &\leq \iiint_{K_{\vec{0}}(R)} \left(\|\vec{h}\|^2 + \underbrace{q \text{div } \vec{h}}_{=0} \right) dV \\ &= \iint_{S_{\vec{0}}(R)} q \cdot \vec{h} \cdot d\vec{F} && \text{nach GREEN} \\ &= \iint_{S_{\vec{0}}(R)} q \cdot \vec{h} \cdot \vec{n} dF && \text{Definition von } d\vec{F} \\ &\leq \iint_{S_{\vec{0}}(R)} |q| \cdot \|\vec{h}\| dF && \text{nach CAUCHY-SCHWARZ} \\ &\leq \iint_{S_{\vec{0}}(R)} B \cdot \frac{C}{r^{2+\varepsilon}} dF && \text{nach Voraussetzung} \\ &= \frac{BC}{R^{2+\varepsilon}} 4\pi R^2 \longrightarrow 0 && \text{für } R \rightarrow \infty \end{aligned}$$

Damit haben wir einen Widerspruch.

6.5.4 Wie man ein Feld mit gegebener Divergenz und Rotation findet

Wir betrachten wieder nur (Skalar- und Vektor-) Felder, die auf ganz \mathbb{R}^3 definiert sind und schnell genug abklingen. Die Aufgabe besteht darin, zu einem gegebenen Vektorfeld \vec{f} (mit $\text{div } \vec{f} = \vec{0}$, sonst geht es nicht) und einem ebenfalls gegebenen Skalarfeld g , ein Vektorfeld \vec{v} derart zu finden, daß $\text{rot } \vec{v} = \vec{f}$ und $\text{div } \vec{v} = g$. Es erweist sich, daß ein derartiges Vektorfeld \vec{v} stets gefunden werden kann und zwar in der Form $\vec{v} = \text{rot } \vec{p} + \text{grad } q$, wobei das Vektorpotential \vec{p} zusätzlich noch quellenfrei gewählt werden kann.

⁴⁰Das ist mehr als man wirklich braucht.

Das folgt relativ leicht aus einer Tatsache, die wir hier nicht beweisen können: Wenn das Skalarfeld h schnell genug abklingt, so gibt es eine eindeutig bestimmte Lösung für die POISSON-Gleichung $\Delta p = h$. Derartige Lösungen haben Sie in der Elektrodynamik sicher bereits oft hingeschrieben:

$$p(\vec{x}) = \frac{-1}{4\pi} \iiint_{\mathbb{R}^3} \frac{h(\vec{u})}{\|\vec{x} - \vec{u}\|} d\vec{u}.$$

Im Moment verstehen wir allerdings (jedenfalls offiziell) noch nicht einmal, was ein über den ganzen Raum erstrecktes Integral mit einem zudem unbeschränkten Integranden überhaupt sein soll. Wenn wir von der Lösbarkeit der POISSON-Gleichungen ausgehen, kann man das gesuchte \vec{v} wie folgt finden.

Zunächst liefern $\operatorname{div} \vec{f} = 0$ und die Sternförmigkeit von \mathbb{R}^3 ein Vektorpotential für \vec{f} , also ein Vektorfeld \vec{a} mit $\vec{f} = \operatorname{rot} \vec{a}$. Wegen $\operatorname{rot} \operatorname{grad} = \vec{0}$ gilt dann auch $\operatorname{rot} (\vec{a} + \operatorname{grad} b + \operatorname{grad} q) = \vec{f}$, egal, wie die Skalarfelder b und q gewählt werden.

Zunächst wählen wir b , so daß $\operatorname{div} (\vec{a} + \operatorname{grad} b) = 0$. Dazu muß $\Delta b = -\operatorname{div} \vec{a}$ sein; nach dem zitierten Fakt⁴¹, können wir ein derartiges b finden. Dann ist $\vec{a} + \operatorname{grad} b$ ein quellenfreies Feld auf der sternförmigen Menge \mathbb{R}^3 , also hat es ein Vektorpotential: $\vec{a} + \operatorname{grad} b = \operatorname{rot} \vec{p}$. Damit schließlich $\operatorname{div} (\vec{a} + \operatorname{grad} b + \operatorname{grad} q) = \operatorname{div} (\operatorname{rot} \vec{p} + \operatorname{grad} q) = \Delta q = g$ wird, braucht q aber nur die POISSON-Gleichung $\Delta q = g$ zu erfüllen. Das geht und löst die Aufgabe. Will man noch die Bedingung $\operatorname{div} \vec{p} = 0$ erfüllen, so korrigiert man: $\vec{p} \rightsquigarrow \vec{p} + \operatorname{grad} c$ mit einem Skalarfeld c das $\Delta c = -\operatorname{div} \vec{p}$ erfüllt. Als Folgerung erhalten wir die folgende auch wieder etwas unscharf formulierte Aussage, die auch als

6.5.5 Satz von HELMHOLTZ

gilt. Jedes auf \mathbb{R}^3 definierte C^1 -Vektorfeld, das schnell genug abklingt, läßt sich als $\operatorname{rot} \vec{p} + \operatorname{grad} q$ schreiben, wobei das Vektorpotential \vec{p} zudem noch quellenfrei vorausgesetzt werden kann (sog. Eichbedingung).

Sei nämlich \vec{w} das gegebene Vektorfeld. Wir setzen $\vec{f} := \operatorname{rot} \vec{w}$ und $g := \operatorname{div} \vec{w}$. Dann sagt der vorige Satz, daß es \vec{p} und q derart gibt, daß $\vec{v} = \operatorname{rot} \vec{p} + \operatorname{grad} q$, die Bedingungen $\operatorname{rot} \vec{v} = \vec{f}$ und $\operatorname{div} \vec{v} = g$ erfüllt. Da aber \vec{w} und \vec{v} dann gleiche Rotation und Divergenz haben, stimmen sie überein.

⁴¹Na ja, eigentlich hat $\operatorname{div} \vec{a}$ keinen Grund schnell abzuklingen. Der Beweis ist nicht ganz wasserdicht.