

# Mathematik II für Physiker

Wolf P. Barth

Sommersemester 08

Department Mathematik der Universität Erlangen, Bismarckstraße 1 $\frac{1}{2}$   
e-mail: barth@mi.uni-erlangen.de

Version vom 14. Juli 2008

## Inhaltsverzeichnis

<b>1</b>	<b>Ähnliche Matrizen</b>	<b>2</b>
1.1	Diagonalisierbarkeit . . . . .	2
1.2	Invariante direkte Summenzerlegung . . . . .	9
1.3	Jordansche Normalform . . . . .	14
1.4	Anwendung: Systeme von Differentialgleichungen . . . . .	20
1.5	Bilinearformen . . . . .	30
1.6	Quadriken . . . . .	38
1.7	Innere Produkte und adjungierte Operatoren . . . . .	48
<b>2</b>	<b>Differentialrechnung mit mehreren Veränderlichen</b>	<b>59</b>
2.1	Grundlagen . . . . .	59
2.2	Differenzierbare Abbildungen . . . . .	66
2.3	Rechenregeln und Beispiele . . . . .	73
2.4	Taylor-Entwicklung . . . . .	81
2.5	Umkehrung differenzierbarer Abbildungen . . . . .	85
2.6	Implizite Funktionen . . . . .	95
2.7	Lokale Extrema . . . . .	102
2.8	Die Differentialoperatoren <i>grad</i> , <i>rot</i> , <i>div</i> und $\Delta$ . . . . .	109
<b>3</b>	<b>Gewöhnliche Differentialgleichungen</b>	<b>119</b>
3.1	Elementare Lösungsmethoden . . . . .	120
3.1.1	Trennung der Variablen . . . . .	120
3.1.2	Inhomogene lineare Differentialgleichung . . . . .	121
3.1.3	Bernoullische Differentialgleichung . . . . .	122
3.1.4	Homogene Differentialgleichung . . . . .	123
3.2	Theorie: Existenz und Eindeutigkeit, Systeme . . . . .	127
3.3	Lineare Systeme . . . . .	135
3.4	Autonome Systeme, Stabilität . . . . .	146

# 1 Ähnliche Matrizen

In diesem Kapitel möchte ich die in der Modul-Beschreibung dieser Vorlesung vorgegebenen Inhalte aus der Linearen Algebra zusammenstellen, soweit sie nicht schon im ersten Semester behandelt wurden. Im Wesentlichen handelt es sich immer darum, Matrizen durch Koordinatentransformation zu vereinfachen. Bei einer Transformation

$$T^{-1} \cdot A \cdot T$$

geht man von  $A$  über zu einer *ähnlichen* Matrix. Daher der Name dieses Kapitels. Es werden aber auch Matrizen eine Rolle spielen, die vermöge

$$T^t \cdot A \cdot T$$

transformiert werden. Nur wenn  $T$  orthogonal ist, also gilt  $T^{-1} = T^t$ , ist die transformierte Matrix ähnlich zu  $A$ . Aber unter irgend einer Überschrift muss ich halt diese Inhalte zusammenstellen.

## 1.1 Diagonalisierbarkeit

Wir betrachten hier nur noch reelle oder komplexe Matrizen. Ich werde folgende Notation verwenden:

	Menge der
$M(n \times n, \mathbb{R})$	reellen $n \times n$ -Matrizen
$M(n \times n, \mathbb{C})$	komplexen $n \times n$ -Matrizen
$GL(n, \mathbb{R})$	reellen invertierbaren $n \times n$ -Matrizen
$GL(n, \mathbb{C})$	komplexen invertierbaren $n \times n$ -Matrizen

Weiter bezeichnet  $\mathbb{1}_n$  die  $n \times n$ -Einheitsmatrix.

Meist läuft die Theorie für reelle und komplexe Matrizen völlig parallel. Nur wo es wirklich darauf ankommt, werde ich spezifizieren, ob ich reelle oder komplexe Matrizen meine.

Zwei  $n \times n$  Matrizen  $C$  und  $C'$  heißen *ähnlich* (manchmal sagt man auch äquivalent), wenn es eine invertierbare  $n \times n$ -Matrix  $T$  gibt mit

$$C' = T^{-1} \cdot C \cdot T.$$

Ähnlichkeit von Matrizen ist eine Äquivalenz-Relation, d.h., es gilt

- Satz 1.1** a) *Reflexivität: Jede Matrix ist zu sich selbst ähnlich.*  
 b) *Symmetrie: Ist  $C'$  zu  $C$  ähnlich, dann ist auch umgekehrt  $C$  zu  $C'$  ähnlich.*  
 c) *Transitivität: Ist  $C'$  zu  $C$  ähnlich sowie  $C''$  zu  $C'$ , dann ist auch  $C''$  zu  $C$  ähnlich.*

Beweise. a)  $C = \mathbb{1}_n^{-1} \cdot C \cdot \mathbb{1}_n$ .

b) Aus  $C' = T^{-1} \cdot C \cdot T$  folgt  $C = T \cdot C' \cdot T^{-1}$ .

c) Es sei  $C' = T^{-1} \cdot C \cdot T$  und  $C'' = U^{-1} \cdot C' \cdot U$ . Dann ist

$$C'' = U^{-1} T^{-1} \cdot C \cdot T U = (TU)^{-1} \cdot C \cdot (TU).$$

□

Koordinatentransformationen im  $\mathbb{R}^n$  führen auf ähnliche Matrizen. Ist nämlich  $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$  ein Vektor, so sind die  $x_\nu, \nu = 1, \dots, n$  seine Koordinaten bezüglich der kanonischen Basis  $\mathbf{e}_1, \dots, \mathbf{e}_n$  des  $\mathbb{R}^n$ . Sein Koordinaten  $y_1, \dots, y_n$  bezüglich einer neuen Basis  $\mathbf{t}_1, \dots, \mathbf{t}_n$  des  $\mathbb{R}^n$  erhält man wie folgt: Es ist

$$\mathbf{x} = y_1 \mathbf{t}_1 + \dots + y_n \mathbf{t}_n.$$

Wenn ich Vektoren als Spaltenvektoren meine, kann ich die letzte Gleichung schreiben

$$\mathbf{x} = T \cdot \mathbf{y} \quad \text{bzw.} \quad \mathbf{y} = T^{-1} \cdot \mathbf{x}$$

mit der invertierbaren Transformationsmatrix

$$T = (\mathbf{t}_1, \dots, \mathbf{t}_n),$$

welche die neuen Basisvektoren als Spalten enthält.

Auf Matrizen von linearen Abbildungen  $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}^n$  schlägt der Übergang in neue Koordinaten wie folgt durch: Die Abbildung sei

$$\varphi : \mathbf{x} \mapsto \varphi(\mathbf{x}) = C \cdot \mathbf{x}$$

mit der darstellenden Matrix  $C$ . Die Koordinaten des Bildvektors  $\varphi(\mathbf{x})$  in der neuen Basis bilden den Spaltenvektor

$$T^{-1} \cdot \varphi(\mathbf{x}) = T^{-1} \cdot C \cdot \mathbf{x} = (T^{-1} \cdot C \cdot T) \cdot \mathbf{y}.$$

Im neuen Koordinatensystem wird die Abbildung  $\varphi$  dargestellt durch die zu  $C$  ähnliche Matrix  $T^{-1} \cdot C \cdot T$ .

Viele lineare Probleme, die durch eine Matrix definiert sind, kann man vereinfachen, indem man in ein neues Koordinatensystem übergeht, wo die erwähnte Matrix eine einfachere Form annimmt. Die einfachst denkbare Form ist die Diagonalform.

**Definition 1.1** Die  $n \times n$ -Matrix  $C$  heißt diagonalisierbar, wenn sie ähnlich zu einer Diagonalmatrix  $D$  ist.

Eine lineare Abbildung  $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}^n$  hat genau dann eine Diagonalmatrix

$$D = \begin{pmatrix} d_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & d_n \end{pmatrix}$$

als darstellende Matrix, wenn die kanonischen Basisvektoren  $\mathbf{e}_1, \dots, \mathbf{e}_n$  Eigenvektoren (mit den Eigenwerten  $d_1, \dots, d_n$ ) sind. Deswegen ist die darstellende Matrix  $C$  von  $\varphi$  genau dann diagonalisierbar, wenn es eine Basis  $\mathbf{t}_1, \dots, \mathbf{t}_n \in \mathbb{R}^n$  gibt, die aus Eigenvektoren für  $C$  besteht. Damit haben wir das erste, tautologische, Diagonalisierbarkeitskriterium:

**Satz 1.2** Die  $n \times n$ -Matrix  $C$  ist genau dann diagonalisierbar, wenn es eine Basis  $\mathbf{t}_1, \dots, \mathbf{t}_n$  aus Eigenvektoren für  $C$  gibt.

Dieses Kriterium ist leider weniger physikalisch von Interesse, mehr akademisch. Näher am Kern des Problems ist folgendes Kriterium:

**Satz 1.3 (Diagonalisierbarkeit, notwendig)** Wenn die Matrix  $C$  diagonalisierbar ist, dann zerfällt ihr charakteristisches Polynom

$$\chi_C(\lambda) = (\lambda_1 - \lambda) \cdot \dots \cdot (\lambda_n - \lambda)$$

in Linearfaktoren.

Beweis. Das charakteristische Polynom einer Diagonalmatrix ist ein Produkt von Linearfaktoren. Um den Satz zu beweisen, genügt es zu zeigen: Ähnliche Matrizen haben das gleiche charakteristische Polynom. Sei also  $C$  ähnlich zu  $D = T^{-1} \cdot C \cdot T$ . Dann ist

$$\begin{aligned} \chi_C(\lambda) &= \det(T^{-1}DT - \lambda\mathbb{1}_n) = \det(T^{-1} \cdot (D - \lambda\mathbb{1}_n) \cdot T) = \\ &= \det(T^{-1}) \cdot \det(D - \lambda\mathbb{1}_n) \cdot \det(T) = \det(D - \lambda\mathbb{1}_n) = \chi_D(\lambda). \end{aligned} \quad \square$$

Dies ist jetzt eine Situation, wo es darauf ankommt, ob wir sie reell oder komplex betrachten:

**Beispiel 1.1** Wir betrachten die Drehmatrix

$$C = \begin{pmatrix} \cos(\alpha) & -\sin(\alpha) \\ \sin(\alpha) & \cos(\alpha) \end{pmatrix}.$$

Ihr charakteristisches Polynom ist

$$(\cos(\alpha) - \lambda)^2 + \sin^2(\alpha)$$

mit den komplexen Nullstellen

$$\lambda_{1,2} = \cos(\alpha) \pm i \cdot \sin(\alpha).$$

Hier ist  $\sin(\alpha) = 0$  genau dann, wenn die Matrix  $C = \pm\mathbb{1}_2$  diagonal ist. Und wenn  $\sin(\alpha) \neq 0$  ist, dann hat die Matrix  $C$  keine reellen Eigenwerte. Sie ist nicht reell diagonalisierbar. Aber, wenn  $\sin(\alpha) \neq 0$  ist, dann sind die beiden komplexen Eigenwerte voneinander verschieden, und die Matrix ist komplex diagonalisierbar. Dies folgt aus

**Satz 1.4 (Diagonalisierbarkeit, hinreichend)** Das charakteristische Polynom

$$\chi_C(\lambda) = (\lambda_1 - \lambda) \cdot \dots \cdot (\lambda_n - \lambda)$$

der Matrix  $C$  zerfalle in Linearfaktoren. Sind die Eigenwerte  $\lambda_1, \dots, \lambda_n$  alle (paarweise) voneinander verschieden, dann ist  $C$  diagonalisierbar.

Beweis. Zu jedem Eigenwert  $\lambda_\nu$  gibt es einen Eigenvektor  $\mathbf{t}_\nu$ . Um zu zeigen, dass diese  $n$  Eigenvektoren eine Basis bilden, genügt es zu zeigen,  $m$  Eigenvektoren zu verschiedenen Eigenwerten sind linear unabhängig. Wir tun dies mit Induktion nach  $m$  und verwenden den bekannten Test. Sei etwa

$$\sum_{\nu=1}^m c_\nu \mathbf{t}_\nu = \mathbf{0}.$$

Induktionsanfang ( $m = 1$ ): Weil der Eigenvektor  $\mathbf{t}_1 \neq \mathbf{0}$  ist, folgt aus  $c_1 \mathbf{t}_1 = \mathbf{0}$ , dass  $c_1 = 0$  ist.

Induktionsschluss: Wir multiplizieren die Test-Gleichung mit der Matrix  $C - \lambda_m \mathbb{1}_n$ . Wegen

$$(C - \lambda_m) \cdot \mathbf{t}_m = \mathbf{0}$$

erhalten wir

$$\sum_{\nu=1}^{m-1} c_\nu (\lambda_\nu - \lambda_m) \mathbf{t}_\nu = \mathbf{0}.$$

Nach Induktionsvoraussetzung sind hier alle Koeffizienten

$$c_\nu (\lambda_\nu - \lambda_m) = 0.$$

Weil für  $\nu = 1, \dots, m-1$  vorausgesetzt ist, dass  $\lambda_\nu \neq \lambda_m$ , sehen wir  $c_1 = \dots = c_{m-1} = 0$  und dann auch  $c_m = 0$ .  $\square$

Das ist ja sehr schön, vor allem, wenn alle Eigenwerte tatsächlich voneinander verschieden sind. Aber leider ist das gerade in interessanten Situationen nicht immer der Fall. Physiker haben dafür sogar eine eigene Redeweise (entartete Eigenwerte) eingeführt.

**Definition 1.2** *Das charakteristische Polynom der Matrix  $C$*

$$\chi_C(\lambda) = (\lambda_1 - \lambda) \cdot \dots \cdot (\lambda_n - \lambda)$$

zerfalle in Linearfaktoren. Der Eigenwert  $\lambda_\nu$  heißt entartet, wenn sein Linearfaktor  $(\lambda_\nu - \lambda)$  mehr als einmal vorkommt. Die Anzahl  $r_\nu$  der vorkommenden Faktoren  $(\lambda_\nu - \lambda)$  heißt die algebraische Vielfachheit dieses Eigenwertes. Fasst man die Linearfaktoren zu entarteten Eigenwerten zusammen, so kann man das charakteristische Polynom schreiben als

$$\chi_C(\lambda) = (\lambda_1 - \lambda)^{r_1} \cdot \dots \cdot (\lambda_k - \lambda)^{r_k}, \quad r_1 + \dots + r_k = n, \quad \lambda_1, \dots, \lambda_k \text{ paarweise verschieden.}$$

Zu jedem Eigenwert  $\lambda_\nu$  von  $C$  gibt es immer Eigenvektoren  $\mathbf{v} \neq \mathbf{0}$ . Die Menge aller dieser Eigenvektoren, zusammen mit dem Nullvektor (der ja kein Eigenvektor ist), bildet den Eigenraum zum Eigenwert  $\lambda_\nu$ . Dieser Eigenraum ist der Nullraum

$$E_C(\lambda_\nu) = \ker(C - \lambda_\nu \mathbb{1}_n)$$

der Matrix  $C - \lambda_\nu \mathbb{1}_n$  und damit ein Untervektorraum des  $\mathbb{R}^n$ . Die Dimension dieses Eigenraums berechnet sich nach dem Dimensionssatz für lineare Abbildungen als

$$\dim(E_C(\lambda_\nu)) = n - \text{Rang}(C - \lambda_\nu \mathbb{1}_n).$$

**Definition 1.3** *Es sei  $\lambda_\nu$  ein Eigenwert der Matrix  $C$ . Die Dimension des zugehörigen Eigenraums  $E_C(\lambda_\nu)$  heißt die geometrische Vielfachheit des Eigenwertes  $\lambda_\nu$ .*

**Satz 1.5** *Für jeden Eigenwert ist immer*

$$\text{algebraische Vielfachheit} \geq \text{geometrische Vielfachheit}.$$

Beweis. Es sei  $r := \dim E_C(\lambda_\nu)$  die geometrische Vielfachheit des Eigenwerts  $\lambda_\nu$ . Eine Basis  $\mathbf{t}_1, \dots, \mathbf{t}_r$  des Eigenraums ergänzen wir zu einer Basis  $\mathbf{t}_1, \dots, \mathbf{t}_r, \mathbf{t}_{r+1}, \dots, \mathbf{t}_n$  des ganzen  $n$ -dimensionalen Vektorraums. Der Übergang in diese neue Basis ändert die Matrix  $C$  in die zu  $C$  ähnliche Matrix

$$r \left\{ \begin{pmatrix} \lambda_\nu & & * & \dots & * \\ & \ddots & \vdots & & \vdots \\ & & \lambda_\nu & * & * \\ & & & \vdots & \vdots \\ & & & * & * \end{pmatrix} \right.$$

Aus der Kästchenregel für die Determinante folgt, dass das charakteristische Polynom  $= \chi_C(\lambda)$  dieser Matrix den Faktor  $(\lambda_\nu - \lambda)^r$  enthält.  $\square$

Es gehört zu den Bosheiten des täglichen Lebens, dass die geometrische Vielfachheit eines Eigenwerts tatsächlich echt kleiner sein kann, als dessen algebraische Vielfachheit.

**Beispiel 1.2 (Jordan-Block)** *Wir betrachten eine  $n \times n$ -Matrix*

$$J := \begin{pmatrix} c & 1 & & & \\ & \ddots & \ddots & & \\ & & c & 1 & \\ & & & & c \end{pmatrix}$$

mit dem charakteristischen Polynom

$$\chi_J(\lambda) = (c - \lambda)^n.$$

Die Matrix hat den einzigen Eigenwert  $c$  mit der algebraischen Vielfachheit  $n$ . Nun ist

$$\text{Rang}(J - c\mathbb{1}_n) = n - 1,$$

und deswegen hat dieser Eigenwert die geometrische Vielfachheit 1. Falls  $n > 1$  ist, sind beide Vielfachheiten voneinander verschieden.

**Satz 1.6 (Diagonalisierbarkeit, notwendig und hinreichend)** *Die  $n \times n$ -Matrix  $C$  ist genau dann diagonalisierbar, wenn*

- 1) *das charakteristische Polynom der Matrix  $C$  in Linearfaktoren zerfällt, und*
- 2) *für jeden Eigenwert von  $C$  die algebraische und die geometrische Vielfachheit übereinstimmen.*

Beweis. Wir brauchen nur zu zeigen, dass aus 1) und 2) zusammen die Diagonalisierbarkeit folgt. Dazu wählen wir für jeden Eigenraum eine Basis. Wegen 2) ist die Summe der Dimensionen der Eigenräume gleich  $n$ , der Summe der algebraischen Vielfachheiten aller Eigenwerte. Die gewählten Basen enthalten zusammen also  $n$  Vektoren. Ähnlich wie beim Beweis von 1.3 zeigt man, dass diese alle linear unabhängig sind. Zusammen bilden sie also eine Basis des  $n$ -dimensionalen Vektorraums, bestehend aus Eigenvektoren für  $C$ .  $\square$

Allerdings gibt es auch eine positive Nachricht:

**Satz 1.7 (Hauptachsentransformation für reelle symmetrische Matrizen)** Die reelle  $n \times n$ -Matrix sei symmetrisch. Dann gibt es eine reelle orthogonale  $n \times n$ -Matrix  $U$  derart, dass die zu  $S$  ähnliche Matrix

$$U^{-1} \cdot S \cdot U$$

Diagonalform hat.

Herr Knop hat mir glaubhaft versichert, dass er den Beweis dieses Satzes in seiner Vorlesung 'Lineare Algebra I' noch behandelt hat. Reelle symmetrische Matrizen sind also diagonalisierbar. Weil die Transformationsmatrix  $U$  orthogonal gewählt werden kann, gibt es sogar eine ONB (Ortho-Normal-Basis) aus Eigenvektoren. Diesen Satz muss man kennen, denn nur er garantiert, dass man wirklich eine ONB aus Eigenvektoren findet, wenn man sie sucht. Und man sucht sie so, wie man halt immer Eigenvektoren sucht. Hierzu das allereinfachste Beispiel.

**Beispiel 1.3** Wir betrachten die reelle symmetrische  $2 \times 2$ -Matrix

$$S := \begin{pmatrix} -2 & 1 \\ 1 & -2 \end{pmatrix}.$$

Ihr charakteristisches Polynom

$$(-2 - \lambda)(-2 - \lambda) - 1 = (\lambda + 2)^2 - 1$$

hat die Nullstellen

$$\lambda_{1,2} = \mp 1 - 2 = -3, -1.$$

Wir berechnen zugehörige Eigenvektoren

$\lambda_i$	$S - \lambda_i \mathbb{1}_2$	$\mathbf{v}_i$
$\lambda_1 = -3$	$\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$	$\mathbf{v}_1 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$
$\lambda_2 = -1$	$\begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}$	$\mathbf{v}_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$

Nach Normierung dieser Eigenvektoren erhalten wir die orthogonale Transformationsmatrix

$$U = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}.$$

Zur Sicherheit können wir noch nachrechnen

$$U^{-1} \cdot S \cdot U = \frac{1}{2} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} -2 & 1 \\ 1 & -2 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} = \begin{pmatrix} -3 & 0 \\ 0 & -1 \end{pmatrix}.$$

**Aufgabe 1.1** Der Homomorphismus  $\varphi : \mathbb{R}^3 \rightarrow \mathbb{R}^2$  werde bezüglich der kanonischen Basen durch die Matrix

$$M = \begin{pmatrix} 0 & 2 & 2 \\ 1 & -2 & 2 \end{pmatrix}$$

beschrieben. Man berechne die Matrixdarstellung von  $\varphi$  bezüglich der Basis

$$\mathbf{a}_1 = (0, 1, 1), \quad \mathbf{a}_2 = (1, 0, 3), \quad \mathbf{a}_3 = (1, 0, 1)$$

des  $\mathbb{R}^3$  und der Basis

$$\mathbf{b}_1 = (1, 1), \quad \mathbf{b}_2 = (1, -1)$$

des  $\mathbb{R}^2$ .

**Aufgabe 1.2** Geben Sie die darstellende Matrix der linearen Abbildung

$$f : \mathbb{R}^3 \rightarrow \mathbb{R}^3, \quad \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \mapsto \begin{pmatrix} x_2 \\ x_3 \\ x_1 \end{pmatrix}$$

bezüglich der kanonischen Basis des  $\mathbb{R}^3$  an und bezüglich der Basis

$$\mathbf{a}_1 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \quad \mathbf{a}_2 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \quad \mathbf{a}_3 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \in \mathbb{R}^3.$$

**Aufgabe 1.3** Im  $\mathbb{R}^4$  seien die Vektoren

$$\mathbf{a}_1 = \begin{pmatrix} 1 \\ 2 \\ 0 \\ 0 \end{pmatrix}, \quad \mathbf{a}_2 = \begin{pmatrix} 2 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \quad \mathbf{a}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 2 \end{pmatrix}, \quad \mathbf{a}_4 = \begin{pmatrix} 0 \\ 0 \\ 2 \\ 1 \end{pmatrix}$$

gegeben. Weiter sei  $f : \mathbb{R}^4 \rightarrow \mathbb{R}^4$  eine lineare Abbildung mit

$$f(\mathbf{a}_1) = \mathbf{a}_2, \quad f(\mathbf{a}_2) = \mathbf{a}_1, \quad f(\mathbf{a}_3) = f(\mathbf{a}_4) = \mathbf{a}_3 + \mathbf{a}_4.$$

Geben Sie die darstellende Matrix von  $f$  in der kanonischen Basis des  $\mathbb{R}^4$  an.

**Aufgabe 1.4** Bestimmen Sie alle Eigenwerte und die Dimensionen aller Eigenräume für die Matrizen

$$M = \begin{pmatrix} 3 & -2 & 4 \\ 4 & -3 & 4 \\ -2 & 1 & -3 \end{pmatrix} \quad \text{und} \quad N = \begin{pmatrix} 3 & -2 & 4 \\ 3 & -3 & 2 \\ -2 & 1 & -3 \end{pmatrix}$$

und entscheiden Sie, ob  $M$  und  $N$  ähnlich sind.

**Aufgabe 1.5** In einem vierdimensionalen Vektorraum  $V$  sei  $\{\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3, \mathbf{b}_4\}$  eine Basis. Eine lineare Abbildung  $f : V \rightarrow V$  habe die Eigenschaft

$$\begin{aligned} f(\mathbf{b}_1) &= \mathbf{0}, & f(\mathbf{b}_2) &= \mathbf{b}_1 + 4\mathbf{b}_4, \\ f(\mathbf{b}_3) &= \mathbf{b}_2 - 8\mathbf{b}_4, & f(\mathbf{b}_4) &= \mathbf{b}_3 + 5\mathbf{b}_4. \end{aligned}$$

Ist  $f$  diagonalisierbar?

**Aufgabe 1.6** Entscheiden Sie, welche der folgenden Matrizen zueinander ähnlich sind und welche nicht, und begründen Sie Ihre Antwort:

$$a) \quad A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}, \quad C = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}.$$

$$b) \quad A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}, \quad B = \frac{1}{9} \begin{pmatrix} -1 & 4 & 8 \\ 4 & -7 & 4 \\ 8 & 4 & -1 \end{pmatrix}.$$

**Aufgabe 1.7** Ist die symmetrische Matrix

$$A = \begin{pmatrix} 1 & \sqrt{-1} \\ \sqrt{-1} & 1 \end{pmatrix}$$

diagonalisierbar? Wenn ja, berechne man eine Basis aus Eigenvektoren.

**Aufgabe 1.8** Für die Matrizen

$$\begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix} \quad \text{und} \quad \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ \frac{3}{2} & \frac{1}{2} & 0 & 1 \\ \frac{1}{2} & \frac{3}{2} & 1 & 0 \end{pmatrix}$$

zeige man, dass ihre Eigenwerte samt algebraischer Vielfachheit übereinstimmen, aber nur eine diagonalisierbar ist.

## 1.2 Invariante direkte Summenzerlegung

Hier betrachten wir  $n \times n$ -Matrizen  $C$ , deren charakteristisches Polynom

$$\chi_C(\lambda) = (\lambda_1 - \lambda)^{r_1} \cdot \dots \cdot (\lambda_k - \lambda)^{r_k}$$

in Linearfaktoren mit den algebraischen Vielfachheiten  $r_1, \dots, r_k$  zerfällt. Beim Rechnen mit komplexen Zahlen ist dies wegen des Fundamentalsatzes der Algebra keine Einschränkung. Aber bei reellen Matrizen ist diese Voraussetzung nicht immer erfüllt. Jede reelle Matrix ist jedoch auch eine komplexe Matrix. Also gelten die komplexen Ergebnisse dieses Abschnitts auch für reelle Matrizen. Wenn man aber reelle Ergebnisse wünscht, muss man geeignete Anpassungen vornehmen.

**Satz 1.8** Für eine  $n \times n$ -Matrix  $C$  sind äquivalent:

- Das charakteristische Polynom  $\chi_C(\lambda)$  zerfällt in Linearfaktoren.
- Die Matrix  $C$  ist ähnlich zu einer oberen Dreiecksmatrix

$$\begin{pmatrix} c_1 & * & * \\ 0 & \ddots & * \\ 0 & 0 & c_n \end{pmatrix}.$$

Beweis. Die Richtung b) $\Rightarrow$ a) ist trivial. Denn das charakteristische Polynom von  $C$  ist das charakteristische Polynom der ähnlichen Dreiecksmatrix. Und das charakteristische Polynom der Dreiecksmatrix ist das Produkt der Faktoren  $(c_1 - \lambda) \cdot \dots \cdot (c_n - \lambda)$  auf der Diagonale.

a) $\Rightarrow$ b) Wir verwenden einen Eigenvektor  $\mathbf{v}_1$  zum ersten Eigenwert  $\lambda_1$ . Ihn ergänzen wir zu einer Basis  $\mathbf{v}_1, \dots, \mathbf{v}_n$ . Die durch  $C$  beschriebene lineare Abbildung hat bezüglich dieser Basis eine Matrix der Form

$$\begin{pmatrix} \lambda_1 & * \\ 0 & C' \end{pmatrix}$$

mit einer  $(n - 1) \times (n - 1)$ -Matrix  $C'$ . Ihr charakteristisches Polynom ist

$$\chi_C(\lambda) = (\lambda_1 - \lambda) \cdot \chi_{C'}(\lambda).$$

Deswegen zerfällt auch das charakteristische Polynom von  $C'$  in Linearfaktoren.

Nach Induktionsannahme ist dann diese Matrix  $C'$  ähnlich zu einer oberen Dreiecksmatrix  $D'$ , also ist  $C' = (T')^{-1} \cdot D' \cdot T'$ . Und  $C$  ist ähnlich zur Matrix

$$\begin{pmatrix} \lambda_1 & * \\ 0 & (T')^{-1} D' T' \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & T' \end{pmatrix}^{-1} \cdot \begin{pmatrix} \lambda_1 & * \\ 0 & D' \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 \\ 0 & T' \end{pmatrix}$$

und damit ähnlich zur oberen Dreiecksmatrix

$$\begin{pmatrix} \lambda_1 & * \\ 0 & D' \end{pmatrix}.$$

□

Auf der Diagonale einer zu  $C$  ähnlichen oberen Dreiecksmatrix  $D$  stehen die Eigenwerte von  $C$  in der Reihenfolge, wie wir sie beim Induktionsbeweis verwendet haben. Wir können deswegen auch erreichen, dass gleiche Eigenwerte hintereinander kommen. D.h., auf der Diagonale stehen die Eigenwerte in einer Reihenfolge

$$\underbrace{\lambda_1, \dots, \lambda_1}_{r_1}, \dots, \underbrace{\lambda_k, \dots, \lambda_k}_{r_k}, \quad r_1 + \dots + r_k = n.$$

Die obere Dreiecksmatrix  $D$  hat also links oben ein  $r_1 \times r_1$ -Kästchen der Form

$$\begin{pmatrix} \lambda_1 & * & * \\ 0 & \ddots & * \\ 0 & 0 & \lambda_1 \end{pmatrix} = \lambda_1 \mathbb{1}_{r_1} + N.$$

Dabei ist  $N$  eine obere Dreiecksmatrix, auf deren Diagonale lauter Nulleinträge stehen. Wir analysieren deswegen solche Matrizen näher.

**Satz 1.9 (Nilpotente Matrizen)** Für eine  $r \times r$ -Matrix  $N$  sind äquivalent:

- Ihr charakteristisches Polynom ist  $\lambda^r$ .
- Sie ist ähnlich zu einer oberen Dreiecksmatrix, auf deren Diagonale nur Null-Einträge stehen.
- Es ist  $N^r = 0$ , die Nullmatrix.

Beweis. a) $\Rightarrow$ b) Alle Eigenwerte von  $N$  sind nach Voraussetzung  $= 0$ . Satz 1.8 ist anwendbar und zeigt, dass  $N$  ähnlich zu einer oberen Dreiecksmatrix ist, auf deren Diagonale die Eigenwerte  $= 0$  stehen.

b) $\Rightarrow$ c) Es sei  $D$  die zu  $N$  ähnliche Matrix aus b). Damit ist  $N = T^{-1} \cdot D \cdot T$  und

$$N^r = \underbrace{(T^{-1}DT) \cdot \dots \cdot (T^{-1}DT)}_r = T^{-1} \cdot D^r \cdot T.$$

Somit genügt es zu zeigen, dass  $D^r = 0$  ist. Diese Aussage ist aber klar, wenn man den  $2 \times 2$ -Fall

$$\begin{pmatrix} 0 & a \\ 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} 0 & b \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

verstanden hat.

c) $\Rightarrow$ a) Es genügt, die Aussage für eine komplexe Matrix  $N$  zu beweisen. Sei  $\lambda_1$  ein Eigenwert dieser Matrix und  $\mathbf{v}_1$  ein zugehöriger Eigenvektor. Dann ist

$$N \cdot \mathbf{v}_1 = \lambda_1 \mathbf{v}_1, \quad N^2 \mathbf{v}_1 = \lambda_1^2 \mathbf{v}_1, \quad \dots, \quad N^r \cdot \mathbf{v}_1 = \lambda_1^r \mathbf{v}_1 = \mathbf{0}.$$

Daraus folgt  $\lambda_1 = 0$ . Wir haben gezeigt: Die Matrix  $N$  hat nur 0 als ihre Eigenwerte. Das bedeutet aber  $\chi_N(\lambda) = \lambda^n$ .  $\square$

**Definition 1.4** Eine Matrix heißt nilpotent, wenn sie die Eigenschaften aus Satz 1.9 besitzt.

Kommen wir zurück auf unsere  $n \times n$ -Matrix  $C$  und die dazu ähnliche obere Dreiecksmatrix

$$D = \begin{pmatrix} \lambda_1 \mathbb{1}_{r_1} + N & * \\ 0 & D' \end{pmatrix}.$$

Hier ist  $D'$  eine obere Dreiecksmatrix der Größe  $(n - r_1) \times (n - r_1)$ . Und auf ihrer Diagonale stehen nur Eigenwerte  $\neq \lambda_1$ . Wir betrachten die Matrix

$$D - \lambda_1 \mathbb{1}_n = \begin{pmatrix} N & * \\ 0 & D' - \lambda_1 \mathbb{1}_{n-r_1} \end{pmatrix}.$$

Hier hat die obere Dreiecksmatrix  $D' - \lambda_1 \mathbb{1}_{n-r_1}$  auf ihrer Diagonale lauter Einträge  $\neq 0$ . Deswegen hat sie den Maximalrang  $n - r_1$ , ebenso wie alle ihre Potenzen. Mit Satz 1.9 und kästchenweiser Matrizenmultiplikation finden wir

$$(D - \lambda_1 \mathbb{1}_n)^{r_1} = \begin{pmatrix} 0 & * \\ 0 & (D' - \lambda_1 E_{n-r_1})^{r_1} \end{pmatrix}.$$

Das Resultat ist eine Matrix, deren erste  $r_1$  Spalten Nullspalten sind, und ihre letzten  $n - r_1$  Spalten linear unabhängig.

Wir definieren Untervektorräume wie folgt:

- $V_{\lambda_1}$  sei aufgespannt von den ersten  $r_1$  Basisvektoren  $\mathbf{e}_1, \dots, \mathbf{e}_{r_1}$ . Er ist der Nullraum der Matrix  $(D - \lambda_1 \mathbb{1}_n)^{r_1}$  und hat die Dimension  $r_1$ .

- $V'$  werde aufgespannt von den letzten  $n - r_1$  Spaltenvektoren der obigen Matrix. Er ist der Bildraum der Matrix  $(D - \lambda_1 \mathbb{1}_n)^{r_1}$  und hat die Dimension  $n - r_1$ .

Die Dimensionen der Unterräume  $V_{\lambda_1}$  und  $V'$  ergänzen sich zu  $n$ . Außerdem besteht der Durchschnitt beider Unterräume nur aus dem Nullvektor: Ein Vektor  $\mathbf{v}$  im Durchschnitt gehört zu  $V_{\lambda_1}$  und hat nur Nullen als seine letzten  $n - r_1$  Komponenten. Er gehört aber auch zu  $V'$  und ist eine Linearkombination der letzten  $n - r_1$  Spalten der obigen Matrix. Weil deren letzte  $n - r_1$  Komponenten linear unabhängige Vektoren darstellen, kann diese Linearkombination nur der Nullvektor sein.

**Definition 1.5 (Direkte Summe)** *Es sei  $V$  ein Vektorraum mit Untervektorräumen  $V_1, V_2 \subset V$ . Man sagt:  $V$  ist die direkte Summe der Unterräume  $V_1$  und  $V_2$ , in Zeichen*

$$V = V_1 \oplus V_2,$$

wenn

$$V_1 + V_2 = V \quad \text{und} \quad V_1 \cap V_2 = \{\mathbf{0}\}.$$

Wegen der Dimensionsformel für Durchschnitte von Untervektorräumen ist dazu äquivalent:

$$\dim(V_1) + \dim(V_2) = \dim(V) \quad \text{und} \quad V_1 \cap V_2 = \{\mathbf{0}\}.$$

Es ist leicht zu sehen: Die direkte Summenzerlegung  $V = V_1 \oplus V_2$  ist äquivalent dazu, dass jeder Vektor  $\mathbf{v} \in V$  eine Summe

$$\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2 \quad \text{mit Vektoren} \quad \mathbf{v}_1 \in V_1, \mathbf{v}_2 \in V_2$$

ist, die durch  $\mathbf{v}$  eindeutig bestimmt sind.

Diese direkte Summe von Vektorräumen kann man iterieren und auf offensichtliche Weise die mehrfache direkte Summe

$$V = V_1 \oplus \dots \oplus V_k$$

definieren.

Um unsere Ergebnisse fest zu halten gehen wir noch schnell über zur linearen Abbildung  $\varphi$ , welche durch die Matrix  $D$  beschrieben wird. Dann hat  $\varphi - \lambda_1 \text{id}$  die darstellende Matrix  $D - \lambda_1 \mathbb{1}_n$ . Für diese Abbildung haben wir bewiesen:

**Satz 1.10** *Es sei  $\varphi : V \rightarrow V$  eine lineare Abbildung des Vektorraums  $V$  in sich, deren charakteristisches Polynom den Linearfaktor  $\lambda_1 - \lambda$  mit der algebraischen Vielfachheit  $r_1$  abspaltet. Dann gibt es eine direkte Summenzerlegung*

$$V = \text{Kern}(\varphi - \lambda_1 \text{id})^{r_1} \oplus \text{Bild}(\varphi - \lambda_1 \text{id})^{r_1}.$$

Dabei ist

$$\dim(\text{Kern}(\varphi - \lambda_1 \text{id})^{r_1}) = r_1, \quad \dim(\text{Bild}(\varphi - \lambda_1 \text{id})^{r_1}) = n - r_1.$$

**Definition 1.6** *Es sei  $\varphi : V \rightarrow V$  ein Endomorphismus und  $V = V_1 \oplus V_2$  eine direkte Summenzerlegung. Diese Zerlegung heißt  $\varphi$ -invariant, wenn  $\varphi$  jeden der beiden Unterräume  $V_1$  und  $V_2$  in sich abbildet. D.h.:*

$$\varphi|_{V_1} : V_1 \rightarrow V_1, \quad \varphi|_{V_2} : V_2 \rightarrow V_2.$$

**Satz 1.11** Die direkte Summenzerlegung aus Satz 1.10 ist  $\varphi$ -invariant.

Beweis. a)  $\varphi$  bildet den Unterraum  $\text{Kern}(\varphi - \lambda_1 \text{id})^{r_1}$  in sich ab: Sei dazu  $\mathbf{v}$  ein Vektor aus diesem Kern. Wegen

$$\varphi \circ (\varphi - \lambda_1 \text{id}) = \varphi^2 - \lambda_1 \varphi = (\varphi - \lambda_1 \text{id}) \circ \varphi$$

gilt auch

$$\varphi \circ (\varphi - \lambda_1 \text{id})^{r_1} = (\varphi - \lambda_1 \text{id})^{r_1} \circ \varphi.$$

Damit folgt für  $\varphi(\mathbf{v})$

$$(\varphi - \lambda_1 \text{id})^{r_1}(\varphi(\mathbf{v})) = \varphi((\varphi - \lambda_1 \text{id})^{r_1}(\mathbf{v})) = \mathbf{0}.$$

Also gehört auch  $\varphi(\mathbf{v})$  zum betrachteten Kern.

b)  $\varphi$  bildet den Unterraum  $\text{Bild}(\varphi - \lambda_1 \text{id})^{r_1}$  in sich ab: Sei  $\mathbf{v} = (\varphi - \lambda_1 \text{id})^{r_1} \cdot \mathbf{w}$  ein Vektor aus diesem Bildraum. Mit dem gleichen Vertauschungstrick wie in a) sieht man, auch der Vektor

$$\varphi(\mathbf{v}) = (\varphi - \lambda_1 \text{id})^{r_1} \cdot \varphi(\mathbf{w})$$

gehört zu diesem Bildraum. □

Die Sätze 1.10 und 1.11 kann man iterieren, und mit vollständiger Induktion beweisen:

**Satz 1.12** Der Endomorphismus  $\varphi : V \rightarrow V$  habe die paarweise verschiedenen Eigenwerte  $\lambda_1, \dots, \lambda_k$  und das charakteristische Polynom

$$(\lambda_1 - \lambda)^{r_1} \cdot \dots \cdot (\lambda_k - \lambda)^{r_k}, \quad r_1 + \dots + r_k = \dim(V).$$

Dann gibt es eine  $\varphi$ -invariante direkte Summenzerlegung

$$V = V_1 \oplus \dots \oplus V_k$$

mit

$$V_i = \text{Kern}(\varphi - \lambda_i)^{r_i}, \quad \dim(V_i) = r_i.$$

Außerdem hat der eingeschränkte Endomorphismus  $\varphi|_{V_i}$  das charakteristische Polynom  $(\lambda - \lambda_i)^{r_i}$ .

Nur die Aussage über das charakteristische Polynom ist noch nicht bewiesen. Aber sie folgt aus Satz 1.9. □

Der Sinn einer invarianten direkten Summenzerlegungen erschließt sich, wenn man zu einer Basis übergeht, welche dieser Zerlegung angepasst ist:

**Definition 1.7** Die Basis  $\mathbf{v}_1, \dots, \mathbf{v}_n$  heißt der direkten Summenzerlegung  $V = V_1 \oplus \dots \oplus V_k$  angepasst, wenn

- die ersten  $r_1$  Vektoren der Basis eine Basis von  $V_1$  bilden,
- die nächsten  $r_2$  Vektoren eine Basis von  $V_2$  bilden,
- $\vdots$
- die letzten  $r_k$  Vektoren eine Basis von  $V_k$  bilden.

Aus dem Basis-Ergänzungs-Satz folgt, dass es zu jeder direkten Summenzerlegung solche angepassten Basen gibt. Ist  $\varphi : V \rightarrow V$  linear,  $V = V_1 \oplus \dots \oplus V_k$  eine  $\varphi$ -invariante direkte Summenzerlegung und  $\mathbf{v}_1, \dots, \mathbf{v}_n$  eine der Zerlegung angepasste Basis, so wird  $\varphi$  in dieser Basis durch eine Blockmatrix

$$B = \begin{pmatrix} B_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & B_k \end{pmatrix}$$

beschrieben. Dabei sind die Untermatrizen  $B_i$  quadratisch mit der Größe  $\dim(V_i)$ . Eine solche Matrix nennen wir die *direkte Matrixsumme* der Matrizen  $B_1, \dots, B_k$ .

Geht man jetzt wieder über zu Matrizen so sieht man:

**Satz 1.13 (Hauptsatz)** *Es sei  $C$  eine  $n \times n$ -Matrix mit paarweise verschiedenen Eigenwerten  $\lambda_1, \dots, \lambda_k$  und dem charakteristischen Polynom*

$$(\lambda_1 - \lambda)^{r_1} \cdot \dots \cdot (\lambda_k - \lambda)^{r_k}.$$

*Dann ist  $C$  ähnlich zu einer Blockmatrix  $B$  wie eben angegeben. Dabei haben die Untermatrizen  $B_i$  die Größe  $r_i \times r_i$  und das charakteristische Polynom  $(\lambda_i - \lambda)^{r_i}$ .*

Beweis. Es sei  $\varphi$  die lineare Abbildung mit darstellender Matrix  $C$ . Weiter sei eine  $\varphi$ -invariante direkte Summenzerlegung wie in Satz 1.12 gewählt und eine Basis die dieser Zerlegung angepasst ist. In dieser Basis hat  $\varphi$  eine Matrix der behaupteten Blockform.  $\square$

Ich habe diesen Satz 'Hauptsatz' genannt, weil die invarianten Unterräume  $V_i = \text{Kern}(\varphi - \lambda_i \text{id})^{r_i}$  *Haupträume* zu den Eigenwerten  $\lambda_i$  heißen.

**Aufgabe 1.9** *Bestimmen Sie Basen für die Haupträume der linearen Abbildung  $\varphi : \mathbb{R}^4 \rightarrow \mathbb{R}^3$ , die durch folgende Matrix gegeben wird:*

$$a) \begin{pmatrix} 1 & 1 & 2 & 0 \\ 0 & 1 & 4 & 0 \\ 0 & 0 & -1 & 0 \\ -4 & -2 & -4 & -1 \end{pmatrix}, \quad b) \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}.$$

### 1.3 Jordansche Normalform

Wir betrachten hier  $n \times n$ -Matrizen  $C$  mit dem einzigen Eigenwert  $c$  und einem charakteristischen Polynom  $(c - \lambda)^n$ . Wir versuchen, dazu eine möglichst einfache ähnliche Matrix zu finden. Wenn wir dies dann auf die Block-Matrizen  $B_i$  aus Satz 1.13 anwenden, erhalten wir die Jordansche Normalform.

Nach Satz 1.9 wissen wir,  $C$  ist ähnlich zu einer Matrix

$$c\mathbb{1}_n + N,$$

wobei  $N$  nilpotent mit  $N^n = 0$  ist. Wegen

$$T^{-1}(c\mathbb{1}_n + N)T = cT^{-1}\mathbb{1}_nT + T^{-1}NT = c\mathbb{1}_n + T^{-1}NT$$

genügt es, die nilpotente Matrix  $N$  auf eine möglichst einfache Normalform zu bringen. Das heißt, wir beschränken uns auf den Fall, wo der einzige Eigenwert  $c = 0$  ist.

Ist  $\varphi : V \rightarrow V$  nilpotent, so definiert jeder Vektor  $\mathbf{v} \in V$  eine endliche *Kette* von Bildvektoren

$$\mathbf{v}, \varphi(\mathbf{v}), \varphi^2(\mathbf{v}), \dots, \varphi^{p-1}(\mathbf{v}) \neq \mathbf{0}, \quad \varphi^p(\mathbf{v}) = \mathbf{0}.$$

Spätestens für  $p = n$  ist  $\varphi^p(\mathbf{v}) = \mathbf{0}$  und wir brechen die Kette ab. (Zur Kette sollen nur die Vektoren  $\varphi^k(\mathbf{v})$  gehören, welche ungleich  $\mathbf{0}$  sind.)

**Satz 1.14** *Ist  $\varphi : V \rightarrow V$  eine nilpotente lineare Abbildung des endlich-dimensionalen Vektorraums  $V$  in sich, so gibt es eine Basis von  $V$ , die sich nur aus Ketten für  $\varphi$  zusammensetzt. In einer solchen Basis wird die Abbildung  $\varphi$  durch eine Matrix beschrieben, welche die direkte Matrixsumme von Blöcken*

$$\begin{pmatrix} 0 & 1 & 0 & \dots & \dots \\ \cdot & \ddots & \ddots & \dots & \cdot \\ \cdot & & \ddots & \ddots & 0 \\ \cdot & & & \ddots & 1 \\ 0 & \cdot & \cdot & \cdot & 0 \end{pmatrix} \quad \text{Jordanblock zum Eigenwert } 0$$

ist.

Beweis. Weil  $\varphi$  nilpotent ist, gibt es ein  $p$  mit  $\varphi^p = 0$  aber  $\varphi^{p-1} \neq 0$ . Wir beweisen die Behauptung durch Induktion nach  $p$ . Der Induktionsanfang ist  $p = 1$ . Dann ist  $\varphi = 0$  die Nullabbildung. Jeder Vektor  $\mathbf{0} \neq \mathbf{v} \in V$  stellt eine - wenn auch kurze - Kette dar. Und jede Basis von  $V$  besteht aus derartigen kurzen Ketten.

Sei jetzt  $p \geq 2$ . Wir betrachten den Bildraum  $B := \varphi(V) \subset V$ . Die Einschränkung  $\varphi|_B$  bildet auch  $B$  in das Bild  $\varphi(V)$  ab, definiert also eine Abbildung  $\varphi|_B : B \rightarrow B$ . Für jeden Vektor  $\mathbf{b} = \varphi(\mathbf{v}) \in B$  ist

$$\varphi^{p-1}(\mathbf{b}) = \varphi^{p-1}(\varphi(\mathbf{v})) = \varphi^p(\mathbf{v}) = 0.$$

Deswegen ist  $(\varphi|_B)^{p-1} = 0$ , und wir können auf  $\varphi|_B$  die Induktionsannahme anwenden. Es gibt also eine Basis von  $B$ , die aus Ketten besteht:

$$\begin{array}{ccccccc} \mathbf{b}_1 & \rightarrow & \varphi(\mathbf{b}_1) & \rightarrow & \dots & \rightarrow & \varphi^{r_1}(\mathbf{b}_1), \\ \vdots & & \vdots & & & & \vdots \\ \mathbf{b}_k & \rightarrow & \varphi(\mathbf{b}_k) & \rightarrow & \dots & \rightarrow & \varphi^{r_k}(\mathbf{b}_k). \end{array}$$

Hier soll jeweils  $\varphi^{r_i}(\mathbf{b}_i) \neq \mathbf{0}$  sein, aber  $\varphi^{r_i+1}(\mathbf{b}_i) = \mathbf{0}$ .

Zuerst verlängern wir unsere Ketten etwas: Jedes  $\mathbf{b}_i \in B$  ist ein Bild  $\varphi(\mathbf{v}_i)$ . Wir können zu jeder Kette einen solchen Vektor  $\mathbf{v}_i$  hinzunehmen und ihre Länge um 1 vergrößern:

$$\begin{array}{ccccccc} \mathbf{v}_i & \rightarrow & \varphi(\mathbf{v}_i) & \rightarrow & \varphi^2(\mathbf{v}_i) & \rightarrow & \dots & \rightarrow & \varphi^{r_i+1}(\mathbf{v}_i), \\ & & \parallel & & \parallel & & \dots & & \parallel \\ & & \mathbf{b}_i & \rightarrow & \varphi(\mathbf{b}_i) & \rightarrow & \dots & \rightarrow & \varphi^{r_i}(\mathbf{b}_i). \end{array}$$

Dann vermehren wir unsere Ketten auch noch: Die Vektoren  $\varphi^{r_i}(\mathbf{b}_i) = \varphi^{r_i+1}(\mathbf{v}_i)$ ,  $i = 1, \dots, k$ , gehören zum Kern von  $\varphi$ . Als Teil der gewählten Basis von  $B$  sind sie linear unabhängig. Wir können sie durch Vektoren  $\mathbf{v}_{i+1}, \dots, \mathbf{v}_l$  zu einer Basis des Kerns ergänzen. Jeden dieser ergänzenden Vektoren im Kern fassen wir als eine kurze Kette auf.

Jetzt schreiben wir mal alle unsere Ketten zusammen:

$$\begin{array}{cccc} \mathbf{v}_1 & \varphi(\mathbf{v}_1) & \dots & \varphi^{r_1+1}(\mathbf{v}_1) \\ \vdots & \vdots & & \vdots \\ \mathbf{v}_k & \varphi(\mathbf{v}_k) & \dots & \varphi^{r_k+1}(\mathbf{v}_k) \\ \mathbf{v}_{k+1} & & & \\ \vdots & & & \\ \mathbf{v}_l & & & \end{array}$$

Die Anzahl aller Vektoren in unserer Kettenbasis von  $B$  war  $= \dim(B)$ . Hier haben wir insgesamt  $k$  Vektoren hinzugenommen um jede Kette etwas zu verlängern. Schließlich haben wir unsere Ketten um

$$l - k = \dim(\text{Kern}(\varphi)) - k$$

kurze Ketten vermehrt. Damit ist die Anzahl aller Vektoren in unseren Ketten

$$\dim(\text{Bild}(\varphi)) + \dim(\text{Kern}(\varphi)) = \dim V$$

geworden. Wir müssen also nur noch zeigen, dass alle Vektoren unserer Ketten linear unabhängig sind, dann haben wir eine Kettenbasis von  $V$ .

Wie wir die lineare Unabhängigkeit testen, ist klar, nur etwas mühselig hinzuschreiben: Sei also

$$\sum_{i=1}^k (c_i \mathbf{v}_i + \sum_{j_i=1}^{r_i+1} c_{i,j_i} \varphi^{j_i}(\mathbf{v}_i)) + \sum_{i=k+1}^l c_i \mathbf{v}_i = \mathbf{0}.$$

Auf diese Gleichung wenden wir  $\varphi$  an. Dabei fällt aus jeder Kette der letzte Vektor heraus. Und die ersten Vektoren werden genau die Kettenvektoren der Basis von  $B$ , mit der wir angefangen haben. Weil die linear unabhängig waren, folgt

$$c_i = c_{i,1} = \dots = c_{i,r_i} = 0 \quad \text{für } i = 1, \dots, k.$$

Übrig bleibt eine lineare Relation zwischen den letzten Vektoren aus jeder Kette. Die sind aber linear unabhängig, weil sie eine Basis von  $\text{Kern}(\varphi)$  bilden. Dann müssen auch deren Koeffizienten in der linearen Relation alle  $= 0$  sein.

Ist eine Kette  $\mathbf{v}, \varphi(\mathbf{v}), \dots, \varphi^p(\mathbf{v})$  als Teil der Basis gewählt, so gehört hierzu eine Blockmatrix

$$\begin{pmatrix} 0 & 0 & \dots \\ 1 & 0 & \dots \\ 0 & 1 & \\ \vdots & \vdots & \ddots \end{pmatrix}.$$

Das ist die Transponierte eines Jordan-Blocks. Kehrt man in der Basis die Reihenfolge der Vektoren in der Kette um, so wird daraus ein echter Jordan-Block.  $\square$

Die in Satz 1.14 auftretenden Ketten sind i.a. nicht eindeutig bestimmt, deswegen sind auch die zur direkten Matrixsumme gehörenden Unterräume durch die lineare Abbildung i.a. nicht eindeutig bestimmt. Ihre *Anzahlen und Dimensionen* dagegen sind sehr wohl eindeutig bestimmt.

**Satz 1.15 (Zusatz zu Satz 1.14: Eindeutigkeit)** Die Größen der in Satz 1.14 auftretenden Jordanblöcke und die Anzahl der Blöcke einer festen Größe sind durch die Abbildung  $\varphi$  eindeutig bestimmt. Insbesondere ist die Anzahl aller Ketten gleich der Dimension von  $\text{Kern}(\varphi)$ .

Beweis (Induktion nach  $p$ ). Der letzte Vektor einer jeden Kette liegt im Kern von  $\varphi$ . Wenden wir  $\varphi$  auf eine Kette an, so fällt der erste Vektor weg, die anderen reproduzieren sich. Weil die ursprünglichen Ketten eine Basis von  $V$  bildeten, sind die reproduzierten Kettenreste eine Kettenbasis von  $B = \text{Bild}(\varphi)$ . Die Anzahl aller ursprünglichen Ketten ist gleich der Zahl weggefallener Vektoren und damit gleich

$$\dim(V) - \dim(B) = \dim(\text{Kern}(\varphi)).$$

Durch  $\varphi$  ist die Abbildung  $\varphi|_B : B \rightarrow B$  eindeutig festgelegt. Nach Induktionsannahme hängen damit die Anzahlen und Längen der Ketten in jeder Kettenbasis für  $\varphi|_B$  nur von  $\varphi$  ab. Jede ursprüngliche Kette einer Länge  $\geq 2$  definiert eine Kette für  $B$  mit einer um 1 kleineren Länge. Deswegen hängen auch die Längen aller ursprünglichen Ketten nur von  $\varphi$  ab.  $\square$

**Satz 1.16 (Jordansche Normalform)** Jede  $n \times n$ -Matrix, deren charakteristisches Polynom in Linearfaktoren zerfällt, ist ähnlich zu einer direkten Matrix-Summe von Jordan-Blöcken. Dabei sind die Anzahlen der Blöcke einer festen Größe zu einem festen Eigenwert durch die Matrix eindeutig bestimmt.

Beweis. Nach Satz 1.13 ist die Matrix ähnlich zu einer direkten Matrix-Summe von Matrizen  $C_0$  mit einem charakteristischen Polynom  $(\lambda_0 - \lambda)^k$ . Trigonalisieren wir eine solche Matrix  $C_0$  mit Satz 1.9, so sehen wir, dass  $C_0 - \lambda_0 \mathbb{1}_k$  nilpotent ist. Wegen Satz 1.14 ist die Matrix  $C_0 - \lambda_0 \mathbb{1}_k$  ähnlich zu einer direkten Summe von Jordan-Blöcken mit Eigenwert 0. Dann ist  $C_0$  ähnlich zu einer direkten Summe von Jordan-Blöcken gleicher Größe zum Eigenwert  $\lambda_0$ .

Die Eindeutigkeitsaussage folgt aus der Eindeutigkeit in Satz 1.12 und Satz 1.16.  $\square$

Wegen des Eindeutigkeitsteils in Satz 1.16 können wir von der Jordanschen Normalform einer Matrix sprechen.

**Beispiel 1.4 (n=2)** Die möglichen Jordan-Normalformen für  $2 \times 2$ -Matrizen sind

$$\begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}, \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}, \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix},$$

wo  $\lambda_1 \neq \lambda_2$ . Um zu entscheiden, welche Jordan-Normalform eine reelle Matrix

$$C = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

hat, berechnen wir das charakteristische Polynom

$$\chi_C(\lambda) = (a - \lambda)(d - \lambda) - bc = \delta - \sigma\lambda + \lambda^2,$$

wo wir  $\det(C) = \delta$  und  $\text{sp}(C) = \sigma$  abkürzten. Die beiden Eigenwerte sind dann

$$\lambda_{1,2} = \frac{1}{2}(\sigma \pm \sqrt{\sigma^2 - 4\delta}),$$

und beide Eigenwerte fallen genau dann zusammen, wenn  $\sigma^2 = 4\delta$ . Daraus folgt:  $C$  ist ähnlich zu

$$\begin{aligned} \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} &\Leftrightarrow sp(C)^2 > 4 \cdot det(C), \\ \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix} &\Leftrightarrow sp(C)^2 = 4 \cdot det(C) \text{ und } b = c = 0, \\ \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix} &\Leftrightarrow sp(C)^2 = 4 \cdot det(C) \text{ und } b \neq 0 \text{ oder } c \neq 0, \end{aligned}$$

Wenn  $sp(C)^2 < 4 \cdot det(C)$  ist, dann hat das charakteristische Polynom von  $C$  keine reellen Nullstellen, und die Matrix  $C$  ist reell nicht einmal trigonalisierbar.

**Beispiel 1.5 (n=3)** Die möglichen Jordan-Normalformen für  $3 \times 3$ -Matrizen sind

$$\begin{aligned} &\begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix}, & \begin{pmatrix} \lambda_1 & 1 & 0 \\ 0 & \lambda_1 & 0 \\ 0 & 0 & \lambda_2 \end{pmatrix}, & \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_1 & 0 \\ 0 & 0 & \lambda_2 \end{pmatrix} \\ &\begin{pmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{pmatrix}, & \begin{pmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{pmatrix}, & \begin{pmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{pmatrix} \end{aligned}$$

mit  $\lambda_1 \neq \lambda_2 \neq \lambda_3 \neq \lambda_1$ .

Die Bestimmung der Jordanschen Normalform einer gegebenen Matrix  $C$  ist praktisch wichtig z.B. bei linearen Differentialgleichungen mit konstanten Koeffizienten (s. Abschnitt 1.4). Wie geht man dabei vor? Als erstes muss man natürlich das charakteristische Polynom der Matrix  $C$  berechnen und dann deren Eigenwerte. Will man die Transformation in Jordansche Normalform konkret durchführen, braucht man als nächstes die Haupträume  $V_i$  aus Satz 1.12. Natürlich ist das Berechnen der auftretenden Matrizenprodukte (von Hand) i.a. eine ziemliche Arbeit. Man sollte sich deswegen die Unterstützung durch einen Computer sichern. Aber wenn nur die Jordansche Normalform selbst, und nicht die Koordinatentransformation dazu gefragt ist, genügt es meist den Rang der Matrix  $C - \lambda_i \mathbb{1}$  zu bestimmen. Im Kern dieser Matrix liegt ja genau ein eindimensionaler Unterraum für jeden Jordan-Block. Wenn diese Matrix also Rang  $r$  hat, ist  $n - r$  die Anzahl der Jordan-Blöcke. Weil in konkreten Aufgaben selten Matrizen mit mehr als fünf Zeilen vorkommen, gibt es dann häufig nur noch eine einzige Möglichkeit für die richtige Jordan-Normalform. Und bei Matrizen mit mehr als fünf Zeilen ist die Berechnung der Jordan-Normalform von Hand sowieso ziemlich sinnlos. In jedem Paket für symbolisches Rechnen gibt es einschlägige Routinen.

Anstatt also jetzt noch allgemeine Rezepte anzugeben, möchte ich nur noch ein Beispiel diskutieren:

**Beispiel 1.6** *Es sei*

$$A = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \in M(5 \times 5, \mathbb{C}).$$

*Wir wollen eine Basis bestimmen, in der  $A$  Jordan-Normalform hat. Zuerst bräuchten wir das charakteristische Polynom. Aber wegen der beiden Nullzeilen hat  $A$  den Rang = 3 und damit den Eigenwert*

$\lambda_1 = 0$  mindestens mit der algebraischen Vielfachheit 2. Deswegen berechnen wir einfach mal die Potenzen von  $A = A - \lambda_1 \mathbb{1}_5$ , die wir ja sowieso berechnen müssten:

$$A^2 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \quad A^3 = 0.$$

Mit Satz 1.9 sehen wir, dass  $A$  tatsächlich nur den einzigen Eigenwert 0 und das charakteristische Polynom  $\lambda^5$  hat. Weil  $A^2 \neq 0$  ist, muss es (genau) einen Jordan-Block der Größe 3 geben. Weil  $A$  den Rang = 3 und damit einen Nullraum der Dimension 2 hat, gibt es genau zwei Jordan-Blöcke. Die Matrix  $A$  hat somit die Jordan-Normalform

$$\left( \begin{array}{ccc|cc} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{array} \right).$$

Um eine zugehörige Basis zu bestimmen, brauchen wir zuerst eine Kette der Länge 3. Wir kennen  $A^2$  und sehen  $A^2 \cdot \mathbf{e}_1 = \mathbf{e}_3 \neq \mathbf{0}$ . Damit erzeugt  $\mathbf{e}_1$  eine Kette

$$\mathbf{e}_1, \quad A \cdot \mathbf{e}_1 = \mathbf{e}_4, \quad A^2 \cdot \mathbf{e}_1 = \mathbf{e}_3$$

der Länge 3. Um eine weitere Kette der Länge 2 zu erzeugen, brauchen wir einen Vektor, der nicht in  $\text{span}(\mathbf{e}_1, \mathbf{e}_3, \mathbf{e}_4)$  liegt, und von  $A$  nicht auf  $\mathbf{0}$  abgebildet wird. Ein solcher Vektor ist z.B.  $\mathbf{e}_5$  mit  $A \cdot \mathbf{e}_5 = \mathbf{e}_2 - \mathbf{e}_4$ . Jetzt schreiben wir die beiden Ketten in umgekehrter Reihenfolge hin

$$\mathbf{e}_3, \mathbf{e}_4, \mathbf{e}_1, \mathbf{e}_2 - \mathbf{e}_4, \mathbf{e}_5$$

und erhalten eine Basis, bezüglich der die durch  $A$  beschriebene Abbildung eine darstellende Matrix in der angegebenen Jordan-Normalform hat.

**Aufgabe 1.10** Gegeben sei eine reelle  $n \times n$ -Matrix der Form

$$A = \begin{pmatrix} 0 & & & -a_0 \\ 1 & \ddots & & -a_1 \\ & \ddots & 0 & \vdots \\ & & 1 & -a_{n-1} \end{pmatrix}.$$

i) Man zeige: Das charakteristische Polynom  $\chi(\lambda) = \det(\lambda E_n - A)$  von  $A$  lautet

$$\chi(\lambda) = \lambda^n + a_{n-1} \lambda^{n-1} + \dots + a_1 \lambda + a_0.$$

ii) Man zeige: Für jeden Eigenwert  $\lambda_i$  von  $A$  hat der zugehörige Eigenraum die Dimension 1.

iii) Man bestimme die Jordansche Normalform von  $A$  unter der Annahme, dass  $\chi(\lambda)$  in Linearfaktoren zerfällt.

**Aufgabe 1.11** Sei

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

darstellende Matrix eines Endomorphismus  $f : \mathbb{R}^5 \rightarrow \mathbb{R}^5$  bzgl. der kanonischen Basis des  $\mathbb{R}^5$ . Geben Sie eine Matrix  $M$  in Jordan-Normalform und eine Basis  $B$  des  $\mathbb{R}^5$  an, so dass  $M$  die darstellende Matrix von  $f$  bzgl.  $B$  ist.

**Aufgabe 1.12** Sei

$$A = \begin{pmatrix} -6 & 8 & -8 \\ -4 & 6 & -4 \\ 0 & 0 & 2 \end{pmatrix} \in M(3 \times 3, \mathbb{C}).$$

Bestimmen Sie

a) das charakteristische Polynom von  $A$ , b) die Jordansche Normalform von  $A$ .

**Aufgabe 1.13** Sei  $\varphi$  das Polynom  $(t-1)^3(t+1)^2 \in \mathbb{C}[t]$ . Welche Jordanschen Normalformen treten bei komplexen  $5 \times 5$ -Matrizen mit dem charakteristischen Polynom  $\varphi$  auf?

**Aufgabe 1.14** Sei

$$A = \begin{pmatrix} 2 & 2 & 1 \\ -1 & -1 & -1 \\ 1 & 2 & 2 \end{pmatrix} \in M(3 \times 3, \mathbb{C}).$$

a) Bestimmen Sie die Eigenwerte und Eigenräume von  $A$ .

b) Geben Sie die Jordansche Normalform von  $A$  an.

**Aufgabe 1.15** Wie sieht die Jordansche Normalform der Matrix

$$A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & -3 & 3 \end{pmatrix}$$

aus?

## 1.4 Anwendung: Systeme von Differentialgleichungen

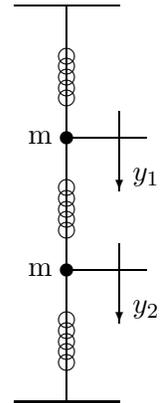
Beginnen wir mit einem einfachen intuitiven Beispiel:

Wir betrachten drei gekoppelte Federn, wie in nebenstehender Zeichnung. Oben und unten seien sie fixiert. Die Ruhelagen der Federn seien  $y_1 = 0$  und  $y_2 = 0$ . Beide Federkonstanten seien  $= k$  und beide Massen seien  $= m$ . Dann lauten die Bewegungsgleichungen

$$\begin{aligned} m\ddot{y}_1 &= -ky_1 + k(y_2 - y_1) = k(y_2 - 2y_1), \\ m\ddot{y}_2 &= -k(y_2 - y_1) - ky_2 = k(y_1 - 2y_2). \end{aligned}$$

Zur Vereinfachung setzen wir  $k = m = 1$  und erhalten ein System von zwei Differentialgleichungen zweiter Ordnung, das in Matrix-Form so aussieht:

$$\begin{pmatrix} \ddot{y}_1 \\ \ddot{y}_2 \end{pmatrix} = \begin{pmatrix} -2 & 1 \\ 1 & -2 \end{pmatrix} \cdot \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}.$$



**Eine** Differentialgleichung zweiter Ordnung, wie in diesem Beispiel, stellt keinen Physiker vor Probleme. Aber **zwei** davon, die gekoppelt sind?

In Beispiel 1.3 habe ich Ihnen für die Koeffizientenmatrix

$$C = \begin{pmatrix} -2 & 1 \\ 1 & -2 \end{pmatrix}$$

vorgerechnet

$$T^{-1} \cdot C \cdot T = D$$

mit

$$T = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}, \quad D = \begin{pmatrix} -3 & 0 \\ 0 & -1 \end{pmatrix},$$

bzw.

$$C = T \cdot D \cdot T^{-1}.$$

Wenn wir die gesuchten Funktionen  $y_1(t)$  und  $y_2(t)$  zu einem Vektor  $\mathbf{y}(t)$  zusammenfassen, sowie ihre zweiten Ableitungen zum Vektor  $\ddot{\mathbf{y}}$ , so schreibt sich das System komprimiert als

$$\ddot{\mathbf{y}} = C \cdot \mathbf{y}, \quad \dot{\mathbf{y}} = T \cdot D \cdot T^{-1} \cdot \mathbf{y}, \quad T^{-1}\ddot{\mathbf{y}} = D \cdot (T^{-1} \cdot \mathbf{y}).$$

Jetzt transformieren wir das Problem via

$$\mathbf{x} = T^{-1} \cdot \mathbf{y}$$

in das Differentialgleichungssystem

$$\ddot{\mathbf{x}} = D \cdot \mathbf{x}.$$

Weil  $D$  eine Diagonalmatrix ist, liegt das System jetzt in entkoppelter Form

$$\ddot{x}_1 = -3x_1, \quad \ddot{x}_2 = -x_2$$

vor. Lösungen dafür sind leicht anzugeben:

$$x_1(t) = c_1 \cos(\sqrt{3}t) + c_2 \sin(\sqrt{3}t), \quad x_2(t) = c_3 \cos(t) + c_4 \sin(t), \quad c_1, \dots, c_4 \in \mathbb{R}.$$

Mit

$$\mathbf{y} = T \cdot \mathbf{x}, \text{ bzw. } \begin{aligned} y_1 &= \frac{1}{\sqrt{2}}(x_1 + x_2) \\ y_2 &= \frac{1}{\sqrt{2}}(-x_1 + x_2) \end{aligned}$$

erhalten wir nach Multiplikation mit  $\sqrt{2}$  für das ursprüngliche System die Lösungen

$$\begin{aligned} y_1 &= c_1 \cos(\sqrt{3}t) + c_2 \sin(\sqrt{3}t) + c_3 \cos(t) + c_4 \sin(t) \\ y_2 &= -c_1 \cos(\sqrt{3}t) - c_2 \sin(\sqrt{3}t) + c_3 \cos(t) + c_4 \sin(t) \end{aligned}$$

Die Lösungen sind kompliziert, aber man kann sie deuten als Überlagerung (= Linearkombination) der beiden harmonischen Schwingungen

$$y_1 = c_1 \cos(\sqrt{3}t) + c_2 \sin(\sqrt{3}t), \quad y_2 = -y_1, \quad (c_3 = c_4 = 0)$$

und

$$y_1 = c_3 \cos(t) + c_4 \sin(t), \quad y_2 = y_1, \quad (c_1 = c_2 = 0)$$

Die erste Schwingung gehört zum Eigenwert  $\lambda_1 = -3$ , die zweite zum Eigenwert  $\lambda_2 = -1$ . Die beiden Schwingungen heißen Eigenschwingungen und ihre Frequenzen Eigenfrequenzen des Systems. Und daher kommt der Name Eigenwert.

Jetzt kommen wir zum eigentlichen Thema dieses Abschnitts. Wir betrachten hier Systeme linearer Differentialgleichungen mit konstanten Koeffizienten. In voller Allgemeinheit sehen sie so aus:

$$\begin{aligned} \dot{y}_1 &= a_{1,1}y_1 + \dots + a_{1,n}y_n \\ &\vdots \\ \dot{y}_n &= a_{n,1}y_1 + \dots + a_{n,n}y_n \end{aligned}$$

Hier ist es wesentlich, dass wir genau so viele Gleichungen wie Unbekannte haben. Die Koeffizientenmatrix

$$A = (a_{i,j})_{i,j=1,\dots,n}$$

ist also immer quadratisch.

Um das System in Matrixform schreiben zu können, fasst man die  $n$  gesuchten Funktionen zu einem Vektor

$$\mathbf{y}(t) = \begin{pmatrix} y_1(t) \\ \vdots \\ y_n(t) \end{pmatrix}$$

zusammen. Dieser Funktionenvektor ist eine Abbildung

$$\mathbf{y} : \mathbb{R} \rightarrow \mathbb{R}^n.$$

Das Differenzieren von Abbildungen werden wir erst im nächsten Kapitel behandeln. Aber schon jetzt definieren wir die Ableitung von  $\mathbf{y}$  als

$$\dot{\mathbf{y}} = \begin{pmatrix} \dot{y}_1 \\ \vdots \\ \dot{y}_n \end{pmatrix}.$$

Dann lautet das System in Matrixschreibweise

$$\dot{\mathbf{y}} = A \cdot \mathbf{y}.$$

Das ist schön, nur: Das eingangs behandelte System hatte zwar konstante Koeffizienten, aber die beiden Differentialgleichungen hatten die Ordnung 2. Durch folgenden Trick kann man auch lineare Differentialgleichungen höherer Ordnung mit konstanten Koeffizienten hier einordnen:

Gegeben sei die Differentialgleichung  $k$ -ter Ordnung

$$y^{(k)} = c_1 y + c_2 \dot{y} + \dots + c_k y^{(k-1)}, \quad c_1, \dots, c_k \in \mathbb{R}.$$

Wir führen  $k$  Funktionen

$$y_1 := y, \quad y_2 = \dot{y}, \quad \dots, \quad y_k := y^{(k-1)}$$

ein. Die erfüllen die  $k - 1$  Differentialgleichungen erster Ordnung

$$\dot{y}_1 = y_2, \quad \dots, \quad \dot{y}_{k-1} = y_k.$$

Und die ursprüngliche Differentialgleichung  $n$ -ter Ordnung wird damit äquivalent zum System

$$\begin{array}{rcl} \dot{y}_1 & = & y_2 \\ \vdots & & \ddots \\ \dot{y}_{k-1} & = & y_k \\ \dot{y}_k & = & c_1 y_1 + c_2 y_2 + \dots + c_k y_k \end{array}$$

Die  $k \times k$ -Koeffizientenmatrix

$$C = \begin{pmatrix} & 1 & & \\ & & \ddots & \\ & & & 1 \\ c_1 & c_2 & \dots & c_k \end{pmatrix}$$

hat zwar ziemlich viele Nullen und wirkt dadurch ziemlich aufgeblasen, aber das ist uns die erreichte Vereinheitlichung wert.

**Beispiel 1.7** *Jetzt möchte ich die beiden gekoppelten Differentialgleichungen zweiter Ordnung der gekoppelten Federn so standardisieren, dass daraus ein  $4 \times 4$ -System von Differentialgleichungen erster Ordnung wird. Ich ergänze das System also um die Funktionen*

$$y_3 := \dot{y}_1, \quad y_4 := \dot{y}_2.$$

*Dann erhalte ich das System*

$$\begin{array}{rcl} \dot{y}_1 & = & y_3 \\ \dot{y}_2 & = & y_4 \\ \dot{y}_3 & = & -2y_1 + y_2 \\ \dot{y}_4 & = & y_1 - 2y_2 \end{array}$$

Der ganze Sinn dieses Abschnitts besteht darin, zu untersuchen, wie sich Basistransformationen auf ein System

$$\dot{\mathbf{y}} = A \cdot \mathbf{y}$$

auswirken. Transformieren wir also

$$A = T^{-1} \cdot D \cdot T$$

(wobei wir insgeheim hoffen, erreichen zu können, dass  $D$  eine Diagonalmatrix wird). Unser System transformiert sich dann wie folgt:

$$\dot{\mathbf{y}} = T^{-1} \cdot D \cdot T \cdot \mathbf{y}, \quad \text{bzw.} \quad T\dot{\mathbf{y}} = D \cdot (T\mathbf{y}).$$

Wenn wir übergehen zu

$$\mathbf{x} = T\mathbf{y} \quad \text{mit} \quad \dot{\mathbf{x}} = T\dot{\mathbf{y}},$$

erhalten wir das System

$$\dot{\mathbf{x}} = D \cdot \mathbf{x}.$$

Jede Lösung  $\mathbf{x}(t)$  des transformierten Systems mit der zu  $A$  ähnlichen Matrix  $D$  liefert durch Rücktransformation  $\mathbf{y} = T^{-1} \cdot \mathbf{x}$  eine Lösung  $\mathbf{y}(t)$  des ursprünglichen Systems.

Der einfachste Fall liegt vor, wenn  $D$  eine Diagonalmatrix

$$D = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}$$

ist. Das System besteht aus den entkoppelten Gleichungen

$$\dot{x}_i = \lambda_i x_i, \quad (i = 1, \dots, n)$$

mit den offensichtlichen Lösungen

$$x_i(t) = c_i \cdot e^{\lambda_i t}, \quad i = 1, \dots, n.$$

In Vektorschreibweise sind das die Lösungen

$$\mathbf{x}(t) = c_i e^{\lambda_i t} \cdot \mathbf{e}_i.$$

Bei der Rücktransformation wird aus dem Koordinatenvektor  $\mathbf{e}_i$  ein Eigenvektor  $\mathbf{a}_i$  zum Eigenwert  $\lambda_i$  für die Matrix  $A$ . Fassen wir zusammen:

**Satz 1.17 (e-hoch-Ansatz)** a) Ist  $\mathbf{a}$  ein Eigenvektor der Matrix  $A$  zum Eigenwert  $\lambda$ , so ist

$$\mathbf{y}(t) = e^{\lambda t} \cdot \mathbf{a}$$

eine Lösungs-Vektor-Funktion des Systems  $\dot{\mathbf{y}} = A \cdot \mathbf{y}$ .

b) Ist  $A$  diagonalisierbar mit den linear unabhängigen Eigenvektoren  $\mathbf{a}_1, \dots, \mathbf{a}_n$  zu Eigenwerten  $\lambda_1, \dots, \lambda_n$ , so gehören dazu  $n$  linear unabhängige Lösungen  $e^{\lambda_i t} \mathbf{a}_i$ .

Die Aussage dieses Satzes kann man direkt nachrechnen, man braucht dazu nicht den Weg über die ähnliche Matrix  $D$  zu gehen.

Falls  $A$  diagonalisierbar ist, hat man damit das System gelöst. Was das genau bedeutet, kann ich aber erst im Kapitel über gewöhnliche Differentialgleichungen präzisieren. Es bleiben die beiden fundamentalen Probleme der Linearen Algebra:

Wir sind nur an reellen Lösungen  $\mathbf{y}(t)$  interessiert. Was passiert, wenn ein Eigenwert  $\lambda$  von  $A$  komplex ist? Dann ist auch der zugehörige Eigenvektor  $\mathbf{a}$  komplex. Aber, wir sind ja auch nur an Systemen mit einer reellen Koeffizientenmatrix  $A$  interessiert. Deren charakteristisches Polynom ist dann auch reell und hat mit jeder komplexen, nicht-reellen Wurzel auch die konjugiert-komplexe Wurzel  $\bar{\lambda}$ . Dazu gehört der komplex-konjugierte Eigenvektor  $\bar{\mathbf{a}}$ . Wir haben also auch zwei konjugiert-komplexe Lösungen

$$\mathbf{y}(t) = e^{\lambda t} \mathbf{a}, \quad \bar{\mathbf{y}}(t) = e^{\bar{\lambda} t} \bar{\mathbf{a}}.$$

Falls etwa

$$\lambda = \mu + i\omega, \quad \bar{\lambda} = \mu - i\omega, \quad \omega \neq 0$$

ist, dann können wir mit der Eulerschen Formel die beiden Lösungen schreiben

$$\mathbf{y}(t) = e^{\mu t} (\cos(\omega t) + i \sin(\omega t)) \mathbf{a}, \quad \bar{\mathbf{y}}(t) = e^{\mu t} (\cos(\omega t) - i \sin(\omega t)) \bar{\mathbf{a}}.$$

Sie lassen sich folgendermaßen zu zwei reellen Lösungen rekombinieren:

$$\begin{aligned} \mathbf{y}_1 &= \operatorname{Re}(\mathbf{y}(t)) = \frac{1}{2} (\mathbf{y}(t) + \bar{\mathbf{y}}(t)) \\ &= e^{\mu t} (\cos(\omega t) \operatorname{Re}(\mathbf{a}) - \sin(\omega t) \operatorname{Im}(\mathbf{a})) \\ \mathbf{y}_2 &= \operatorname{Im}(\mathbf{y}(t)) = \frac{1}{2i} (\mathbf{y}(t) - \bar{\mathbf{y}}(t)) \\ &= e^{\mu t} (\cos(\omega t) \operatorname{Im}(\mathbf{a}) + \sin(\omega t) \operatorname{Re}(\mathbf{a})). \end{aligned}$$

Das sieht gefährlich aus. Aber als Eigenvektoren zu zwei verschiedenen Eigenwerten sind  $\mathbf{a}$  und  $\bar{\mathbf{a}}$  (über  $\mathbb{C}$ ) linear unabhängig. Dann sind auch die beiden reellen Vektoren  $\operatorname{Re}(\mathbf{a})$  und  $\operatorname{Im}(\mathbf{a})$  linear unabhängig. Wählen wir sie als Basisvektoren, so werden unsere Lösungen

$$\mathbf{y}_1 = e^{\mu t} \begin{pmatrix} \cos(t) \\ -\sin(t) \end{pmatrix}, \quad \mathbf{y}_2 = e^{\mu t} \begin{pmatrix} \sin(t) \\ \cos(t) \end{pmatrix}.$$

Und auch alle Linearkombinationen

$$c_1 \mathbf{y}_1 + c_2 \mathbf{y}_2 = e^{\mu t} \cdot \begin{pmatrix} \cos(t) & \sin(t) \\ -\sin(t) & \cos(t) \end{pmatrix} \cdot \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}, \quad c_1, c_2 \in \mathbb{R},$$

sind Lösungen. Für  $\mu = 0$  handelt es sich um (in unserem Koordinatensystem rechtsdrehende) Kreise, für  $\mu \neq 0$  um Spiralen.

Zweitens können Eigenwerte entarten. Betrachten wir etwa ein System

$$\dot{\mathbf{y}} = J_\lambda \cdot \mathbf{y}$$

mit einem  $n \times n$ -Jordanblock

$$J = \begin{pmatrix} \lambda & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & \lambda \end{pmatrix}$$

als Koeffizientenmatrix. Wir modifizieren den e-hoch-Ansatz

$$\frac{d}{dt}(e^{\lambda t} \mathbf{y}) = \lambda e^{\lambda t} \mathbf{y} + e^{\lambda t} \dot{\mathbf{y}} = e^{\lambda t} (\lambda \mathbf{y} + \dot{\mathbf{y}})$$

und sehen: Die Vektorfunktion  $e^{\lambda t} \mathbf{y}$  ist genau dann eine Lösung des Systems mit Koeffizientenmatrix  $J_\lambda$ , wenn  $\mathbf{y}$  eine Lösung des Systems

$$\dot{\mathbf{y}} = N \cdot \mathbf{y}$$

mit der nilpotenten Koeffizientenmatrix  $N = J_0$  ist. Es genügt also, ein System mit Koeffizientenmatrix  $N$  zu betrachten. Das System sieht dann so aus:

$$\begin{array}{rcl} \dot{y}_1 & = & y_2 \\ \dot{y}_2 & = & y_3 \\ & \vdots & \ddots \\ \dot{y}_{n-1} & = & y_n \\ \dot{y}_n & = & 0 \end{array}$$

Für  $\nu \geq 2$  ist

$$y_\nu = \frac{d^{\nu-1}}{dt^{\nu-1}} y_1$$

mit

$$\dot{y}_n = \frac{d^n}{dt^n} y_1 = 0.$$

Wir sind gelandet bei einer ziemlich banalen Differentialgleichung  $n$ -ter Ordnung für  $y_1$ . Ihre Lösungen sind die Polynome

$$y_1 = c_1 + c_2 t + \dots + c_n t^{n-1}, \quad c_1, \dots, c_n \in \mathbb{R}.$$

Zu einem solchen Polynom  $y_1$  gehören die Komponenten

$$y_\nu = \frac{d^{\nu-1}}{dt^{\nu-1}} y_1 = \sum_{k=\nu}^{n-1} c_k \cdot k \cdot \dots \cdot (k - \nu + 1) t^{k-\nu}.$$

Etwas schematischer schreiben wir die erhaltenen Lösungen als

$$\begin{pmatrix} 1 & t & t^2 & \dots & t^{n-1} \\ & 1 & 2t & & (n-1)t^{n-2} \\ & & 2 & & (n-1)(n-2)t^{n-3} \\ & & & \ddots & \vdots \\ & & & (n-2)! & (n-1)!t \\ & & & & (n-1)! \end{pmatrix} \cdot \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ \vdots \\ \vdots \\ c_n \end{pmatrix}.$$

Multipliziert man sie mit dem Faktor  $e^{\lambda t}$  bekommt man Lösungen des ursprünglichen Systems mit  $\lambda \neq 0$ .

Richtig kompliziert wird es, wenn beide Komplikationen gleichzeitig auftreten: Ein Jordanblock zu einem komplexen Eigenwert  $\lambda$ . Da muss man die Polynomlösungen zum Jordanblock mit der komplexen Exponentialfunktion  $e^{\lambda t}$  multiplizieren. Aber mit jeder komplexen Lösung ist auch die komplex-konjugierte Funktion eine Lösung. Und ähnlich wie oben, kann man beide zu zwei reellen Lösungen rekombinieren, welche dann Winkelfunktionen als Faktor enthalten.

Auf die Details möchte ich hier nicht mehr eingehen. Stattdessen möchte ich die für  $n = 2$  möglichen Fälle

$$\dot{\mathbf{y}} = A \cdot \mathbf{y}$$

diskutieren und zeichnen.

**Fall 1:** Die Matrix  $A$  ist diagonalisierbar mit den reellen Eigenwerten  $\lambda_1$  und  $\lambda_2$ . Nachdem wir in ein Koordinatensystem übergehen, das von Eigenvektoren aufgespannt wird, sieht unser System so aus:

$$\dot{y}_1 = \lambda_1 y_1, \quad \dot{y}_2 = \lambda_2 y_2.$$

Es hat die Lösungskurven

$$y_1 = c_1 \cdot e^{\lambda_1 t}, \quad y_2 = c_2 \cdot e^{\lambda_2 t}.$$

Zunächst sortieren wir einige Spezialfälle aus:

	Lösungskurven
$\lambda_1 = \lambda_2 = 0$	Punkte $(c_1, c_2)$
$\lambda_1 = 0, \lambda_2 \neq 0$	Geraden $y_1 = c_1$
$\lambda_1 \neq 0, \lambda_2 = 0$	Geraden $y_2 = c_2$
$\lambda_1 = \lambda_2 \neq 0$	Geraden durch den Nullpunkt.

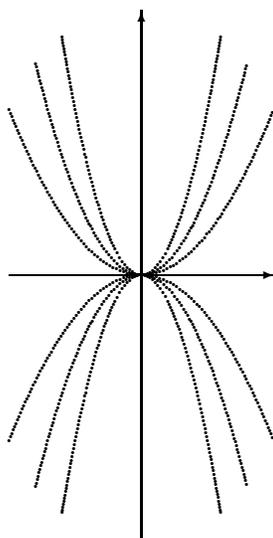
Es bleiben die Fälle

$$0 \neq \lambda_1 \neq \lambda_2 \neq 0.$$

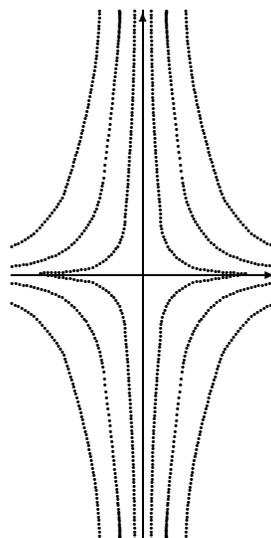
Für  $c_1 = 0$ , bzw.  $c_2 = 0$  haben wir als Lösungskurve die positiven oder negativen Halbachsen auf der  $y_1$ -, bzw.  $y_2$ -Achse. Für  $c_1 \cdot c_2 \neq 0$  eliminieren wir  $t$  vermöge

$$\left(\frac{y_1}{c_1}\right)^{\lambda_2} = \left(\frac{y_2}{c_2}\right)^{\lambda_1}, \quad \frac{y_2}{c_2} = \left(\frac{y_1}{c_1}\right)^{\lambda_2/\lambda_1}, \quad y_2 = c \cdot y_1^{\lambda_2/\lambda_1} \text{ mit } c \neq 0.$$

Einen wesentlichen Unterschied macht es jetzt noch, ob die Eigenwerte  $\lambda_1$  und  $\lambda_2$  das gleiche oder verschiedene Vorzeichen haben.



$\lambda_1 = 1, \lambda_2 = 2$



$\lambda_1 = -1, \lambda_2 = 2$

**Fall 2:** Beide Eigenwerte  $\lambda_1 = \lambda_2 = \lambda$  stimmen überein, aber die Matrix  $A$  ist nicht diagonalisierbar. Wir wählen das Koordinatensystem so, dass  $A$  in Jordan-Normalform

$$A = \begin{pmatrix} \lambda & 1 \\ & \lambda \end{pmatrix}$$

vorliegt. Damit sind die Spalten der Matrix

$$\begin{pmatrix} e^{\lambda t} & te^{\lambda t} \\ & e^{\lambda t} \end{pmatrix}$$

Lösungen. Lösungen sind also alle Vektoren

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} c_1 e^{\lambda t} + c_2 t e^{\lambda t} \\ c_2 e^{\lambda t} \end{pmatrix}.$$

Hier möchte ich die Koordinaten  $y_1$  und  $y_2$ , sowie  $c_1$  und  $c_2$  vertauschen, weil ich diesen Fall schon gezeichnet habe. Unsere Lösungen haben dann die Form

$$y_1 = c_1 \cdot e^{\lambda t}, \quad y_2 = (c_1 t + c_2) \cdot e^{\lambda t}.$$

Spezielle Lösungskurven bildet für  $c_1 = 0$  die  $y_2$ -Achse, schön unterteilt in ihren positiven, ihren negativen Abschnitt und den Nullpunkt.

Für  $c_1 \neq 0$  haben wir

$$y_2 = \left(t + \frac{c_2}{c_1}\right) y_1$$

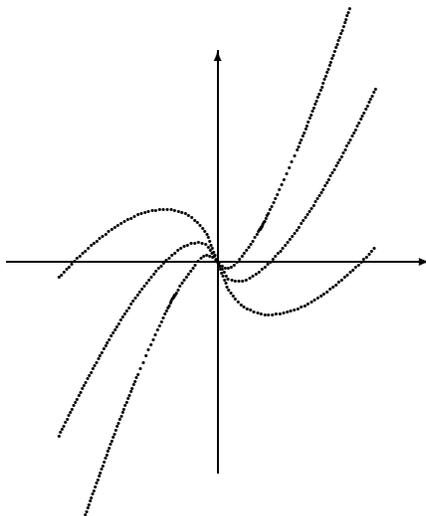
mit

$$e^{\lambda t} = \frac{y_1}{c_1}, \quad \lambda t = \ln\left(\frac{y_1}{c_1}\right), \quad t = \frac{1}{\lambda} \ln\left(\frac{y_1}{c_1}\right).$$

In der rechten Halbebene ist  $y_1 > 0$  und  $c_1 > 0$ . Hier können wir schreiben

$$y_2 = \frac{c_2}{c_1} y_1 + \frac{1}{\lambda} y_1 \ln\left(\frac{y_1}{c_1}\right) = \left(\frac{c_2}{c_1} - \frac{1}{\lambda} \ln(c_1)\right) \cdot y_1 + \frac{1}{\lambda} y_1 \ln(y_1) = c y_1 + \frac{1}{\lambda} y_1 \ln(y_1), \quad c \in \mathbb{R}.$$

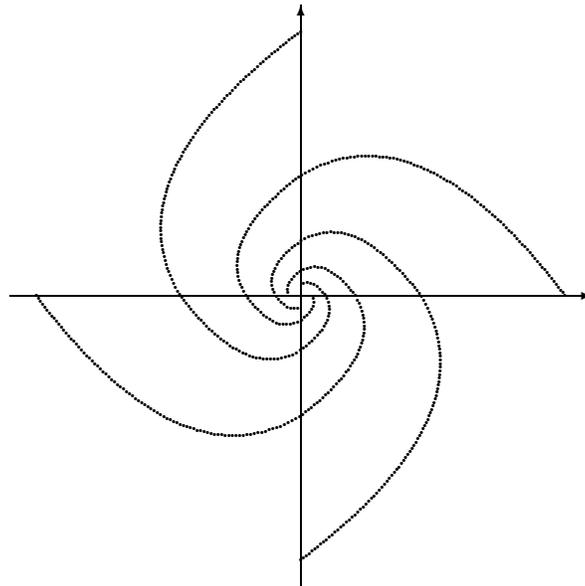
(Die Lösungskurven in der linken Halbebene liegen punktsymmetrisch zu denen in der rechten Halbebene.)



**Fall 3:** Es gibt zwei konjugiert-komplexe Eigenwerte

$$\lambda_{1,2} = \mu \pm i \cdot \omega, \quad \omega > 0.$$

Diesen Fall habe ich oben diskutiert. In einem geeigneten Koordinatensystem sind es rechtsdrehende Spiralen, die so aussehen:



**Aufgabe 1.16** Finden Sie je zwei linear unabhängige Lösungen für die Systeme

$$a) \quad \dot{\mathbf{y}} = \begin{pmatrix} 1 & 12 \\ 3 & 1 \end{pmatrix} \cdot \mathbf{y} \quad b) \quad \dot{\mathbf{y}} = \begin{pmatrix} -3 & 5 \\ -2 & 4 \end{pmatrix} \cdot \mathbf{y}.$$

**Aufgabe 1.17** Finden Sie drei linear unabhängige Lösungen für

$$\dot{\mathbf{y}} = \begin{pmatrix} 0 & -1 & 1 \\ 1 & 0 & -1 \\ -1 & 1 & 0 \end{pmatrix} \cdot \mathbf{y}.$$

**Aufgabe 1.18** Betrachtet werde ein System wie am Anfang dieses Abschnitts, aber bestehend aus drei Massenpunkten (mit den Koordinaten  $y_1, y_2, y_3$ ) und vier Federn. Die zugehörige Bewegungsgleichung ist

$$\ddot{\mathbf{y}} = \begin{pmatrix} -2 & 1 & 0 \\ 1 & -2 & 1 \\ 0 & 1 & -2 \end{pmatrix} \cdot \mathbf{y}.$$

Bestimmen Sie seine Eigenfrequenzen und Eigenschwingungen.

**Aufgabe 1.19** Bestimmen Sie die Eigenvektoren der Matrix

$$A_4 := \begin{pmatrix} -2 & 1 & & \\ 1 & -2 & 1 & \\ & 1 & -2 & 1 \\ & & 1 & -2 \end{pmatrix}.$$

*Hinweis: Wie könnten die Eigenschwingungen des zugehörigen Systems aus vier Massenpunkten, die durch fünf Federn gekoppelt sind, wohl aussehen?*

## 1.5 Bilinearformen

Eine reelle  $n \times n$ -Matrix  $A$  kann auch für etwas anderes verwendet werden, als zur Beschreibung einer linearen Abbildung

$$\mathbf{w} \mapsto A \cdot \mathbf{w}.$$

Der Ausdruck

$$\mathbf{v}^t \cdot A \cdot \mathbf{w}$$

ist ein Skalar, der linear von  $\mathbf{v}$  und von  $\mathbf{w}$  abhängt. Dadurch wird eine bilineare Abbildung

$$\mathbb{R}^n \times \mathbb{R}^n \ni (\mathbf{v}, \mathbf{w}) \mapsto \mathbf{v}^t \cdot A \cdot \mathbf{w} \in \mathbb{R}$$

definiert.

**Definition 1.8 (Bilinearform)** *Es sei  $V$  ein reeller Vektorraum. Eine Bilinearform  $\varphi$  auf  $V$  ist eine Abbildung*

$$\varphi : V \times V \ni (\mathbf{v}, \mathbf{w}) \mapsto \varphi(\mathbf{v}, \mathbf{w}) \in \mathbb{R},$$

*die in  $\mathbf{v}$  und  $\mathbf{w}$  linear ist. D.h. explizit: Für alle  $c, c' \in \mathbb{R}, \mathbf{v}, \mathbf{v}', \mathbf{w}, \mathbf{w}' \in V$  gelten*

$$\varphi(c\mathbf{v} + c'\mathbf{v}', \mathbf{w}) = c\varphi(\mathbf{v}, \mathbf{w}) + c'\varphi(\mathbf{v}', \mathbf{w}),$$

$$\varphi(\mathbf{v}, c\mathbf{w} + c'\mathbf{w}') = c\varphi(\mathbf{v}, \mathbf{w}) + c'\varphi(\mathbf{v}, \mathbf{w}').$$

**Beispiel 1.8** *a) Bilinearformen kommen auch in der Analysis vor: Sei etwa  $V$  der  $\mathbb{R}$ -Vektorraum der auf einem Intervall  $[a, b]$  stetigen Funktionen. Dann ist die Abbildung*

$$V \times V \ni (f, g) \mapsto \int_a^b f(x)g(x)dx \in \mathbb{R}$$

*eine Bilinearform auf  $V$*

*b) Es seien  $f, g : V \rightarrow \mathbb{R}$  Linearformen. Dann heißt die Bilinearform*

$$f \otimes g : V \times V \ni (\mathbf{v}, \mathbf{w}) \mapsto f(\mathbf{v}) \cdot g(\mathbf{w})$$

*das Tensorprodukt der beiden Linearformen  $f$  und  $g$ .*

*c) Das euklidische Skalarprodukt*

$$(\mathbf{v}, \mathbf{w}) = \sum_{\nu=1}^n v_{\nu}w_{\nu}$$

*ist eine Bilinearform auf dem  $\mathbb{R}^n$ .*

**Definition 1.9** *Es sei  $\varphi$  eine Bilinearform auf dem reellen Vektorraum  $V$  und  $\mathbf{u}_1, \dots, \mathbf{u}_n \in V$  eine Basis. Dann heißt die  $n \times n$ -Matrix*

$$G := (\varphi(\mathbf{u}_{\mu}, \mathbf{u}_{\nu}))_{\mu, \nu=1, \dots, n}$$

*die darstellende Matrix von  $\varphi$  bezüglich der gegebenen Basis.*

Sind  $\mathbf{v} = \sum v_\mu \mathbf{u}_\mu$  und  $\mathbf{w} = \sum w_\nu \mathbf{u}_\nu$  zwei Vektoren in  $V$ , so ist wegen der Bilinearität

$$\begin{aligned}\varphi(\mathbf{v}, \mathbf{w}) &= \varphi\left(\sum_{\mu} v_{\mu} \mathbf{u}_{\mu}, \sum_{\nu} w_{\nu} \mathbf{u}_{\nu}\right) \\ &= \sum_{\mu} v_{\mu} \varphi\left(\mathbf{u}_{\mu}, \sum_{\nu} w_{\nu} \mathbf{u}_{\nu}\right) \\ &= \sum_{\mu, \nu} v_{\mu} \varphi(\mathbf{u}_{\mu}, \mathbf{u}_{\nu}) w_{\nu}\end{aligned}$$

Mit der darstellenden Matrix  $G$  kann man das Ergebnis

$$(v_1, \dots, v_n) \cdot G \cdot \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix}$$

schreiben. Der Sinn dieser Übung ist: Mit der darstellenden Matrix der Bilinearform  $\varphi$  bezüglich der Basis  $\mathbf{u}_1, \dots, \mathbf{u}_n$  kann man den Wert  $\varphi(\mathbf{v}, \mathbf{w})$  aus den Koordinaten der Vektoren  $\mathbf{v}$  und  $\mathbf{w}$  bezüglich der Basis  $\mathbf{u}_1, \dots, \mathbf{u}_n$  ausrechnen. Das ist ganz ähnlich, wie mit der darstellenden Matrix einer linearen Abbildung. Wesentlich verschieden ist dagegen das Transformationsverhalten:

Die invertierbare Matrix  $T$  beschreibe die Transformation von der Basis  $\mathbf{u}_1, \dots, \mathbf{u}_n$  in eine neue Basis  $\mathbf{t}_1, \dots, \mathbf{t}_n$ . D.h., die Spaltenvektoren der Matrix  $T$  sind die Koordinatenvektoren der Vektoren  $\mathbf{t}_\nu$  bezüglich der Basis  $\mathbf{u}_1, \dots, \mathbf{u}_n$ , d.h.,

$$\mathbf{t}_\nu = \sum_l t_{l,\nu} \mathbf{u}_l.$$

Die darstellende Matrix von  $\varphi$  in der neuen Basis hat dann die Einträge

$$\varphi(\mathbf{t}_\mu, \mathbf{t}_\nu) = \varphi\left(\sum_k t_{k,\mu} \mathbf{u}_k, \sum_l t_{l,\nu} \mathbf{u}_l\right) = \sum_{k,l} t_{k,\mu} \varphi(\mathbf{u}_k, \mathbf{u}_l) t_{l,\nu}.$$

In Matrix-Schreibweise ist dies die Zahl

$$(t_{1,\mu}, \dots, t_{n,\mu}) \cdot G \cdot \begin{pmatrix} t_{1,\nu} \\ \vdots \\ t_{n,\nu} \end{pmatrix}.$$

Und damit wird die neue darstellende Matrix

$$T^t \cdot G \cdot T.$$

Das ist zwar ähnlich, wie bei der Matrix einer linearen Abbildung. Aber wo bei der linearen Abbildung  $T^{-1}$  steht, da steht bei einer Bilinearform  $T^t$ . Man sagt: Das Transformationsverhalten ist zweifach kovariant.

Weil die Transformationsmatrix  $T$  invertierbar ist, hat die Matrix  $T^t \cdot G \cdot T$  den gleichen Rang wie  $G$ . Der Rang der darstellenden Matrix ändert sich nicht bei einer Basistransformation.

**Definition 1.10** Der Rang einer Bilinearform ist der Rang der darstellenden Matrix bezüglich einer (und damit jeder) Basis. Eine Bilinearform heißt nicht ausgeartet, wenn ihr Rang maximal ist.

**Beispiel 1.9** Das euklidische Skalarprodukt auf dem  $\mathbb{R}^n$  hat bezüglich der kanonischen Basis  $\mathbf{e}_1, \dots, \mathbf{e}_n$  die darstellende Matrix

$$((\mathbf{e}_\mu, \mathbf{e}_\nu)) = (\delta_{\mu,\nu}) = \mathbb{1}_n$$

und ist deswegen nicht ausgeartet.

**Beispiel 1.10 (Hyperbolische Ebene)** Die Bilinearform auf dem  $\mathbb{R}^2$  mit darstellender Matrix

$$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

ist nicht ausgeartet. Ihr Wert auf dem Vektor  $(1, 1)$  ist

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

Deswegen ist ihre Einschränkung auf den Unterraum  $\mathbb{R} \cdot (1, 1)$  identisch  $= 0$  und damit ausgeartet.

**Satz 1.18** Für eine Bilinearform  $\varphi$  auf dem endlich-dimensionalen Vektorraum  $V$  sind äquivalent:

- a)  $\varphi$  ist nicht ausgeartet.
- b) Zu jedem Vektor  $\mathbf{0} \neq \mathbf{v} \in V$  existiert ein Vektor  $\mathbf{w} \in V$  mit  $\varphi(\mathbf{v}, \mathbf{w}) \neq 0$ .

Beweis. Es sei  $G$  die darstellende Matrix von  $\varphi$  bezüglich einer Basis von  $V$ . Diese Matrix hat genau dann den Maximalrang, wenn für jeden Vektor  $\mathbf{0} \neq \mathbf{v} \in \mathbb{R}^n$  gilt  $\mathbf{v}^t \cdot G \neq \mathbf{0}$ . Das ist genau dann der Fall, wenn es einen Vektor  $\mathbf{w} \in \mathbb{R}^n$  gibt mit  $\mathbf{v}^t \cdot G \cdot \mathbf{w} \neq 0$ .  $\square$

**Definition 1.11** Es sei  $\varphi$  eine Bilinearform auf dem Vektorraum  $V$  und  $U \subset V$  ein Untervektorraum. Das orthogonale Komplement von  $U$  bezüglich  $\varphi$  ist der Untervektorraum

$$U^\perp := \{\mathbf{v} \in V : \varphi(\mathbf{v}, \mathbf{u}) = 0 \text{ für alle } \mathbf{u} \in U\}.$$

**Beispiel 1.11** a) Für das euklidische Skalarprodukt auf dem  $\mathbb{R}^n$  ist dieses orthogonale Komplement genau das orthogonale Komplement aus dem ersten Semester.

- b) Auch für  $U = V$  ist

$$V^\perp = \{\mathbf{v} \in V : \varphi(\mathbf{v}, \mathbf{u}) = 0 \text{ für alle } \mathbf{u} \in V\}$$

definiert. Mit Satz 1.18 sehen wir:  $\varphi$  auf  $V$  ist genau dann nicht ausgeartet, wenn  $V^\perp = \{\mathbf{0}\}$ .

**Satz 1.19 (Orthogonale direkte Summe)** Es sei  $\varphi$  eine Bilinearform auf dem endlich-dimensionalen Vektorraum  $V$  und  $U \subset V$  ein Untervektorraum, auf dem  $\varphi$  nicht ausgeartet ist. Dann ist

$$V = U \oplus U^\perp$$

eine direkte Summe.

Beweis. Wir nehmen an  $V = \mathbb{R}^n$ . Weiter sei  $G$  die darstellende Matrix von  $\varphi$ . Wir wählen eine Basis  $\mathbf{u}_1, \dots, \mathbf{u}_k$  für  $U$ . Bezüglich dieser Basis ist die darstellende Matrix von  $\varphi|_U$

$$\begin{pmatrix} \mathbf{u}_1^t \\ \vdots \\ \mathbf{u}_k^t \end{pmatrix} \cdot G \cdot (\mathbf{u}_1, \dots, \mathbf{u}_k).$$

Nach Voraussetzung hat diese den Maximalrang  $k$ . Dann hat auch die Matrix

$$\begin{pmatrix} \mathbf{u}_1^t \\ \vdots \\ \mathbf{u}_k^t \end{pmatrix} \cdot G$$

diesen Rang  $k$ , sowie die zugehörige lineare Abbildung

$$\mathbb{R}^n \rightarrow \mathbb{R}^k, \quad \mathbf{x} \mapsto \begin{pmatrix} \mathbf{u}_1^t \\ \vdots \\ \mathbf{u}_k^t \end{pmatrix} \cdot G \cdot \mathbf{x}.$$

$U^\perp$  ist genau der Kern dieser linearen Abbildung und hat die Dimension

$$\dim(U^\perp) = n - k = n - \dim(U).$$

Es bleibt zu zeigen:  $U \cap U^\perp = \{\mathbf{0}\}$ . Das folgt aber aus Satz 1.18, weil  $\varphi$  auf  $U$  nicht ausgeartet ist.  $\square$

**Definition 1.12** Die Bilinearform  $\varphi$  auf dem Vektorraum  $V$  heißt *symmetrisch* (bzw. *antisymmetrisch*), wenn für alle  $\mathbf{v}, \mathbf{w} \in V$  gilt

$$\varphi(\mathbf{v}, \mathbf{w}) = \varphi(\mathbf{w}, \mathbf{v}) \quad \text{bzw.} \quad \varphi(\mathbf{v}, \mathbf{w}) = -\varphi(\mathbf{w}, \mathbf{v}).$$

Eine Bilinearform mit darstellender Matrix  $G$  ist genau dann symmetrisch wenn  $G = G^t$  symmetrisch ist, und sie ist genau dann antisymmetrisch, wenn  $G = -G^t$  alternierend ist.

**Beispiel 1.12** Das euklidische Skalarprodukt auf dem  $\mathbb{R}^n$  ist symmetrisch. Die zur Relativitätstheorie gehörende Minkowski-Form auf dem  $\mathbb{R}^4$  hat die darstellende Matrix

$$\begin{pmatrix} 1 & & & \\ & 1 & & \\ & & 1 & \\ & & & -1 \end{pmatrix}$$

und ist auch symmetrisch.

Die Polarisationsformel für das euklidische Skalarprodukt verallgemeinert sich wörtlich auf symmetrische Bilinearformen.

**Satz 1.20 (Polarisationsformel)** Die Bilinearform auf dem Vektorraum  $V$  sei symmetrisch. Dann gilt für alle  $\mathbf{v}, \mathbf{w} \in V$

$$\varphi(\mathbf{v}, \mathbf{w}) = \frac{1}{2} (\varphi(\mathbf{v} + \mathbf{w}, \mathbf{v} + \mathbf{w}) - \varphi(\mathbf{v}, \mathbf{v}) - \varphi(\mathbf{w}, \mathbf{w})).$$

Der Beweis besteht aus der gleichen Rechnung, wie im ersten Semester.

Die Bilinearform  $\varphi$  ist nicht die Nullform, wenn es Vektoren  $\mathbf{v}, \mathbf{w} \in V$  gibt mit  $\varphi(\mathbf{v}, \mathbf{w}) \neq 0$ . Satz 1.20 zeigt: Wenn die symmetrische Bilinearform nicht die Nullform ist, dann gibt es einen Vektor  $\mathbf{v}$  mit  $\varphi(\mathbf{v}, \mathbf{v}) \neq 0$ .

**Satz 1.21 (Diagonalisierung)** *Die Bilinearform auf dem  $n$ -dimensionalen Vektorraum  $V$  sei symmetrisch. Dann gibt es eine Basis von  $V$ , bezüglich der  $\varphi$  durch eine Diagonalmatrix dargestellt wird.*

Beweis (Induktion nach  $n$ ). Für  $n = 1$  ist nichts zu zeigen. Sei also  $n \geq 2$ . Wenn  $\varphi$  die Nullform ist, dann ist auch nichts zu zeigen. Sei also  $\varphi$  nicht die Nullform. Wie eben angemerkt, gibt es dann einen Vektor  $\mathbf{v} \in V$  mit  $\varphi(\mathbf{v}, \mathbf{v}) \neq 0$ . Sei  $U = \mathbb{R} \cdot \mathbf{v}$  der von  $\mathbf{v}$  aufgespannte eindimensionale Unterraum. Auf diesem ist dann  $\varphi$  nicht entartet. Nach Satz 1.19 ist  $V = U \oplus U^\perp$  eine orthogonale direkte Summe. Die Behauptung folgt mit der Induktionsannahme für den  $n - 1$ -dimensionalen Raum  $U^\perp$ .  $\square$

Weil die darstellende Matrix einer symmetrischen Bilinearform symmetrisch ist, folgt Satz 1.21 auch aus dem Satz über die Hauptachsen-Transformation. Dort geschieht die Transformation ja mit einer orthogonalen Matrix  $T$ . Für die ist  $T^{-1} = T^t$  und

$$T^{-1} \cdot G \cdot T = T^t \cdot G \cdot T.$$

Die Unterschiede sind:

	Hauptachsen-Transformation	Satz 1.21
Transformation	orthogonal	invertierbar
Diagonaleinträge	Eigenwerte	?
Beweis	lang	kurz

Die erhaltene Diagonalisierung der darstellenden Matrix  $G$  reicht mir noch nicht. Ich will sie noch weiter standardisieren:

**Satz 1.22** *Zu jeder reellen symmetrischen  $n \times n$ -Matrix  $G$  gibt es eine invertierbare reelle Matrix  $T$  so, dass  $T^t \cdot G \cdot T$  eine Diagonalmatrix ist, welche auf ihrer Diagonale nur die Einträge  $\pm 1$  und  $0$  enthält. Deren Reihenfolge können wir wie folgt normalisieren:*

$$T^t \cdot G \cdot T = \begin{pmatrix} \mathbb{1}_p & & \\ & -\mathbb{1}_m & \\ & & 0 \end{pmatrix}.$$

Beweis. Wegen Satz 1.21 können wir o.B.d.A. annehmen, dass

$$G = \begin{pmatrix} g_1 & & \\ & \ddots & \\ & & g_n \end{pmatrix}$$

in Diagonalform vorliegt. Wenn  $g_1 > 0$  ist, ersetzen wir den ersten Basisvektor durch den Vektor

$$\mathbf{t}_1 := \frac{1}{\sqrt{g_1}} \mathbf{e}_1.$$

Für die durch  $G$  definierte Bilinearform  $\varphi$  ist dann  $\varphi(\mathbf{t}_1) = 1$ . Nachdem wir  $\mathbf{t}_1$  statt  $\mathbf{e}_1$  als ersten Basisvektor nehmen, ändert sich der erste Diagonaleintrag ab zu 1. Wenn  $g_1 < 0$  ist wählen wir

$$\mathbf{t}_1 := \frac{1}{\sqrt{-g_1}} \mathbf{e}_1$$

um den ersten Diagonaleintrag in  $-1$  zu verändern.

Analog behandeln wir jeden Diagonaleintrag  $\neq 0$ . Danach haben wir auf der Diagonale nur Einträge  $\pm 1$  oder  $0$ . Die Reihenfolge der Basisvektoren können wir dann noch so verändern, dass zuerst die Einträge  $= +1$ , dann die Einträge  $= -1$  und zuletzt die Nuller kommen.  $\square$

Die Zahl  $p + m$  der Diagonaleinträge  $\neq 0$  ist der Rang von  $G$ . Deswegen ist die Summe  $p + m$  unabhängig von der gewählten Diagonalisierung. Dies gilt aber auch für die Zahlen  $p$  und  $m$  selbst:

**Satz 1.23 (Sylvesterscher Trägheitssatz)** *Die symmetrische reelle Matrix  $G$  werde wie in Satz 1.22 diagonalisiert. Dann sind die Zahlen  $p$  und  $m$  unabhängig von der gewählten Basistransformation.*

Beweis. Weil  $p + m$  von der Diagonalisierung unabhängig ist, genügt es, zu zeigen, dass  $p$  nicht von der Diagonalisierung abhängt. Dazu betrachten wir den Untervektorraum  $U \subset \mathbb{R}^n$ , der von den letzten  $n - p$  Basisvektoren aufgespannt wird. Dieser hat die Eigenschaft: Für jeden Vektor  $\mathbf{u} \in U$  ist

$$\varphi(\mathbf{u}, \mathbf{u}) \leq 0.$$

Dabei ist  $\varphi$  die durch  $G$  definierte Bilinearform.

Jeder Untervektorraum  $U'$  einer Dimension  $> n - p$  schneidet (nach der Dimensionsformel für Untervektorräume) den Untervektorraum  $V$ , der von den ersten  $p$  Basisvektoren aufgespannt wird, in einem Vektorraum der Dimension  $> 0$ . Deswegen enthält  $U'$  einen Vektor  $\mathbf{0} \neq \mathbf{v} \in V$ . Für diesen ist  $\varphi(\mathbf{v}, \mathbf{v}) > 0$ . Wir haben gezeigt: Die Zahl  $n - p$  ist die größtmögliche Dimension eines Unterraums  $W \subset \mathbb{R}^n$  mit der Eigenschaft

$$\varphi(\mathbf{u}, \mathbf{u}) \leq 0 \text{ für alle } \mathbf{u} \in W.$$

Diese Zahl ist unabhängig von der gewählten Diagonalisierung.  $\square$

**Definition 1.13** *Das Zahlenpaar  $(p, m)$  heißt Signatur der reellen symmetrischen Matrix  $G$  (bzw., der durch  $G$  definierten symmetrischen Bilinearform). Die Differenz  $p - m$  heißt Trägheitsindex*

Die Zahlen  $p$  und  $m$  sind charakteristisch für reelle symmetrische Matrizen. Denn für zwei reelle symmetrische Matrizen  $G$  und  $H$  sind äquivalent:

- a) Die Matrizen  $G$  und  $H$  haben die gleiche Signatur.
- b) Es gibt eine invertierbare reelle Matrix  $T$  mit  $H = T^t \cdot G \cdot T$ .

**Definition 1.14 (Definitheit)** *Die symmetrische Bilinearform  $\varphi$  auf dem reellen Vektorraum  $V$  der Dimension  $n$  heißt*

	<i>falls</i>	<i>für alle</i>	<i>bzw. äquivalent</i>
<i>positiv definit</i>	$\varphi(\mathbf{v}, \mathbf{v}) > 0$	$\mathbf{0} \neq \mathbf{v} \in V$	$p = n$
<i>positiv semidefinit</i>	$\varphi(\mathbf{v}, \mathbf{v}) \geq 0$	$\mathbf{v} \in V$	$m = 0$
<i>negativ definit</i>	$\varphi(\mathbf{v}, \mathbf{v}) < 0$	$\mathbf{0} \neq \mathbf{v} \in V$	$m = n$
<i>negativ semidefinit</i>	$\varphi(\mathbf{v}, \mathbf{v}) \leq 0$	$\mathbf{v} \in V$	$p = 0$

Die Form  $\varphi$  heißt indefinit, wenn sie weder positiv noch negativ semidefinit ist, d.h., wenn  $p > 0$  und  $m > 0$ .

Die gleichen Bezeichnungen verwendet man auch für symmetrische Matrizen.

**Beispiel 1.13** Das euklidische Skalarprodukt  $(\mathbf{v}, \mathbf{w})$  auf dem  $\mathbb{R}^n$  ist positiv definit. Die Minkowski-Form hat die Signatur  $(3, 1)$  und ist deswegen indefinit.

Die Definitheit symmetrischer Matrizen ist wesentlich bei der Untersuchung differenzierbarer Funktionen  $f(x_1, \dots, x_n)$  mehrerer Veränderlicher auf lokale Extrema. Deswegen ist ein handliches Kriterium zum Erkennen der positiv-Definitheit von beachtlichem praktischem Interesse.

**Definition 1.15** Unter-Determinanten einer Matrix  $G$  heißen auch Minoren. Die Determinanten der linken oberen Untermatrizen

$$G_k := (g_{\mu,\nu})_{\mu,\nu=1,\dots,k}$$

heißen Hauptminoren.

**Satz 1.24 (Haupt-Minoren-Kriterium, Hurwitz-Kriterium)** Eine symmetrische reelle Matrix  $G$  ist genau dann positiv-definit, wenn alle ihre Hauptminoren  $\det(G_k) > 0$  sind.

Beweis.  $\Rightarrow$ : Wenn  $G$  positiv definit ist, gibt es nach Satz 1.22 eine invertierbare reelle Matrix  $T$  mit

$$T^t \cdot G \cdot T = \mathbb{1}_n, \quad \det(G) \cdot \det(T^2) = 1, \quad \det(G) = \frac{1}{\det(T)^2} > 0.$$

Die Untermatrizen  $G_k$  sind die darstellenden Matrizen der durch  $G$  definierten Bilinearform auf dem Untervektorraum, der von den ersten  $k$  Einheitsvektoren aufgespannt wird. Mit  $G$  sind auch sie positiv definit.

$\Leftarrow$ : Jetzt sind alle Hauptminoren  $\det(G_k) > 0$  vorausgesetzt. Wir beweisen die positiv-Definitheit von  $G$  durch Induktion nach  $n$ . Nach Induktionsannahme gibt es eine invertierbare  $(n-1) \times (n-1)$ -Matrix  $T'$  mit

$$(T')^t \cdot G_{n-1} \cdot T' = \mathbb{1}_{n-1}.$$

Mit

$$T := \begin{pmatrix} T' & \\ & 1 \end{pmatrix}$$

wird

$$T^t \cdot G \cdot T = \begin{pmatrix} 1 & & & * \\ & 1 & & * \\ & & \ddots & \vdots \\ & & & * \dots * \end{pmatrix}.$$

Dabei ist nach Voraussetzung

$$\det(T^t \cdot G \cdot T) = \det(T)^2 \cdot \det(G) > 0.$$

Wir können deswegen o.b.d.A. annehmen, dass  $G$  selbst in obiger Form vorliegt.

Es sei  $U \subset \mathbb{R}^n$  der von den ersten  $n-1$  Basisvektoren aufgespannte Unterraum. Auf ihm wird die Bilinearform durch die Einheitsmatrix beschrieben und ist nicht-entartet. Nach Satz 1.19 ist  $\mathbb{R}^n =$

$U \oplus U^\perp$ . Wir ändern den letzten Basisvektor jetzt ab in einen erzeugenden Vektor für  $U^\perp$ . Nach dieser Basistransformation ändert sich  $G$  in eine Diagonalmatrix

$$\begin{pmatrix} 1 & & & \\ & \ddots & & \\ & & 1 & \\ & & & a \end{pmatrix}.$$

Weil sich das Vorzeichen von  $\det(G) > 0$  dabei nicht ändert, muss  $a > 0$  sein. Die neue Matrix ist positiv definit, und damit auch die Ausgangsmatrix  $G$ .  $\square$

**Aufgabe 1.20** Zeigen Sie für die reelle Matrix

$$S = \begin{pmatrix} a & b \\ b & c \end{pmatrix}$$

dass

$S$ ist	genau dann, wenn
positiv definit	$a > 0$ und $\det(S) > 0$ ,
negativ definit	$a < 0$ und $\det(S) > 0$ ,
indefinit	$\det S < 0$ .

**Aufgabe 1.21** Zeigen Sie, die Matrix

$$D_4 := \begin{pmatrix} -2 & 1 & 1 & 1 \\ 1 & -2 & & \\ 1 & & -2 & \\ 1 & & & -2 \end{pmatrix}$$

ist negativ definit.

**Aufgabe 1.22** Betrachtet werde die  $n \times n$ -Matrix

$$A_n := \begin{pmatrix} -2 & 1 & & & \\ 1 & -2 & \ddots & & \\ & \ddots & \ddots & 1 & \\ & & & 1 & -2 \end{pmatrix}$$

Zeigen Sie:

a)  $\det(A_n) = (-1)^n(n+1)$ .

b) Die Matrix  $A_n$  ist negativ definit.

## 1.6 Quadriken

Eine *Quadrik* im affinen Raum  $\mathbb{R}^n$  ist eine Teilmenge  $Q \subset \mathbb{R}^n$ , welche durch Polynome

$$\text{vom Grad 2 } \sum_{k \leq l} a_{k,l} x_k x_l \quad \text{und} \quad \text{vom Grad 1 } \sum_{k=1}^n b_k x_k$$

und eine Konstante  $c \in \mathbb{R}$  beschrieben werden kann:

$$Q = \{ \mathbf{x} \in \mathbb{R}^n : \sum a_{k,l} x_k x_l + \sum b_k x_k + c = 0 \}$$

Die Koeffizienten  $a_{k,l}$  wollen wir zu einer quadratischen  $n \times n$ -Matrix  $A$  zusammenfassen. Dann ist der quadratische Anteil der obigen Gleichung

$$\sum_{k,l} a_{k,l} x_k x_l = \mathbf{x}^t \cdot A \cdot \mathbf{x}$$

die *quadratische Form* zur Bilinearform mit Matrix  $A$ .

Diese Matrix  $A$  braucht zunächst nicht symmetrisch zu sein. Aber wegen

$$\mathbf{x}^t \cdot A \cdot \mathbf{x} = (\mathbf{x}^t \cdot A \cdot \mathbf{x})^t = \mathbf{x}^t \cdot A^t \cdot \mathbf{x}$$

ist

$$\mathbf{x}^t \cdot A \cdot \mathbf{x} = \mathbf{x}^t \cdot S \cdot \mathbf{x}$$

mit der symmetrischen Matrix  $S = (A + A^t)/2$ . Wir können (und wollen) deswegen immer  $A = S$  symmetrisch annehmen.

Den linearen Anteil  $\sum b_k x_k$  können wir als Produkt mit dem Vektor  $\mathbf{b} = (b_k)$  auffassen. Dabei ist es rechentechnisch vorteilhaft, einen Faktor 2 vorwegzunehmen. Dann sieht unsere Quadriken-gleichung in Matrizenform also so aus:

$$\mathbf{x}^t \cdot A \cdot \mathbf{x} + 2\mathbf{b}^t \cdot \mathbf{x} + c = 0.$$

**Definition 1.16** Wir nennen die Matrix  $A$  die Koeffizientenmatrix und die symmetrische  $(n+1) \times (n+1)$ -Matrix

$$A' := \begin{pmatrix} A & \mathbf{b} \\ \mathbf{b}^t & c \end{pmatrix}$$

die erweiterte (oder geränderte) Koeffizientenmatrix.

Mit dieser erweiterten Koeffizientenmatrix können wir die Gleichung der Quadrik etwas kompakter schreiben, wenn wir noch einen *erweiterten Vektor*

$$\mathbf{x}' := \begin{pmatrix} \mathbf{x} \\ 1 \end{pmatrix} \in \mathbb{R}^{n+1}$$

eingeführen. Die Gleichung ist ja

$$(\mathbf{x}')^t \cdot A' \cdot \mathbf{x}' = (\mathbf{x}^t, 1) \cdot \begin{pmatrix} A & \mathbf{b} \\ \mathbf{b}^t & c \end{pmatrix} \cdot \begin{pmatrix} \mathbf{x} \\ 1 \end{pmatrix} = \mathbf{x}^t \cdot A \cdot \mathbf{x} + 2\mathbf{b}^t \cdot \mathbf{x} + c = 0.$$

Unser Ziel in diesem Abschnitt ist die Klassifikation der Quadriken. Das bedeutet Folgendes: Durch eine geeignete Transformation  $\mathbf{y} = F(\mathbf{x})$  wollen wir die Gleichung  $q(\mathbf{x}) = 0$  der Quadrik  $Q$  auf eine möglichst einfache Form bringen. Das einfachste Beispiel was mir hier einfällt, ist eine Parabel, in Koordinaten  $(x_1, x_2)$  etwa

$$x_1^2 + 2b_1x_1 + 2b_2x_2 + c = 0, \quad b_2 \neq 0.$$

Mit der quadratischen Ergänzung  $y_1 := x_1 + b_1$  wird daraus

$$y_1^2 + 2b_2x_2 + c - b_1^2,$$

und mit

$$y_2 := -2b_2 \cdot \left(x_2 + \frac{c - b_1^2}{2b_2}\right)$$

schließlich

$$y_2 = y_1^2.$$

Das sieht doch ganz manierlich aus, wir kennen die Gleichung aus dem Gymnasium.

So schön das schließlich auch funktionieren wird, es gibt zwei etwas unangenehmere Details. Wir müssen verschiedene Dinge unterscheiden, und uns überlegen, was wir eigentlich wollen:

1) Wir müssen zwischen der transformierten Quadrik  $F(Q)$  und der transformierten Gleichung unterscheiden. Das hat folgenden Grund, den ich mir nie merken kann, und immer wieder mühselig herleiten muss:

$$F(Q) = \{F(\mathbf{x}) : \mathbf{x} \in Q\} = \{\mathbf{y} \in \mathbb{R}^n : F^{-1}(\mathbf{y}) \in Q\} = \{\mathbf{y} \in \mathbb{R}^n : q(F^{-1}(\mathbf{y})) = 0\}.$$

Das wollen wir hier nur einmal anmerken, und später schlicht verdrängen.

2) Wir müssen uns genau überlegen, welche Art von Transformationen  $F$  wir verwenden wollen. Das oben kurz vorgestellte Beispiel der Parabel zeigt, dass wir um Translationen nicht herumkommen werden. Wir könnten also z.B. die Bewegungen

$$\mathbf{y} = F(\mathbf{x}) = \mathbf{t} + U \cdot \mathbf{x}, \quad \mathbf{t} \in \mathbb{R}^n, U \in O(n)$$

verwenden. Vorher wollen wir allerdings erst noch etwas Allgemeineres probieren.

**Definition 1.17** Eine Abbildung  $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$  heißt affin, wenn es einen Translationsvektor  $\mathbf{t} \in \mathbb{R}^n$  und eine invertierbare  $n \times n$ -Matrix  $C \in GL(n, \mathbb{R})$  gibt so, dass

$$F(\mathbf{x}) = \mathbf{t} + C \cdot \mathbf{x}.$$

Wahrscheinlich heißen diese Abbildungen so, weil sie affine Unterräume (Lösungsräume eines nicht notwendig homogenen linearen Gleichungssystems) wieder auf affine Unterräume abbilden. (Oder ist es umgekehrt?)

Wir transformieren die Quadrik

$$Q : \mathbf{x}^t \cdot A \cdot \mathbf{x} + 2\mathbf{b}^t \cdot \mathbf{x} + c = 0$$

unter einer affinen Transformation

$$\mathbf{x} = C \cdot \mathbf{y} + \mathbf{t}.$$

Die transformierte Gleichung ist dann

$$\begin{aligned}
 & \mathbf{x}^t \cdot A \cdot \mathbf{x} + 2\mathbf{b}^t \cdot \mathbf{x} + c \\
 = & (C\mathbf{y} + \mathbf{t})^t \cdot A \cdot (C\mathbf{y} + \mathbf{t}) + 2\mathbf{b}^t \cdot (C\mathbf{y} + \mathbf{t}) + c \\
 = & \mathbf{y}^t \cdot (C^t A C) \cdot \mathbf{y} + 2(\mathbf{A}\mathbf{t} + \mathbf{b})^t C \cdot \mathbf{y} + \mathbf{t}^t A \mathbf{t} + 2\mathbf{b}^t \mathbf{t} + c \\
 = & 0.
 \end{aligned}$$

Das ziemlich unübersichtliche Resultat unserer Rechnung wird etwas übersichtlicher wenn wir uns die erweiterte Koeffizientenmatrix der transformierten Quadrik ansehen. Das ist die Matrix

$$\begin{pmatrix} C^t A C & C^t(\mathbf{A}\mathbf{t} + \mathbf{b}) \\ (\mathbf{A}\mathbf{t} + \mathbf{b})^t C & c' \end{pmatrix}$$

mit

$$c' := \mathbf{t}^t A \mathbf{t} + 2\mathbf{b}^t \mathbf{t} + c.$$

Es ist eine Produktmatrix

$$\begin{pmatrix} C & \mathbf{t} \\ \mathbf{0}^t & 1 \end{pmatrix}^t \cdot \begin{pmatrix} A & \mathbf{b} \\ \mathbf{b}^t & c \end{pmatrix} \cdot \begin{pmatrix} C & \mathbf{t} \\ \mathbf{0}^t & 1 \end{pmatrix}.$$

**Definition 1.18** Die erweiterte Matrix zur affinen Transformation

$$F : \mathbf{t} + C \cdot \mathbf{x}$$

ist die  $(n+1) \times (n+1)$ -Matrix

$$C' := \begin{pmatrix} C & \mathbf{t} \\ \mathbf{0}^t & 1 \end{pmatrix}$$

Sie heißt so, weil für die erweiterten Vektoren  $\mathbf{x}'$  und  $\mathbf{y}'$  zu  $\mathbf{x}$  und  $\mathbf{y} = F(\mathbf{x})$  gilt

$$\mathbf{y}' = C' \cdot \mathbf{x}'.$$

In der obigen Produktmatrix wurde die erweiterte Koeffizientenmatrix  $A'$  der Quadrik von rechts mit der erweiterten Matrix  $C'$  für die affine Transformation  $F$  multipliziert, und von links mit deren Transponierter. Mit  $C$  ist auch  $C'$  invertierbar. Insbesondere finden wir: Bei einer affinen Transformation der Quadrik  $Q$  bleiben Rang und Signatur der eigentlichen Koeffizientenmatrix  $A$ , sowie auch der erweiterten Koeffizientenmatrix  $A'$  unverändert.

Unser Ziel ist es, die Gleichung einer Quadrik, durch affine Transformationen auf eine möglichst einfache Normalform zu bringen. Wenn wir von solchen Normalformen einen Katalog aufstellen können, dann haben wir die Quadriken in Bezug auf affine Transformationen *klassifiziert*.

Nun transformierte sich die Koeffizientenmatrix  $A$  genau wie die Matrix einer symmetrischen Bilinearform unter der linearen Transformation mit Matrix  $C$ . Nach Satz 1.21 gibt es eine solche Matrix  $C$  derart, dass die transformierte Koeffizientenmatrix

$$\tilde{A} = C^t A C = \begin{pmatrix} a_1 & & \\ & \ddots & \\ & & a_n \end{pmatrix}$$



Oder das ist nicht der Fall. Dann können wir im Unterraum  $x_1 = \dots = x_r = 0$  eine Transformation durchführen, die die Linearform  $\mathbf{x} \mapsto \mathbf{b}^t \cdot \mathbf{x}$  auf die Linearform  $\mathbf{x} \mapsto \mathbf{e}_{r+1}^t \cdot \mathbf{x}$  transformiert. Die Quadrikgleichung wird

$$\sum_{k=1}^r a_k x_k^2 + x_{r+1} + c = 0.$$

Und wenn wir schließlich noch  $x_{r+1}$  durch  $x_{r+1} + c$  ersetzen (Translation), so nimmt die Gleichung folgende Form an:

$$\sum_{k=1}^r a_k x_k^2 + x_{r+1} = 0.$$

Fassen wir zusammen:

**Satz 1.25 (Affine Normalform)** *Die Gleichung einer Quadrik kann durch eine affine Transformation entweder auf eine Form ohne linearen Anteil*

$$\sum_{k=1}^r a_k x_k^2 + c = 0$$

oder auf die Form

$$\sum_{k=1}^r a_k x_k^2 + x_{r+1} = 0$$

gebracht werden. Hat die Koeffizientenmatrix  $A$  den Maximalrang  $n$ , so liegt der erste Fall vor. Hat die Matrix  $A$  einen Rang  $r < n$ , so hängt die Situation vom Rang der erweiterten Koeffizientenmatrix  $A'$  ab:

$$\begin{aligned} \text{Rang}(A') = r & \quad \text{Fall 1), } c = 0, \\ \text{Rang}(A') = r + 1 & \quad \text{Fall 1), } c \neq 0, \\ \text{Rang}(A') = r + 2 & \quad \text{Fall 2).} \end{aligned}$$

Weiter können wir die Diagonaleinträge  $\neq 0$  auf  $\pm 1$  normalisieren. In der Tabelle auf der nächsten Seite habe ich die Normalformen reeller Quadriken im  $\mathbb{R}^n$  für  $n \leq 3$  zusammengestellt, die man auf diese Weise bekommt. Ich orientiere mich dabei primär an der Signatur  $\text{Sign}(A)$  der Koeffizientenmatrix  $A$ . Allerdings kann man jede Gleichung mit  $-1$  durchmultiplizieren, das ändert die Signatur, aber nicht die Quadrik. Zwei Gleichungen, die sich so unterscheiden, gebe ich nicht zweimal an. Außerdem nehme ich immer an, dass  $\text{Rang}(A) > 0$  ist, weil es sich sonst nicht um die Gleichung einer Quadrik handelt. Die Koordinaten habe ich dabei  $x, y, z$  genannt.

$n$	$Rg(A)$	$Sign(A)$	$Rg(A')$	$Sign(A')$	Gleichung	Quadrik
1	1	(1, 0)	2	(2, 0)	$x^2 + 1 = 0$	$\emptyset$
				(1, 1)	$x^2 - 1 = 0$	zwei Punkte
			1	(1, 0)	$x^2 = 0$	ein Punkt
2	2	(2, 0)	3	(3, 0)	$x^2 + y^2 + 1 = 0$	$\emptyset$
				(2, 1)	$x^2 + y^2 - 1 = 0$	Kreis
			2	(2, 0)	$x^2 + y^2 = 0$	Punkt
			3	(2, 1)	$x^2 - y^2 + 1 = 0$	Hyperbel
			2	(1, 1)	$x^2 - y^2 = 0$	schneidendes Geradenpaar
	1	(1, 0)	3	(2, 1)	$x^2 + y = 0$	Parabel
			2	(2, 0)	$x^2 + 1 = 0$	$\emptyset$
				(1, 1)	$x^2 - 1 = 0$	paralleles Geradenpaar
			1	(1, 0)	$x^2 = 0$	Gerade
3	3	(3, 0)	4	(4, 0)	$x^2 + y^2 + z^2 + 1 = 0$	$\emptyset$
				(3, 1)	$x^2 + y^2 + z^2 - 1 = 0$	Sphäre
			3	(3, 0)	$x^2 + y^2 + z^2 = 0$	Punkt
			4	(3, 1)	$x^2 + y^2 - z^2 + 1 = 0$	zweischaliges Hyperboloid
				(2, 2)	$x^2 + y^2 - z^2 - 1 = 0$	einschaliges Hyperboloid
			3	(2, 1)	$x^2 + y^2 - z^2 = 0$	Kegel
			4	(3, 1)	$x^2 + y^2 - z = 0$	Paraboloid
	2	(2, 0)	3	(3, 0)	$x^2 + y^2 + 1 = 0$	$\emptyset$
				(2, 1)	$x^2 + y^2 - 1 = 0$	Kreiszyylinder
			2	(2, 0)	$x^2 + y^2 = 0$	Gerade
			4	(2, 2)	$x^2 - y^2 + z = 0$	Sattelfläche
			3	(2, 1)	$x^2 - y^2 + 1 = 0$	hyperbolischer Zylinder
			2	(1, 1)	$x^2 - y^2 = 0$	schneidendes Ebenenpaar
			3	(2, 1)	$x^2 + y = 0$	parabolischer Zylinder
	1	(1, 0)	2	(2, 0)	$x^2 + 1 = 0$	$\emptyset$
				(1, 1)	$x^2 - 1 = 0$	paralleles Ebenenpaar
			1	(1, 0)	$x^2 = 0$	Ebene

Das ist doch eine recht eindrucksvolle Tabelle: in einer Dimension drei, in zwei Dimensionen neun, in drei Dimensionen 17 Fälle. Mich beeindruckt am meisten, dass man alle diese Normalformen alleine durch den Rang und Index der Koeffizientenmatrix und der geränderten Matrix unterscheiden kann. Natürlich sind fast alle Fälle Schrott (= Entartungsfälle).

**Definition 1.20** Eine Quadrik  $Q$  heißt nicht-entartet, wenn die erweiterte Koeffizientenmatrix invertierbar ist.

Lassen wir die leeren Mengen beiseite, so gibt es also die folgenden nicht-entarteten Quadriken:

$n$	Quadrik
1	zwei Punkte
2	Kreis Hyperbel Parabel
3	Sphäre zweischaliges Hyperboloid einschaliges Hyperboloid Paraboloid Sattelfläche

**Definition 1.21** Eine Quadrik  $Q$  in der affinen Ebene  $\mathbb{R}^2$  heißt Kegelschnitt.

Kegelschnitte (natürlich im Reellen) haben schon die alten Griechen gekannt und ausgiebig untersucht. Sie haben sie definiert als den Durchschnitt eines Kegels mit einer Ebene, daher der Name.

Den Durchschnitt einer Quadrik

$$\mathbf{x}^t \cdot A \cdot \mathbf{x} + 2\mathbf{b}^t \cdot \mathbf{x} + c = 0$$

mit einer Geraden berechnet man, indem man die Parametrisierung der Geraden

$$\mathbf{x} = \mathbf{v} + s\mathbf{w}, \quad s \in \mathbb{R},$$

in die Quadrikgleichung einsetzt:

$$\begin{aligned} & (\mathbf{v} + s\mathbf{w})^t \cdot A \cdot (\mathbf{v} + s\mathbf{w}) + 2\mathbf{b}^t \cdot (\mathbf{v} + s\mathbf{w}) + c \\ &= \mathbf{v}^t A \mathbf{v} + 2\mathbf{v}^t A \mathbf{w} \cdot s + \mathbf{w}^t A \mathbf{w} \cdot s^2 + 2\mathbf{b}^t \mathbf{v} + 2\mathbf{b}^t \mathbf{w} \cdot s + c \\ &= 0. \end{aligned}$$

Das ist eine quadratische Gleichung in  $s$ . Natürlich kann es passieren, dass der Koeffizient bei  $s^2$  verschwindet, und sich der Grad der Gleichung erniedrigt. Aber i.A. ist es eine echte quadratische Gleichung in  $s$ . Die braucht keine reellen Lösungen zu haben, dann schneidet die Gerade die Quadrik nicht. Oder sie hat zwei reelle Lösungen, dann schneidet die Gerade die Quadrik in zwei Punkten, oder die beiden reellen Lösungen fallen zusammen, dann berührt die Gerade die Quadrik in einem Punkt, und heißt *Tangente*.

Jetzt wollen wir nicht mehr beliebige affine Transformationen benutzen können, um eine Quadrik in eine Normalform zu transformieren, sondern nur Bewegungen. Die Normalform einer Quadrik, die dabei herauskommt, heißt deren *metrische*, oder *euklidische Normalform*.

Betrachten wir eine Quadrik mit erweiterter Koeffizientenmatrix

$$A' = \begin{pmatrix} A & \mathbf{b} \\ \mathbf{b}^t & c \end{pmatrix}$$

und gehen wir die Transformationen im Beweis von Satz 1.25 nochmal durch:

- 1) Als erstes wurde die Koeffizientenmatrix  $A$  mit Satz 1.21 durch eine lineare Transformation in Diagonalform transformiert. Wir können aber auch die Hauptachsentransformation verwenden, und dasselbe mit einer *orthogonalen* Transformation erreichen. Ganz genau dasselbe natürlich nicht: Durch lineare Transformationen bekommt man eine Diagonalmatrix mit den Einträgen  $\pm 1$  und  $0$ . Nach einer orthogonalen Transformation stehen auf der Diagonale die Eigenwerte von  $A$ . Die Anzahlen der positiven, negativen, oder Null-Einträge ist dieselbe, wie in der affinen Normalform, den Wert der Einträge  $a_k \neq 0$  können wir jetzt aber nicht mehr auf  $\pm 1$  normieren.
- 2) Durch eine Translation können wir, ganz genau so wie in 1.25 die Gleichung der Quadrik in eine Form

$$\sum_{k=1}^r a_k x_k^2 + \sum_{k=r+1}^n b_k x_k + c = 0$$

transformieren.

- 3) Durch eine orthogonale Transformation kann man jetzt die Linearform  $\mathbf{b}^t \cdot \mathbf{x}$  nicht mehr auf  $\mathbf{e}_{r+1}^t \cdot \mathbf{x}$  transformieren, sondern nur noch auf

$$\mathbf{x} \mapsto b \cdot \mathbf{e}_{r+1}^t \cdot \mathbf{x} = b \cdot x_{r+1} \quad \text{mit} \quad b = \|\mathbf{b}\|.$$

Wenn  $b \neq 0$  ist, kann man die Gleichung natürlich durch  $b$  teilen, und damit diese Konstante auf 1 normieren.

- 4) Wenn  $b \neq 0$  ist, kann man durch eine abschließende Translation  $x_{r+1} \mapsto x_{r+1} - c/b$  noch  $bx_{r+1} + c$  in  $bx_{r+1}$  transformieren.

**Satz 1.26 (Metrische Normalform)** *Die Gleichung einer Quadrik  $Q \subset \mathbb{R}^n$  kann durch eine Bewegung entweder auf eine Form ohne linearen Anteil*

$$\sum_{k=1}^r a_k x_k^2 + c = 0$$

oder auf eine Form

$$\sum_{k=1}^r a_k x_k^2 + bx_{r+1} = 0$$

gebracht werden. Die möglichen Fälle hängen wie in Satz 1.25 von der Beziehung zwischen dem Rang der Koeffizientenmatrix  $A$  und dem Rang der erweiterten Koeffizientenmatrix  $A'$  ab.

Es lohnt sich natürlich nicht, für alle 29 Fälle aus der Tabelle bei Satz 1.25 nochmal die metrische Normalform hinzuschreiben. Ich möchte diese Normalformen nur für die nichtentarteten, nichtleeren Quadriken in Dimension zwei und drei angeben. Dabei schreibe ich eine reelle Zahl  $> 0$  als Quadrat  $a^2$ ,  $a \in \mathbb{R}$ , eine Zahl  $< 0$  als  $-a^2$ ,  $a \in \mathbb{R}$ . Weil man die Gleichung mit einer beliebigen Konstante  $\neq 0$  durchmultiplizieren darf, kann man das konstante Glied, wenn vorhanden, immer auf 1 normieren.

$n$	$Sign(A)$	$Sign(A')$	Gleichung	Quadrik
2	(2, 0)	(2, 1)	$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$	Ellipse
	(1, 1)	(2, 1)	$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$	Hyperbel
	(1, 0)	(2, 1)	$y = p \cdot x^2$	Parabel
3	(3, 0)	(3, 1)	$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1$	Ellipsoid
	(2, 1)	(3, 1)	$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = -1$	zweischaliges Hyperboloid
	(2, 1)	(2, 2)	$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 1$	einschaliges Hyperboloid
	(2, 0)	(3, 1)	$z = \frac{x^2}{a^2} + \frac{y^2}{b^2}$	Paraboloid
	(1, 1)	(2, 2)	$z = \frac{x^2}{a^2} - \frac{y^2}{b^2}$	Sattelfläche

Die Achsen eines Koordinatensystems, in dem die Quadrik  $Q$  eine der angegebenen Normalformen annimmt, heißen die *Hauptachsen* der Quadrik. Daher kommt überhaupt der Name Hauptachsentransformation. Ihre Richtungen sind die Richtungen der Eigenvektoren der symmetrischen Matrix  $A$ . Manchmal nennt man auch die Länge, welche die Quadrik auf einer dieser Achsen ausschneidet, die Hauptachsen. Ist  $\lambda > 0$  der Eigenwert zum Eigenvektor in Richtung einer dieser Achsen, und ist die Konstante in der Gleichung auf 1 normiert, so ist diese Strecke

$$a = \frac{1}{\sqrt{\lambda}}.$$

Eine Bewegung bildet eine Ellipse mit den Hauptachsen  $a$  und  $b$  immer auf eine Ellipse mit denselben Hauptachsen  $a$  und  $b$  ab. Und die Bewegung führt auch die Richtungen der Hauptachsen ineinander über. Bei einer affinen Transformation ist das nicht so. So ist etwa das Bild des Kreises

$$x^2 + y^2 = 1$$

unter der affinen Transformation

$$\xi = a \cdot x, \eta = b \cdot y,$$

die Ellipse

$$\frac{\xi^2}{a^2} + \frac{\eta^2}{b^2} = 1.$$

So sieht man: *Jede Ellipse ist das affine Bild eines Kreises*. Das kann man ausnutzen, um Aussagen für Ellipsen zu beweisen. Etwa: Eine Gerade schneidet eine Ellipse in zwei Punkten, in einem Punkt (und berührt sie dann), oder überhaupt nicht. Oder: Durch einen Punkt  $\mathbf{p}$  außerhalb einer Ellipse gibt es zwei Tangenten an diese Ellipse.

**Aufgabe 1.23 a)** Bestimmen Sie eine invertierbare  $3 \times 3$ -Matrix  $M$  mit reellen Koeffizienten und einen Vektor  $b \in \mathbb{R}^3$  so, dass die quadratische Gleichung

$$x^2 + y^2 + z^2 + xy + xz + yz + x + y + z = 1$$

mit den Variablen  $x, y, z$  die Form

$$\alpha u^2 + \beta v^2 + \gamma w^2 = \rho \text{ mit } \alpha, \beta, \gamma, \rho \in \mathbb{R}$$

erhält, falls man  $(x, y, z)$  durch  $(u, v, w) \cdot M + b$  ersetzt.

b) Um welche Quadrik handelt es sich unter a)?

**Aufgabe 1.24** Man bestimme alle in

$$\{x \in \mathbb{R}^3 \mid x_1^2 - x_2^2 - x_3^2 + 2x_2x_3 + 6x_1 + 2x_2 + 1 = 0\}$$

enthaltenen Geraden durch  $(-1, 2, 2)$ .

**Aufgabe 1.25** Im  $\mathbb{R}^3$  mit dem üblichen Skalarprodukt sei  $E$  das Ellipsoid, das durch die Gleichung

$$\left(\frac{x}{2}\right)^2 + \left(\frac{y}{2}\right)^2 + \left(\frac{z}{3}\right)^2 = 1$$

definiert ist. Man berechne die Hauptachsen der Ellipse, die als Schnitt der Ebene  $x - y + 2z = 0$  mit dem Ellipsoid  $E$  entsteht.

**Aufgabe 1.26** Man transformiere folgende Kurve zweiter Ordnung im  $\mathbb{R}^2$  auf ihre euklidische Normalform:

$$\{x \in \mathbb{R}^2 \mid 2x_1^2 - 2x_1x_2 + 2x_2^2 + 2x_1 - 3x_2 - 5 = 0\}.$$

**Aufgabe 1.27** Sei

$$Q = \{(x, y, z) \in \mathbb{R}^3 \mid \frac{5}{16}x^2 + y^2 + \frac{5}{16}z^2 - \frac{3}{8}xz - \frac{1}{2}x - \frac{1}{2}z = 0\}.$$

a) Man zeige, dass  $Q$  ein Ellipsoid ist, und bestimme dessen Mittelpunkt und Hauptachsen.

b) Man gebe eine affine Abbildung  $f : \mathbb{R}^3 \rightarrow \mathbb{R}^3$  an, so dass  $f$  eine Bijektion der Einheitssphäre  $S^2 = \{(x, y, z) \in \mathbb{R}^3 \mid x^2 + y^2 + z^2 = 1\}$  auf  $Q$  induziert.

**Aufgabe 1.28** Man zeige, dass durch die Gleichung

$$5x^2 - 2xy + 5y^2 + 10x - 2y - 6 = 0$$

im  $\mathbb{R}^2$  eine Ellipse definiert ist. Ferner bestimme man ihren Mittelpunkt, ihre Hauptachsen, die Länge der Achsenabschnitte und skizziere die Ellipse.

**Aufgabe 1.29** Gegeben sei die Matrix

$$C := \begin{pmatrix} 7 & 0 & 3 \\ 0 & -2 & 0 \\ 3 & 0 & -1 \end{pmatrix}.$$

a) Bestimmen Sie alle Eigenwerte, die zugehörigen Eigenräume und ON-Basen in den Eigenräumen.

b) Geben Sie eine orthogonale Matrix an, die  $C$  diagonalisiert.

c) Für einen Vektor  $x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$  bezeichne  $x^t$  den zugehörigen Zeilenvektor. Geben Sie an, welches geometrische Gebilde die Quadrik

$$Q := \{x \mid x \in \mathbb{R}^3, x^t C x = 1\}$$

darstellt (Skizze!).

**Aufgabe 1.30** Transformieren Sie die folgende Quadrik auf Hauptachsenform:

$$5x_1^2 + 7x_2^2 + 6x_3^2 - 4x_1x_3 - 4x_2x_3 - 2x_1 + 20x_2 - 8x_3 + 9 = 0$$

**Aufgabe 1.31** Gegeben sei die quadratische Form

$$\Phi(x_1, x_2, x_3) = 2 \sum_{i < k, i, k=1}^3 x_i x_k.$$

Transformieren Sie  $\Phi$  auf Hauptachsen, und geben Sie die (orthogonale) Transformationsmatrix, sowie den Trägheitsindex (d.h., die Signatur) von  $\Phi$  an.

**Aufgabe 1.32** Im euklidischen Raum  $\mathbb{R}^3$  sei eine Quadrik  $Q$  durch die Gleichung

$$x_1^2 + 2x_2x_3 - 2x_1 + 2 = 0$$

gegeben. Bestimmen Sie eine eigentliche Bewegung des  $\mathbb{R}^3$  der Form  $x = Sy + t$ , so dass  $Q$  in den Koordinaten  $y_1, y_2, y_3$  euklidische Normalform annimmt. Welcher Quadriktyp liegt vor? Zeigen Sie weiter, dass  $Q$  rotationssymmetrisch ist, und bestimmen Sie im ursprünglichen Koordinatensystem die Rotationsachse.

**Aufgabe 1.33** Im  $\mathbb{R}^4$  sei  $\langle, \rangle: \mathbb{R}^4 \rightarrow \mathbb{R}$  die Minkowski-Metrik

$$\left\langle \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix}, \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} \right\rangle := x_1y_1 - x_2y_2 - x_3y_3 - x_4y_4,$$

und  $U$  sei der lineare Unterraum  $\{x \mid x_1 + x_2 + x_3 + x_4 = 0\}$ . Beschreiben Sie die Quadrik  $Q_U := \{x \in U \mid \langle x, x \rangle = 1\}$  in  $U$  durch Koordinaten in  $U$ . Von welchem affinen Typ ist  $Q_U$ ?

**Aufgabe 1.34** Sei  $Q \subset \mathbb{R}^3$  das hyperbolische Paraboloid gegeben durch die Gleichung  $z = x^2 - y^2$ . Zeigen Sie: Durch jeden Punkt von  $Q$  gehen genau zwei Geraden, die ganz in  $Q$  verlaufen.

## 1.7 Innere Produkte und adjungierte Operatoren

**Definition 1.22** Ein inneres Produkt auf dem reellen Vektorraum  $V$  ist eine symmetrische, positiv-definite Bilinearform

Der Raum  $V$  zusammen mit dem inneren Produkt heißt euklidischer Raum.

**Beispiel 1.14** Das euklidische Skalarprodukt

$$(\mathbf{x}, \mathbf{y}) = \sum_{\nu=1}^n x_\nu y_\nu$$

auf dem  $\mathbb{R}^n$  ist ein inneres Produkt.

**Beispiel 1.15** Es sei  $V$  der Vektorraum aller stetigen Funktionen auf dem Intervall  $[a, b]$ . Dann ist

$$\langle f, g \rangle := \int_a^b f(x)g(x)dx$$

ein inneres Produkt auf  $V$ .

Ist  $\langle \mathbf{v}, \mathbf{w} \rangle$  ein inneres Produkt auf dem endlich-dimensionalen Vektorraum  $V$ , so kann man wie im Gram-Schmidtschen Orthonormalisierungsverfahren eine Orthonormalbasis  $\mathbf{v}_1, \dots, \mathbf{v}_n$  finden. Es ist also  $\langle \mathbf{v}_\mu, \mathbf{v}_\nu \rangle = \delta_{\mu, \nu}$ . Geht man in diese Orthonormalbasis über, so wird das innere Produkt das standardmäßige euklidische Skalarprodukt.

Für endlich-dimensionale Vektorräume ist also der Begriff des inneren Produkts absolut überflüssig. Anders ist es bei Vektorräumen unendlicher Dimension. Deswegen stelle ich hier schnell nochmal die Eigenschaften eines inneren Produkts  $\langle \mathbf{v}, \mathbf{w} \rangle$  und der damit definierten Norm  $\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$  zusammen:

- 1)  $\|c\mathbf{v}\| = |c| \cdot \|\mathbf{v}\|$  für  $c \in \mathbb{R}$ .
- 2)  $\|\mathbf{v}\| \geq 0$  und  $\|\mathbf{v}\| = 0$  nur dann, wenn  $\mathbf{v} = \mathbf{0}$ .
- 3)  $|\langle \mathbf{v}, \mathbf{w} \rangle| \leq \|\mathbf{v}\| \cdot \|\mathbf{w}\|$  (Cauchy-Schwarz-Ungleichung)
- 4)  $\|\mathbf{v} + \mathbf{w}\| \leq \|\mathbf{v}\| + \|\mathbf{w}\|$  (Dreiecksungleichung)

**Beispiel 1.16 (Hilbert-Raum  $l^2(\mathbb{R})$ )** Man definiert den Folgenraum

$$l^2(\mathbb{R}) := \{(a_\nu)_{\nu \in \mathbb{N}} : \sum_0^\infty a_\nu^2 \text{ ist konvergent}\}.$$

Darauf definiert man das innere Produkt

$$\langle (a_\nu), (b_\nu) \rangle := \sum_{\nu=0}^\infty a_\nu b_\nu.$$

Zunächst ist weder klar, dass  $l^2(\mathbb{R})$  ein Vektorraum ist, noch, dass die Reihe in der Definition des inneren Produkts konvergiert. Aber für Folgen  $(a_\nu), (b_\nu) \in l^2(\mathbb{R})$  gibt es Schranken  $A, B \in \mathbb{R}$  mit

$$\sum_{\nu=0}^N a_\nu^2 < A^2 \text{ und } \sum_{\nu=0}^N b_\nu^2 < B^2$$

für alle  $N$ . Mit der Dreiecksungleichung und der Cauchy-Schwarz-Ungleichung im  $\mathbb{R}^{N+1}$  folgt daraus

$$\left| \sum_{\nu=0}^N a_\nu b_\nu \right| \leq \sum_{\nu=0}^N |a_\nu| \cdot |b_\nu| \leq A \cdot B.$$

Die Reihe  $\sum a_\nu b_\nu$  konvergiert also absolut. Und damit sieht man dann, dass die Reihe

$$\sum_{\nu=0}^\infty (a_\nu + b_\nu)^2 = \sum_{\nu=0}^\infty a_\nu^2 + 2 \sum_{\nu=1}^\infty a_\nu b_\nu + \sum_{\nu=0}^\infty b_\nu^2$$

konvergiert. Mit  $(a_\nu)$  und  $(b_\nu)$  gehört also auch die Summe  $(a_\nu + b_\nu)$  zu  $l^2(\mathbb{R})$ .

Was eine Basis in einem unendlich-dimensionalen Vektorraum ist, darüber kann man streiten. Aber wie im Endlichdimensionalen kann man definieren:

**Definition 1.23** Es sei  $V$  ein Vektorraum mit dem inneren Produkt  $\langle \mathbf{v}, \mathbf{w} \rangle$ . Eine Folge  $\mathbf{v}_1, \mathbf{v}_2, \dots$  von Vektoren in  $V$  heißt

<table style="border-collapse: collapse; width: 100%;"> <tr> <td style="padding: 5px;">Orthogonalsystem (OS)</td> <td style="padding: 5px;">wenn <math>\langle \mathbf{v}_\mu, \mathbf{v}_\nu \rangle =</math></td> </tr> <tr> <td style="padding: 5px;">Orthonormalsystem (ONS)</td> <td style="padding: 5px;"><math>= 0</math> für <math>\mu \neq \nu</math></td> </tr> <tr> <td style="padding: 5px;"></td> <td style="padding: 5px;"><math>= \delta_{\mu, \nu}</math></td> </tr> </table>	Orthogonalsystem (OS)	wenn $\langle \mathbf{v}_\mu, \mathbf{v}_\nu \rangle =$	Orthonormalsystem (ONS)	$= 0$ für $\mu \neq \nu$		$= \delta_{\mu, \nu}$	
Orthogonalsystem (OS)	wenn $\langle \mathbf{v}_\mu, \mathbf{v}_\nu \rangle =$						
Orthonormalsystem (ONS)	$= 0$ für $\mu \neq \nu$						
	$= \delta_{\mu, \nu}$						

**Beispiel 1.17** Wir betrachten den Vektorraum  $V$  der stetigen Funktionen auf dem Intervall  $[0, 2\pi]$  mit dem Inneren Produkt aus Beispiel 1.15. In  $V$  betrachten wir die Funktionen

$$c_\nu(x) := \frac{1}{\sqrt{\pi}} \cos(\nu x), \quad s_\nu(x) := \frac{1}{\sqrt{\pi}} \sin(\nu x).$$

Die Funktionen

$$\frac{1}{\sqrt{2\pi}}, c_1, s_1, c_2, s_2, \dots$$

bilden ein ONS in  $V$ . Der Beweis besteht darin, die vorkommenden Integrale explizit auszurechnen.

**Satz 1.27** Es sei  $\mathbf{v}_1, \mathbf{v}_2, \dots$  ein ONS in  $V$ .

- a) Die Vektoren des ONS sind linear unabhängig.
- b) Jeder Vektor  $\mathbf{v} \in \text{span}\{\mathbf{v}_1, \mathbf{v}_2, \dots\}$  ist eine endliche Linearkombination

$$\mathbf{v} = \sum_{\text{endlich}} \langle \mathbf{v}, \mathbf{v}_\nu \rangle \mathbf{v}_\nu.$$

- c) (Besselsche Ungleichung) Für jeden Vektor  $\mathbf{v}$  und jedes  $k$  gelten

$$\sum_1^k \langle \mathbf{v}, \mathbf{v}_\nu \rangle^2 \leq \|\mathbf{v}\|^2$$

und

$$\sum_1^k \langle \mathbf{v}, \mathbf{v}_\nu \rangle^2 = \|\mathbf{v}\|^2 \Leftrightarrow \mathbf{v} = \sum_1^k \langle \mathbf{v}, \mathbf{v}_\nu \rangle \mathbf{v}_\nu.$$

Beweis a) Es sei  $\sum_1^k c_\nu \mathbf{v}_\nu = \mathbf{0}$ . Für  $\mu = 1, \dots, k$  folgt daraus

$$c_\mu = \sum_{\nu=1}^k c_\nu \langle \mathbf{v}_\nu, \mathbf{v}_\mu \rangle = \langle \sum_{\nu=1}^k c_\nu \mathbf{v}_\nu, \mathbf{v}_\mu \rangle = 0.$$

b) Der Vektor  $\mathbf{v}$  ist eine endliche Linearkombination  $\sum c_\nu \mathbf{v}_\nu$ . Die Koeffizienten sind hier wegen a) durch  $\mathbf{v}$  eindeutig bestimmt. Und es ist

$$\langle \mathbf{v}, \mathbf{v}_\mu \rangle = \langle \sum_{\nu} c_\nu \mathbf{v}_\nu, \mathbf{v}_\mu \rangle = c_\mu.$$

c) Wir definieren  $\mathbf{w} \in V$  durch

$$\mathbf{v} = \sum_1^k \langle \mathbf{v}, \mathbf{v}_\nu \rangle \mathbf{v}_\nu + \mathbf{w}.$$

Offensichtlich ist  $\langle \mathbf{w}, \mathbf{v}_\nu \rangle = 0$  für alle  $\nu$ . Mit Pythagoras folgt daraus

$$\|\mathbf{v}\|^2 = \sum_1^k \langle \mathbf{v}, \mathbf{v}_\nu \rangle^2 + \|\mathbf{w}\|^2.$$

□

**Definition 1.24** Es sei  $U \subset V$  ein endlich-dimensionaler Untervektorraum und  $\mathbf{v}_1, \dots, \mathbf{v}_n$  eine Orthonormal-Basis von  $U$ . Für jeden Vektor  $\mathbf{v} \in V$  hat

$$P(\mathbf{v}) := \sum_{\nu=1}^n \langle \mathbf{v}, \mathbf{v}_\nu \rangle \mathbf{v}_\nu \in U$$

die Eigenschaft

$$\mathbf{w} := \mathbf{v} - P(\mathbf{v}) \perp U.$$

Der Vektor  $P(\mathbf{v})$  heißt die Orthogonalprojektion von  $\mathbf{v}$  in  $U$ .

**Beispiel 1.18** Es sei  $V$  der Vektorraum der stetigen Funktionen auf  $[0, 2\pi]$  mit dem inneren Produkt aus Beispiel 1.15. Weiter sei  $N \in \mathbb{N}$  und  $U$  der  $2N + 1$ -dimensionale Untervektorraum, der von den Funktionen

$$1, c_1(x) = \cos(x), s_1(x) = \sin(x), \dots, c_N(x) = \cos(Nx), s_N(x) = \sin(Nx)$$

aufgespannt wird. Eine Orthonormalbasis von  $U$  bilden die Funktionen

$$\frac{1}{\sqrt{2\pi}}, \frac{c_1}{\sqrt{\pi}}, \frac{s_1}{\sqrt{\pi}}, \dots, \frac{c_N}{\sqrt{\pi}}, \frac{s_N}{\sqrt{\pi}}.$$

Ist  $f \in V$ , so heißen die Zahlen ( $\nu = 1, \dots, N$ )

$$\begin{aligned} a_0 &:= \left\langle f, \frac{1}{\sqrt{2\pi}} \right\rangle = \frac{1}{\sqrt{2\pi}} \int_0^{2\pi} f(x) dx \\ a_\nu &:= \left\langle f, \frac{c_\nu}{\sqrt{\pi}} \right\rangle = \frac{1}{\sqrt{\pi}} \int_0^{2\pi} f(x) \cos(\nu x) dx \\ b_\nu &:= \left\langle f, \frac{s_\nu}{\sqrt{\pi}} \right\rangle = \frac{1}{\sqrt{\pi}} \int_0^{2\pi} f(x) \sin(\nu x) dx \end{aligned}$$

die Fourier-Koeffizienten von  $f$ . Das trigonometrische Polynom

$$\frac{a_0}{\sqrt{2\pi}} + a_1 \cos(x) + b_1 \sin(x) + \dots + a_N \cos(Nx) + b_N \sin(Nx)$$

heißt das  $N$ -te Fourier-Polynom von  $f$ . Es ist die Orthogonalprojektion von  $f$  in den Unterraum  $U$ .

Jetzt ist es an der Zeit, zu verraten, dass es ein komplexes Analogon zu der bisher dargestellten Theorie der inneren Produkte gibt. Sei also jetzt  $V$  ein  $\mathbb{C}$ -Vektorraum. Ist etwa  $\mathbb{C}^n = \mathbb{R}^n \oplus \mathbb{R}^n$  mit

$$\mathbf{v} = (z_1, \dots, z_n) = (x + iy_1, \dots, x_n + iy_n) \in V,$$

so ist die reelle euklidische Norm von  $\mathbf{v}$

$$\|\mathbf{v}\| = \sqrt{\sum_{\nu=1}^n x_\nu^2 + y_\nu^2} = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$$

mit

$$\langle \mathbf{v}, \mathbf{v} \rangle = \sum_{\nu=1}^n \bar{z}_\nu \cdot z_\nu.$$

**Definition 1.25** *Das Standard-hermitesche innere Produkt auf  $\mathbb{C}^n$  ist die Abbildung*

$$\mathbb{C}^n \times \mathbb{C}^n \rightarrow \mathbb{C}, \quad (\mathbf{v}, \mathbf{w}) \mapsto \sum_{\nu=1}^n \bar{v}_\nu \cdot w_\nu.$$

*Mathematiker machen den Querstrich immer über dem zweiten Faktor, aber ich folge hier der Konvention der Physiker.*

**Definition 1.26 (Komplexes inneres Produkt)** *Es sei  $V$  ein  $\mathbb{C}$ -Vektorraum. Ein inneres Produkt auf  $V$  ist eine Abbildung*

$$V \times V \rightarrow \mathbb{C}, \quad (\mathbf{v}, \mathbf{w}) \mapsto \langle \mathbf{v}, \mathbf{w} \rangle$$

*mit folgenden Eigenschaften:*

- *Der Wert  $\langle \mathbf{v}, \mathbf{w} \rangle$  ist linear bezüglich  $\mathbf{w}$ , aber im ersten Argument gilt*

$$\langle c_1 \mathbf{v}_1 + c_2 \mathbf{v}_2, \mathbf{w} \rangle = \bar{c}_1 \langle \mathbf{v}_1, \mathbf{w} \rangle + \bar{c}_2 \langle \mathbf{v}_2, \mathbf{w} \rangle.$$

- *(Hermite-Symmetrie) Für alle  $\mathbf{v}, \mathbf{w} \in V$  ist*

$$\langle \mathbf{w}, \mathbf{v} \rangle = \overline{\langle \mathbf{v}, \mathbf{w} \rangle}.$$

- *(Positiv-Definitheit) Für alle  $\mathbf{v} \in V$  ist*

$$\langle \mathbf{v}, \mathbf{v} \rangle \geq 0 \quad \text{und} \quad \langle \mathbf{v}, \mathbf{v} \rangle = 0 \quad \text{nur für } \mathbf{v} = 0.$$

*(Dies ist sinnvoll. Aus der Hermite-Symmetrie folgt nämlich  $\langle \mathbf{v}, \mathbf{v} \rangle \in \mathbb{R}$  für alle  $\mathbf{v} \in V$ .)*

Wie im Reellen ist auch im Komplexen

$$\|\mathbf{v}\| := \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$$

eine Norm auf  $V$ . Für diese Norm gelten die oben festgehaltenen Eigenschaften 1),...,4) wörtlich wie im Reellen. Auch die Sache mit den Ortho-Normalsystemen ist wörtlich identisch.

Die folgende Definition ist nicht besonders aufregend, aber nützlich als Abkürzung.

**Definition 1.27** Ein euklidischer Vektorraum ist ein  $\mathbb{R}$ -Vektorraum mit einem inneren Produkt. Ein unitärer Vektorraum ist ein  $\mathbb{C}$ -Vektorraum mit einem komplexen inneren Produkt.

Jetzt werde ich lineare Abbildungen  $\varphi : V \rightarrow V$  untersuchen, die sich in Bezug auf ein inneres Produkt speziell verhalten. Als erster speziellen Sorte von Matrizen begegnet man wohl orthogonalen Matrizen.

Das euklidische Standard-Produkt auf dem  $\mathbb{R}^n$  kann man schreiben als

$$\langle \mathbf{v}, \mathbf{w} \rangle = \mathbf{v}^t \cdot \mathbf{w}.$$

Für orthogonale Matrizen  $T$  gilt  $T^t = T^{-1}$ . Daraus folgt bekanntlich

$$\langle T\mathbf{v}, T\mathbf{w} \rangle = (T\mathbf{v})^t \cdot T\mathbf{w} = \mathbf{v}^t \cdot T^t \cdot T \cdot \mathbf{w} = \mathbf{v}^t \cdot \mathbf{w} = \langle \mathbf{v}, \mathbf{w} \rangle.$$

Analog kann man das hermitesche Standard-Produkt auf dem  $\mathbb{C}^n$  schreiben als

$$\langle \mathbf{v}, \mathbf{w} \rangle = \bar{\mathbf{v}}^t \cdot \mathbf{w}.$$

Deswegen gilt die entsprechende Formel

$$\langle U\mathbf{v}, U\mathbf{w} \rangle = \langle \mathbf{v}, \mathbf{w} \rangle$$

für komplexe Matrizen  $U$  mit

$$\bar{U}^t \cdot U = \mathbb{1}, \quad \text{bzw.} \quad U^{-1} = \bar{U}^t.$$

**Definition 1.28** Eine unitäre Matrix ist eine invertierbare komplexe  $n \times n$ -Matrix mit  $U^{-1} = \bar{U}^t$ .

Ist die unitäre Matrix reell, so ist sie orthogonal.

Wenn man nicht im  $\mathbb{R}^n$ , bzw.  $\mathbb{C}^n$  mit dem inneren Standard-Produkt ist, sondern in allgemeineren Vektorräumen, so definiert man:

**Definition 1.29** Es sei  $V$  ein Vektorraum mit einem inneren Produkt. Eine lineare Abbildung  $\varphi : V \rightarrow V$  mit

$$\langle \varphi(\mathbf{v}), \varphi(\mathbf{w}) \rangle = \langle \mathbf{v}, \mathbf{w} \rangle \quad \text{für alle } \mathbf{v}, \mathbf{w} \in V$$

heißt

$$\begin{array}{ll} \text{orthogonal} & \text{im reellen Fall,} \\ \text{unitär} & \text{im komplexen Fall.} \end{array}$$

Eine andere Art von Matrizen, die sich gegenüber dem euklidischen Skalarprodukt im  $\mathbb{R}^n$  speziell verhalten, sind die reellen symmetrische Matrizen. Aus  $S = S^t$  folgt

$$\langle S\mathbf{v}, \mathbf{w} \rangle = (S\mathbf{v})^t \cdot \mathbf{w} = \mathbf{v}^t \cdot S^t \cdot \mathbf{w} = \mathbf{v}^t \cdot S\mathbf{w} = \langle \mathbf{v}, S\mathbf{w} \rangle.$$

Welche Eigenschaft muss eine komplexe  $n \times n$ -Matrix  $H$  besitzen, damit beim hermiteschen Standard-Produkt für alle  $\mathbf{v}$  und  $\mathbf{w} \in \mathbb{C}^n$  gilt

$$\langle H\mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{v}, H\mathbf{w} \rangle \quad ?$$

Die linke Seite dieser Gleichung ist als Matrizenprodukt

$$\overline{(H\mathbf{v})}^t \cdot \mathbf{w} = \bar{\mathbf{v}}^t \cdot \bar{H}^t \cdot \mathbf{w}.$$

Und die rechte Seite ist

$$\bar{\mathbf{v}}^t \cdot H \cdot \mathbf{w}.$$

Beide Seiten stimmen genau dann für alle  $\mathbf{v}, \mathbf{w} \in \mathbb{C}^n$  überein, wenn

$$H = \bar{H}^t.$$

**Definition 1.30** Eine komplexe  $n \times n$ -Matrix  $H$  heißt hermitesch, wenn  $H = H^\dagger := \bar{H}^t$ .

Eine reelle Matrix ist (als komplexe Matrix) genau dann hermitesch, wenn sie symmetrisch ist. Schreibt man die komplexe Matrix  $H$  als

$$H = \operatorname{Re}(H) + i \cdot \operatorname{Im}(H),$$

so ist

$$H \text{ hermitesch} \Leftrightarrow \operatorname{Re}(H) \text{ symmetrisch und } \operatorname{Im}(H) \text{ schiefsymmetrisch.}$$

Um den Begriff der symmetrischen, bzw. hermiteschen Matrizen auf lineare Abbildungen zu übertragen, muss man etwas ausholen.

**Definition 1.31 (Adjungierte Abbildung)** Es seien  $V$  und  $W$  Vektorräume (beide über  $\mathbb{R}$  oder  $\mathbb{C}$ ) mit inneren Produkten, die wir in beiden Fällen als  $\langle \cdot, \cdot \rangle$  schreiben wollen. Ist  $\varphi : V \rightarrow W$  linear, so heißt eine lineare Abbildung  $\psi : W \rightarrow V$  adjungiert zu  $\varphi$ , wenn

$$\langle \varphi(\mathbf{v}), \mathbf{w} \rangle = \langle \mathbf{v}, \psi(\mathbf{w}) \rangle \quad \text{für alle } \mathbf{v} \in V, \mathbf{w} \in W.$$

**Beispiel 1.19** Ist  $V = \mathbb{R}^m$ ,  $W = \mathbb{R}^n$  mit dem Standard-Skalarprodukt und  $\varphi : V \rightarrow W$  durch die  $n \times m$ -Matrix  $A$  gegeben, so definiert  $A^t$  eine adjungierte Abbildung. Im komplexen Fall definiert  $A^\dagger = \bar{A}^t$  eine adjungierte Abbildung.

**Beispiel 1.20** Es sei  $V$  der  $\mathbb{R}$ -Vektorraum der unendlich oft differenzierbaren Funktionen  $f : \mathbb{R} \rightarrow \mathbb{R}$  mit kompaktem Träger. D.H., zu jedem  $f$  gibt es eine Schranke  $C$  mit  $f(x) = 0$  für  $|x| > C$ . Solche Funktionen sind nicht einfach hinzuschreiben, aber  $V$  hat trotzdem unendliche Dimension. Auf  $V$  ist

$$\langle f, g \rangle := \int_{-\infty}^{\infty} f(x)g(x)dx$$

ein inneres Produkt. Aus

$$\int_{-\infty}^{\infty} (xf(x)) \cdot g(x) = \int_{-\infty}^{\infty} f(x) \cdot (xg(x))dx$$

folgt, dass die Abbildung

$$V \rightarrow V, \quad f(x) \mapsto x \cdot f(x)$$

zu sich selbst adjungiert ist. Für die Abbildung

$$\frac{d}{dx} : V \rightarrow V, \quad f \mapsto \frac{df}{dx}$$

folgt aus der Formel für partielle Integration

$$\int_{-\infty}^{\infty} \frac{df}{dx}(x) \cdot g(x) dx = - \int_{-\infty}^{\infty} f(x) \cdot \frac{dg}{dx}(x) dx,$$

dass die Abbildung  $-d/dx$  eine adjungierte Abbildung ist.

Betrachtet man komplex-wertige Funktionen mit dem inneren Produkt

$$\langle f, g \rangle := \int_{-\infty}^{\infty} \bar{f}(x)g(x) dx,$$

so ist die Abbildung  $i \cdot d/dx$  zu sich selbst adjungiert.

**Satz 1.28** a) (Eindeutigkeit) Die adjungierte Abbildung  $\psi$  zu einer Abbildung  $\varphi : V \rightarrow W$  ist, falls sie existiert, durch  $\varphi$  eindeutig bestimmt.

b) (Existenz) Sind  $V$  und  $W$  endlich-dimensional, so gibt es zu jeder linearen Abbildung  $\varphi : V \rightarrow W$  die adjungierte Abbildung.

c) Ist  $\psi$  adjungiert zu  $\varphi$ , so ist umgekehrt auch  $\varphi$  adjungiert zu  $\psi$ .

Beweis. a) Es seien  $\psi_1$  und  $\psi_2$  zwei adjungierte Abbildungen zu  $\varphi$ . Dann gilt also für alle  $\mathbf{v} \in V, \mathbf{w} \in W$

$$\langle \mathbf{v}, (\psi_1 - \psi_2)(\mathbf{w}) \rangle = \langle \varphi(\mathbf{v}), \mathbf{w} \rangle - \langle \varphi(\mathbf{v}), \mathbf{w} \rangle = 0.$$

Weil das innere Produkt auf  $V$  nicht-entartet ist, folgt  $(\psi_1 - \psi_2)(\mathbf{w}) = \mathbf{0}$  für alle  $\mathbf{w} \in W$ . Die Abbildung  $\psi_1 - \psi_2$  ist die Null-Abbildung.

b) folgt aus Beispiel 1.19.

c) Aus der Adjungiertheit von  $\psi$  zu  $\varphi$  folgt

$$\langle \psi(\mathbf{w}), \mathbf{v} \rangle = \overline{\langle \mathbf{v}, \psi(\mathbf{w}) \rangle} = \overline{\langle \varphi(\mathbf{v}), \mathbf{w} \rangle} = \langle \mathbf{w}, \varphi(\mathbf{v}) \rangle. \quad \square$$

**Satz 1.29** Es sei  $\varphi : V \rightarrow W$  adjungiert zu  $\psi : W \rightarrow V$ .

a) Es ist stets  $\text{Ker}(\psi) = \text{Bild}(\varphi)^\perp$ .

b) Wenn  $V$  endlich-dimensional ist, dann gilt auch  $\text{Bild}(\psi) = \text{Ker}(\varphi)^\perp$ .

Beweis. a) Der Vektor  $\mathbf{w} \in W$  gehört genau dann zu  $\text{Ker}(\psi)$ , wenn  $\langle \mathbf{v}, \psi(\mathbf{w}) \rangle = 0$  ist für alle  $\mathbf{v} \in V$ . Äquivalent dazu ist  $\langle \varphi(\mathbf{v}), \mathbf{w} \rangle = 0$  für alle  $\mathbf{v} \in V$ , und das heißt  $\mathbf{w} \in \text{Bild}(\varphi)^\perp$ .

b) Nach a) ist  $\text{Ker}(\varphi) = \text{Bild}(\psi)^\perp$ . Wenn  $V$  endlich-dimensional ist, gilt  $(U^\perp)^\perp = U$  für jeden Untervektorraum  $U \subset V$ . Dann ist also

$$\text{Bild}(\psi) = (\text{Bild}(\psi)^\perp)^\perp = \text{Ker}(\varphi)^\perp. \quad \square$$

**Beispiel 1.21** Es sei  $\varphi : \mathbb{R}^m \rightarrow \mathbb{R}^n$  gegeben durch die  $n \times m$ -Matrix  $A$ . Ein Vektor  $\mathbf{b} \in \mathbb{R}^n$  gehört genau dann zu  $\text{Bild}(\varphi)$ , wenn das Lineare Gleichungssystem  $A \cdot \mathbf{x} = \mathbf{b}$  lösbar ist. Nach Satz 1.29 a) ist das genau dann der Fall, wenn  $(\mathbf{b}, \mathbf{y}) = 0$  für alle  $\mathbf{y}$  mit  $A^t \cdot \mathbf{y} = \mathbf{0}$ .

**Definition 1.32** Eine lineare Abbildung  $\varphi : V \rightarrow V$  heißt selbst-adjungiert, wenn  $\varphi$  seine eigene adjungierte Abbildung ist.

Für den Rest dieses Abschnitts setzen wir voraus, dass  $V$  ein endlich-dimensionaler (reeller oder komplexer) Vektorraum mit dem inneren Produkt  $\langle \cdot, \cdot \rangle$  ist.

**Definition 1.33** Eine lineare Abbildung  $\varphi : V \rightarrow V$  heißt normal, wenn sie mit ihrer adjungierten Abbildung  $\psi$  kommutiert, d.h.,

$$\psi \circ \varphi = \varphi \circ \psi.$$

**Beispiel 1.22** a) Es sei  $V = \mathbb{R}^n$  mit dem euklidischen Standard-Skalarprodukt. Weiter sei  $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}^n$  linear mit darstellender Matrix  $A$ . Die folgenden Fälle normaler Abbildungen sind uns bekannt:

$A$	$A^t$	adjungierte Abbildung zu $\varphi$
symmetrisch	$A$	$\varphi$
antisymmetrisch	$-A$	$-\varphi$
orthogonal	$A^{-1}$	$\varphi^{-1}$

b) Nun sei  $V = \mathbb{C}^n$  mit dem hermiteschen Standard-Skalarprodukt. Wieder sei  $\varphi : \mathbb{C}^n \rightarrow \mathbb{C}^n$  linear mit darstellender Matrix  $A$ . In den folgenden Fällen ist  $\varphi$  normal:

$A$	$A^\dagger$	adjungierte Abbildung zu $\varphi$
hermitesch	$A$	$\varphi$
antihermitesch	$-A$	$-\varphi$
unitär	$A^{-1}$	$\varphi^{-1}$

Es sei  $\mathbf{v} \in V$  ein Eigenvektor von  $\varphi$  zum Eigenwert  $\lambda \in \mathbb{C}$ . Die zu  $\varphi$  adjungierte Abbildung sei  $\varphi^\dagger$ . Wir betrachten die bekannte Gleichungskette

$$\lambda \langle \mathbf{v}, \mathbf{v} \rangle = \langle \mathbf{v}, \varphi(\mathbf{v}) \rangle = \langle \varphi^\dagger(\mathbf{v}), \mathbf{v} \rangle.$$

In den Fällen aus Beispiel 1.22 haben wir

$\varphi^\dagger$	$\varphi^\dagger(\mathbf{v})$	$\lambda =$
$\varphi$	$\lambda \mathbf{v}$	$\bar{\lambda}$
$-\varphi$	$-\lambda \mathbf{v}$	$-\bar{\lambda}$
$\varphi^{-1}$	$\lambda^{-1} \mathbf{v}$	$\bar{\lambda}^{-1}$

Damit finden wir:

**Satz 1.30** Alle Eigenwerte

einer hermiteschen Matrix sind reell  
 einer antihermiteschen Matrix sind rein imaginär  
 einer unitären Matrix haben Betrag 1

**Satz 1.31** Für eine lineare Abbildung  $\varphi : V \rightarrow V$  des endlich-dimensionalen unitären  $\mathbb{C}$ -Vektorraums in sich sind äquivalent:

- a) Es gibt eine Orthonormalbasis von  $V$  aus Eigenvektoren für  $\varphi$ .
- b) Die Abbildung  $\varphi$  ist normal.

Beweis a) $\Rightarrow$ b): Die darstellende Matrix von  $\varphi$  bezüglich der Basis aus Eigenvektoren ist eine Diagonalmatrix  $D$ . Weil die Basis eine Orthonormalbasis ist, hat die zu  $\varphi$  adjungierte Abbildung bezüglich dieser Basis die darstellende Matrix  $D^\dagger$ . Aus  $D \cdot D^\dagger = D^\dagger \cdot D$  folgt, dass  $\varphi$  mit seiner adjungierten Abbildung kommutiert.

b) $\Rightarrow$ a): Weil wir  $V$  als komplexen Vektorraum vorausgesetzt haben, gibt es mindestens einen Eigenwert  $\lambda \in \mathbb{C}$ , und der zugehörige Eigenraum  $U \subset V$  hat eine Dimension  $> 0$ . Die wesentliche Beweislast besteht darin, zu zeigen, dass  $\varphi$  dessen orthogonales Komplement  $U^\perp$  in sich abbildet. Dann gibt es nämlich wegen Induktion nach  $\dim(V)$  eine Orthonormalbasis aus Eigenvektoren für  $\varphi|_{U^\perp}$ . Durch eine Orthonormalbasis von  $U$  ergänzen wir sie zu einer Orthonormalbasis aus Eigenvektoren für  $\varphi$  von ganz  $V$ .

Zunächst folgt aus der Kommutativität  $\varphi \circ \varphi^\dagger = \varphi^\dagger \circ \varphi$ , dass für jeden Eigenvektor  $\mathbf{u} \in U$  gilt:

$$\varphi(\varphi^\dagger(\mathbf{u})) = \varphi^\dagger(\varphi(\mathbf{u})) = \varphi^\dagger(\lambda\mathbf{u}) = \lambda\varphi^\dagger(\mathbf{u}).$$

also gehört auch  $\varphi^\dagger(\mathbf{u})$  zu  $U$ . Die Abbildung  $\varphi^\dagger$  bildet also den Unterraum  $U$  in sich ab.

Sei nun  $\mathbf{v} \in U^\perp$ . Für alle  $\mathbf{u} \in U$  gilt dann

$$\langle \varphi(\mathbf{v}), \mathbf{u} \rangle = \langle \mathbf{v}, \varphi^\dagger(\mathbf{u}) \rangle = 0,$$

weil  $\varphi^\dagger(\mathbf{u})$  wieder zu  $U$  gehört. Es ist also tatsächlich  $\varphi(\mathbf{v}) \in U^\perp$ . □

**Aufgabe 1.35** a) Zeigen Sie: die Pauli-Matrizen

$$\sigma_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

bilden eine Basis des  $\mathbb{C}$ -Vektorraums der komplexen  $2 \times 2$ -Matrizen.

b) Sind die Pauli-Matrizen hermitesch? Sind sie unitär?

**Aufgabe 1.36** Zeigen Sie: Es gibt genau zwei unitäre Matrizen der Form

$$\frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 1 & 1 \\ 1 & a & b \\ 1 & 1 & b \end{pmatrix}$$

mit komplexen Zahlen  $a, b$  gibt, und in beiden Fällen ist  $b = \bar{a}$ .

**Aufgabe 1.37** Es sei  $V$  ein endlich-dimensionaler euklidischer Vektorraum. Die Abbildung  $\varphi : V \rightarrow V$  sei linear und  $\varphi^\dagger$  sei ihre adjungierte Abbildung. Zeigen Sie:

$$\text{Rang}(\varphi) = \text{Rang}(\varphi^\dagger \circ \varphi) = \text{Rang}(\varphi \circ \varphi^\dagger).$$

**Aufgabe 1.38** Zeigen Sie: Das Produkt zweier selbstadjungierter Abbildungen ist genau dann selbstadjungiert, wenn diese Abbildungen kommutieren.

**Aufgabe 1.39** Es sei  $V$  der  $\mathbb{R}$ -Vektorraum der reellen Polynome vom Grad  $\leq 3$  mit dem inneren Produkt  $\langle f, g \rangle = \int_0^1 f(x)g(x)dx$ . Finden Sie die adjungierte Abbildung zur Differentiation  $d : f \mapsto f'$ .

**Aufgabe 1.40** Zeigen Sie, dass die Matrix

$$\begin{pmatrix} 1 & i & 0 \\ i & 1 & i \\ 0 & i & 1 \end{pmatrix}$$

normal ist, und bestimmen sie für diese Matrix eine Basis aus Eigenvektoren.

## 2 Differentialrechnung mit mehreren Veränderlichen

Stoff der Analysis I ist Differential- und Integralrechnung einer reellen Veränderlichen. Dies sind die drei großen Problemkreise

1. Konvergenz (=Folgen und Reihen), stetige Funktionen,
2. Differenzieren,
3. Integrieren.

Und genau diese Inhalte muss man auch in Zusammenhang mit mehreren Veränderlichen behandeln. Inhalt dieses Kapitels ist die Differentialrechnung. Bevor ich mit der Behandlung des Stoffes beginne, muss ich erst einiges zur Notation und zur Sprachregelung vorausschicken.

### 2.1 Grundlagen

Im ersten Semester behandelt man *Funktionen*, d.h. Abbildungen

$$f : D \rightarrow \mathbb{R}, \quad D \subset \mathbb{R}.$$

In diesem Semester betrachten wir allgemeiner *Abbildungen*

$$F : D \rightarrow \mathbb{R}^k, \quad D \subset \mathbb{R}^n$$

mit  $n \geq 1$  und  $k \geq 1$ . Eine solche Abbildung ordnet also jedem Vektor  $\mathbf{x} \in \mathbb{R}^n$  einen neuen Vektor  $\mathbf{y} = F(\mathbf{x}) \in \mathbb{R}^k$  zu. Den können wir in Komponenten schreiben

$$\mathbf{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_k \end{pmatrix} = \begin{pmatrix} F_1(\mathbf{x}) \\ \vdots \\ F_k(\mathbf{x}) \end{pmatrix}$$

und sehen: Die Abbildung  $F : D \rightarrow \mathbb{R}^k$  ist dasselbe wie das  $k$ -tupel

$$F_1 : D \rightarrow \mathbb{R}, \dots, F_k : D \rightarrow \mathbb{R}$$

ihrer *Komponenten*  $F_\kappa$ ,  $\kappa = 1, \dots, k$ .

Abbildungen  $F : D \rightarrow \mathbb{R}$  (Spezialfall  $k = 1$ ) nennt man *Funktionen*. Eine Abbildung  $F : D \rightarrow \mathbb{R}^k$  ist also dasselbe, wie das  $k$ -tupel  $(F_1, \dots, F_k)$  ihrer Komponentenfunktionen.

Je nach den Werten der Dimensionen  $n$  und  $k$  benutzt man Abbildungen  $F : D \rightarrow \mathbb{R}^k$ ,  $D \subset \mathbb{R}^n$ , zur Beschreibung verschiedener Situationen, und hat manchmal auch spezielle Namen dafür:

- $k = 1$ : Das sind die gewöhnlichen Funktionen  $f(x_1, \dots, x_n)$ . Im Fall  $n = 3$  etwa benutzt man sie zur Beschreibung skalarer Felder, wie es z.B. eine Temperaturverteilung ist.
- $n = 1$ : Hier wird jeder reellen Zahl  $t \in D \subset \mathbb{R}$  (jedem *Parameter*) ein Vektor  $\mathbf{y}(t) \in \mathbb{R}^k$  zugeordnet. Für  $k = 3$  tritt so etwas auf bei der Beschreibung der Bahn eines punktförmigen Teilchens im Raum. Der Parameter  $t$  ist meist die Zeit. Und eine solche Abbildung  $D \rightarrow \mathbb{R}^k$  mit  $D \subset \mathbb{R}$  nennt man eine *Kurve* im  $\mathbb{R}^k$ .

- $n = k$ : Solche Abbildungen kommen z.B. vor als Koordinatentransformationen. Die wichtigsten davon sind die Transformation in Polarkoordinaten

$$F : \begin{pmatrix} r \\ \varphi \end{pmatrix} \mapsto \begin{pmatrix} r \cdot \cos(\varphi) \\ r \cdot \sin(\varphi) \end{pmatrix},$$

und die Transformation in Kugelkoordinaten

$$F : \begin{pmatrix} r \\ \varphi \\ \theta \end{pmatrix} \mapsto \begin{pmatrix} r \cdot \cos(\varphi)\cos(\theta) \\ r \cdot \sin(\varphi)\cos(\theta) \\ r \cdot \sin(\theta) \end{pmatrix}.$$

Aber Abbildungen mit  $n = k$ , zumindest in Dimension 3, benutzt man auch zur Beschreibung von physikalischen Situationen: Jedem Vektor  $\mathbf{x}$  wird eine Bildvektor  $\mathbf{y}(\mathbf{x})$  zugeordnet. Stellt man sich diesen Bildvektor  $\mathbf{y}(\mathbf{x})$  im Punkt  $\mathbf{x}$  angeheftet vor, so nennt man das ein *Vektorfeld*. Damit beschreibt man etwa die Geschwindigkeitsverteilung von Flüssigkeitsströmungen oder elektrische oder magnetische Felder.

- $n = 2, k = 3$ : Solche Abbildungen benutzt man zur Beschreibung von *Flächen* (oder spezieller: Oberflächen) im Raum. So wird etwa die Oberfläche der Einheits-Kugel (die Einheits-Sphäre) durch die Abbildung

$$\begin{pmatrix} \varphi \\ \theta \end{pmatrix} \mapsto \begin{pmatrix} \cos(\varphi)\cos(\theta) \\ \sin(\varphi)\cos(\theta) \\ \sin(\theta) \end{pmatrix}$$

beschrieben. Wenn man jeden Punkt der Kugeloberfläche nur einmal beschreiben möchte, muss man sich auf den Bereich

$$D : \{(\varphi, \theta) \in \mathbb{R}^2 : 0 \leq \varphi < 2\pi, -\pi/2 \leq \theta \leq \pi/2\}$$

beschränken.

- $n = k - 1$ : So etwas nennt man - in Analogie zum Fall  $k = 3$  - eine *Hyperfläche* im  $\mathbb{R}^k$ .

Bevor wir anfangen zu differenzieren, noch einige Definitionen:

**Definition 2.1** Die euklidische Länge (oder Norm) eines Vektors  $\mathbf{x} \in \mathbb{R}^n$  bezeichnen wir mit  $\|\mathbf{x}\|$ , also

$$\|\mathbf{x}\| = \sqrt{x_1^2 + \dots + x_n^2} = \sqrt{\sum_{\nu=1}^n x_\nu^2}.$$

Die Länge  $\|\mathbf{x} - \mathbf{y}\|$  eines Differenzvektors  $\mathbf{x} - \mathbf{y}$  heißt der Abstand dieser Vektoren.

Ich möchte hier auch kurz erklären, was diese Norm für  $n \times n$ -Matrizen bedeutet. Eine  $n \times n$ -Matrix

$$A = (a_{i,j})_{i,j=1,\dots,n}$$

kann man ja auch auffassen als einen Vektor

$$(a_{1,1}, a_{1,2}, \dots, a_{1,n}, a_{2,1}, \dots, a_{n,n}) \in \mathbb{R}^{n^2}.$$

Die Norm dieses Vektors nennt man *die Norm der Matrix A*:

$$\| A \| := \sqrt{\sum_{i,j=1}^n a_{i,j}^2}$$

**Beispiel 2.1** Betrachten wir etwa eine  $2 \times 2$ -Matrix

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

Ihre Matrix-Norm ist

$$\| A \| = \sqrt{a^2 + b^2 + c^2 + d^2}.$$

Nun ist

$$A \cdot A^t = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot \begin{pmatrix} a & c \\ b & d \end{pmatrix} = \begin{pmatrix} a^2 + b^2 & * \\ * & c^2 + d^2 \end{pmatrix}$$

mit der Spur

$$\text{tr}(A \cdot A^t) = a^2 + b^2 + c^2 + d^2 = \| A \|^2.$$

Genauso, nur mit Verwendung von vielen Indizes, zeigt man für beliebige  $n \times n$ -Matrizen  $A$ , dass

$$\| A \| = \sqrt{\text{tr}(A \cdot A^t)}.$$

Mit der Matrix-Norm kann man ganz schön rechnen. Es gilt

- a)  $\| A + B \| \leq \| A \| + \| B \|$  für je zwei  $n \times n$ -Matrizen  $A, B$ .
- b)  $\| A \cdot \mathbf{v} \| \leq \| A \| \cdot \| \mathbf{v} \|$  für jeden Vektor  $\mathbf{v} \in \mathbb{R}^n$ .
- c)  $\| A \cdot B \| \leq \| A \| \cdot \| B \|$  für je zwei  $n \times n$ -Matrizen  $A, B$ .

Beweise: Eigenschaft a) ist nichts anderes als die Dreiecksungleichung aus der Linearen Algebra im Vektorraum  $\mathbb{R}^{n^2}$  der  $n \times n$ -Matrizen.

b) Der Vektor  $\mathbf{w} = A \cdot \mathbf{v}$  hat die Komponenten

$$w_i = \sum_{j=1}^n a_{i,j} v_j = (A_i, \mathbf{v}),$$

wo  $A_i$  den  $i$ -ten Zeilenvektor der Matrix  $A$  bezeichnet. Aus der Cauchy-Schwarz-Ungleichung der Linearen Algebra folgt

$$|w_i| \leq \| A_i \| \cdot \| \mathbf{v} \| = \sqrt{\sum_{j=1}^n A_{i,j}^2} \cdot \| \mathbf{v} \|,$$

und damit

$$\| \mathbf{w} \|^2 = \sum_{i=1}^n w_i^2 \leq \sum_{i,j=1}^n A_{i,j}^2 \cdot \| \mathbf{v} \|^2 = \| A \|^2 \cdot \| \mathbf{v} \|^2.$$

c) Bezeichnet  $B^j$  den  $j$ -ten Spaltenvektor der Matrix  $B$ , so hat die Produktmatrix  $A \cdot B$  die Spaltenvektoren  $A \cdot B^j$ . Nach b) ist die Norm eines jeden dieser Vektoren

$$\|A \cdot B^j\| \leq \|A\| \cdot \|B^j\|,$$

und damit

$$\|A \cdot B\| = \sqrt{\sum_{j=1}^n \|A \cdot B^j\|^2} \leq \|A\| \cdot \sqrt{\sum_{j=1}^n \|B^j\|^2} = \|A\| \cdot \|B\|. \quad \square$$

Genauso, wie es konvergente Folgen von Zahlen gibt, so gibt es auch konvergente Folgen von Vektoren:

**Definition 2.2** Die Folge  $(\mathbf{x}_\mu)_{\mu \in \mathbb{N}}$  von Vektoren  $\mathbf{x}_\mu \in \mathbb{R}^n$  heißt konvergent gegen den Vektor  $\mathbf{x} \in \mathbb{R}^n$ , wenn zu jedem  $0 < \epsilon \in \mathbb{R}$  ein  $M(\epsilon)$  existiert mit

$$\|\mathbf{x}_\mu - \mathbf{x}\| < \epsilon \quad \text{für} \quad \mu > M(\epsilon).$$

Für jeden Vektor  $\mathbf{v} = (v_i)$  ist

$$|v_i| \leq \|\mathbf{v}\| \leq \sum_{i=1}^n |v_i|.$$

Für  $\mathbf{v} = \mathbf{x}_\mu - \mathbf{x}$  bedeutet das

$$|x_{\mu,i} - x_i| \leq \|\mathbf{x}_\mu - \mathbf{x}\| \leq \sum_{i=1}^n |x_{\mu,i} - x_i|,$$

und damit sieht man: Die Folge  $(x_\mu)$  konvergiert genau dann gegen den Vektor  $\mathbf{x}$ , wenn für  $i = 1, \dots, n$  die  $i$ -te Komponentenfolge  $x_{\mu,i}$  gegen  $x_i$  konvergiert.

Genau wie für reelle Zahlen hat man auch für Vektoren den Begriff der Cauchy-Folge:

**Definition 2.3** Die Folge  $(\mathbf{x}_\mu)_{\mu \in \mathbb{N}}$  von Vektoren  $\mathbf{x}_\mu \in \mathbb{R}^n$  heißt Cauchy-Folge, wenn zu jedem  $0 < \epsilon \in \mathbb{R}$  ein  $M(\epsilon)$  existiert mit

$$\|\mathbf{x}_\mu - \mathbf{x}_\nu\| < \epsilon \quad \text{für} \quad \mu, \nu > M(\epsilon).$$

Wegen

$$|\mathbf{x}_{\mu,i} - \mathbf{x}_{\nu,i}| \leq \|\mathbf{x}_\mu - \mathbf{x}_\nu\|$$

ist dann für  $i = 1, \dots, n$  die Folge  $\mathbf{x}_{\mu,i}$  der  $i$ -ten Komponenten eine Cauchy-Folge reeller Zahlen. Die konvergiert. Daraus folgt, dass jede Cauchy-Folge von Vektoren im  $\mathbb{R}^n$  konvergiert.

Wir erinnern uns noch an folgende Definitionen, die fast wörtlich so auch im  $\mathbb{R}^1$  benutzt werden: Eine Teilmenge  $M \subset \mathbb{R}^n$

heißt	falls
Kugel	$M = \{\mathbf{x} \in \mathbb{R}^n : \ \mathbf{x} - \mathbf{x}_0\  < r\}$
offen	mit jedem Punkt $\mathbf{x}_0 \in M$ auch eine ganze Kugel um $\mathbf{x}_0$ zu $M$ gehört
abgeschlossen	$\mathbb{R}^n \setminus M = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{x} \notin M\}$ offen ist
Abschluss $\bar{N}$ von $N$	$M$ Durchschnitt aller abgeschlossenen Mengen ist, die $N$ enthalten
Rand $\partial N$ von $N$	$M = \bar{N} \cap (\mathbb{R}^n \setminus \bar{N})$
beschränkt	$M$ in einer Kugel enthalten ist
kompakt	$M$ beschränkt und abgeschlossen ist.

Äquivalent dazu, dass  $M$  offen ist, ist die Eigenschaft: Ist  $\mathbf{x} \in M$ , und ist  $\mathbf{x}_\mu$  eine gegen  $\mathbf{x}$  konvergente Folge von Vektoren, so gibt es immer ein  $m \in \mathbb{N}$  mit  $\mathbf{x}_\mu \in M$  für  $\mu > m$ .

Und äquivalent dazu, dass  $M$  abgeschlossen ist, ist die Eigenschaft: Für jede Folge  $\mathbf{x}_\mu$  von Vektoren  $\mathbf{x}_\mu \in M$ , die gegen einen Vektor  $\mathbf{x} \in \mathbb{R}^n$  konvergiert, gilt  $\mathbf{x} \in M$ .

Wir brauchen noch zwei Eigenschaften, die kompakte Mengen charakterisieren. Aber weil wir sie anwenden und keine Theorie machen wollen, interessieren uns nur die folgenden Implikationen:

Wenn  $M \subset \mathbb{R}^n$  kompakt ist, dann gelten:

- a) Jede Folge von Vektoren  $\mathbf{x}_\mu \in M$  besitzt eine in  $M$  konvergente Teilfolge.
- b) Ist  $M$  enthalten in einer Vereinigung  $\bigcup U_\mu$  von offenen Mengen  $U_\mu \subset \mathbb{R}^n$ , so ist  $M$  auch schon in einer Vereinigung  $\bigcup_1^m U_\mu$  von endlich vielen dieser offenen Mengen enthalten.

Beweis. a) Für  $\nu = 1, \dots, n$  sind die Komponentenfolgen  $(x_{\mu,\nu})_{\mu \in \mathbb{N}}$  beschränkt, weil  $M$  beschränkt ist. Nach dem Satz von Bolzano-Weierstraß in einer Variablen, besitzen sie alle konvergente Teilfolgen. Geht man zu diesen Teilfolgen über, erst in der Folge  $(x_{\mu,1})$ , dann in der Folge  $(x_{\mu,2})$ , usw., so sieht man, dass die Folge  $\mathbf{x}_\mu$  eine konvergente Teilfolge besitzt. Weil  $M$  abgeschlossen ist, muss der Grenzwert dieser Folge zu  $M$  gehören.

b) Wenn es kein  $m$  gibt, mit  $M \subset \bigcup_1^m U_\mu$ , dann gibt es also zu jedem  $m \in \mathbb{N}$  einen Punkt  $\mathbf{x}_m \in M$  mit  $\mathbf{x}_m \notin \bigcup_1^m U_\mu$ . Nach a) hat die Folge  $\mathbf{x}_m$  eine konvergente Teilfolge, wir können gleich annehmen, dass sie selbst konvergiert, etwa gegen  $\mathbf{x} \in M$ . Wegen  $\mathbf{x} \in M \subset \bigcup U_\mu$  gibt es ein  $U_{\mu_0}$  mit  $\mathbf{x} \in U_{\mu_0}$ . Weil  $U_{\mu_0}$  offen ist, gibt es ein  $m$  mit  $\mathbf{x}_\mu \in U_{\mu_0}$  für  $\mu > m$ . Dann ist die ganze Folge enthalten in  $(\bigcup_1^m U_\mu) \cup U_{\mu_0}$ , Widerspruch!

**Beispiel 2.2 (Exponentialfunktion für Matrizen)** Für jede  $n \times n$ -Matrix  $A$  ist die Matrizen-Folge  $(A^\mu)_{\mu \in \mathbb{N}}$  im Vektorraum  $\mathbb{R}^{n^2}$  wohldefiniert, und damit auch die Matrizen-Reihe

$$\sum_{\mu=0}^{\infty} \frac{1}{\mu!} A^\mu.$$

Die Matrix-Norm der Partialsummen

$$\sum_{\mu=m_1}^{m_2} \frac{1}{\mu!} A^\mu$$

dieser Reihe kann man abschätzen durch

$$\left\| \sum_{\mu=m_1}^{m_2} \frac{1}{\mu!} A^\mu \right\| \leq \sum_{\mu=m_1}^{m_2} \frac{1}{\mu!} \|A\|^\mu,$$

d.h., durch die Partialsummen der konvergenten Exponentialreihe

$$e^{\|A\|} = \sum_{\mu=0}^{\infty} \frac{1}{\mu!} \|A\|^\mu.$$

Deswegen bilden die Partialsummen der Matrizen-Exponentialreihe eine Cauchy-Folge, und die Matrizen-Exponentialreihe konvergiert gegen eine Matrix

$$e^A := \sum_{\nu=0}^{\infty} \frac{1}{\nu!} A^\nu.$$

Eine Abbildung  $F : M \rightarrow \mathbb{R}^p$ ,  $M \subset \mathbb{R}^n$  ist in einem Punkt  $\mathbf{x}_0 \in M$  stetig, wenn sie das  $(\epsilon, \delta)$ -Kriterium aus dem ersten Semester wörtlich erfüllt:

Zu jedem  $0 < \epsilon \in \mathbb{R}$   
gibt es ein  $\delta = \delta(\epsilon)$  derart, dass  
für alle  $\mathbf{x} \in \mathbb{R}^n$  mit  $\|\mathbf{x} - \mathbf{x}_0\| < \delta$   
gilt:  $\|F(\mathbf{x}) - F(\mathbf{x}_0)\| < \epsilon$ .

Praktisch bedeutet dies für uns, dass eine Abbildung  $F(\mathbf{x}) = (F_1(\mathbf{x}), \dots, F_k(\mathbf{x}))$  stetig ist, wenn ihre  $k$  Komponentenfunktionen  $F_k$  stetig sind. Und damit kann man dann auch beweisen, dass Stetigkeit äquivalent zur *Folgenstetigkeit* ist:  $F$  ist stetig in  $\mathbf{x}_0 \in \mathbb{R}^n$  genau dann, wenn für jede Folge  $(\mathbf{x}_\nu)$  von Vektoren  $\mathbf{x}_\nu \in \mathbb{R}^n$  gilt:

$$\lim_{\nu \rightarrow \infty} \mathbf{x}_\nu = \mathbf{x}_0 \implies \lim_{\nu \rightarrow \infty} F(\mathbf{x}_\nu) = F(\mathbf{x}_0).$$

**Beispiel 2.3** Die Norm-Funktion  $\mathbf{x} \mapsto \|\mathbf{x}\|$  ist trivialerweise stetig. (Man setze  $\delta := \epsilon$  in der Definition der Stetigkeit. Daraus erhält man

$$\|\lim \mathbf{x}_\nu\| = \lim \|\mathbf{x}_\nu\|$$

für jede konvergente Folge  $\mathbf{x}_\nu$  von Vektoren, und

$$\left\| \sum_{\nu=0}^{\infty} \mathbf{x}_\nu \right\| = \lim_{N \rightarrow \infty} \left\| \sum_{\nu=1}^N \mathbf{x}_\nu \right\| \leq \lim_{N \rightarrow \infty} \sum_{\nu=0}^N \|\mathbf{x}_\nu\| \leq \sum_{\nu=0}^{\infty} \|\mathbf{x}_\nu\|$$

für jede Reihe von Vektoren, bei der die Reihe  $\sum \|\mathbf{x}_\nu\|$  konvergiert.

**Beispiel 2.4** Für jede  $n \times n$ -Matrix  $A$  ist die lineare Abbildung

$$\mathbb{R}^n \ni \mathbf{v} \mapsto A \cdot \mathbf{v}$$

stetig.

*Beweis:* Natürlich können wir  $A \neq 0$  annehmen. Dann ist

$$\|A \cdot \mathbf{v} - A \cdot \mathbf{v}_0\| = \|A \cdot (\mathbf{v} - \mathbf{v}_0)\| \leq \|A\| \cdot \|\mathbf{v} - \mathbf{v}_0\| < \epsilon,$$

falls

$$\|\mathbf{v} - \mathbf{v}_0\| < \frac{\epsilon}{\|A\|}.$$

Eine einfache Folgerung aus der Definition der Stetigkeit ist

**Satz 2.1 (Urbild)** Es sei  $U \subset \mathbb{R}^n$  offen und  $F : U \rightarrow \mathbb{R}^k$  stetig. Für jede offene Menge  $V \subset \mathbb{R}^k$  ist dann die Urbildmenge

$$F^{-1}(V) = \{\mathbf{x} \in U : F(\mathbf{x}) \in V\} \subset \mathbb{R}^n$$

offen.

Beweis. Es sei  $\mathbf{x}_0 \in F^{-1}(V)$ , also  $\mathbf{y}_0 := F(\mathbf{x}_0) \in V$ . Weil  $V$  offen ist, gibt es ein  $\epsilon > 0$  mit

$$\mathbf{y} \in \mathbb{R}^k \text{ und } \|\mathbf{y} - \mathbf{y}_0\| < \epsilon \Rightarrow \mathbf{y} \in V.$$

Sei  $\delta_1 = \delta(\epsilon) > 0$  wie in der Definition der Stetigkeit gewählt. Weil  $U$  offen ist, gibt es außerdem ein  $\delta_2 > 0$  mit

$$\mathbf{x} \in \mathbb{R}^n \text{ und } \|\mathbf{x} - \mathbf{x}_0\| < \delta_2 \Rightarrow \mathbf{x} \in U.$$

Sei  $\delta := \min\{\delta_1, \delta_2\}$ . Für  $\mathbf{x} \in \mathbb{R}^n$  mit  $\|\mathbf{x} - \mathbf{x}_0\| < \delta$  ist dann also  $\mathbf{x} \in U$  und

$$\|F(\mathbf{x}) - F(\mathbf{x}_0)\| < \epsilon.$$

Daraus folgt  $F(\mathbf{x}) \in V$  und  $\mathbf{x} \in F^{-1}(V)$ . Mit  $\mathbf{x}_0$  enthält die Menge  $F^{-1}(V)$  also auch die ganze Kugel um  $\mathbf{x}_0$  vom Radius  $\delta$ . Also ist die Menge  $F^{-1}(V)$  offen im  $\mathbb{R}^n$ .  $\square$

Wir brauchen noch den *Satz vom Maximum*, der in einer Variablen so lautet:

Eine auf dem Intervall  $[a, b]$  stetige Funktion nimmt dort ihr Maximum an.

Und in mehreren Variablen:

Eine auf der kompakten Menge  $M \subset \mathbb{R}^n$  stetige Funktion nimmt in einem Punkt von  $M$  ihr Maximum an.

**Aufgabe 2.1** Es seien  $A$  und  $B$  zwei kommutierende  $n \times n$ -Matrizen (d.h.,  $A \cdot B = B \cdot A$ ). Zeigen Sie

$$e^{A+B} = e^A \cdot e^B$$

und folgern Sie daraus, dass für jede Matrix  $A$  die Matrix  $e^A$  invertierbar ist.

**Aufgabe 2.2** a) Es sei  $a \in \mathbb{R}$  und

$$A = \begin{pmatrix} a & 1 \\ 0 & a \end{pmatrix}.$$

Zeigen Sie für alle  $n \in \mathbb{N}$ :

$$A^n = \begin{pmatrix} a^n & na^{n-1} \\ 0 & a^n \end{pmatrix}.$$

b) Zeigen Sie für  $t \in \mathbb{R}$ :

$$e^{tA} = \begin{pmatrix} e^{at} & te^{at} \\ 0 & e^{at} \end{pmatrix}.$$

**Aufgabe 2.3** Es sei

$$A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

a) Zeigen Sie für  $k \in \mathbb{N}$ :

$$\begin{aligned} n = 4k &\Rightarrow A^n = \mathbb{1} \\ n = 4k + 1 &\Rightarrow A^n = A \\ n = 4k + 2 &\Rightarrow A^n = -\mathbb{1} \\ n = 4k + 3 &\Rightarrow A^n = -A \end{aligned}$$

b) Zeigen Sie für  $t \in \mathbb{R}$ :

$$e^{tA} = \begin{pmatrix} \cos(t) & -\sin(t) \\ \sin(t) & \cos(t) \end{pmatrix}.$$

## 2.2 Differenzierbare Abbildungen

Der Differentialquotient

$$\frac{F(\mathbf{x}) - F(\mathbf{x}_0)}{\mathbf{x} - \mathbf{x}_0}$$

hat für  $\mathbf{x} \in \mathbb{R}^n, n > 1$ , keinen Sinn, weil man durch den Vektor  $\mathbf{x} - \mathbf{x}_0$  nicht dividieren kann. Verallgemeinern lässt sich dagegen die Differenzierbarkeit als lineare Approximierbarkeit. In einer Variablen sah das so aus:  $f$  ist differenzierbar in  $x$ , wenn

$$f(x+h) = f(x) + a \cdot h + \varphi(h) \quad \text{mit} \quad \lim_{h \rightarrow 0} \frac{\varphi(h)}{h} = 0.$$

Dabei ist die Konstante  $a$  die Ableitung  $f'(x)$ . Man verallgemeinert diese Approximations-Eigenschaft folgendermaßen:

**Definition 2.4** Es sei  $M \in \mathbb{R}^n$  offen.  $F : M \rightarrow \mathbb{R}^k$  heißt in  $\mathbf{x} \in M$  total differenzierbar, wenn eine lineare Abbildung

$$A : \begin{cases} \mathbb{R}^n \rightarrow \mathbb{R}^k \\ \mathbf{h} \mapsto A \cdot \mathbf{h} \end{cases}$$

existiert, so, dass für alle  $\mathbf{h} \in \mathbb{R}^n$  (für die  $F(\mathbf{x} + \mathbf{h})$  definiert ist), gilt

$$F(\mathbf{x} + \mathbf{h}) = F(\mathbf{x}) + A \cdot \mathbf{h} + \Phi(\mathbf{h}).$$

Dabei soll

$$\Phi(\mathbf{h}) = F(\mathbf{x} + \mathbf{h}) - F(\mathbf{x}) - A \cdot \mathbf{h}$$

eine Abbildung sein, welche die Eigenschaft

$$\lim_{\mathbf{h} \rightarrow \mathbf{0}} \frac{\Phi(\mathbf{h})}{\|\mathbf{h}\|} = \mathbf{0}$$

besitzt. Die lineare Abbildung  $A$  heißt die (totale) Ableitung von  $F$  in  $\mathbf{x}$ , in Zeichen  $A = F'(\mathbf{x})$ .

Die Bedingung an  $\Phi$  in dieser Definition bedeutet, dass  $\Phi(\mathbf{h})$  schneller gegen Null geht, als  $\mathbf{h}$ , also etwa „von zweiter Ordnung“. Die Abbildung  $F : \mathbf{x}' \mapsto F(\mathbf{x}')$  wird somit durch die (affin-) lineare Abbildung  $F(\mathbf{x}) + A \cdot (\mathbf{x}' - \mathbf{x})$  von höherer als erster Ordnung approximiert. Diese lineare Abbildung ist dabei (im Falle ihrer Existenz) durch die Abbildung  $F$  eindeutig bestimmt: Sind  $A$  und  $B$  zwei lineare Abbildungen wie in der Definition, so gilt

$$\mathbf{0} = \lim_{\mathbf{h} \rightarrow \mathbf{0}} \frac{(A - B) \cdot \mathbf{h}}{\|\mathbf{h}\|} = \lim_{\mathbf{h} \rightarrow \mathbf{0}} (A - B) \cdot \frac{\mathbf{h}}{\|\mathbf{h}\|}.$$

Die Abbildung  $\mathbf{h} \mapsto (A - B) \cdot (\mathbf{h} / \|\mathbf{h}\|)$  ist konstant auf jedem Strahl  $t \cdot \mathbf{h}_0, 0 < t \in \mathbb{R}$ , in der Richtung  $\mathbf{h}_0, \|\mathbf{h}_0\| = 1$ . Da sie für  $t \rightarrow 0$  gegen Null geht, ist sie konstant gleich Null.  $\square$

**Beispiel 2.5** Die konstante Abbildung  $F(\mathbf{x}) = \mathbf{y}_0$  ist differenzierbar mit Ableitung  $F' = \mathbf{0}$ , denn

$$F(\mathbf{x} + \mathbf{h}) = \mathbf{y}_0 = F(\mathbf{x}) + \mathbf{0} \cdot \mathbf{h}.$$

**Beispiel 2.6** Jede (affin-) lineare Abbildung  $F(\mathbf{x}) = \mathbf{y}_0 + A \cdot \mathbf{x}$  ist differenzierbar mit Ableitung  $A$ . Der Beweis ist von ähnlicher Schwierigkeit wie bei Beispiel 2.3.

Natürlich heißt die Abbildung  $F$  auf der ganzen offenen Menge  $M$  differenzierbar, wenn sie in jedem Punkt  $\mathbf{x} \in M$  differenzierbar ist.

Noch ein Wort zur Notation: Eine lineare Abbildung  $A : V \rightarrow W$  von einem Vektorraum  $V$  in einen Vektorraum  $W$  schreibt man im allgemeinen  $A : \mathbf{h} \mapsto A(\mathbf{h})$ . Bei uns kommen die Vektorräume aber immer als  $V = \mathbb{R}^n, W = \mathbb{R}^k$  o.ä. vor. Diese Vektorräume haben die kanonische Standardbasis

$$\mathbf{e}_1 = (1, 0, \dots, 0), \dots, \mathbf{e}_n = (0, \dots, 0, 1).$$

Und bezüglich dieser Standardbasen hat jede lineare Abbildung eine kanonische beschreibende Matrix. Deswegen denken wir uns die lineare Abbildung  $A$  immer als Matrixmultiplikation  $\mathbf{h} \mapsto A \cdot \mathbf{h}$  und machen keinen großen Unterschied zwischen der linearen Abbildung und ihrer beschreibenden Matrix.

Ein ganz anderes Problem ist, dass die lineare Abbildung  $A = F'(\mathbf{x})$  vom Aufpunkt  $\mathbf{x}$  abhängt. Andererseits kann man sie wieder auf Vektoren  $\mathbf{h}$  anwenden. Schreibt man das Resultat  $A(\mathbf{x}) \cdot \mathbf{h} = F'(\mathbf{x})(\mathbf{h})$ , so werden die Formeln sehr schnell unübersichtlich. Deswegen lässt man oft eines der Argumente,  $\mathbf{x}$  oder  $\mathbf{h}$ , weg.

**Satz 2.2 (Triviales Differenzierbarkeitskriterium)** Eine Abbildung  $F = (F_1, \dots, F_k)$  mit den Komponentenfunktionen  $F_1, \dots, F_k$  ist genau dann differenzierbar, wenn alle Funktionen  $F_\kappa$  differenzierbar sind. Die lineare Abbildung

$$F'(\mathbf{x}) : \mathbb{R}^n \rightarrow \mathbb{R}^k$$

ist dann gerade die Abbildung

$$\mathbb{R}^n \ni \mathbf{h} \mapsto \begin{pmatrix} F'_1(\mathbf{x}) \cdot \mathbf{h} \\ \vdots \\ F'_k(\mathbf{x}) \cdot \mathbf{h} \end{pmatrix}.$$

Beweis. Wir bezeichnen die Komponentenfunktionen der linearen Abbildung  $A$  mit  $A_1, \dots, A_k$ , also

$$A \cdot \mathbf{h} = \begin{pmatrix} A_1 \cdot \mathbf{h} \\ \vdots \\ A_k \cdot \mathbf{h} \end{pmatrix}.$$

Dann ist

$$\lim_{\mathbf{0} \neq \mathbf{h} \rightarrow \mathbf{0}} \frac{1}{\|\mathbf{h}\|} (F(\mathbf{x} + \mathbf{h}) - F(\mathbf{x}) - A \cdot \mathbf{h}) = \mathbf{0}$$

$$\iff \lim_{\mathbf{0} \neq \mathbf{h} \rightarrow \mathbf{0}} \frac{1}{\|\mathbf{h}\|} (F_\kappa(\mathbf{x} + \mathbf{h}) - F_\kappa(\mathbf{x}) - A_\kappa \cdot \mathbf{h}) = 0 \text{ für alle } \kappa = 1, \dots, k$$

$$\iff \text{alle Funktionen } F_\kappa, \kappa = 1, \dots, k, \text{ sind differenzierbar in } \mathbf{x} \text{ mit Ableitung } F'_\kappa(\mathbf{x}) = A_\kappa.$$

□

Mit Satz 2.2 ist die Differenzierbarkeit einer Abbildung  $F$  auf die Differenzierbarkeit ihrer Komponentenfunktionen  $F_\kappa$  zurückgeführt. Wir beschäftigen uns deswegen jetzt mit Funktionen.

**Satz 2.3 (Notwendiges Differenzierbarkeitskriterium)** Die Funktion  $f : D \rightarrow \mathbb{R}$  sei differenzierbar in  $\mathbf{x} \in D$ ,  $D \subset \mathbb{R}^n$  offen. Dann sind alle Funktionen

$$\varphi_\nu(t) := f(\mathbf{x} + t \cdot \mathbf{e}_\nu), \quad \mathbf{e}_\nu = \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} \quad (\nu\text{-ter Basisvektor}),$$

differenzierbar in  $t = 0$  und die Ableitung ist

$$f'(\mathbf{x}) \cdot \mathbf{h} = \sum_{\nu=1}^n \left. \frac{d\varphi_\nu}{dt} \right|_{t=0} \cdot h_\nu.$$

Der Beweis von Satz 2.3 ergibt sich sofort, wenn wir in die Definition der Differenzierbarkeit von  $f$  für die Auslenkung  $\mathbf{h} \in \mathbb{R}^n$  speziell nur Vektoren  $t \cdot \mathbf{e}_\nu$  einsetzen.

Die Funktion  $\varphi_\nu(t)$  ist genau die Änderung der Funktion  $f$  in der  $\nu$ -ten Koordinatenrichtung, während die übrigen Koordinaten festgehalten werden:

$$\varphi_\nu(t) = f(x_1, \dots, x_{\nu-1}, x_\nu + t, x_{\nu+1}, \dots, x_n).$$

Ihre Ableitung heißt *partielle Ableitung* von  $f$  in Richtung der Koordinate  $x_\nu$ :

$$\frac{\partial f}{\partial x_\nu}(\mathbf{x}) := \left. \frac{d}{dt} f(x_1, \dots, x_\nu + t, \dots, x_n) \right|_{t=0}.$$

Eine Funktion heißt *partiell differenzierbar*, wenn sie nach allen  $n$  Variablen partiell differenzierbar ist.

**Definition 2.5 (Gradient)** Der Vektor der partiellen Ableitungen einer Funktion  $f$  heißt der Gradient dieser Funktion. In Zeichen:

$$\text{grad}(f) = \nabla f = \left( \frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n} \right).$$

Damit wird der Wert der Ableitung  $f'(\mathbf{x})$  auf einem Vektor  $\mathbf{h}$

$$f'(\mathbf{x})(\mathbf{h}) = \text{grad}(f)(\mathbf{x}) \cdot \mathbf{h} = \nabla f(\mathbf{x}) \cdot \mathbf{h}.$$

Während  $\mathbf{h}$  ein Spaltenvektor ist, ist der Gradient ein Zeilenvektor.

Für eine total differenzierbare Funktion  $f$  ist also

$$f(\mathbf{x} + \mathbf{h}) = f(\mathbf{x}) + \sum_{\nu=1}^n \frac{\partial f}{\partial x_\nu} \cdot h_\nu + \varphi(\mathbf{h})$$

mit einer Störfunktion  $\varphi$ , welche ziemlich klein wird, wenn  $\mathbf{h} \rightarrow \mathbf{0}$  geht. Physiker drücken diesen Sachverhalt meist intuitiver wie folgt aus:

$$df = \frac{\partial f}{\partial x_1} dx_1 + \dots + \frac{\partial f}{\partial x_n} dx_n.$$

Hier sind  $dx_1 = h_1, \dots, dx_n = h_n$  die Komponenten des Auslenkungsvektors  $\mathbf{h}$ , mathematisch völlig exakt. Aber  $df$ , die 'infinitesimale Änderung'  $df$  von  $f$  hat es in sich. Man stellt sich näherungsweise die Änderung  $f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x})$  vor. Exakt ist das nur bis auf die kleine Störfunktion  $\varphi(\mathbf{h})$ . Das bedeutet die Formulierung 'näherungsweise' oder 'infinitesimal'.

Die Sätze 2.2 und 2.3 zusammen zeigen, dass die Ableitungs-Matrix

$$F'(\mathbf{x}) = \begin{pmatrix} \frac{\partial F_1}{\partial x_1}(\mathbf{x}) & \dots & \frac{\partial F_1}{\partial x_n}(\mathbf{x}) \\ \vdots & & \vdots \\ \frac{\partial F_k}{\partial x_1}(\mathbf{x}) & \dots & \frac{\partial F_k}{\partial x_n}(\mathbf{x}) \end{pmatrix}$$

als ihre Einträge die partiellen Ableitungen der Komponentenfunktionen  $F_\nu$  enthält.

**Definition 2.6 (Funktionalmatrix)** Die Ableitungsmatrix  $F'(\mathbf{x})$  der differenzierbaren Abbildung  $F$  heißt auch Funktionalmatrix.

Leider ist Satz 2.3 *nicht umkehrbar*. Eine partiell differenzierbare Funktion von  $n > 1$  Variablen braucht nicht total differenzierbar zu sein. Dafür kann man Beispiele konstruieren, die aber sehr konstruiert sind, und deswegen in der Natur kaum vorkommen. Dagegen gilt

**Satz 2.4 (Hinreichendes Differenzierbarkeitskriterium)** Es sei  $D \subset \mathbb{R}^n$  offen, die Funktion  $f$  sei partiell differenzierbar auf  $D$  und die partiellen Ableitungen  $\partial f / \partial x_\nu$  seien stetig auf  $D$ . Dann ist  $f$  total differenzierbar.

Beweis. Sei  $\mathbf{x} \in D$ . Da  $D$  offen ist, gibt es ein  $r > 0$  derart, dass die Kugel um  $\mathbf{x}$  von diesem Radius  $r$  noch ganz zu  $D$  gehört. Für alle Vektoren  $\mathbf{h} = (h_1, \dots, h_n) \in \mathbb{R}^n$  mit  $\|\mathbf{h}\| < r$  gehören also die Vektoren

$$\mathbf{x} + (h_1, \dots, h_\nu, t, 0, \dots, 0), \quad 0 \leq t \leq h_{\nu+1},$$

auch zu  $D$ . Wir kürzen ab

$$\mathbf{h}_0 := \mathbf{0}, \dots, \mathbf{h}_\nu := (h_1, \dots, h_\nu, 0, \dots, 0), \dots, \mathbf{h}_n = \mathbf{h} = (h_1, \dots, h_n)$$

und folgern aus dem Mittelwertsatz der Differentialrechnung, angewendet auf die Funktion

$$g_\nu(t) := f(x_1 + h_1, \dots, x_{\nu-1} + h_{\nu-1}, x_\nu + t, x_{\nu+1}, \dots, x_n),$$

dass es ein  $\theta_\nu$ ,  $0 \leq \theta_\nu \leq 1$ , gibt mit

$$f(\mathbf{x} + \mathbf{h}_\nu) - f(\mathbf{x} + \mathbf{h}_{\nu-1}) = g_\nu(h_\nu) - g_\nu(0) = g'_\nu(\theta_\nu h_\nu) \cdot h_\nu = \frac{\partial f}{\partial x_\nu}(\mathbf{x}_\nu) \cdot h_\nu,$$

wobei  $\mathbf{x}_\nu = \mathbf{x} + (h_1, \dots, h_{\nu-1}, \theta_\nu h_\nu, 0, \dots, 0)$ . Es folgt

$$f(\mathbf{x} + \mathbf{h}) = f(\mathbf{x}) + \sum_{\nu=1}^n (f(\mathbf{x} + \mathbf{h}_\nu) - f(\mathbf{x} + \mathbf{h}_{\nu-1})) = f(\mathbf{x}) + \sum_{\nu=1}^n \frac{\partial f}{\partial x_\nu}(\mathbf{x}_\nu) \cdot h_\nu,$$

$$\frac{1}{\|\mathbf{h}\|} \left[ f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x}) - \sum_{\nu=1}^n \frac{\partial f}{\partial x_\nu}(\mathbf{x}) \cdot h_\nu \right] = \sum_{\nu=1}^n \frac{h_\nu}{\|\mathbf{h}\|} \left[ \frac{\partial f}{\partial x_\nu}(\mathbf{x}_\nu) - \frac{\partial f}{\partial x_\nu}(\mathbf{x}) \right] \rightarrow 0 \text{ für } \|\mathbf{h}\| \rightarrow 0,$$

denn

$$\frac{|h_\nu|}{\|\mathbf{h}\|} \leq 1, \quad \mathbf{x}_\nu \rightarrow \mathbf{x}, \quad \text{und} \quad \frac{\partial f}{\partial x_\nu} \text{ ist stetig in } \mathbf{x}.$$

□

Fassen wir zusammen:

$\text{stetig partiell differenzierbar} \implies \text{total differenzierbar} \implies \text{partiell differenzierbar}$
---

Wichtiger als diese etwas spröden Definitionen ist eine anschauliche Vorstellung der Ableitung. Was hat beispielsweise die Ableitung einer Funktion  $f(x)$  einer reellen Variablen aus dem ersten Semester mit einer linearen Abbildung  $A : \mathbb{R}^1 \rightarrow \mathbb{R}^1$  zu tun? Differenzierbarkeit bedeutet

$$f(x+h) = f(x) + f'(x) \cdot h + \varphi(x)$$

mit einer ziemlich kleinen Störfunktion  $\varphi$ . Hier ist  $f'(x)$  die Steigung der Tangente an den Graphen der Funktion  $f$ . Diese Tangente wird durch

$$h \mapsto f(x) + f'(x) \cdot h$$

parametrisiert, und die lineare Abbildung ist gerade

$$h \mapsto f'(x) \cdot h,$$

nichts anderes als die Multiplikation mit dem Skalar  $f'(x)$ .

Ähnlich kann man noch die Ableitung einer Funktion  $f(x, y)$  von zwei reellen Variablen  $x, y$  verstehen: Der Graph der Funktion ist eine Fläche im  $x, y, z$ -Raum, die über (einem Teil) der  $x, y$ -Ebene ausgebreitet liegt. Die Approximationseigenschaft

$$f\left(\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} h_1 \\ h_2 \end{pmatrix}\right) = f(x, y) + \frac{\partial f}{\partial x}(x, y) \cdot h_1 + \frac{\partial f}{\partial y}(x, y) \cdot h_2 + \varphi(h_1, h_2)$$

bedeutet die Approximierung des Graphen von  $f$  durch die Ebene mit der affinen Parametrisierung

$$\mathbb{R}^2 \ni \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} \mapsto \begin{pmatrix} x' = x + h_1 \\ y' = y + h_2 \\ z' = f(x, y) + \frac{\partial f}{\partial x}(x, y) \cdot h_1 + \frac{\partial f}{\partial y}(x, y) \cdot h_2 \end{pmatrix}$$

und die lineare Abbildung  $f'(x, y)$  ist gerade die dritte Komponente dieser affinen Abbildung. Die affine Ebene, welche durch

$$\mathbf{h} \mapsto (\mathbf{x} + \mathbf{h}, f(\mathbf{x}) + f'(\mathbf{x}) \cdot \mathbf{h})$$

parametrisiert wird, heißt die *Tangentialebene* im Punkt  $(\mathbf{x}, f(\mathbf{x})) \in \mathbb{R}^3$  an den Graphen der Funktion  $f$ .

Sonst kann man sich (aus Dimensionsgründen) nur noch Abbildungen  $\mathbb{R} \ni t \mapsto (x(t), y(t)) \in \mathbb{R}^2$  (Kurven in der Ebene) oder  $\mathbb{R} \ni t \mapsto (x(t), y(t), z(t)) \in \mathbb{R}^3$  (Kurven im Raum) vorstellen. Der Vektor

$$\dot{\mathbf{x}}(t_0) = \begin{pmatrix} \frac{dx}{dt}(t_0) \\ \frac{dy}{dt}(t_0) \end{pmatrix} \quad \text{bzw.} \quad \dot{\mathbf{x}}(t_0) = \begin{pmatrix} \frac{dx}{dt}(t_0) \\ \frac{dy}{dt}(t_0) \\ \frac{dz}{dt}(t_0) \end{pmatrix}$$

heißt *Geschwindigkeitsvektor* der Kurve. Heftet man ihn im Kurvenpunkt  $(x(t_0), y(t_0)) \in \mathbb{R}^2$ , bzw.  $(x(t_0), y(t_0), z(t_0)) \in \mathbb{R}^3$  an, so spannt er die *Kurventangente* in diesem Punkt auf. Die lineare Abbildung ist dann eine Parametrisierung dieser Kurventangente.

Allgemeiner nennt man jede Abbildung

$$\mathbb{R} \ni t \mapsto \mathbf{x}(t) = \begin{pmatrix} x_1(t) \\ \vdots \\ x_n(t) \end{pmatrix} \in \mathbb{R}^n$$

eine *Kurve* im  $\mathbb{R}^n$ . Sie ist differenzierbar, wenn alle ihre Komponentenfunktionen  $x_\nu(t)$  differenzierbar sind. Ihre Ableitung im Punkt  $t$

$$\dot{\mathbf{x}}(t) = \begin{pmatrix} \dot{x}_1(t) \\ \vdots \\ \dot{x}_n(t) \end{pmatrix} = \lim_{0 \neq h \rightarrow 0} \begin{pmatrix} (x_1(t+h) - x_1(t))/h \\ \vdots \\ (x_n(t+h) - x_n(t))/h \end{pmatrix} = \lim_{0 \neq h \rightarrow 0} \frac{1}{h} (\mathbf{x}(t+h) - \mathbf{x}(t))$$

ist der *Geschwindigkeitsvektor* an die Kurve zum Parameter  $t$ .

**Beispiel 2.7** Die *Helix* (Wendeltreppe) ist die Kurve

$$t \mapsto \begin{pmatrix} \cos(t) \\ \sin(t) \\ t \end{pmatrix}$$

im  $\mathbb{R}^3$ . Ihr *Geschwindigkeitsvektor* ist

$$\dot{\mathbf{x}}(t) = \begin{pmatrix} -\sin(t) \\ \cos(t) \\ 1 \end{pmatrix}.$$

Und damit wird die *Geschwindigkeit*

$$\dot{\mathbf{x}}(t) = \sqrt{\sin^2(t) + \cos^2(t) + 1} = \sqrt{2}.$$

**Beispiel 2.8** Jede  $n \times n$ -Matrix  $A$  definiert die Kurve

$$t \mapsto e^{tA}$$

im Raum  $\mathbb{R}^{n^2}$  der  $n \times n$ -Matrizen. Mit Aufgabe 2.1 findet man

$$\lim_{0 \neq h \rightarrow 0} \frac{1}{h} (e^{(t+h)A} - e^{tA}) = e^{tA} \lim_{h \rightarrow 0} \frac{1}{h} (e^{hA} - \mathbb{1}).$$

Hier ist

$$\frac{1}{h} (e^{hA} - \mathbb{1}) = A + \sum_{\nu=2}^{\infty} \frac{h^{\nu-1}}{\nu!} A^{\nu}.$$

Mit Beispiel 2.3 sieht man

$$\left\| \sum_{\nu=2}^{\infty} \frac{h^{\nu-1} A^{\nu}}{\nu!} \right\| \leq \sum_{\nu=2}^{\infty} \frac{\|A\|^{\nu}}{\nu!} |h|^{\nu-1}.$$

Diese Potenzreihe in  $|h|$  konvergiert und stellt eine in  $|h|$  stetige Funktion dar. Für  $|h| = 0$  hat diese Funktion den Wert  $= 0$ . Daraus folgt

$$\lim_{0 \neq h \rightarrow 0} \frac{1}{h} (e^{hA} - \mathbb{1}) = A.$$

Wir haben bewiesen: Die Funktion

$$\mathbb{R} \rightarrow \mathbb{R}^{n^2}, \quad t \mapsto e^{tA},$$

ist differenzierbar mit Ableitung

$$\frac{d}{dt} e^{tA} = A \cdot e^{tA}.$$

Insbesondere sind alle Spalten der Matrix  $e^{tA}$  Lösungen des Systems

$$\dot{\mathbf{y}} = A \cdot \mathbf{y}$$

linearer Differentialgleichungen.

**Aufgabe 2.4** Bestimmen Sie die Tangentialebene im Punkt  $(1, -2, 2)$  der Fläche im  $\mathbb{R}^3$ , die durch die Gleichung

$$z = 3x^2y + 2xy^2$$

gegeben ist.

**Aufgabe 2.5** Bestimmen Sie die Tangentialebene an den Graphen der Funktion  $f: \mathbb{R}^2 \rightarrow \mathbb{R}$  mit  $f(x, y) = x^3 + y^3 - 3x - 12y$  im Punkt  $(2/0/2)$ .

**Aufgabe 2.6** Gegeben seien zwei differenzierbare Abbildungen

$$\mathbf{a}, \mathbf{b}: \mathbb{R}^3 \rightarrow \mathbb{R}^3 \quad (\text{Felder}),$$

sowie daraus gebildet die skalare Funktion  $\varphi(\mathbf{x}) = (\mathbf{a}(\mathbf{x}) \cdot \mathbf{b}(\mathbf{x}))$  und das Vektorfeld  $\mathbf{A}(\mathbf{x}) = \mathbf{a}(\mathbf{x}) \times \mathbf{b}(\mathbf{x})$ .

Zeigen Sie

$$a) \frac{\partial \varphi}{\partial x_{\nu}} = \left( \frac{\partial \mathbf{a}}{\partial x_{\nu}} \cdot \mathbf{b} \right) + \left( \mathbf{a} \cdot \frac{\partial \mathbf{b}}{\partial x_{\nu}} \right), \quad b) \frac{\partial \mathbf{A}}{\partial x_{\nu}} = \frac{\partial \mathbf{a}}{\partial x_{\nu}} \times \mathbf{b} + \mathbf{a} \times \frac{\partial \mathbf{b}}{\partial x_{\nu}}.$$

**Aufgabe 2.7** Für  $\mathbf{x} \in \mathbb{R}^n$  setzt man oft

$$r = r(\mathbf{x}) = \|\mathbf{x}\| = \sqrt{x_1^2 + \dots + x_n^2}.$$

Ist  $f$  eine differenzierbare Funktion einer Variablen, so bezeichnen wir (etwas salopp) mit  $f(r)$  die Funktion  $\mathbf{x} \mapsto f(r(\mathbf{x}))$  auf  $\mathbb{R}^n$ .

a) Berechnen Sie für  $r \neq 0$  die partiellen Ableitungen  $\partial r / \partial x_\nu$ .

b) Zeigen Sie:  $\frac{\partial f(r)}{\partial x_\nu} = \frac{x_\nu}{r} \cdot f'(r)$ .

c) Zeigen Sie:  $\sum_{\nu=1}^n \frac{\partial^2 f(r)}{\partial x_\nu^2} = f''(r) + \frac{n-1}{r} f'(r)$ .

**Aufgabe 2.8** Berechnen Sie  $\partial \varphi / \partial x_\nu$  für

a)  $\varphi(\mathbf{x}) = \frac{e^{-\alpha \|\mathbf{x}\|}}{\|\mathbf{x}\|}, \quad \alpha \in \mathbb{R}, \quad$  b)  $\varphi(\mathbf{x}) = \frac{(\mathbf{a} \cdot \mathbf{x})}{\|\mathbf{x}\|}, \quad \mathbf{a} \in \mathbb{R}^n.$

### 2.3 Rechenregeln und Beispiele

Es gibt keine anderen Rechenregeln als für Funktionen einer Variablen, sie sehen allerdings gelegentlich etwas anders aus.

**Differenzierbarkeit und Stetigkeit.** Jede differenzierbare Abbildung ist stetig: Wenn  $F$  in  $\mathbf{x}$  differenzierbar ist, dann gilt

$$\lim_{\mathbf{h} \rightarrow \mathbf{0}} \frac{\Phi(\mathbf{h})}{\|\mathbf{h}\|} = \lim_{\mathbf{h} \rightarrow \mathbf{0}} \frac{1}{\|\mathbf{h}\|} (F(\mathbf{x} + \mathbf{h}) - F(\mathbf{x}) - F'(\mathbf{x}) \cdot \mathbf{h}) = 0.$$

Nach Multiplikation mit  $\|\mathbf{h}\|$  folgt daraus

$$\lim_{\mathbf{h} \rightarrow \mathbf{0}} (F(\mathbf{x} + \mathbf{h}) - F(\mathbf{x}) - F'(\mathbf{x}) \cdot \mathbf{h}) = \lim_{\mathbf{h} \rightarrow \mathbf{0}} \|\mathbf{h}\| \cdot \lim_{\mathbf{h} \rightarrow \mathbf{0}} \frac{\Phi(\mathbf{h})}{\|\mathbf{h}\|} = 0.$$

Weil aber

$$\lim_{\mathbf{h} \rightarrow \mathbf{0}} F'(\mathbf{x}) \cdot \mathbf{h} = \mathbf{0}$$

gilt (lineare Abbildungen sind stetig), folgt daraus

$$\lim_{\mathbf{h} \rightarrow \mathbf{0}} F(\mathbf{x} + \mathbf{h}) = F(\mathbf{x}).$$

**Linearität.** Sind  $F, G : D \rightarrow \mathbb{R}^k$  differenzierbar,  $\alpha, \beta \in \mathbb{R}$ , so ist auch  $\alpha F + \beta G : D \rightarrow \mathbb{R}^k$  differenzierbar mit Ableitung

$$(\alpha F + \beta G)'(\mathbf{x}) = \alpha F'(\mathbf{x}) + \beta G'(\mathbf{x}).$$

Dies ergibt sich unmittelbar aus der Definition, genau wie im ersten Semester.

Weil man Abbildungen i.A. nicht multiplizieren kann, gibt es hier keine allgemeine Produktregel. Aber es gibt Spezialfälle:

a) Sind  $t \mapsto \mathbf{a}(t) \in \mathbb{R}^n$  und  $t \mapsto \mathbf{b}(t) \in \mathbb{R}^n$  differenzierbare Abbildungen, so ist auch

$$t \mapsto (\mathbf{a}(t) \cdot \mathbf{b}(t)) = \sum_{\nu=1}^n a_{\nu}(t) \cdot b_{\nu}(t)$$

differenzierbar mit Ableitung

$$\frac{d}{dt}(\mathbf{a}(t) \cdot \mathbf{b}(t)) = \sum_{\nu=1}^n \frac{d}{dt}(a_{\nu}(t) \cdot b_{\nu}(t)) = \sum_{\nu=1}^n \dot{a}_{\nu} \cdot b_{\nu} + a \cdot \dot{b}_{\nu} = (\dot{\mathbf{a}} \cdot \mathbf{b}) + (\mathbf{a} \cdot \dot{\mathbf{b}}).$$

b) Ist  $A(t)$  eine  $n \times n$ -Matrix, die differenzierbar von  $t \in \mathbb{R}$  abhängt, so gilt

$$\frac{d}{dt}(A \cdot \mathbf{a}) = \dot{A} \cdot \mathbf{a} + A \cdot \dot{\mathbf{a}}.$$

c) Sind  $A(t)$  und  $B(t)$  zwei  $n \times n$ -Matrizen, die differenzierbar von  $t \in \mathbb{R}$  abhängen, so gilt

$$\frac{d}{dt}(A \cdot B) = \dot{A} \cdot B + A \cdot \dot{B}.$$

**Kettenregel.** Es sei  $D_n \subset \mathbb{R}^n$  offen und  $F : D_n \rightarrow \mathbb{R}^k$  differenzierbar in  $\mathbf{x} \in D_n$ . Weiter sei  $D_k \subset \mathbb{R}^k$  offen mit  $F(\mathbf{x}) \in D_k$ ,  $G : D_k \rightarrow \mathbb{R}^l$  differenzierbar in  $F(\mathbf{x})$ . Dann ist auch  $G \circ F$  in  $\mathbf{x}$  differenzierbar mit Ableitung

$$(G \circ F)'(\mathbf{x}) = G'(F(\mathbf{x})) \cdot F'(\mathbf{x}).$$

Beweis. Wir setzen  $A := F'(\mathbf{x})$ ,  $B := G'(F(\mathbf{x}))$  und

$$\begin{array}{ccc} C & := & B \cdot A \\ \mathbb{R}^n \rightarrow \mathbb{R}^l & & \mathbb{R}^k \rightarrow \mathbb{R}^l \quad \mathbb{R}^n \rightarrow \mathbb{R}^k. \end{array}$$

Dann haben wir

$$\begin{aligned} & \frac{1}{\|\mathbf{h}\|} [G(F(\mathbf{x} + \mathbf{h})) - G(F(\mathbf{x})) - C \cdot \mathbf{h}] \\ = & \frac{1}{\|\mathbf{h}\|} [G(F(\mathbf{x} + \mathbf{h})) - G(F(\mathbf{x})) - G'(F(\mathbf{x})) \cdot (F(\mathbf{x} + \mathbf{h}) - F(\mathbf{x}))] \\ & + \frac{1}{\|\mathbf{h}\|} [G'(F(\mathbf{x})) \cdot (F(\mathbf{x} + \mathbf{h}) - F(\mathbf{x})) - C \cdot \mathbf{h}] \\ = & \frac{\|F(\mathbf{x} + \mathbf{h}) - F(\mathbf{x})\|}{\|\mathbf{h}\|} \cdot \frac{[G(F(\mathbf{x} + \mathbf{h})) - G(F(\mathbf{x})) - G'(F(\mathbf{x})) \cdot (F(\mathbf{x} + \mathbf{h}) - F(\mathbf{x}))]}{\|F(\mathbf{x} + \mathbf{h}) - F(\mathbf{x})\|} \\ & + \frac{1}{\|\mathbf{h}\|} \cdot G'(F(\mathbf{x})) [F(\mathbf{x} + \mathbf{h}) - F(\mathbf{x}) - F'(\mathbf{x}) \cdot \mathbf{h}], \end{aligned}$$

falls  $F(\mathbf{x} + \mathbf{h}) \neq F(\mathbf{x})$ . Da  $F$  in  $\mathbf{x}$  stetig ist, geht  $F(\mathbf{x} + \mathbf{h})$  gegen  $F(\mathbf{x})$  für  $\mathbf{h} \rightarrow \mathbf{0}$ . Da  $G$  in  $F(\mathbf{x})$  differenzierbar ist, folgt daraus

$$\lim_{\mathbf{0} \neq \mathbf{h} \rightarrow \mathbf{0}} \frac{[G(F(\mathbf{x} + \mathbf{h})) - G(F(\mathbf{x})) - G'(F(\mathbf{x})) \cdot (F(\mathbf{x} + \mathbf{h}) - F(\mathbf{x}))]}{\|F(\mathbf{x} + \mathbf{h}) - F(\mathbf{x})\|} = \mathbf{0}.$$

Das Produkt

$$F'(\mathbf{x}) \cdot \frac{\mathbf{h}}{\|\mathbf{h}\|}$$

ist eine stetige Funktion des Einheitsvektors  $\mathbf{h}/\|\mathbf{h}\|$  und bleibt deswegen beschränkt, wenn  $\mathbf{h} \rightarrow \mathbf{0}$ . Da  $F$  in  $\mathbf{x}$  differenzierbar ist, gilt

$$\lim_{\mathbf{0} \neq \mathbf{h} \rightarrow \mathbf{0}} \left( \frac{F(\mathbf{x} + \mathbf{h}) - F(\mathbf{x})}{\|\mathbf{h}\|} - F'(\mathbf{x}) \cdot \frac{\mathbf{h}}{\|\mathbf{h}\|} \right) = \mathbf{0},$$

und deswegen bleibt auch

$$\frac{1}{\|\mathbf{h}\|} (F(\mathbf{x} + \mathbf{h}) - F(\mathbf{x}))$$

beschränkt für  $\mathbf{h} \rightarrow \mathbf{0}$ . Der erste Summand in unserem obigen Ausdruck geht also gegen  $\mathbf{0}$  wenn  $\mathbf{h} \rightarrow \mathbf{0}$ . (Dies gilt natürlich auch falls  $F(\mathbf{x} + \mathbf{h}) = F(\mathbf{x})$ .)

Der zweite Summand geht aber ebenfalls gegen Null, wegen der Differenzierbarkeit von  $F$  in  $\mathbf{x}$  und der Stetigkeit der linearen Abbildung  $G'(F(\mathbf{x}))$ .  $\square$

Die Kettenregel besagt, dass die linearen Abbildungen  $F'(\mathbf{x})$  und  $G'(F(\mathbf{x}))$  in dieser Reihenfolge hintereinander ausgeführt werden, um die Ableitung von  $G \circ F$  in  $\mathbf{x}$  zu bekommen. Hintereinander Ausführen linearer Abbildungen ist aber Multiplikation der beschreibenden Matrizen. So ist das Produkt der Ableitungen in der Kettenregel das Matrizenprodukt der Funktionalmatrizen. In dieser Form ist die Kettenregel also eine ziemlich wörtliche Verallgemeinerung der Kettenregel aus dem ersten Semester. Nur das Produkt von Skalaren ist ersetzt durch das Produkt von Matrizen.

**Beispiel 2.9 (Richtungsableitung)** Die Funktion  $f(x_1, \dots, x_n)$  sei differenzierbar. Wir betrachten die Funktion auf der Geraden

$$\mathbf{x} = \mathbf{x}_0 + t \cdot \mathbf{v}, \quad t \in \mathbb{R}$$

mit Anfangsvektor  $\mathbf{x}_0$  und Richtungsvektor  $\mathbf{v}$ . Schreiben wir die Parametrisierungsabbildung

$$g: \begin{cases} \mathbb{R} \rightarrow \mathbb{R}^n \\ t \mapsto \mathbf{x}_0 + t \cdot \mathbf{v}, \end{cases}$$

so ist die auf die Gerade eingeschränkte Funktion  $f(g(t))$ . Wir differenzieren sie nach der Kettenregel

$$\begin{aligned} (f \circ g)'(0) &= f'(g(0)) \cdot g'(0) \\ &= \left( \frac{\partial f}{\partial x_1}(\mathbf{x}_0), \dots, \frac{\partial f}{\partial x_n}(\mathbf{x}_0) \right) \cdot \begin{pmatrix} \frac{dg_1}{dt}(0) \\ \vdots \\ \frac{dg_n}{dt}(0) \end{pmatrix} \\ &= \left( \frac{\partial f}{\partial x_1}(\mathbf{x}_0), \dots, \frac{\partial f}{\partial x_n}(\mathbf{x}_0) \right) \cdot \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} \end{aligned}$$

$$= \sum_{\nu=1}^n \frac{\partial f}{\partial x_{\nu}}(\mathbf{x}_0) \cdot v_{\nu}.$$

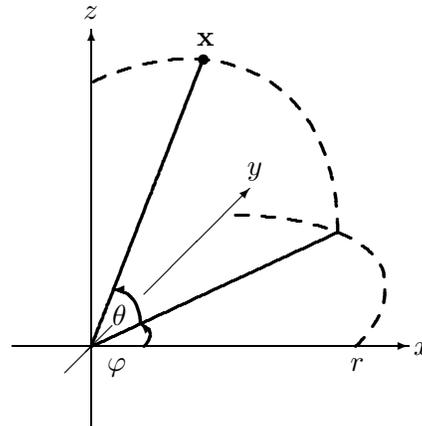
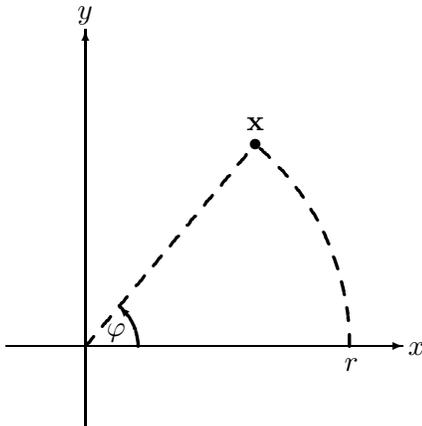
Dieses Produkt nennt man auch die Richtungsableitung von  $f$  in Richtung des Vektors  $\mathbf{v}$ .

**Beispiel 2.10 (Transformation in Polarkoordinaten)** Zwischen den euklidischen Koordinaten  $x, y$  der Ebene und den Polarkoordinaten besteht die Beziehung

$$\begin{aligned} x &= r \cdot \cos(\varphi) \\ y &= r \cdot \sin(\varphi). \end{aligned}$$

Die Zuordnung  $(r, \varphi) \mapsto (x, y)$  ist eine differenzierbare Abbildung. Wir berechnen ihre Funktionalmatrix

$$\begin{pmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \varphi} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \varphi} \end{pmatrix} = \begin{pmatrix} \cos(\varphi) & -r \cdot \sin(\varphi) \\ \sin(\varphi) & r \cdot \cos(\varphi) \end{pmatrix}.$$



**Beispiel 2.11 (Transformation in Kugelkoordinaten)** Die Beziehung lautet jetzt

$$\begin{aligned} x &= r \cdot \cos(\varphi) \cdot \cos(\theta) \\ y &= r \cdot \sin(\varphi) \cdot \cos(\theta) \\ z &= r \cdot \sin(\theta). \end{aligned}$$

Die Funktionalmatrix ist

$$\begin{pmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \varphi} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \varphi} & \frac{\partial y}{\partial \theta} \\ \frac{\partial z}{\partial r} & \frac{\partial z}{\partial \varphi} & \frac{\partial z}{\partial \theta} \end{pmatrix} = \begin{pmatrix} \cos(\varphi)\cos(\theta) & -r\sin(\varphi)\cos(\theta) & -r\cos(\varphi)\sin(\theta) \\ \sin(\varphi)\cos(\theta) & r\cos(\varphi)\cos(\theta) & -r\sin(\varphi)\sin(\theta) \\ \sin(\theta) & 0 & r\cos(\theta) \end{pmatrix}.$$

Ähnlich wie in der Differentialrechnung einer Variablen kann man auch hier die Differentiation iterieren: Man sagt, die Abbildung  $F$  ist

- stetig differenzierbar, wenn  $F$  differenzierbar ist und alle partiellen Ableitungen  $\partial F_\kappa / \partial x_\nu$  wieder stetig sind;
- zweimal differenzierbar, wenn die partiellen Ableitungen wieder differenzierbar sind. In diesem Fall gibt es dann zweite partielle Ableitungen

$$\frac{\partial^2 F_\kappa}{\partial x_{\nu_1} \partial x_{\nu_2}} = \frac{\partial}{\partial x_{\nu_1}} \left( \frac{\partial F_\kappa}{\partial x_{\nu_2}} \right).$$

Wenn diese höheren partiellen Ableitungen stetig sind, dann sind sie unabhängig von der Reihenfolge, in der man differenziert:

**Satz 2.5 (Symmetrie der zweiten Ableitung)** Die Funktion  $f : D \rightarrow \mathbb{R}$  sei zweimal stetig differenzierbar. Dann gilt

$$\frac{\partial^2 f}{\partial x_{\nu_1} \partial x_{\nu_2}} = \frac{\partial^2 f}{\partial x_{\nu_2} \partial x_{\nu_1}}.$$

Beweis. O.B.d.A. beweisen wir die Behauptung nur für den Fall  $n = 2$ . Die Variablen nennen wir dann  $x$  und  $y$ . Da es sich um zweite Ableitungen dreht, bilden wir vom Differenzenquotienten bezüglich  $x$

$$\frac{1}{h}(f(x+h, y) - f(x, y))$$

den Differenzenquotient bezüglich  $y$

$$\begin{aligned} \frac{1}{hk} [(f(x+h, y+k) - f(x, y+k)) - (f(x+h, y) - f(x, y))] &= \\ \frac{1}{hk} [f(x+h, y+k) - f(x, y+k) - f(x+h, y) + f(x, y)]. \end{aligned}$$

Dieser iterierte Differenzenquotient ist symmetrisch bezüglich  $x$  und  $y$ , die Reihenfolge, in der wir ihn bildeten, sieht man ihm nicht mehr an.

Den Differenzenquotienten schreiben wir nach dem Mittelwertsatz der Differentialrechnung als Differentialquotient an einer Zwischenstelle, zuerst bezüglich  $x$

$$\frac{\partial}{\partial x}(f(x + \theta_1 h, y + k) - f(x + \theta_1 h, y))$$

und dann bezüglich  $y$

$$\frac{\partial^2}{\partial y \partial x} f(x + \theta_1 h, y + \theta_2 k)$$

mit  $\theta_1, \theta_2 \in \mathbb{R}$  zwischen 0 und 1.

Wenn wir den Mittelwertsatz in der umgekehrten Reihenfolge anwenden, erst auf  $y$  und dann auf  $x$ , so wird der gleiche Differentialquotient

$$\frac{\partial^2}{\partial x \partial y} f(x + \theta_3 h, y + \theta_4 k)$$

mit neuen Werten  $\theta_3, \theta_4 \in \mathbb{R}$  zwischen 0 und 1. Wir haben also die Gleichung

$$\frac{\partial^2}{\partial y \partial x} f(x + \theta_1 h, y + \theta_2 k) = \frac{\partial^2}{\partial x \partial y} f(x + \theta_3 h, y + \theta_4 k)$$

für die zweiten partiellen Ableitungen, in vertauschter Reihenfolge, in verschiedenen Punkten. Nun lassen wir aber  $h$  und  $k$  gegen 0 gehen, dann gehen beide Punkte gegen  $(x, y)$  und aus der Stetigkeit der zweiten partiellen Ableitungen folgt Gleichheit der Werte

$$\frac{\partial^2 f}{\partial y \partial x}(x, y) = \frac{\partial^2 f}{\partial x \partial y}(x, y).$$

□

Eine der wichtigsten Anwendungen der Differentialrechnung, jedenfalls historisch die wichtigste, ist das Ermitteln von lokalen Extrema: Maxima und Minima. Auch bei Funktionen  $f(\mathbf{x}) = f(x_1, \dots, x_n)$  mehrerer Variablen sagt man, die Funktion nimmt auf einer Menge  $D$  in einem Punkt  $\mathbf{x}_0 \in M$  ein Maximum (Minimum) an, wenn  $f(\mathbf{x}) \leq f(\mathbf{x}_0)$ , bzw.  $f(\mathbf{x}) \geq f(\mathbf{x}_0)$  für alle  $\mathbf{x} \in D$ . Und  $\mathbf{x}_0$  ist ein *lokales Maximum*, bzw. *Minimum*, wenn es ein  $r > 0$  gibt, so, dass diese Ungleichungen für alle  $\mathbf{x} \in D$  mit  $\|\mathbf{x} - \mathbf{x}_0\| < r$  gelten.

**Satz 2.6 (Notwendige Bedingung für lokale Extrema)** Die Funktion  $f$  sei differenzierbar auf der offenen Menge  $D \subset \mathbb{R}^n$ . Nimmt  $f$  in einem Punkt  $\mathbf{x}_0 \in D$  ein lokales Extremum an, so verschwinden dort die partiellen Ableitungen:

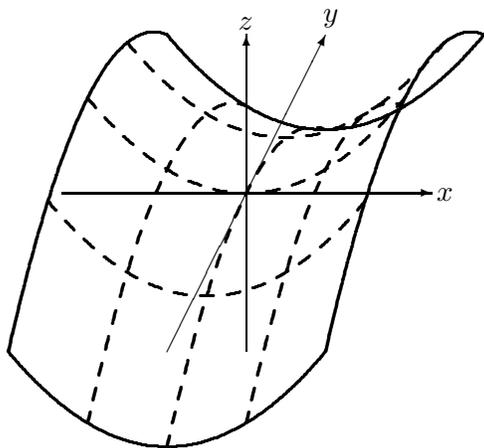
$$\frac{\partial f}{\partial x_1}(\mathbf{x}_0) = \dots = \frac{\partial f}{\partial x_n}(\mathbf{x}_0) = 0.$$

Beweis. Es gibt ein  $r > 0$  derart, dass alle  $n$  Funktionen

$$f(\mathbf{x}_0 + t \cdot \mathbf{e}_k), \quad k = 1, \dots, n,$$

auf dem offenen Intervall  $|t| < r$  differenzierbar sind. Für  $t = 0$  nehmen sie ein lokales Extremum an, ihre Ableitungen verschwinden für  $t = 0$ , dies sind aber gerade die partiellen Ableitungen von  $f$  in  $\mathbf{x}_0$ .

□



**Beispiel 2.12** Für die Funktion  $f(x, y) = x^2 - y^2$  verschwinden beide partiellen Ableitungen

$$\frac{\partial f}{\partial x} = 2x, \quad \frac{\partial f}{\partial y} = 2y,$$

im Nullpunkt. Hier liegt aber kein lokales Extremum, weil  $f(0, 0) = 0$  und  $f(x, 0) > 0$  für alle  $x \neq 0$ , während  $f(0, y) < 0$  für alle  $y \neq 0$ . Einen derartigen Punkt nennt man Sattelpunkt.

Für Abschätzungen wichtig ist der Mittelwertsatz. In einer Veränderlichen hat er die Form

$$f(y) - f(x) = f'(\xi) \cdot (y - x)$$

mit einem Zwischenpunkt

$$\xi = x + \theta(y - x), \quad 0 < \theta < 1.$$

So ist er nicht direkt auf Abbildungen verallgemeinerbar, und zwar aus zwei Gründen:

1) Es sei  $M \subset \mathbb{R}^n$  offen und  $F : M \rightarrow \mathbb{R}^k$  differenzierbar. Um für zwei Punkte  $\mathbf{x}, \mathbf{y} \in M$  die Werte  $F(\mathbf{x})$  und  $F(\mathbf{y})$  zu vergleichen, betrachtet man die Abbildung

$$F : t \mapsto F(\mathbf{x} + t \cdot (\mathbf{y} - \mathbf{x})),$$

d.h., die Abbildung  $F$  eingeschränkt auf die Verbindungsgerade der Punkte  $\mathbf{x}$  und  $\mathbf{y}$ . Man braucht, dass die Strecke

$$\{\mathbf{x} + t(\mathbf{y} - \mathbf{x}), \quad 0 \leq t \leq 1\}$$

zwischen  $\mathbf{x}$  und  $\mathbf{y}$  ganz in der Definitionsmenge  $M$  liegt. Dieses Problem löst man durch eine Definition: Die Menge  $M \subset \mathbb{R}^n$  heißt *konvex*, wenn mit je zwei Punkten auch ihre Verbindungsstrecke zu  $M$  gehört. Z.B. sind Kugeln konvexe Mengen.

2) Für Abbildungen kann der Zwischenwertsatz prinzipiell nicht gelten. Als Gegenbeispiel betrachten wir die Abbildung

$$F : \mathbb{R} \rightarrow \mathbb{R}^2, \quad F : t \mapsto \begin{pmatrix} \cos(t) \\ \sin(t) \end{pmatrix}.$$

Hier ist

$$F(0) = F(2\pi) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \text{also} \quad F(2\pi) - F(0) = \mathbf{0}.$$

Aber für alle  $\xi \in \mathbb{R}$  ist

$$F'(\xi) = \begin{pmatrix} -\sin(\xi) \\ \cos(\xi) \end{pmatrix} \neq \mathbf{0}.$$

Also gibt es kein  $\xi$  mit der Zwischenwerteigenschaft

$$F(2\pi) - F(0) = F'(\xi) \cdot (2\pi - 0).$$

Trotzdem kann man die aus dem Zwischenwertsatz folgende Abschätzung retten:

**Satz 2.7** Die Menge  $M \subset \mathbb{R}^n$  sei offen und konvex. Die Abbildung  $F : M \rightarrow \mathbb{R}^k$  sei differenzierbar mit

$$\|F'(\xi)\| \leq L$$

für alle  $\xi \in M$ . Dann gilt für alle  $\mathbf{x}, \mathbf{y} \in M$

$$\|F(\mathbf{y}) - F(\mathbf{x})\| \leq L \cdot \|\mathbf{y} - \mathbf{x}\|.$$

Beweis. Es sei  $\mathbf{v} := F(\mathbf{y}) - F(\mathbf{x}) \in \mathbb{R}^k$ . Wenn  $\mathbf{v} = \mathbf{0}$  gelten sollte, sind wir trivialerweise fertig. Andernfalls betrachten wir die Funktion

$$f : M \rightarrow \mathbb{R}, \quad f : \mathbf{x} \mapsto (\mathbf{v}, F(\mathbf{x})).$$

Ihre Einschränkung auf die Verbindungsstrecke  $f(\mathbf{x} + t(\mathbf{y} - \mathbf{x}))$  hat die Ableitung

$$(\text{grad } f(\mathbf{x} + t(\mathbf{y} - \mathbf{x})), (\mathbf{y} - \mathbf{x})).$$

Das folgt wie in Beispiel 2.9. Mit dem Zwischenwertsatz in einer Veränderlichen erhalten wir

$$f(\mathbf{y}) - f(\mathbf{x}) = (\text{grad } f(\xi), (\mathbf{y} - \mathbf{x})), \quad \xi = \mathbf{x} + \theta(\mathbf{y} - \mathbf{x}), \quad 0 < \theta < 1.$$

Und aus der Cauchy-Schwarz-Ungleichung folgt

$$|f(\mathbf{y}) - f(\mathbf{x})| \leq \| \text{grad } f(\xi) \| \cdot \| \mathbf{y} - \mathbf{x} \|.$$

Hier ist

$$\begin{aligned} \text{grad } f(\xi) &= \left( \frac{\partial}{\partial x_\nu} \sum_{\mu=1}^n v_\mu F_\mu(\xi) \right)_{\nu=1, \dots, n} = \left( \sum_{\mu=1}^n v_\mu \frac{\partial F_\mu}{\partial x_\nu}(\xi) \right)_{\nu=1, \dots, n} = \mathbf{v}^t \cdot F'(\xi), \\ \| \text{grad } f(\xi) \| &\leq \| \mathbf{v} \| \cdot \| F'(\xi) \| \leq L \| \mathbf{v} \|. \end{aligned}$$

Insgesamt sehen wir

$$\| \mathbf{v} \|^2 = \mathbf{v}^t \cdot (F(\mathbf{y}) - F(\mathbf{x})) = |f(\mathbf{y}) - f(\mathbf{x})| \leq \| \mathbf{v} \| \cdot L \cdot \| \mathbf{y} - \mathbf{x} \|.$$

Division durch  $\| \mathbf{v} \| \neq 0$  ergibt schließlich die behauptete Ungleichung. □

**Aufgabe 2.9** Warum gibt es keine partiell differenzierbare Funktion  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$  mit

$$\frac{\partial f}{\partial x} = \arctan(xy), \quad \frac{\partial f}{\partial y} = e^x \sin y$$

für alle  $(x, y) \in \mathbb{R}^2$ ?

**Aufgabe 2.10** Es seien  $F$  und  $G$  zweimal stetig partiell differenzierbare Funktionen von  $\mathbb{R}$  in  $\mathbb{R}$  und  $u : \mathbb{R}^2 \rightarrow \mathbb{R}$  sei durch  $u(t, x) = F(x + t) + G(x - t)$  definiert. Beweisen Sie

$$\frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 u}{\partial x^2}.$$

**Aufgabe 2.11** Berechnen Sie alle partiellen Ableitungen 2. Ordnung für die Funktionen

$$f(x, y) = 2x^3 + 3xy^2, \quad g(x, y) = e^{2x-y} \cdot \sin(x + y).$$

**Aufgabe 2.12** Es sei  $F : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  definiert durch

$$(r, \varphi) \mapsto (r \cos(\varphi), r \sin(\varphi)).$$

a) Bestimmen Sie  $F'(r, \varphi)^{-1}$ , wo es existiert.

b) Es sei  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$  differenzierbar und  $g = f \circ F$ . Zeigen Sie für  $r \neq 0$

$$\begin{aligned} \frac{\partial f}{\partial x} F(r, \varphi) &= \cos(\varphi) \cdot \frac{\partial g}{\partial r}(r, \varphi) - \frac{1}{r} \sin(\varphi) \cdot \frac{\partial g}{\partial \varphi}(r, \varphi), \\ \frac{\partial f}{\partial y} F(r, \varphi) &= \sin(\varphi) \cdot \frac{\partial g}{\partial r}(r, \varphi) + \frac{1}{r} \cos(\varphi) \cdot \frac{\partial g}{\partial \varphi}(r, \varphi). \end{aligned}$$

**Aufgabe 2.13 (Formel von Euler)** Es sei  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  differenzierbar und  $0 < p \in \mathbb{N}$ . Die Funktion  $f$  heißt homogen vom Grad  $p$ , wenn für alle  $\mathbf{x} \in \mathbb{R}^n$ ,  $t \in \mathbb{R}$  gilt

$$f(t \cdot \mathbf{x}) = t^p \cdot f(\mathbf{x}).$$

Zeigen Sie für jede solche Funktion

$$\sum_{\nu=1}^n x_{\nu} \cdot \frac{\partial f}{\partial x_{\nu}}(\mathbf{x}) = p \cdot f(\mathbf{x}).$$

**Aufgabe 2.14** Es sei  $\Phi$  eine symmetrische Bilinearform auf  $\mathbb{R}^n$  und  $q(\mathbf{x}) = \Phi(\mathbf{x}, \mathbf{x})$  die zugehörige quadratische Form. Zeigen Sie, dass  $q$  in jedem Punkt  $\mathbf{x} \in \mathbb{R}^n$  differenzierbar ist, und berechnen Sie  $q'(\mathbf{x})$ .

**Aufgabe 2.15** Für festes  $(x, y, z) \in \mathbb{R}^3$  werde eine Abbildung  $\mathbf{r} : \mathbb{R} \rightarrow \mathbb{R}^3$  definiert durch

$$\mathbf{r}(t) = \begin{pmatrix} \cos(\omega t) & -\sin(\omega t) & 0 \\ \sin(\omega t) & \cos(\omega t) & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} x \\ y \\ z \end{pmatrix}.$$

Zeigen Sie

$$\frac{d\mathbf{r}}{dt} = \omega \mathbf{e}_3 \times \mathbf{r}(t).$$

**Aufgabe 2.16** In jedem Punkt des Einheitskreises  $\{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$  berechne man die Ableitung von

$$f(x, y) = (x^2 - 2y^2) \exp(x^2 + y^2)$$

in Richtung der (positiv orientierten) Kreistangente. In welchen Punkten nimmt die Richtungsableitung Extremwerte an und wie groß sind diese? (Hinweis:  $\sin(2t) = 2\sin(t)\cos(t)$ .)

## 2.4 Taylor-Entwicklung

Jede genügend oft differenzierbare Funktion  $f$  von mehreren Variablen kann man genauso, wie eine Funktion einer Variablen, in eine Taylor-Reihe entwickeln. Als ich diese Vorlesung das vorletzte mal las, habe ich diesem Thema nicht die geringste Beachtung geschenkt. Ich dachte, wenn man die Taylor-Entwicklung in einer Variablen verstanden hat, dann kann man die Taylor-Entwicklung in mehreren Variablen im Schlaf. Und außerdem, wo soll das in der Physik vorkommen? Inzwischen hat mich aber eine ziemlich gut informierte Studentin der Mathematik und Physik davon überzeugt, dass ich dieses Thema doch besprechen sollte. Also, was gibt es dazu zu sagen?

Erinnern wir uns - ach es ist so lange her - was die Taylor-Reihe einer Funktion  $f(x)$  einer reellen Variablen  $x$  ist: Man fixiert einen Wert  $x \in \mathbb{R}$ , um den man entwickelt, und eine Auslenkung  $h \in \mathbb{R}$ , und schreibt

$$f(x+h) = f(x) + f'(x) \cdot h + \frac{1}{2} f''(x) \cdot h^2 + \dots + \frac{1}{n!} f^{(n)}(x) h^n + \dots$$

Der Sinn dieser Zeile besteht darin, den Wert  $f(x+h)$  der Funktion  $f$  im ausgelenkten Punkt  $x+h$  durch die Ableitungen  $f^{(\nu)}(x)$  der Funktion  $f$  im fixierten Punkt  $x$  auszudrücken. Dabei gibt es drei Probleme von ganz verschiedenen Schwierigkeitsgraden:

**Problem 1:** Bei der  $\nu$ -ten Ableitung  $f^{(\nu)}(x)$  steht die  $\nu$ -te Potenz  $h^\nu$  der Auslenkung. Das ist ganz einfach zu merken. Und es ist auch sehr praktisch, denn für kleine Auslenkung  $h$  wird  $h^\nu$  sehr klein, umso kleiner, je größer  $\nu$  ist. Es kommt also nur auf die ersten Potenzen an, etwa  $\nu = 0, 1, 2, 3$ , je nach dem, wie genau man es nimmt.

**Problem 2:** Bei der  $\nu$ -ten Ableitung  $f^{(\nu)}(x)$  steht nicht nur die Potenz  $h^\nu$ , sondern noch ein Zahlenfaktor, nämlich

$$\frac{1}{\nu!}.$$

Den muss man sich merken. Aber das geht einigermaßen, wenn man an die Exponentialreihe

$$e^x = 1 + x + \frac{1}{2}x^2 + \frac{1}{6}x^3 + \dots = \sum \frac{1}{\nu!}x^\nu$$

denkt. Und diese Exponentialreihe muss man einfach wissen. An der führt im Leben eines Naturwissenschaftlers kein Weg vorbei. Und wenn man sich an sie erinnert, erinnert man sich auch an den Zahlenfaktor  $1/\nu!$ .

**Problem 3:** Beim Hinschreiben der Taylor-Reihe muss jeder Mensch irgendwann aufhören. Menschen sind beschränkt, und können nicht alle unendlich vielen Summanden der Taylor-Reihe hinschreiben. In der Praxis ist also eine Taylor-Reihe dazu da, abgebrochen zu werden. Und wenn man abbricht, macht man einen Fehler. Den kann man Restglied  $R_n(h)$  nennen. Und dann sieht die abgebrochene Taylor-Reihe so aus:

$$f(x+h) = \sum_{\nu=0}^n \frac{1}{\nu!} f^{(\nu)}(x) h^\nu + R_n(h).$$

Und für einen Mathematiker ist die Kontrolle über das Restglied das Wesentliche. Damit kann er Abschätzungen beweisen. Für einen Physiker dagegen ist das Restglied klein, kleiner als  $const \cdot h^{n+1}$ . Damit kann er Trends feststellen.

Das Schwierigste an der Taylor-Formel ist, sich eine Formel für das Restglied zu merken. Das ist schon in einer Variablen schwierig, in mehreren erst recht. Und deswegen wollen wir uns nicht hierauf konzentrieren, sondern auf die ersten Summanden der Taylor-Reihe.

Betrachten wir also eine Funktion  $f(x, y)$  von zwei Variablen, genügend oft differenzierbar. Wir fixieren einen Aufpunkt  $(x, y)$  und wollen den Wert  $f(x+h, y+k)$  in einem Punkt mit dem Auslenkungsvektor  $(h, k)$  durch eine Reihe ausdrücken. Dazu fixieren wir zuerst noch die Auslenkung  $y+k$  im zweiten Argument, und entwickeln für festes  $y+k$  die Funktion  $f(x+h, y+k)$  der einen Variablen  $h$  in eine Taylor-Reihe. Da kommen dann alle partiellen Ableitungen von  $f$  nach  $x$  vor. Es ist mir zu viel Schreib-Arbeit, die ordentlich hinzuschreiben. Deswegen möchte ich abkürzen:

$$f_x = \frac{\partial f}{\partial x}, \quad f_{xx} = \frac{\partial^2 f}{\partial x^2}, \quad f_{xxx} = \frac{\partial^3 f}{\partial x^3}, \quad \text{usw.}$$

Dann wird also

$$f(x+h, y+k) = f(x, y+k) + f_x(x, y+k) \cdot h + \frac{1}{2} f_{xx}(x, y+k) \cdot h^2 + \dots + \frac{1}{\mu!} f_{x\dots x}(x, y+k) \cdot h^\mu + \dots$$

(Wie gesagt, das Restglied lasse ich weg, mir kommt es hier nur auf die Taylor-Koeffizienten an.)  
Damit haben wir also die Funktion  $f$  für festes  $y+k$  in eine Taylor-Reihe nach  $x$  entwickelt. Und jetzt entwickeln wir die Koeffizienten in Taylor-Reihen nach  $y$ :

$$\begin{aligned} f(x, y+k) &= f(x, y) + f_y(x, y) \cdot k + \frac{1}{2} f_{yy}(x, y) \cdot k^2 + \dots + \frac{1}{\nu!} f_{yy\dots y}(x, y) \cdot k^\nu + \dots, \\ f_x(x, y+k) &= f_x(x, y) + f_{xy}(x, y) \cdot k + \frac{1}{2} f_{xyy}(x, y) \cdot k^2 + \dots + \frac{1}{\nu!} f_{xyy\dots y}(x, y) \cdot k^\nu + \dots, \\ f_{xx}(x, y+k) &= f_{xx}(x, y) + f_{xxy}(x, y) \cdot k + \frac{1}{2} f_{xxyy}(x, y) \cdot k^2 + \dots + \frac{1}{\nu!} f_{xxyy\dots y}(x, y) \cdot k^\nu + \dots, \end{aligned}$$

und so weiter. Setzen wir diese Koeffizienten in die eindimensionale Entwicklung nach  $x$  ein, so finden wir

$$\begin{aligned} &f(x+h, y+k) \\ = &f(x, y) + f_y(x, y) \cdot k + \frac{1}{2} f_{yy}(x, y) \cdot k^2 + \frac{1}{3!} f_{yyy}(x, y) \cdot k^3 + \dots \\ &+ f_x(x, y) \cdot h + f_{xy}(x, y) \cdot hk + \frac{1}{2} f_{xyy}(x, y) \cdot hk^2 + \frac{1}{3!} f_{xyyy}(x, y) \cdot hk^3 + \dots \\ &+ \frac{1}{2} f_{xx}(x, y) \cdot h^2 + \frac{1}{2} f_{xxy}(x, y) \cdot h^2k + \frac{1}{2 \cdot 2} f_{xxyy}(x, y) \cdot h^2k^2 + \frac{1}{2 \cdot 3!} f_{xxyyy}(x, y) \cdot h^2k^3 + \dots \\ &+ \frac{1}{3!} f_{xxx}(x, y) \cdot h^3 + \frac{1}{3!} f_{xxxy}(x, y) \cdot h^3k + \frac{1}{3! \cdot 2} f_{xxxyy}(x, y) \cdot h^3k^2 + \frac{1}{3! \cdot 3!} f_{xxxyyy}(x, y) \cdot h^3k^3 + \dots \end{aligned}$$

Das ist ja eine schöne Bescherung! Wie soll man sich diesen Formelkram merken? Naja, was kommt alles vor? Alle partiellen Ableitungen

$$\frac{\partial^{\mu+\nu} f}{\partial x_\mu \partial y_\nu}(x, y)$$

im Entwicklungspunkt. Und die kommen zusammen mit der Potenz

$$h^\mu k^\nu$$

und dem numerischen Faktor

$$\frac{1}{\mu! \cdot \nu!}.$$

Die Taylor-Reihe in zwei Variablen ist also

$$f(x+h, y+k) = \sum_{\mu, \nu=0}^{\infty} \frac{1}{\mu! \cdot \nu!} \frac{\partial^{\mu+\nu} f}{\partial x_\mu \partial y_\nu}(x, y) \cdot h^\mu k^\nu.$$

Wenn man ein wenig nachdenkt, merkt man, es kann eigentlich gar nicht anders sein. Also ist es doch nicht so schwer zu merken.

**Beispiel 2.13** Wir betrachten die Funktion  $r(x, y) = \sqrt{x, y}$ . Ihre partiellen Ableitungen bis zur dritten Ordnung sind

$$\begin{aligned} r_x &= \frac{x}{r}, & r_y &= \frac{y}{r}, \\ r_{xx} &= \frac{1}{r} - \frac{x^2}{r^3}, & r_{xy} &= -\frac{xy}{r^3}, & r_{yy} &= \frac{1}{r} - \frac{y^2}{r^3}, \end{aligned}$$

$$r_{xxx} = -\frac{3x}{r^3} + \frac{3x^2}{r^5}, \quad r_{xxy} = -\frac{y}{r^3} + \frac{3x^2y}{r^5}, \quad r_{xyy} = -\frac{x}{r^3} + \frac{3xy^2}{r^5}, \quad r_{yyy} = -\frac{3y}{r^3} + \frac{3y^3}{r^5}.$$

Etwa im Punkt  $(x, y) = (3, 4)$  mit  $r = 5$  haben diese Ableitungen die Werte

$$r_x = \frac{3}{5}, \quad r_y = \frac{4}{5},$$

$$r_{xx} = \frac{16}{125}, \quad r_{xy} = -\frac{12}{125}, \quad r_{yy} = \frac{9}{125},$$

$$r_{xxx} = -\frac{144}{3125}, \quad r_{xxy} = \frac{8}{3125}, \quad r_{xyy} = \frac{69}{3125}, \quad r_{yyy} = -\frac{108}{3125}.$$

Die Taylorentwicklung um diesen Punkt bis zur dritten Ordnung ist also

$$\begin{aligned} \sqrt{(3+h)^2 + (4+k)^2} &= 5 + \frac{1}{5}(3h + 4k) + \\ &\quad \frac{1}{250}(16h^2 - 24hk + 9k^2) + \\ &\quad \frac{1}{18750}(-144h^3 + 24h^2k + 207hk^2 - 108k^3) + \dots \end{aligned}$$

**Aufgabe 2.17** Es sei wie immer  $r(x, y) = \sqrt{x^2 + y^2}$ . Bestimmen Sie die Taylor-Entwicklung dieser Funktion  $r$  im Punkt  $(x, y) = (1, 0)$  bis zur zweiten Ordnung.

**Aufgabe 2.18** Berechnen Sie die Taylorreihe der Funktion  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ ,

$$f(x, y) = \sin(x^2 + y^2),$$

mit dem Ursprung des  $\mathbb{R}^2$  als Entwicklungspunkt.

**Aufgabe 2.19** Berechnen Sie die Taylorreihe von  $f(x, y) = \frac{1}{x+y}$  zum Entwicklungspunkt  $(1, 1) \in \mathbb{R}^2$ .

**Aufgabe 2.20** Bestimmen Sie die Taylor-Reihe der Funktion

$$F(x, y) = \frac{2xy}{x+y}, \quad x, y > 0;$$

im Punkt  $(x, y) = (1, 1)$  bis einschließlich Glieder zweiter Ordnung.

## 2.5 Umkehrung differenzierbarer Abbildungen

Hier untersuchen wir, wann eine differenzierbare Abbildung  $F$  eine differenzierbare Umkehrabbildung besitzt. Das heißt folgendes: Es gibt offene Mengen  $U \subset \mathbb{R}^n$  und  $V \subset \mathbb{R}^k$ , differenzierbare Abbildungen

$$F : U \rightarrow V \quad G : V \rightarrow U$$

mit

$$G(F(\mathbf{x})) = \mathbf{x} \text{ für alle } \mathbf{x} \in U, \quad F(G(\mathbf{y})) = \mathbf{y} \text{ für alle } \mathbf{y} \in V.$$

Eine notwendige Bedingung dafür folgt sofort aus der Kettenregel:

$$\mathbb{1}_n = (G \circ F)'(\mathbf{x}) = G'(F(\mathbf{x})) \cdot F'(\mathbf{x}) \quad \text{und} \quad \mathbb{1}_k = (F \circ G)'(\mathbf{y}) = F'(G(\mathbf{y})) \cdot G'(\mathbf{y}).$$

Das geht nur, wenn  $k = n$  und die  $n \times n$ -Matrix  $F'(\mathbf{x})$  invertierbar ist. Inhalt dieses Abschnitts ist der Beweis der entsprechenden hinreichenden Bedingung:

**Satz 2.8 (Umkehrsatz)** *Es sei  $U \subset \mathbb{R}^n$  offen,  $F : U \rightarrow \mathbb{R}^n$  stetig differenzierbar mit  $\det(F'(\mathbf{x}_0)) \neq 0$  für ein  $\mathbf{x}_0 \in U$ . Dann gibt es offene Mengen*

$$U_0 \subset U \text{ mit } \mathbf{x}_0 \in U, \quad V_0 \subset \mathbb{R}^n \text{ mit } \mathbf{y}_0 := F(\mathbf{x}_0) \in V_0$$

so, dass  $F : U_0 \rightarrow V_0$  bijektiv und die Umkehrabbildung  $G : V_0 \rightarrow U_0$  differenzierbar ist.

Diese Aussage ist nur lokal, ganz anders als die globale Aussage in der Theorie einer Veränderlichen: Ist  $U \subset \mathbb{R}$  ein offenes Intervall,  $f : U \rightarrow \mathbb{R}$  stetig differenzierbar mit  $f'(x) \neq 0$  für alle  $x \in U$ , dann ist  $V := f(U) \subset \mathbb{R}$  ein offenes Intervall, und es gibt eine differenzierbare Umkehrabbildung  $g : V \rightarrow U$ . Ein konkretes Beispiel dafür, dass der Umkehrsatz schon in zwei Veränderlichen nur lokal gilt, liefert die Polarkoordinatenabbildung.

**Beispiel 2.14** *Wir betrachten die stetig differenzierbare Abbildung*

$$F : ]0, \infty[ \times \mathbb{R} \rightarrow \mathbb{R}^2, \quad \begin{pmatrix} r \\ \varphi \end{pmatrix} \mapsto \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} r \cdot \cos(\varphi) \\ r \cdot \sin(\varphi) \end{pmatrix}.$$

Ihre Funktionalmatrix ist

$$F'(r, \varphi) = \begin{pmatrix} \cos(\varphi) & -r \sin(\varphi) \\ \sin(\varphi) & r \cos(\varphi) \end{pmatrix}$$

mit

$$\det(F') = r(\cos^2(\varphi) + \sin^2(\varphi)) = r \neq 0.$$

Lokal ist diese Abbildung umkehrbar, z.B. auf der rechten Halbebene durch

$$r = \sqrt{x^2 + y^2}, \quad \varphi = \operatorname{arctg}(y/x).$$

Aber weil der Winkel  $\varphi$  durch  $(\mathbf{x}, \mathbf{y}) \neq \mathbf{0}$  nur bis auf ganzzahlige Vielfache von  $2\pi$  bestimmt ist, ist die Abbildung nicht global umkehrbar.

**Definition 2.7** *Die Determinante der  $n \times n$ -Funktionalmatrix  $F'(\mathbf{x})$  heißt Funktionaldeterminante.*

**Beispiel 2.15** Die Funktionaldeterminante für den Übergang

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} r \cdot \cos(\varphi)\cos(\theta) \\ r \cdot \sin(\varphi)\cos(\theta) \\ r \cdot \sin(\theta) \end{pmatrix}$$

in Kugelkoordinaten ist

$$\det \frac{\partial(x, y, z)}{\partial(r, \varphi, \theta)} = \begin{pmatrix} \cos(\varphi)\cos(\theta) & -r\sin(\varphi)\cos(\theta) & -r\cos(\varphi)\sin(\theta) \\ \sin(\varphi)\cos(\theta) & r\cos(\varphi)\cos(\theta) & -r\sin(\varphi)\sin(\theta) \\ \sin(\theta) & 0 & r\cos(\theta) \end{pmatrix} = r^2 \cdot \cos(\theta).$$

Den Beweis des Umkehrsatzes 2.8 geben wir in mehreren Schritten. Es genügt, die folgende abgepeckte Version des Umkehrsatzes zu beweisen:

**Satz 2.9** Unter der Voraussetzung von Satz 2.8 gibt es Radien  $r > 0$  und  $s > 0$  derart, dass

1) die abgeschlossene Kugel

$$\bar{U} = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x} - \mathbf{x}_0\| \leq r\}$$

ganz in  $U$  enthalten ist und

2) die abgeschlossene Kugel

$$\bar{V} = \{\mathbf{y} \in \mathbb{R}^n : \|\mathbf{y} - \mathbf{y}_0\| \leq s\}$$

ganz in der Bildmenge  $F(\bar{U})$  enthalten ist,

sowie eine stetige Abbildung  $G: \bar{V} \rightarrow \bar{U}$  mit

3)  $F(G(\mathbf{y})) = \mathbf{y}$  für alle  $\mathbf{y} \in \bar{V}$ .

Zunächst beweisen wir, dass der Umkehrsatz 2.8 aus Satz 2.9 folgt: Dazu ersetzen wir die abgeschlossenen Kugeln  $\bar{U}, \bar{V}$  durch die offenen Kugeln

$$U_1 := \{\|\mathbf{x} - \mathbf{x}_0\| < r\}, \quad V_1 := \{\|\mathbf{y} - \mathbf{y}_0\| < s\}.$$

Die Urbildmenge

$$V_0 := \{\mathbf{y} \in V_1 : G(\mathbf{y}) \in U_1\}$$

enthält  $\mathbf{y}_0 = F(\mathbf{x}_0)$  und ist nach Satz 2.1 offen. Die Urbildmenge

$$U_0 := \{\mathbf{x} \in U_1 : F(\mathbf{x}) \in V_0\}$$

enthält  $\mathbf{x}_0 = G(\mathbf{y}_0)$  und ist nach Satz 2.1 ebenfalls offen. Dann sind die Abbildungen  $F|U_0$  und  $G|V_0$  Umkehrabbildungen voneinander.

Es bleibt zu beweisen, dass  $G|V_0$  nicht nur stetig, sondern sogar differenzierbar ist. Die Differenzierbarkeit von  $G$  ist in jedem Punkt  $\mathbf{y} \in V_0$  zu zeigen. Aber mit  $\mathbf{x} := G(\mathbf{y}) \in U_0$  haben wir für  $\mathbf{x}$  und  $\mathbf{y}$  genau die Situation der Punkte  $\mathbf{x}_0$  und  $\mathbf{y}_0$ . Es genügt also zu zeigen, dass  $G$  in  $\mathbf{y}_0$  differenzierbar ist.

Für  $\mathbf{x} \in U_0$  setzen wir  $\mathbf{y} = F(\mathbf{x})$ , bzw.  $\mathbf{x} = G(\mathbf{y})$  für  $\mathbf{y} \in V_0$ . Die Differenzierbarkeit von  $F$  in  $\mathbf{x}_0$  bedeutet

$$\mathbf{y} = \mathbf{y}_0 + F'(\mathbf{x}_0) \cdot (\mathbf{x} - \mathbf{x}_0) + \Phi(\mathbf{x} - \mathbf{x}_0) \quad \text{mit} \quad \lim_{\mathbf{x}_0 \neq \mathbf{x} \rightarrow \mathbf{x}_0} \frac{\Phi(\mathbf{x} - \mathbf{x}_0)}{\|\mathbf{x} - \mathbf{x}_0\|} = \mathbf{0},$$

bzw.

$$\mathbf{x} = \mathbf{x}_0 + F'(\mathbf{x}_0)^{-1}(\mathbf{y} - \mathbf{y}_0) - F'(\mathbf{x}_0)^{-1} \cdot \Phi(\mathbf{x} - \mathbf{x}_0).$$

Wegen

$$\| F'(\mathbf{x}_0)^{-1} \cdot \Phi(\mathbf{x} - \mathbf{x}_0) \| \leq \| F'(\mathbf{x}_0)^{-1} \| \cdot \| \Phi(\mathbf{x} - \mathbf{x}_0) \|^2$$

genügt es zu zeigen, dass

$$\lim_{\mathbf{y}_0 \neq \mathbf{y} \rightarrow \mathbf{y}_0} \frac{\Phi(\mathbf{x} - \mathbf{x}_0)}{\| \mathbf{y} - \mathbf{y}_0 \|} = \mathbf{0}.$$

Aber mit  $\mathbf{y} \neq \mathbf{y}_0$  ist auch  $\mathbf{x} \neq \mathbf{x}_0$  und wir können schreiben

$$\frac{\Phi(\mathbf{x} - \mathbf{x}_0)}{\| \mathbf{y} - \mathbf{y}_0 \|} = \frac{\Phi(\mathbf{x} - \mathbf{x}_0)}{\| \mathbf{x} - \mathbf{x}_0 \|} \cdot \frac{\| \mathbf{x} - \mathbf{x}_0 \|}{\| \mathbf{y} - \mathbf{y}_0 \|}.$$

Für  $\mathbf{x}_0 \neq \mathbf{x} \rightarrow \mathbf{x}_0$  geht hier der erste Faktor gegen 0. Es genügt also, zu zeigen, dass dabei der zweite Faktor nach oben beschränkt, bzw.

$$\frac{\| F(\mathbf{x}) - F(\mathbf{x}_0) \|}{\| \mathbf{x} - \mathbf{x}_0 \|}$$

nach unten beschränkt bleibt, d.h.  $> q > 0$ .

Mit der Minus-Dreiecksungleichung ist

$$\| F(\mathbf{x}) - F(\mathbf{x}_0) \| \geq \| F'(\mathbf{x}_0) \cdot (\mathbf{x} - \mathbf{x}_0) \| - \| \Phi(\mathbf{x} - \mathbf{x}_0) \|.$$

Auf der kompakten Menge  $\{\mathbf{h} \in \mathbb{R}^n : \|\mathbf{h}\| = 1\}$  nimmt die stetige Funktion  $\| F'(\mathbf{x}_0) \cdot \mathbf{h} \|^2$  ihr Minimum

$$m := \min_{\|\mathbf{h}\|=1} \| F'(\mathbf{x}_0) \cdot \mathbf{h} \|^2$$

an. Weil  $F'(\mathbf{x}_0)$  invertierbar vorausgesetzt ist, gilt  $m > 0$ . Für alle  $\mathbf{x} \neq \mathbf{x}_0$  folgt daraus

$$\| F'(\mathbf{x}_0) \cdot (\mathbf{x} - \mathbf{x}_0) \| \geq m \cdot \| \mathbf{x} - \mathbf{x}_0 \|$$

und

$$\frac{\| F'(\mathbf{x}_0) \cdot (\mathbf{x} - \mathbf{x}_0) \|}{\| \mathbf{x} - \mathbf{x}_0 \|} \geq m.$$

Ist außerdem  $\mathbf{x} - \mathbf{x}_0$  klein genug, so ist

$$\frac{\| \Phi(\mathbf{x} - \mathbf{x}_0) \|}{\| \mathbf{x} - \mathbf{x}_0 \|} < \frac{m}{2}.$$

Daraus folgt die Beschränktheit nach unten mit

$$q = m - \frac{m}{2} = \frac{m}{2}. \quad \square$$

Zum Beweis von Satz 2.9 möchte ich hier völlig neues Beweisprinzip einführen, das sich z.B. auch auf Differentialgleichungen anwenden lässt. Ausgangspunkt ist das Newton-Verfahren zur numerischen Approximation von Nullstellen differenzierbarer Funktionen  $f(x)$  in einer Veränderlichen.

Sei etwa  $f$  stetig differenzierbar auf einem offenen Intervall, welches das Intervall  $[a, b] \subset \mathbb{R}$  enthält. Es gelte  $f(a) < 0$  und  $f(b) > 0$ . Weil  $f$  stetig ist, gibt es nach dem Zwischenwertsatz im Intervall  $[a, b]$  eine Nullstelle von  $f$ . Wenn wir zusätzlich  $f' > 0$  voraussetzen, ist  $f$  streng monoton und die

Nullstelle eindeutig bestimmt. Die Idee ist, mit einem geeigneten Punkt  $x_0 \in [a, b]$  zu beginnen und dort die Funktion linear zu approximieren. D.h., wir ersetzen die Funktion  $f$  durch ihre Tangente im Punkt  $(x_0, f(x_0))$  mit der Gleichung

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

Diese Tangente schneidet die  $x$ -Achse im Punkt  $x_1$  mit

$$y(x_1) = f(x_0) + f'(x_0) \cdot (x_1 - x_0) = 0.$$

Für dieses  $x_1$  gilt

$$f'(x_0) \cdot (x_1 - x_0) = -f(x_0), \quad \text{bzw.} \quad x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}.$$

Vermöge des Iterationsverfahrens

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}$$

hofft man, die Nullstelle von  $f$  zu approximieren. Natürlich kann man numerische Bedingungen an  $f$  formulieren, die dies garantieren. Ich möchte dies hier nicht tun, sondern das einfachste nicht ganz triviale Beispiel numerisch vorstellen.

**Beispiel 2.16 (Approximation von  $\sqrt{2}$ )** *Betrachtet werde die Funktion*

$$f(x) = x^2 - 2 \quad \text{mit} \quad f'(x) = 2x.$$

*Als Startpunkt wählen wir  $x_0 = 2$ . Die Iterationsformel liefert mit dem Taschenrechner*

$k$	$x_k$	$f(x_k)$	$f'(x_k)$	$f(x_k)/f'(x_k)$
0	2	2	4	0.5
1	1.5	0.25	3	0.083333
2	1.416667	0.006945	2.833334	0.002451
3	1.414215			

*Nach drei Iterationsschritten haben wir die Nullstelle*

$$\sqrt{2} = 1.4142135\dots$$

*auf fünf Stellen genau berechnet.*

Für meine Zwecke ist aber nicht die Geschwindigkeit der Approximation entscheidend, sondern die Übersichtlichkeit des Verfahrens. Deswegen verwende ich das *vereinfachte Newton-Verfahren*. Dabei wird  $f'(x_k)$  nicht bei jedem Schritt neu berechnet, sondern stets durch  $f'(x_0)$  ersetzt. Die Iterationsformel verändert sich dann in

$$x_{k+1} := x_k - \frac{f(x_k)}{f'(x_0)}.$$

**Beispiel 2.17** Wie eben betrachten wir die Funktion  $f(x) = x^2 - 2$  mit dem Startpunkt  $x_0 = 2$ . Jetzt verwenden wir  $f'(x_0) = 4$  und berechnen

$k$	$x_k$	$f(x_k)$	$f(x_k)/4$
0	2	2	0.5
1	1.5	0.25	0.0625
2	1.4375	0.066406	0.016602
3	1.420898	0.018951	0.004738
4	1.416160	0.005509	0.001377
5	1.414783		

Auch nach fünf Iterationsschritten ist das Ergebnis nur auf drei Stellen genau.

Wesentlich ist für meine Zwecke allerdings eine Garantie dafür, dass das Verfahren konvergiert. Dazu ist ein gewisser Aufwand an neuen Begriffen erforderlich.

**Definition 2.8 (Metrischer Raum)** Ein metrischer Raum ist eine Menge  $X$  zusammen mit einer Funktion (Metrik)

$$d : X \times X \rightarrow \mathbb{R},$$

welche folgende Eigenschaften hat:

M1: Es ist stets  $d(x, y) \geq 0$  und  $d(x, y) = 0$  nur, wenn  $x = y$ .

M2: Es ist  $d(x, y) = d(y, x)$  (Symmetrie).

M3: Es gilt die Dreiecksungleichung

$$d(x, z) \leq d(x, y) + d(y, z) \quad \text{für alle } x, y, z \in X.$$

**Beispiel 2.18** Auf dem Raum  $\mathbb{R}^n$  definiert die euklidische Norm eine Metrik vermöge

$$d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|.$$

Aber diese Metrik versieht auch jede Teilmenge  $A \subset \mathbb{R}^n$  mit der Struktur eines metrischen Raums.

Wegen Beispiel 2.18 führt man natürlich nicht den Begriff des metrischen Raums ein, sondern vor allem wegen

**Beispiel 2.19** Es sei  $A \subset \mathbb{R}^n$  kompakt. Wir betrachten den  $\mathbb{R}$ -Vektorraum  $C^0(A)$  der stetigen Funktionen  $f : A \rightarrow \mathbb{R}$ . Jede stetige Funktion  $f$  auf  $A$  nimmt auf  $A$  ihr Maximum an und deswegen ist

$$\|f\|_A := \max_{\mathbf{x} \in A} |f(\mathbf{x})|$$

wohldefiniert. Wir setzen

$$d(f, g) := \|f - g\|_A.$$

Damit wird  $C^0(A)$  ein metrischer Raum. Der Beweis der Eigenschaften M1, M2, M3 ist eine langweilige Übungsaufgabe und nur Mathematikern zumutbar.

Mit  $C^0(A, \mathbb{R}^k)$  bezeichnet man den Vektorraum der stetigen Abbildungen  $F : A \rightarrow \mathbb{R}^k$ . Eine Norm auf diesem Vektorraum ist

$$\|F\|_A := \max_{\mathbf{x} \in A} \|F(\mathbf{x})\|.$$

Die zugehörige Metrik macht  $C^0(A, \mathbb{R}^k)$  zu einem metrischen Raum.

In einem metrischen Raum  $(X, d)$  kann man wie üblich Konvergenz definieren: Eine Folge  $x_\nu$  in  $X$  konvergiert gegen  $x \in X$ , wenn gilt: Zu jedem  $\epsilon > 0$  existiert ein  $N(\epsilon) \in \mathbb{N}$  so, dass für alle  $\nu > N(\epsilon)$  gilt:  $d(x_\nu, x) < \epsilon$ .

Und auch den Begriff der Cauchy-Folge kann man wie üblich erklären: Die Folge  $x_\nu$  heißt Cauchy-Folge, wenn zu jedem  $\epsilon > 0$  ein  $N(\epsilon) \in \mathbb{N}$  existiert so, dass für  $\mu, \nu > N(\epsilon)$  gilt:  $d(x_\mu, x_\nu) < \epsilon$ .

**Definition 2.9** *Der metrische Raum  $X, d$  heißt vollständig, wenn jede Cauchy-Folge in  $X$  konvergiert.*

**Beispiel 2.20** *Der Raum  $\mathbb{R}^n$  mit der Metrik  $d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|$  ist vollständig. Auch jede abgeschlossene Menge  $A \subset \mathbb{R}^n$  mit dieser Metrik ist vollständig.*

Die einzige nicht-triviale Tatsache in dieser abstrakt-leeren Umgebung ist folgender Satz:

**Satz 2.10** *Es sei  $A \subset \mathbb{R}^n$  kompakt und  $d$  die in Beispiel 2.19 definierte Metrik auf  $C^0(A)$ . Dann ist  $C^0(A)$  mit der Metrik  $d$  vollständig.*

Beweis. Wir betrachten eine Cauchyfolge  $f_\nu$  in  $C^0(A)$ . Für jeden Punkt  $\mathbf{x} \in A$  bilden die Zahlen  $f_\nu(\mathbf{x})$  eine Cauchy-Folge in  $\mathbb{R}$  und konvergieren. Wir definieren eine Funktion  $f : A \rightarrow \mathbb{R}$  durch

$$f(\mathbf{x}) := \lim_{\nu \rightarrow \infty} f_\nu(\mathbf{x}).$$

Es sei  $N(\epsilon)$  für die Cauchy-Folge  $f_\nu$  wie in der obigen Definition gewählt. Wenn  $\nu \geq N(\epsilon)$  ist, dann gilt für alle  $\mu \geq \nu$  und für alle  $\mathbf{x} \in A$

$$|f_\nu(\mathbf{x}) - f_\mu(\mathbf{x})| \leq \|f_\nu - f_\mu\| < \epsilon.$$

Daraus folgt

$$|f_\nu(\mathbf{x}) - f(\mathbf{x})| = |f_\nu(\mathbf{x}) - \lim_{\mu} f_\mu(\mathbf{x})| \leq \epsilon.$$

Die Funktionenfolge  $f_\nu$  konvergiert *gleichmäßig* gegen die Funktion  $f$ . Wie in der Analysis I folgt daraus die Stetigkeit der Funktion  $f$ . Also gehört die Grenzfunktion  $f$  zu  $C^0(A)$ . Und für  $\nu \geq N(\epsilon)$  ist

$$\|f_\nu - f\|_A \leq \epsilon. \quad \square$$

Mit  $C^0(A)$  ist auch  $C^0(A, \mathbb{R}^k)$  vollständig, ebenso wie für jede abgeschlossene Teilmenge  $B \subset \mathbb{R}^k$  der Unterraum (kein Vektorraum!)

$$C^0(A, B) = \{F : A \rightarrow B \text{ stetig}\}.$$

**Definition 2.10** *Es sei  $(M, d)$  ein metrischer Raum. Eine Abbildung  $P : M \rightarrow M$  heißt kontrahierend, wenn eine Konstante  $L < 1$  existiert mit*

$$d(P(u), P(v)) \leq L \cdot d(u, v)$$

für alle  $u, v \in M$ .

**Beispiel 2.21** *Es sei  $U \subset \mathbb{R}^n$  offen und  $F : U \rightarrow U$  differenzierbar. Es gebe ein  $L < 1$  mit  $\|F'(\mathbf{x})\| \leq L$  für alle  $\mathbf{x} \in U$ . Nach Satz 2.7 ist eine solche Abbildung kontrahierend.*

**Definition 2.11** Ein Punkt  $\mathbf{x} \in M$  heißt Fixpunkt einer Abbildung  $P : M \rightarrow M$ , wenn  $P(\mathbf{x}) = \mathbf{x}$ .

**Beispiel 2.22** Die Funktion

$$f : \mathbb{R} \rightarrow \mathbb{R}, \quad x \mapsto x^2$$

hat genau die beiden Fixpunkte  $x = 0$  und  $x = 1$ .

**Satz 2.11 (Fixpunktsatz von Banach)** Es sei  $(M, d)$  ein vollständiger metrischer Raum und  $P : M \rightarrow M$  kontrahierend. Dann besitzt  $P$  genau einen Fixpunkt  $u \in M$ .

Beweis. a) Eindeutigkeit: Es seien  $u \neq v \in M$  zwei Fixpunkte von  $P$ . Dann ist also

$$d(u, v) = d(P(u), P(v)) \leq L \cdot d(u, v) < d(u, v)$$

wegen  $d(u, v) > 0$ . Das ist ein Widerspruch.

b) Existenz: Wir wählen einen beliebigen Punkt  $u_0 \in M$  und betrachten die rekursiv definierte Folge

$$u_1 := P(u_0), u_2 := P(u_1), \dots, u_{n+1} := P(u_n), \dots$$

in  $M$ . Wir kürzen ab  $d := d(u_1, u_0)$ . Durch Induktion zeigt man

$$d(u_2, u_1) \leq L \cdot d, \dots, d(u_{n+1}, u_n) \leq L^n \cdot d.$$

Mit der geometrischen Summenformel folgt daraus für  $m < n$

$$d(u_n, u_m) \leq d(u_{m+1}, u_m) + \dots + d(u_n, u_{n-1}) \leq (L^m + \dots + L^{n-1}) \cdot d = L^m \cdot \frac{1 - L^{n-m}}{1 - L} \cdot d.$$

Wegen  $0 \leq L < 1$  ist  $0 < 1 - L^{n-m} \leq 1$  und

$$L^m \cdot \frac{1 - L^{n-m}}{1 - L} \leq \frac{L^m}{1 - L}.$$

Wir haben bewiesen:

$$d(u_n, u_m) \leq L^m \cdot \frac{d}{1 - L}.$$

Deswegen ist die Folge  $(u_n)$  eine Cauchyfolge in  $M$  und konvergiert gegen einen Grenzwert  $u \in M$ .

Bei vorgegebenem  $\epsilon > 0$  ist  $d(u_n, u) < \epsilon$ , wenn  $n$  groß genug ist. Aber dann ist auch

$$d(P(u_n), P(u)) \leq L \cdot d(u_n, u) < L \cdot \epsilon < \epsilon.$$

Die Folge  $P(u_n)$  konvergiert gegen  $P(u) \in M$ . Aber als Teilfolge der Folge  $(u_n)$  hat die Folge  $(P(u_n))$  den gleichen Grenzwert  $u$ . Damit ist

$$P(u) = \lim P(u_n) = \lim u_n = u$$

der gesuchte Fixpunkt. □

Dieser Banachsche Fixpunktsatz wird auf das vereinfachte Newtonverfahren folgendermaßen angewendet, um die Umkehrfunktion zu konstruieren: Es seien

$$\bar{U} = [x_0 - r, x_0 + r], \bar{V} = [y_0 - s, y_0 + s] \subset \mathbb{R}$$

abgeschlossene endliche Intervalle und  $f : \bar{U} \rightarrow \bar{V}$  stetig differenzierbar (auf einem offenen Intervall, welches  $\bar{U}$  enthält). Es gelte  $f(x_0) = y_0$  mit  $f'(x_0) \neq 0$ . Die Umkehrfunktion  $g = f^{-1}$  ordnet jedem  $y \in \bar{V}$  den Wert  $x$  mit  $f(x) = y$  zu. Diesen Wert  $y$  approximieren wir mit dem vereinfachten Newton-Verfahren, indem wir die Nullstelle der Funktion  $f - y$  suchen. Dazu starten wir mit der Funktion  $g_0(y) := x_0$  und iterieren mit der Formel

$$g_{k+1}(y) := g_k(y) - \frac{f(g_k(y)) - y}{f'(x_0)}.$$

Diese Iterationsformel definiert die Abbildung

$$P : C^0(\bar{V}) \rightarrow C^0(\bar{V}), \quad P(g)(y) := g(y) - \frac{(f \circ g)(y) - y}{f'(x_0)}.$$

Unter geeigneten Voraussetzungen gelten:

- 1) Die Abbildung  $P$  ist kontrahierend (bezüglich der Norm  $\|g\|_{\bar{V}}$ ),
- 2) Wenn  $g$  das Intervall  $\bar{V}$  nach  $\bar{U}$  abbildet, dann tut dies  $P(g)$  ebenso.

Diese geeigneten Voraussetzungen möchte ich erst später bei der Formulierung der allgemeinen Aussage im  $\mathbb{R}^n$  angeben. Hier zunächst noch ein einfaches eindimensionales Beispiel.

**Beispiel 2.23** Wir approximieren die Umkehrfunktion der Funktion  $f(x) = x^2 - 1$ , d.h., die Funktion  $g(y) = \sqrt{y+1}$ . Wir wählen  $x_0 = 1$  mit  $y_0 = 0$  und  $f'(x_0) = 2$ . Dann haben wir also

$$g_0 \equiv x_0 = 1$$

und die Rekursionsformel

$$g_{k+1}(y) := g_k(y) - \frac{g_k^2(y) - 1 - y}{2}.$$

Für  $k = 1, 2, 3$  erhalten wir so die Funktionen

$$\begin{aligned} g_1(y) &= 1 - \frac{1^2 - 1 - y}{2} \\ &= 1 + \frac{y}{2} \\ g_2(y) &= 1 + \frac{y}{2} - \frac{(1 + \frac{y}{2})^2 - 1 - y}{2} \\ &= 1 + \frac{y}{2} - \frac{y^2}{8} \\ g_3(y) &= 1 + \frac{y}{2} - \frac{y^2}{8} - \frac{1}{2} \cdot \left( \left( 1 + \frac{y}{2} - \frac{y^2}{8} \right)^2 - 1 - y \right) \\ &= 1 + \frac{y}{2} - \frac{y^2}{8} + \frac{y^3}{16} - \frac{y^4}{128} \end{aligned}$$

Zum Vergleich: Die Taylorentwicklung der Umkehrfunktion um  $y = 0$  lautet

$$\sqrt{1+y} = \sum_k \binom{1/2}{k} y^k = 1 + \frac{y}{2} - \frac{y^2}{8} + \frac{y^3}{16} - \frac{5y^4}{128} \pm \dots$$

Nun zum Beweis von Satz 2.9.

Präzisieren wir zunächst die Definitionsbereiche unserer Abbildungen.

1) Wir wählen  $r > 0$  so klein, dass die abgeschlossene Kugel

$$\bar{U} = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x} - \mathbf{x}_0\| \leq r\}$$

ganz im Definitionsbereich von  $F$  enthalten ist, und

$$\|\mathbb{1}_n - F'(\mathbf{x}_0)^{-1}F'(\mathbf{x})\| < \frac{1}{2}$$

gilt für alle  $\mathbf{x} \in \bar{U}$ . Das ist möglich, weil für  $\mathbf{x} = \mathbf{x}_0$  gilt  $\mathbb{1}_n = F'(\mathbf{x}_0)^{-1}F'(\mathbf{x})$  und  $F'(\mathbf{x})$  stetig ist. Weil außerdem auch  $\det F'(\mathbf{x})$  stetig ist mit  $\det F'(\mathbf{x}_0) \neq 0$ , können wir  $r$  auch noch so klein wählen, dass  $\det F'(\mathbf{x}) \neq 0$  ist für alle  $\mathbf{x} \in \bar{U}$ .

2) Jetzt setzen wir

$$s := \frac{r}{2 \|F'(\mathbf{x}_0)\|} \quad \text{und} \quad \bar{V} := \{\mathbf{y} \in \mathbb{R}^n : \|\mathbf{y} - \mathbf{y}_0\| \leq s\}.$$

Nach Satz 2.10 ist der metrische Raum

$$C^0(\bar{V}, \bar{U}) = \{G : \bar{V} \rightarrow \bar{U}, G \text{ stetig}\}$$

vollständig. Wir definieren die Abbildung  $P : C^0(\bar{V}, \bar{U}) \rightarrow C^0(\bar{V}, \mathbb{R}^n)$  durch

$$P(G)(\mathbf{y}) := G(\mathbf{y}) - F'(\mathbf{x}_0)^{-1} \cdot (F(G(\mathbf{y})) - \mathbf{y}).$$

Als erstes zeigen wir, dass  $P$  kontrahierend ist. Seien dazu  $G_1, G_2 \in C^0(\bar{V}, \bar{U})$ . Für alle  $\mathbf{y} \in \bar{V}$  haben wir

$$\begin{aligned} P(G_1)(\mathbf{y}) - P(G_2)(\mathbf{y}) &= G_1(\mathbf{y}) - G_2(\mathbf{y}) - F'(\mathbf{x}_0)^{-1} \cdot (F(G_1(\mathbf{y})) - F(G_2(\mathbf{y}))) \\ &= \left( id_{\bar{V}} - F'(\mathbf{x}_0)^{-1}F \right) (G_1(\mathbf{y})) - \left( id_{\bar{V}} - F'(\mathbf{x}_0)^{-1}F \right) (G_2(\mathbf{y})). \end{aligned}$$

Die Ableitung der Abbildung

$$id_U - F'(\mathbf{x}_0)^{-1}F : U \rightarrow \mathbb{R}^n$$

in einem Punkt  $\mathbf{x} \in \bar{U}$  ist

$$\mathbb{1}_n - F'(\mathbf{x}_0)^{-1}F'(\mathbf{x}).$$

Mit Satz 2.7 folgt daraus

$$\|P(G_1)(\mathbf{y}) - P(G_2)(\mathbf{y})\| \leq \|\mathbb{1}_n - F'(\mathbf{x}_0)^{-1}F'(\mathbf{x})\|_{\bar{U}} \cdot \|G_1(\mathbf{y}) - G_2(\mathbf{y})\| \leq \frac{1}{2} \|G_1(\mathbf{y}) - G_2(\mathbf{y})\|.$$

Also ist  $P$  kontrahierend mit  $L = \frac{1}{2}$ .

Wir müssen noch zeigen, dass  $P$  eine Abbildung  $C^0(\bar{V}, \bar{U}) \rightarrow C^0(\bar{V}, \bar{U})$  definiert. Sei also  $G \in C^0(\bar{V}, \bar{U})$ , d.h.,  $G : \bar{V} \rightarrow \bar{U}$  ist stetig mit  $\|G(\mathbf{y}) - \mathbf{x}_0\| \leq r$  für alle  $\mathbf{y} \in \bar{V}$ . Für die folgende Abschätzung fassen wir  $\mathbf{x}_0$  auf als konstante Abbildung in  $C^0(\bar{V}, \bar{U})$ . Dann gilt für alle  $\mathbf{y} \in \bar{V}$

$$\|P(G)(\mathbf{y}) - \mathbf{x}_0\| \leq \|P(G)(\mathbf{y}) - P(\mathbf{x}_0)\| + \|P(\mathbf{x}_0) - \mathbf{x}_0\|.$$

Mit der Kontraktionseigenschaft können wir den ersten Summanden abschätzen

$$\|P(G)(\mathbf{y}) - P(\mathbf{x}_0)\| \leq \frac{1}{2} \|G(\mathbf{y}) - \mathbf{x}_0\| \leq \frac{r}{2}.$$

Und im zweiten Summanden haben wir

$$(P(\mathbf{x}_0) - \mathbf{x}_0)(\mathbf{y}) = \mathbf{x}_0 - F'(\mathbf{x}_0)^{-1}(F(\mathbf{x}_0) - \mathbf{y}) - \mathbf{x}_0 = F'(\mathbf{x}_0)^{-1}(\mathbf{y}_0 - \mathbf{y}).$$

Daraus folgt

$$\|P(\mathbf{x}_0)(\mathbf{y}) - \mathbf{x}_0\| \leq \|F'(\mathbf{x}_0)^{-1}\| \cdot s = \frac{r}{2}.$$

Insgesamt haben wir

$$\|P(G)(\mathbf{y}) - \mathbf{x}_0\| \leq \frac{r}{2} + \frac{r}{2} = r.$$

Der Banachsche Fixpunktsatz liefert eine stetige Abbildung  $G: \bar{V} \rightarrow \bar{U}$  mit  $F(G(\mathbf{y})) = \mathbf{y}$  für alle  $\mathbf{y} \in \bar{V}$ .  $\square$

Im Beweis haben wir gesehen

$$G'(\mathbf{y}_0) = F'(\mathbf{x}_0)^{-1}.$$

Die Matrix-Einträge der Umkehrmatrix

$$F'(\mathbf{x}_0)^{-1} = \frac{1}{\det F'(\mathbf{x}_0)} \cdot F'(\mathbf{x}_0)^{adj}$$

erhält man aus denen der Matrix  $F'(\mathbf{x}_0)$  durch Bildung von Unterdeterminanten und anschließender Division durch  $\det(F'(\mathbf{x}_0)) \neq 0$ . Dies zeigt

**Satz 2.12 (Zusatz)** Die Umkehrabbildung  $G$  einer  $k$ -mal (stetig) differenzierbaren Abbildung  $F$  ist auch wieder  $k$ -mal (stetig) differenzierbar.

**Aufgabe 2.21** Zeigen Sie, dass die Abbildung

$$F: \mathbb{R}^2 \rightarrow \mathbb{R}^2, \quad \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \frac{1}{2} \begin{pmatrix} 1 - y \\ 2 + x \end{pmatrix}$$

kontrahierend ist und berechnen Sie ihren Fixpunkt.

**Aufgabe 2.22** Bestimmen Sie maximale Teilintervalle  $I \subset \mathbb{R} \setminus \{0\}$ , auf denen die Funktion

$$f(x) := 1 + \frac{1}{x}$$

kontrahierend ist. In welchem dieser Intervalle besitzt  $f$  welchen Fixpunkt?

**Aufgabe 2.23** Wo ist die Abbildung

$$F: \mathbb{R}^2 \rightarrow \mathbb{R}^2, \quad \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} x^2 - y^2 \\ 2xy \end{pmatrix}$$

lokal umkehrbar? Bestimmen Sie eine lokale Umkehrung  $F^{-1}: x = x(u, v), y = y(u, v)$  bei  $(u, v) = (1, 0)$  derart, dass  $F^{-1}(1, 0) = (1, 0)$ .

**Aufgabe 2.24** Es sei

$$F : \mathbb{R}^2 \rightarrow \mathbb{R}^2, \quad \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} e^x \cos(y) \\ e^x \sin(y) \end{pmatrix}.$$

a) Berechnen Sie die Funktionalmatrix  $F'(x, y)$ .

b) Zeigen Sie:  $F$  ist bei  $(x, y)$  lokal umkehrbar genau dann, wenn  $y \notin \{\pi/2 + \pi\mathbb{Z}\}$ .

c) Bestimmen Sie die Ableitung der lokalen Inversen von  $F$  um  $(0, 0)$  im Punkt  $F(0, 0)$ .

**Aufgabe 2.25** Es sei

$$F : \mathbb{R}^2 \rightarrow \mathbb{R}^2, \quad \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} (x+y)^3 \\ (x-y)^7 \end{pmatrix}.$$

a) Zeigen Sie, dass  $F$  injektiv ist.

b) Bei welchen Punkten  $(u, v)$  existiert lokal eine differenzierbare Umkehrabbildung zu  $F$ ?

## 2.6 Implizite Funktionen

Ein Spezialfall des Umkehrsatzes ist ein lineares Gleichungssystem

$$A \cdot \mathbf{x} = \mathbf{b}$$

mit invertierbarer  $n \times n$ -Koeffizientenmatrix  $A$ . Zu jedem  $\mathbf{b} \in \mathbb{R}^n$  hat dieses System die eindeutig bestimmte Lösung  $\mathbf{x} = A^{-1} \cdot \mathbf{b}$ . Die Abbildung  $F : \mathbf{x} \mapsto A \cdot \mathbf{x}$  hat die Umkehrabbildung  $G : \mathbf{b} \mapsto A^{-1} \cdot \mathbf{b}$ . Im Unterschied zu Satz 2.8 ist hier allerdings die Umkehrabbildung global definiert.

In der Linearen Algebra betrachtet man auch die allgemeinere Situation, wo die Anzahl der Unbekannten größer ist als die Anzahl der Gleichungen. Betrachten wir etwa eine  $p \times (n+p)$ -Matrix  $A$  und das System

$$A \cdot \mathbf{x} = \mathbf{0}.$$

Dieses System löst man, indem man die Koeffizientenmatrix auf Zeilenstufenform bringt. In der Linearen Algebra versucht man, die Stufen immer möglichst weit links zu erzeugen. Aber wenn  $A$  den maximalen Rang  $p$  besitzt, dann kann man die Stufen innerhalb jeder  $p \times p$ -Teilmatrix erzeugen, welche diesen maximalen Rang  $p$  hat. Normieren wir die Situation folgendermaßen:

Es sei  $A = (A_1, A_2)$  eine  $p \times (n+p)$ -Matrix mit einer  $p \times n$ -Matrix  $A_1$  und einer invertierbaren  $p \times p$ -Matrix  $A_2$ . Schreiben wir den Vektor der Unbekannten als

$$(\mathbf{x}, \mathbf{y}) \quad \text{mit} \quad \mathbf{x} \in \mathbb{R}^n, \mathbf{y} \in \mathbb{R}^p,$$

so lautet das Gleichungssystem

$$A_1 \cdot \mathbf{x} + A_2 \cdot \mathbf{y} = \mathbf{0}.$$

Es wird gelöst durch

$$\mathbf{y} = -A_2^{-1} \cdot A_1 \cdot \mathbf{x}, \quad \mathbf{x} \in \mathbb{R}^n \text{ beliebig.}$$

Dies können wir auch so formulieren:

Gegeben sei eine Abbildung

$$F : \mathbb{R}^{n+p} \rightarrow \mathbb{R}^p, \quad (\mathbf{x}, \mathbf{y}) \mapsto A_1 \cdot \mathbf{x} + A_2 \cdot \mathbf{y}$$

mit invertierbarer  $p \times p$ -Matrix

$$A_2 = \left( \frac{\partial F_i}{\partial y_j} \right)_{i,j=1,\dots,p}.$$

Dann gibt es eine Abbildung

$$G : \mathbb{R}^n \rightarrow \mathbb{R}^p, \quad \mathbf{x} \mapsto -A_2^{-1} \cdot A_1 \cdot \mathbf{x}$$

so, dass die Lösungen des Systems genau die Vektoren  $(\mathbf{x}, G(\mathbf{x}))$  mit  $\mathbf{x} \in \mathbb{R}^n$  sind. Man sagt: Mit der Abbildung  $\mathbf{y} = G(\mathbf{x})$  haben wir das System nach  $\mathbf{y}$  aufgelöst.

Der folgende Satz verallgemeinert diese Situation *lokal* auf nicht-lineare Gleichungssysteme. Dabei schreiben wir wie eben die Vektoren im  $\mathbb{R}^{n+p}$  als  $(\mathbf{x}, \mathbf{y})$  mit  $\mathbf{x} \in \mathbb{R}^n$  und  $\mathbf{y} \in \mathbb{R}^p$ .

**Satz 2.13** *Es sei  $M \subset \mathbb{R}^{n+p}$  offen und  $F : M \rightarrow \mathbb{R}^p$  stetig differenzierbar. Gegeben sei ein Punkt  $(\mathbf{x}_0, \mathbf{y}_0) \in \mathbb{R}^{n+p}$  mit*

- 1)  $F(\mathbf{x}_0, \mathbf{y}_0) = \mathbf{0}$ ,
- 2)  $\det \left( \frac{\partial F_i}{\partial y_j}(\mathbf{x}_0, \mathbf{y}_0) \right)_{i,j=1,\dots,p} \neq 0$ .

*Dann existieren*

*eine offene Menge  $U \subset M \subset \mathbb{R}^n$  mit  $(\mathbf{x}_0, \mathbf{y}_0) \in U$ ,*

*eine offene Menge  $V \subset \mathbb{R}^n$  mit  $\mathbf{x}_0 \in V$ ,*

*und eine stetig differenzierbare Abbildung  $G : V \rightarrow \mathbb{R}^p$*

*so, dass gilt:*

$$(\mathbf{x}, \mathbf{y}) \in U \text{ mit } F(\mathbf{x}, \mathbf{y}) = \mathbf{0} \quad \Leftrightarrow \quad \mathbf{x} \in V \text{ und } \mathbf{y} = G(\mathbf{x}).$$

Vor dem Beweis erst

**Beispiel 2.24** *Wir betrachten die Funktion*

$$f : \mathbb{R}^2 \rightarrow \mathbb{R}, \quad f(x, y) = x^2 + y^2 - 1.$$

*sie ist stetig differenzierbar und hat als Nullstellenmenge*

$$\{(x, y) \in \mathbb{R}^2 : f(x, y) = 0\}$$

*den Einheitskreis. Ihre partielle Ableitung*

$$\frac{\partial f}{\partial y} = 2y$$

*ist  $\neq 0$  für  $y \neq 0$ . Fixieren wir einen Punkt  $(x_0, y_0)$  auf dem Einheitskreis mit  $y_0 > 0$ . Dort kann man die Gleichung  $f(x, y) = 0$  lokal nach  $y$  auflösen vermöge*

$$y = \sqrt{1 - x^2}.$$

*Mit der Notation von Satz 2.13 haben wir hier:*

Die Dimensionen sind  $n = p = 1$ .  
 Die offene Menge  $V \subset \mathbb{R}$  ist das Intervall  $] - 1, 1[$ .  
 Die offene Menge  $U \subset \mathbb{R}^2$  ist

$$\{(x, y) \in \mathbb{R}^2 : x \in V, y > 0\}.$$

Die Abbildung  $G : V \rightarrow \mathbb{R}^2$  ist  $x \mapsto (x, \sqrt{1 - x^2})$ .

Beweis des Satzes: Wir betrachten die Abbildung

$$\Phi : M \rightarrow \mathbb{R}^{n+p}, \quad \Phi(\mathbf{x}, \mathbf{y}) = \begin{pmatrix} \mathbf{x} \\ F(\mathbf{x}, \mathbf{y}) \end{pmatrix}.$$

Diese Abbildung ist stetig differenzierbar mit der Funktionalmatrix

$$\Phi' = \begin{pmatrix} \mathbb{1}_n & \mathbf{0} \\ \frac{\partial F}{\partial \mathbf{x}} & \frac{\partial F}{\partial \mathbf{y}} \end{pmatrix}.$$

Nach Voraussetzung 2) hat im Punkt  $(\mathbf{x}_0, \mathbf{y}_0)$  die rechte untere Teilmatrix den Maximalrang  $= p$ . Dann hat in diesem Punkt die ganze Funktionalmatrix  $\Phi'(\mathbf{x}_0, \mathbf{y}_0)$  den Rang  $n + p$ . Damit ist der Umkehrsatz anwendbar. Es gibt also offene Mengen  $U \subset M \subset \mathbb{R}^{n+p}$  mit  $(\mathbf{x}_0, \mathbf{y}_0) \in U$ , sowie  $V' \subset \mathbb{R}^{n+p}$  mit

$$\Phi(\mathbf{x}_0, \mathbf{y}_0) = (\mathbf{x}_0, F(\mathbf{x}_0, \mathbf{y}_0)) \in V',$$

und eine stetig differenzierbare Umkehrabbildung  $\Psi : V' \rightarrow U$  für  $\Phi : U \rightarrow V'$ .

Dass  $\Psi$  die Abbildung  $\Phi$  umkehrt bedeutet explizit

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \\ y_1 \\ \vdots \\ y_p \end{pmatrix} = \begin{pmatrix} \Psi_1(\mathbf{x}, F(\mathbf{x}, \mathbf{y})) \\ \vdots \\ \Psi_n(\mathbf{x}, F(\mathbf{x}, \mathbf{y})) \\ \Psi_{n+1}(\mathbf{x}, F(\mathbf{x}, \mathbf{y})) \\ \vdots \\ \Psi_{n+p}(\mathbf{x}, F(\mathbf{x}, \mathbf{y})) \end{pmatrix}.$$

Wir haben also

$$\Psi_1(\mathbf{x}, F(\mathbf{x}, \mathbf{y})) = x_1, \dots, \Psi_n(\mathbf{x}, F(\mathbf{x}, \mathbf{y})) = x_n$$

und

$$\Psi_{n+1}(\mathbf{x}, F(\mathbf{x}, \mathbf{y})) = y_1, \dots, \Psi_{n+p}(\mathbf{x}, F(\mathbf{x}, \mathbf{y})) = y_p.$$

Unter  $\Phi$  wird die Nullstellenmenge

$$\{(\mathbf{x}, \mathbf{y}) \in U : F_1(\mathbf{x}, \mathbf{y}) = \dots = F_p(\mathbf{x}, \mathbf{y}) = 0\}$$

bijektiv auf die Menge

$$V := V' \cap \{(\mathbf{x}, \mathbf{0}) \in \mathbb{R}^{n+p}\}$$

abgebildet. Wir fassen  $V$  auf als offene Menge im  $\mathbb{R}^n \subset \mathbb{R}^{n+p}$  der ersten  $n$  Koordinaten.

Die Abbildung  $G : V \rightarrow U$  definieren wir durch

$$G(\mathbf{x}) = \begin{pmatrix} G_1(\mathbf{x}) \\ \vdots \\ G_p(\mathbf{x}) \end{pmatrix} := \begin{pmatrix} \Psi_{n+1}(\mathbf{x}, \mathbf{0}) \\ \vdots \\ \Psi_{n+p}(\mathbf{x}, \mathbf{0}) \end{pmatrix}.$$

Diese Abbildung  $G$  ist stetig differenzierbar, weil  $\Psi$  dies ist. Und sie bildet  $V$  bijektiv auf die Nullstellenmenge  $\{(\mathbf{x}, \mathbf{y}) \in U : F(\mathbf{x}, \mathbf{y}) = \mathbf{0}\}$  ab.  $\square$

Wir wollen die Aussage von Satz 2.13 noch etwas weiter analysieren. Mit den Komponentenfunktionen  $F_1, \dots, F_k$  von  $F$  ist die Bedingung  $F(\mathbf{x}, \mathbf{y}) = \mathbf{0}$  das *nicht-lineare* Gleichungssystem

$$\begin{aligned} F_1(x_1, \dots, x_n, y_1, \dots, y_p) &= 0 \\ &\vdots \\ F_p(x_1, \dots, x_n, y_1, \dots, y_p) &= 0 \end{aligned}$$

Durch die Abbildung  $\mathbf{y} = G(\mathbf{x})$ , bzw. durch deren Komponentenfunktionen

$$\begin{aligned} y_1 &= G_1(x_1, \dots, x_n) \\ &\vdots \\ y_p &= G_p(x_1, \dots, x_n) \end{aligned}$$

wird dieses System lokal *nach  $y$  aufgelöst*. Die Funktionen  $y_1(\mathbf{x}) = G_1(\mathbf{x}), \dots, y_p(\mathbf{x}) = G_p(\mathbf{x})$  sind implizit durch die Gleichungen  $F_1(\mathbf{x}, \mathbf{y}) = \dots = F_p(\mathbf{x}, \mathbf{y}) = 0$  definiert.

Der einfachste Fall liegt natürlich dann vor, wenn wir eine Funktion  $f(x, y)$  von zwei Variablen haben.

**Definition 2.12** Für  $c \in \mathbb{R}$  heißen die Mengen

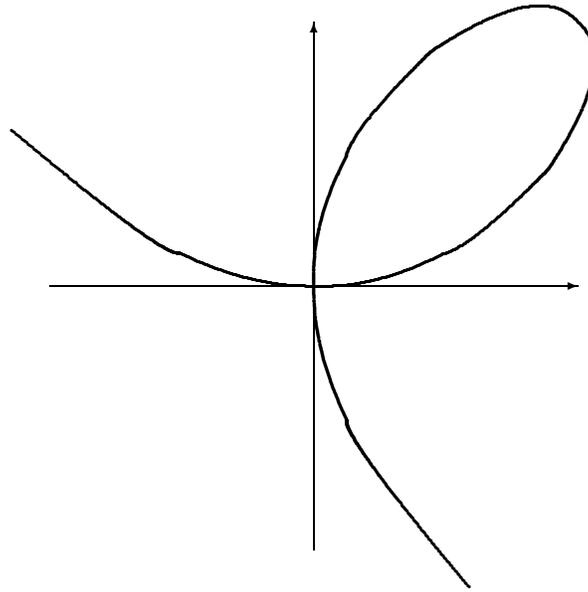
$$\{(x, y) \in \mathbb{R}^2 : f(x, y) = c\}$$

Niveaulinien der Funktion  $f$ .

Wo  $f_y \neq 0$  ist, dort können wir die Gleichung  $f(x, y) = c$  lokal nach  $y$  auflösen. Die Niveaulinie  $f(x, y) = c$  sieht dort aus wie der Graph der Funktion  $y(x)$ . Die Tangente an die Niveaulinie im Punkt  $(x, y)$  wird aufgespannt vom Vektor  $(1, y'(x)) \in \mathbb{R}^2$ . Weil  $f(x, y(x)) \equiv 0$  ist, folgt aus der Kettenregel

$$0 = \frac{d}{dx} f(x, y(x)) = f_x + f_y \frac{dy}{dx} = \left( \text{grad } f(x, y), \begin{pmatrix} 1 \\ dy/dx \end{pmatrix} \right).$$

Der Gradient der Funktion  $f$  steht senkrecht auf den (Tangenten an die) Niveaulinien!



**Beispiel 2.25 (Cartesisches Blatt)** Es sei  $p = n = 1$  und

$$f(x, y) := x^3 + y^3 - xy.$$

Satz 2.13 ist überall dort anwendbar, wo

$$\frac{\partial f}{\partial y} = 3y^2 - x \neq 0,$$

und zeigt, dass in solchen Punkten die Gleichung  $f(x, y) = 0$  lokal nach  $y$  auflösbar ist. Punkte, wo das nicht geht, erfüllen die Gleichungen

$$x^3 + y^3 - xy = 0 \quad \text{und} \quad x = 3y^2.$$

Setzen wir  $x = 3y^2$  in die erste Gleichung ein, so erhalten wir die Bedingung

$$27y^6 + y^3 - 3y^3 = y^3(27y^3 - 2) = 0.$$

Dies führt auf zwei Werte für  $y$ , nämlich

- 1)  $y = 0$ . Hier ist auch  $x = 0$ . Die Kurve, welche durch die Gleichung  $f(x, y) = 0$  gegeben ist, überkreuzt sich hier selbst.
- 2)  $y = \frac{1}{3}\sqrt[3]{2}$ . Hier findet man  $x = \frac{1}{3}\sqrt[3]{4}$ . In diesem Punkt  $(x, y)$  hat die Kurve eine senkrechte Tangente.

Auch ohne die Auflösung  $\mathbf{y} = G(\mathbf{x})$  explizit zu kennen, kann man deren Ableitung ausrechnen.

**Satz 2.14** Unter der Voraussetzung von Satz 2.13 sei  $\mathbf{y} = \mathbf{y}(\mathbf{x}) = G(\mathbf{x})$  eine lokale Auflösung des Gleichungssystems  $F(\mathbf{x}, \mathbf{y}) = \mathbf{0}$ . Für die Ableitung dieser Auflösung gilt

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = - \left( \frac{\partial F}{\partial \mathbf{y}} \right)^{-1} \cdot \frac{\partial F}{\partial \mathbf{x}}.$$

Beweis. Wir spalten die  $p \times (n + p)$ -Funktionalmatrix  $F'$  auf in die beiden partiellen Ableitungen nach  $\mathbf{x}$  und  $\mathbf{y}$ :

$$F' = \left( \underbrace{\frac{\partial F}{\partial \mathbf{x}}}_{p \times n}, \underbrace{\frac{\partial F}{\partial \mathbf{y}}}_{p \times p} \right).$$

Mit  $F(\mathbf{x}, \mathbf{y}(\mathbf{x})) \equiv \mathbf{0}$  folgt aus der Kettenregel

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

Die behauptete Formel ergibt sich daraus nach Multiplikation mit der Matrix  $(\partial F / \partial \mathbf{y})^{-1}$ .  $\square$

**Beispiel 2.26** Die Kreisgleichung  $f(x, y) = x^2 + y^2 - 1 = 0$  besitzt für  $y \neq 0$  die explizite lokale Auflösung

$$y = \pm \sqrt{1 - x^2}.$$

damit können wir

$$y'(x) = \pm \frac{-x}{\sqrt{1 - x^2}} = -\frac{x}{y}$$

berechnen. Die Formel aus Satz 2.14 wird

$$f_x = 2x, f_y = 2y, \quad y' = -\frac{f_x}{f_y} = -\frac{x}{y},$$

das gleiche Ergebnis.

**Beispiel 2.27** Für die Funktion  $f(x, y) = x^3 + y^3 - xy$  haben wir die Auflösung  $y = y(x)$  nicht explizit ausgerechnet. Aber mit Satz 2.14 können wir ihre Ableitung berechnen. Die partiellen Ableitungen von  $f$  sind

$$f_x = 3x^2 - y, \quad f_y = 3y^2 - x.$$

Daraus folgt

$$y'(x) = -\frac{f_x}{f_y} = -\frac{3x^2 - y}{3y^2 - x}.$$

Gelegentlich wird ein physikalisches System durch drei Zustandsgrößen beschrieben, ich nenne sie einfach  $x, y, z$ , zwischen denen eine Zustandsgleichung  $f(x, y, z) = 0$  besteht. Ein konkretes Beispiel dafür ist die ideale Gasgleichung

$$p \cdot V - \text{const} \cdot T = 0.$$

Hält man eine dieser drei Größen fest, so kann man (unter der Voraussetzung des Satzes über implizite Funktionen) die resultierende Gleichung zwischen den beiden anderen Funktionen auflösen lokal bei  $(x_0, y_0, z_0)$  mit  $f(x_0, y_0, z_0) = 0$ , etwa

$$\begin{aligned} y &= y(x)_{z_0} & \text{aus} & f(x, y, z_0) = 0 \\ x &= x(z)_{y_0} & \text{aus} & f(x, y_0, z) = 0 \\ z &= z(y)_{x_0} & \text{aus} & f(x_0, y, z) = 0 \end{aligned}$$

Mit Satz 2.14 erhält man für deren Ableitungen (ich verwende hier die Notation der Physiker)

$$\begin{aligned} \left(\frac{\partial y}{\partial x}\right)_{z_0}(x_0) &= -\frac{f_x(x_0, y_0, z_0)}{f_y(x_0, y_0, z_0)} \\ \left(\frac{\partial x}{\partial z}\right)_{y_0}(z_0) &= -\frac{f_z(x_0, y_0, z_0)}{f_x(x_0, y_0, z_0)} \\ \left(\frac{\partial z}{\partial y}\right)_{x_0}(y_0) &= -\frac{f_y(x_0, y_0, z_0)}{f_z(x_0, y_0, z_0)} \end{aligned}$$

Daraus folgt die bemerkenswerte Formel

$$\left(\frac{\partial y}{\partial x}\right)_{z_0} \cdot \left(\frac{\partial x}{\partial z}\right)_{y_0} \cdot \left(\frac{\partial z}{\partial y}\right)_{x_0} = (-1)^3 \cdot \frac{f_x \cdot f_z \cdot f_y}{f_y \cdot f_x \cdot f_z} = -1.$$

Wenn ich in der idealen Gasgleichung der Einfachheit halber  $c = 1$  setze, wird sie  $f(p, V, T) = p \cdot V - T = 0$ . Auflösungen sind

$$V = \frac{T}{p}, \quad p = \frac{T}{V}, \quad T = p \cdot V.$$

Damit erhalten wir die Ableitungen

$$\left(\frac{\partial V}{\partial p}\right)_T = -\frac{T}{p^2}, \quad \left(\frac{\partial p}{\partial T}\right)_V = \frac{1}{V}, \quad \left(\frac{\partial T}{\partial V}\right)_p = p.$$

Ihr Produkt ist

$$\left(\frac{\partial V}{\partial p}\right)_T \cdot \left(\frac{\partial p}{\partial T}\right)_V \cdot \left(\frac{\partial T}{\partial V}\right)_p = -\frac{T}{p \cdot V} = -1.$$

**Aufgabe 2.26** Wo ist

$$e^{x-y} - (x+y)^2 = 0$$

lokal nach  $y$  auflösbar? Berechnen Sie dort die Ableitung der Auflösung  $y(x)$ .

**Aufgabe 2.27** Es sei  $0 < c \in \mathbb{R}$ . Bei welchen Punkten  $(x, y) \in \mathbb{R}^2$  ist die Gleichung

$$x^2 + xy^2 + y^4 = c$$

lokal nach  $y$  auflösbar? Berechnen Sie dort die Ableitung  $dy/dx$  der lokalen Auflösung  $y(x)$ .

**Aufgabe 2.28** Wo ist die Gleichung

$$x^{2/3} + y^{2/3} = 1$$

lokal nach  $y$  auflösbar? Berechnen Sie dort die Ableitung der implizit definierten Funktion  $y(x)$ . Skizzieren Sie mit dieser Information die Gestalt der Kurve im  $\mathbb{R}^2$ , welche durch diese Gleichung definiert wird.

**Aufgabe 2.29** Die van der Waalssche Gasgleichung lautet etwas vereinfacht

$$f(p, V, T) = \left(p + \frac{a}{V^2}\right)(V - b) - NT = 0, \quad a, b, N \in \mathbb{R}.$$

Berechnen Sie

$$\left(\frac{\partial V}{\partial p}\right)_T, \quad \left(\frac{\partial p}{\partial T}\right)_V, \quad \left(\frac{\partial T}{\partial V}\right)_p.$$

**Aufgabe 2.30** Die Lemniskate ist die Kurve im  $\mathbb{R}^2$  mit der Gleichung

$$(x^2 + y^2)^2 - 2(x^2 - y^2) = 0.$$

- a) Bei welchen Kurvenpunkten ist die Kurvengleichung lokal nach  $y$  auflösbar? Bestimmen Sie dort die Ableitung  $y'$  der lokalen Auflösung  $y(x)$ .  
 b) Bestimmen Sie alle Punkte der Lemniskate, in denen die Kurve eine waagrechte Tangente besitzt.

## 2.7 Lokale Extrema

Bei differenzierbaren Funktionen  $f(x)$  einer reellen Variablen  $x$  gibt es für lokale Extrema zwei Kriterien:

$$\begin{aligned} \text{Notwendiges Kriterium: } & x_0 \text{ ist lokales Extremum} \Rightarrow f'(x_0) = 0. \\ \text{Hinreichendes Kriterium: } & f'(x_0) = 0, f''(x_0) \neq 0 \Rightarrow x_0 \text{ ist lokales Extremum.} \end{aligned}$$

Das notwendige Kriterium für ein Extremum einer Funktion  $f(x_1, \dots, x_n)$  von mehreren Variablen haben wir bereits kennengelernt:

$$f \text{ hat in } \mathbf{x}_0 = (x_{0,1}, \dots, x_{0,n}) \text{ ein lokales Extremum} \Rightarrow \frac{\partial f}{\partial x_1}(\mathbf{x}_0) = \dots = \frac{\partial f}{\partial x_n}(\mathbf{x}_0) = 0.$$

Zur Formulierung eines hinreichenden Kriteriums brauchen wir alle zweiten partiellen Ableitungen  $\partial^2 f / \partial x_\mu \partial x_\nu$ . Diese zweiten Ableitungen bilden eine  $n \times n$ -Matrix

$$H_f(x) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \dots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \dots & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix},$$

die sogenannte *Hesse-Matrix*. Wenn  $f$  zweimal stetig differenzierbar ist, d.h., wenn die zweiten partiellen Ableitungen stetig sind, hängen diese nicht von der Reihenfolge ab, in der sie gebildet werden. Wir brauchen uns also nicht darum zu kümmern, ob es  $\partial^2 f / \partial x_\mu \partial x_\nu$  oder  $\partial^2 f / \partial x_\nu \partial x_\mu$  heißt. Das ist schön, weil dies bedeutet, dass die Hesse-Matrix symmetrisch ist. Symmetrische Matrizen werden in der linearen Algebra ausführlich untersucht. (Hauptachsentransformation!) Dies werden wir jetzt anwenden.

Schränken wir  $f$  auf eine Gerade

$$\mathbf{x} = \mathbf{x}_0 + t \cdot \mathbf{v}$$

durch  $\mathbf{x}_0$  mit Richtungsvektor  $\mathbf{v}$  ein, so ist die zweite Ableitung der eingeschränkten Funktion

$$\begin{aligned} \frac{d^2}{dt^2} f(\mathbf{x}_0 + t \cdot \mathbf{v}) &= \frac{d}{dt} \sum_{\nu=1}^n \frac{\partial f}{\partial x_\nu}(\mathbf{x}_0 + t \cdot \mathbf{v}) \cdot v_\nu \\ &= \sum_{\mu, \nu=1}^n \frac{\partial^2 f}{\partial x_\mu \partial x_\nu}(\mathbf{x}_0 + t \cdot \mathbf{v}) \cdot v_\mu v_\nu. \end{aligned}$$

Im Punkt  $\mathbf{x}_0$  selbst ( $t = 0$ ) erhalten wir für diese zweite Ableitung

$$\sum_{\mu, \nu=1}^n H_{\mu, \nu} \cdot v_{\mu} v_{\nu} = \mathbf{v}^t \cdot H \cdot \mathbf{v},$$

den Wert der durch die Hessematrix gegebenen quadratischen Form auf dem Richtungsvektor  $\mathbf{v}$ .

Eine quadratische Form  $\mathbf{v} \mapsto \mathbf{v}^t \cdot H \cdot \mathbf{v}$  heißt *positiv definit*, wenn für alle Vektoren  $0 \neq \mathbf{v} \in \mathbb{R}^n$  der Wert  $\mathbf{v}^t \cdot H \cdot \mathbf{v}$  positiv, d.h. echt  $> 0$  ist. Dies ist äquivalent dazu, dass alle Eigenwerte der Matrix  $H$  positiv sind. Die Form heißt *negativ definit*, wenn  $\mathbf{v}^t \cdot H \cdot \mathbf{v} < 0$  für alle  $\mathbf{v} \neq \mathbf{0}$ .

In der linearen Algebra beweist man aber auch ein Kriterium, bei dem man die Eigenwerte nicht zu kennen braucht (Satz 1.24):

**Satz 2.15 (Hurwitz-Kriterium)** *Eine symmetrische reelle Matrix*

$$H = (h_{\mu, \nu})_{\mu, \nu=1, \dots, n}$$

*ist genau dann positiv definit, wenn ihre  $n$  Hauptminoren  $D_k$ ,  $k = 1, \dots, n$ , positiv sind:*

$$D_k := \det(h_{\mu, \nu})_{\mu, \nu=1, \dots, k} > 0.$$

**Satz 2.16 (Hinreichendes Kriterium für lokales Minimum)** *Die Funktion  $f$  sei zweimal stetig differenzierbar auf der offenen Menge  $D \subset \mathbb{R}^n$ . In dem Punkt  $\mathbf{x}_0 \in D$  gelte*

- $\frac{\partial f}{\partial x_1} = \dots = \frac{\partial f}{\partial x_n} = 0$ ,
- *die Hesse-Matrix  $H_f(\mathbf{x}_0)$  ist positiv definit.*

*Dann hat  $f$  in  $\mathbf{x}_0$  ein lokales Minimum.*

Beweis. Wenn  $\mathbf{x}_0$  kein lokales Minimum für  $f$  wäre, dann gäbe es eine Folge von Punkten  $\mathbf{x}_0 \neq \mathbf{x}_k \in D$ , die gegen  $\mathbf{x}_0$  konvergiert, mit  $f(\mathbf{x}_k) \leq f(\mathbf{x}_0)$  für alle  $k$ . Sei  $\mathbf{v}_k := \mathbf{x}_k - \mathbf{x}_0 \in \mathbb{R}^n$ . Dann wird durch

$$\mathbb{R} \ni t \mapsto \mathbf{x}_0 + t \cdot \mathbf{v}_k$$

die Gerade durch  $\mathbf{x}_0$  und  $\mathbf{x}_k$  parametrisiert. Die auf diese Gerade eingeschränkte Funktion

$$\varphi(t) := f(\mathbf{x}_0 + t \cdot \mathbf{v}_k)$$

der reellen Variablen  $t$  hat bei  $t = 0$

- verschwindende Ableitung  $\varphi'(0) = \sum_{\nu=1}^n \frac{\partial f}{\partial x_{\nu}}(\mathbf{x}_0) \cdot v_{k, \nu}$
- und positive zweite Ableitung  $\varphi''(0) = \sum_{\mu, \nu=1}^n \frac{\partial^2 f}{\partial x_{\mu} \partial x_{\nu}}(\mathbf{x}_0) \cdot v_{k, \mu} v_{k, \nu}$ .

Wäre die zweite Ableitung von  $\varphi$

$$\varphi''(t) = \sum_{\mu, \nu=1}^n \frac{\partial^2 f}{\partial x_{\mu} \partial x_{\nu}}(\mathbf{x}_0 + t \cdot \mathbf{v}_k) \cdot v_{k, \mu} v_{k, \nu}$$

auf dem ganzen  $t$ -Intervall  $[0, 1]$  positiv, so hätte  $\varphi$  auf diesem ganzen Intervall monoton wachsende erste Ableitung, die Funktion wäre echt monoton wachsend. Da

$$\varphi(1) = f(\mathbf{x}_k) \leq f(\mathbf{x}_0) = \varphi(0),$$

kann dies nicht der Fall sein. Es gibt also einen Zwischenwert  $0 < \theta < 1$  mit  $\varphi''(\theta) = 0$ . Setzen wir  $\mathbf{y}_k := \mathbf{x}_0 + \theta \mathbf{v}_k$ , so haben wir deswegen für die Einheitsvektoren  $\mathbf{w}_k := \mathbf{v}_k / \|\mathbf{v}_k\|$

$$H_f(\mathbf{y}_k)(\mathbf{w}_k) = \frac{1}{\|\mathbf{v}_k\|^2} \sum_{\mu, \nu=1}^n \frac{\partial^2 f}{\partial x_\mu \partial x_\nu}(\mathbf{y}_k) v_{k,\mu} v_{k,\nu} = 0.$$

Da die Einheitskugel im  $\mathbb{R}^n$  kompakt ist, besitzt die Folge der Vektoren  $\mathbf{w}_k$  eine konvergente Teilfolge. Nachdem wir zu dieser Teilfolge übergehen können wir o.B.d.A. annehmen  $\mathbf{w}_k \rightarrow \mathbf{w}_0$  für  $k \rightarrow \infty$ . Die Funktion

$$D \times \mathbb{R}^n \ni (\mathbf{x}, \mathbf{v}) \mapsto H_f(\mathbf{x})(\mathbf{v})$$

ist stetig, da die zweiten partiellen Ableitungen von  $f$  stetig vorausgesetzt sind. Es folgt

$$H_f(\mathbf{x}_0)(\mathbf{w}_0) = \lim_{k \rightarrow \infty} H_f(\mathbf{x}_k)(\mathbf{w}_k) = 0.$$

Das geht aber nicht, weil  $H_f(\mathbf{x}_0)$  positiv definit und der Vektor  $\mathbf{w}_0 \neq \mathbf{0}$  ist. □

Geht man von  $f$  über zu  $-f$ , so erhält man aus Satz 2.16 ein hinreichendes Kriterium für das Vorhandensein eines lokalen Maximum: Die Funktion  $f$  sei zweimal stetig differenzierbar auf der offenen Menge  $D \subset \mathbb{R}^n$ . In einem Punkt  $\mathbf{x}_0 \in D$  sollen alle ersten partiellen Ableitungen verschwinden und die Hesse-Matrix negativ-definit sein. Dann ist  $\mathbf{x}_0$  ein lokales Maximum.

Die symmetrische Matrix  $H$  heißt *indefinit*, wenn es Vektoren  $\mathbf{v}$  und  $\mathbf{w}$  gibt mit

$$\mathbf{v}^t \cdot H \cdot \mathbf{v} > 0 \quad \text{und} \quad \mathbf{w}^t \cdot H \cdot \mathbf{w} < 0.$$

Ist die Hesse-Matrix  $H_f(\mathbf{x})$  indefinit in einem Punkt  $\mathbf{x}$ , wo  $\text{grad}(f) = \mathbf{0}$  ist, so liegt dort bestimmt kein Extremum vor: Auf der Geraden  $\mathbf{x} + t \cdot \mathbf{v}$  hat  $f$  ein lokales Minimum, und auf der Geraden  $\mathbf{x} + t \cdot \mathbf{w}$  ein lokales Maximum. Und beliebig nahe bei  $\mathbf{x}$  gibt es Punkte  $\mathbf{x}_1, \mathbf{x}_2$  mit

$$f(\mathbf{x}_1) < f(\mathbf{x}) < f(\mathbf{x}_2).$$

Ein Beispiel dafür ist der Sattelpunkt  $f(x, y) = x^2 - y^2$ . Die Hesse-Matrix im Nullpunkt

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

ist indefinit.

Es gibt auch noch *semidefinite* Matrizen  $H$ , wo also entweder stets  $\mathbf{v}^t \cdot H \cdot \mathbf{v} \geq 0$  oder  $\mathbf{v}^t \cdot H \cdot \mathbf{v} \leq 0$  gilt. Ist die Hesse-Matrix  $H_f(\mathbf{x})$  semidefinit, so kann man mit unseren Methoden über ein eventuelles Extremum nichts aussagen.

Im  $n$ -dimensionalen Raum,  $n \geq 2$ , gibt es noch ein Phänomen, das in einer Dimension nicht vorkommt: *Extrema unter Nebenbedingungen*. Dabei wird nicht ein Extremum auf einer offenen Teilmenge  $D \subset \mathbb{R}^n$  gesucht, sondern unter einer (oder mehreren) Nebenbedingung(en). Physiker nennen soetwas

häufig auch etwas plastischer *Zwangsbedingungen*. So eine Bedingung wird durch eine Funktion  $g(\mathbf{x})$  gegeben, die den Wert 0 haben soll. Die Bedingung etwa, dass sich ein Punkt  $\mathbf{x}$  auf der Einheitskugel aufhalten soll, wird durch die Funktion  $g(\mathbf{x}) = \sum x_\nu^2 - 1 = 0$  beschrieben.

Die Methode, lokale Extrema unter Nebenbedingungen zu suchen, beruht auf dem Satz über implizite Funktionen. Wir wollen sie in Dimension  $n = 2$  beweisen, weiter unten aber noch ganz allgemein formulieren.

**Satz 2.17 (Lagrange-Multiplikator,  $n = 2$ )** Die Funktionen  $f(x, y)$  und  $g(x, y)$  seien differenzierbar,  $g$  stetig differenzierbar, auf der offenen Menge  $D \subset \mathbb{R}^2$ . Für  $\mathbf{x}_0 = (x_0, y_0) \in D$  gelte

$$g(\mathbf{x}_0) = 0, \quad \text{grad}(g)[\mathbf{x}_0] \neq 0.$$

Die Funktion  $f$  nehme im Punkt  $\mathbf{x}_0$  ein Extremum unter der Nebenbedingung  $g(\mathbf{x}) = 0$  an. Dann existiert ein Skalar  $\lambda \in \mathbb{R}$  (Lagrange-Multiplikator) so, dass

$$\begin{aligned} \frac{\partial f}{\partial x}(\mathbf{x}_0) &= \lambda \frac{\partial g}{\partial x}(\mathbf{x}_0) \\ \frac{\partial f}{\partial y}(\mathbf{x}_0) &= \lambda \frac{\partial g}{\partial y}(\mathbf{x}_0) \end{aligned}$$

Beweis. O.B.d.A. sei  $\partial g / \partial y(\mathbf{x}_0) \neq 0$ . Wegen des Satzes über implizite Funktionen können wir die Gleichung  $g(x, y) = 0$  bei  $\mathbf{x}_0$  lokal nach  $y$  auflösen. Sei  $y = h(x)$ ,  $|x - x_0| < r$ , eine derartige Auflösung. Dass  $f$  ein lokales Extremum unter der Nebenbedingung  $g(x, y) = 0$  besitzt, das heißt, dass die Funktion  $x \mapsto f(x, h(x))$  bei  $x_0$  ein lokales Extremum besitzt. Es folgt

$$0 = \left. \frac{d}{dx} f(x, h(x)) \right|_{x=x_0} = \frac{\partial f}{\partial x}(\mathbf{x}_0) + \frac{\partial f}{\partial y}(\mathbf{x}_0) \cdot \frac{dh}{dx}(x_0).$$

Andererseits wissen wir aber auch (Zusatz zum Satz über implizite Funktionen)

$$0 = \left. \frac{d}{dx} g(x, h(x)) \right|_{x=x_0} = \frac{\partial g}{\partial x}(\mathbf{x}_0) + \frac{\partial g}{\partial y}(\mathbf{x}_0) \cdot \frac{dh}{dx}(x_0).$$

Damit haben wir ein lineares Gleichungssystem

$$\begin{pmatrix} \frac{\partial f}{\partial x}(\mathbf{x}_0) & \frac{\partial f}{\partial y}(\mathbf{x}_0) \\ \frac{\partial g}{\partial x}(\mathbf{x}_0) & \frac{\partial g}{\partial y}(\mathbf{x}_0) \end{pmatrix} \cdot \begin{pmatrix} 1 \\ \frac{dh}{dx}(x_0) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

mit einer nichttrivialen Lösung. Dann müssen die Zeilen der Koeffizienten-Matrix linear abhängig sein. Es besteht also eine nichttriviale lineare Relation

$$\begin{aligned} \lambda \frac{\partial g}{\partial x}(\mathbf{x}_0) + \mu \frac{\partial f}{\partial x}(\mathbf{x}_0) &= 0, \\ \lambda \frac{\partial g}{\partial y}(\mathbf{x}_0) + \mu \frac{\partial f}{\partial y}(\mathbf{x}_0) &= 0, \end{aligned}$$

Weil  $\partial g / \partial y(\mathbf{x}_0) \neq 0$ , kann  $\mu$  nicht 0 sein. Wir können dann beispielsweise  $\mu = -1$  annehmen und erhalten die Behauptung.  $\square$

**Beispiel 2.28** Wir untersuchen die Funktion  $f(x, y) = x^3 + y^3$  unter der Nebenbedingung  $x^2 + y^2 = 1$  auf lokale Extrema. Wir haben also  $g(x, y) = x^2 + y^2 - 1$  und

$$\begin{aligned} \frac{\partial f}{\partial x} &= 3x^2, & \frac{\partial g}{\partial x} &= 2x, \\ \frac{\partial f}{\partial y} &= 3y^2, & \frac{\partial g}{\partial y} &= 2y. \end{aligned}$$

Aus Satz 2.17 erhalten wir die Bedingungen

$$3x^2 = \lambda \cdot 2x, \quad 3y^2 = \lambda \cdot 2y.$$

Um  $\lambda$  zu eliminieren, fassen wir dies als ein lineares Gleichungssystem mit nicht-trivialer Lösung  $(1, -\lambda)$  und Determinante

$$3x^2 \cdot 2y - 3y^2 \cdot 2x = 6xy \cdot (x - y) = 0$$

auf. Dies ergibt die Bedingungen

$$\begin{aligned} x = 0 &\implies y = \pm 1 &\implies f(x, y) = \pm 1 \\ y = 0 &\implies x = \pm 1 &\implies f(x, y) = \pm 1 \\ x = y &\implies x = y = \pm \frac{1}{\sqrt{2}} &\implies f(x, y) = \pm \frac{1}{\sqrt{2}}. \end{aligned}$$

Damit haben wir sechs Punkte auf dem Kreis ausgerechnet. Zwischen diesen Punkten hat die Funktion  $f$  auf dem Kreis kein lokales Extremum, ist also monoton. Daraus können wir schließen, dass alle sechs Punkte lokale Extrema unter der Nebenbedingung sind, und zwar

$(x, y)$	$f(x, y)$	
$(1, 0)$	1	Maximum
$(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$	$\frac{1}{\sqrt{2}}$	lokales Minimum
$(0, 1)$	1	Maximum
$(-1, 0)$	-1	Minimum
$(-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}})$	$-\frac{1}{\sqrt{2}}$	lokales Maximum
$(0, -1)$	-1	Minimum.

Wir wollen noch das allgemeine Kriterium formulieren.

**Satz 2.18 (Lagrange-Multiplikatoren)** Es sei  $D \subset \mathbb{R}^n$  offen,  $f : D \rightarrow \mathbb{R}$  differenzierbar und  $G = (G_1, \dots, G_k) : D \rightarrow \mathbb{R}^k$  stetig differenzierbar. Der Rang der Funktionalmatrix von  $G$  in  $\mathbf{x}_0 \in D$

$$\left( \frac{\partial G_\kappa}{\partial x_\nu}(\mathbf{x}_0) \right)_{\nu=1, \dots, n, \kappa=1, \dots, k}$$

sei  $= k$ . Nimmt  $f$  in  $\mathbf{x}_0 \in D$  ein lokales Extremum unter den Nebenbedingungen  $G_1 = \text{const}_1, \dots, G_k = \text{const}_k$  an, dann gibt es Skalare  $\lambda_1, \dots, \lambda_k \in \mathbb{R}$  (Lagrange-Multiplikatoren), so, dass gilt

$$\begin{aligned} \frac{\partial f}{\partial x_1}(\mathbf{x}_0) &= \lambda_1 \frac{\partial G_1}{\partial x_1}(\mathbf{x}_0) + \dots + \lambda_k \frac{\partial G_k}{\partial x_1}(\mathbf{x}_0) \\ &\vdots \\ \frac{\partial f}{\partial x_n}(\mathbf{x}_0) &= \lambda_1 \frac{\partial G_1}{\partial x_n}(\mathbf{x}_0) + \dots + \lambda_k \frac{\partial G_k}{\partial x_n}(\mathbf{x}_0). \end{aligned}$$

Dies sind  $n$  Gleichungen. Zusammen mit den  $k$  Nebenbedingungen haben wir  $n + k$  Gleichungen für die  $n + k$  Unbekannten  $x_{0,1}, \dots, x_{0,n}$  und  $\lambda_1, \dots, \lambda_k$ . Obwohl diese Gleichungen i.a. nicht-linear sind, besteht doch eine gewisse Hoffnung, dass man sie auflösen und so das Extremum  $\mathbf{x}_0$  ermitteln kann.

**Aufgabe 2.31** Sei

$$f(x, y) := \frac{1}{x} - \frac{1}{y} + 4x - 9y \quad (x \neq 0, y \neq 0).$$

Bestimmen Sie alle lokalen Extrema von  $f$ .

**Aufgabe 2.32** Gegeben sei die Funktion  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$  durch

$$f(x, y) = (x^2 + 2y^2 - 1)(x^2 + y^2 - 1).$$

a) Skizzieren Sie die Menge  $\{(x, y) \in \mathbb{R}^2 : f(x, y) = 0\}$  und markieren Sie die Bereiche, wo  $f(x, y)$  positiv, bzw. negativ ist.

b) Berechnen Sie die Ableitung  $f'$  und die Hesse-Matrix von  $f$ .

c) Bestimmen Sie alle lokalen Extrema von  $f$ .

**Aufgabe 2.33** Es sei  $A = (a_{i,j})_{i,j=1,\dots,n}$  eine positiv definite symmetrische  $n \times n$ -Matrix und  $\mathbf{b} \in \mathbb{R}^n$ . Wo besitzt

$$f(\mathbf{x}) = \sum_{i,j=1}^n a_{i,j} x_i x_j + \sum_{i=1}^n b_i x_i$$

ein lokales Minimum oder Maximum?

**Aufgabe 2.34** Es seien

$$L_1 : \mathbf{x}(s) = \mathbf{w}_1 + s\mathbf{v}_1, \quad L_2 : \mathbf{y}(t) = \mathbf{w}_2 + t\mathbf{v}_2$$

zwei Geraden im  $\mathbb{R}^n$  mit  $\mathbf{v}_1, \mathbf{v}_2, \mathbf{w}_1, \mathbf{w}_2 \in \mathbb{R}^n$  und  $\mathbf{v}_1, \mathbf{v}_2$  linear unabhängig. Zeigen Sie: Es gibt eindeutig bestimmte Punkte  $\mathbf{x}_0 \in L_1$  und  $\mathbf{y}_0 \in L_2$ , für die der Abstand  $\|\mathbf{x}(s) - \mathbf{y}(t)\|$  minimal ist.

**Aufgabe 2.35** a) Bestimmen Sie die lokalen Extrema der Funktion

$$h(x, y) = x^3 + x^2y + xy^2 + y^3 - 6x - 6y.$$

b) Bestimmen Sie das Maximum der Funktion  $h$  auf der Kreisscheibe  $x^2 + y^2 \leq 16$  vom Radius 4.

**Aufgabe 2.36** Bestimmen Sie die Extrema der Funktion

$$f : \mathbb{R}^n \rightarrow \mathbb{R}, \quad (x_1, \dots, x_n) \mapsto x_1^2 \cdot \dots \cdot x_n^2,$$

unter der Nebenbedingung

$$g(x_1, \dots, x_n) := x_1^2 + \dots + x_n^2 - 1 = 0.$$

**Aufgabe 2.37** Welche Punkte der Menge

$$M = \{(x, y, z) \mid x^4 + y^4 + z^4 = 3\} \subset \mathbb{R}^3$$

haben minimalen Abstand vom Ursprung des  $\mathbb{R}^3$ ? (Rechnung!)

**Aufgabe 2.38** Seien  $p_1, \dots, p_n > 0$  und  $y_1, \dots, y_n \in \mathbb{R}^3$ . Die Funktion  $f : \mathbb{R}^3 \rightarrow \mathbb{R}$  sei gegeben durch

$$f(x) := \sum_{j=1}^n p_j \|x - y_j\|^2,$$

wobei  $\|\cdot\|$  die euklidische Norm im  $\mathbb{R}^3$  bezeichnet. Man zeige, dass es genau eine Minimalstelle von  $f$  gibt, und gebe diese an.

**Aufgabe 2.39** Durch die Gleichung

$$\frac{x^2}{9} + \frac{y^2}{16} + \frac{z^2}{36} = 1$$

ist ein Ellipsoid  $E$  im  $\mathbb{R}^3$  gegeben. Die  $E$  einbeschriebenen Quader  $Q$  sind von der Form  $Q = [-x, x] \times [-y, y] \times [-z, z]$  mit  $(x, y, z) \in E$  und  $x, y, z > 0$ . Zeigen Sie, dass es unter diesen Quadern genau einen mit größtem Volumen gibt, und geben Sie diesen an.

**Aufgabe 2.40** Sei  $K$  die Kreisfläche  $K = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 \leq 4\}$  und  $A = (0, 0)$ ,  $B = (1, -1)$ ,  $C = (2, 5)$ . Man bestimme diejenigen Punkte von  $K$ , für welche die Summe der Abstandsquadrate von den Punkten  $A, B$  und  $C$  am kleinsten, bzw. am größten ist.

**Aufgabe 2.41** Es sei  $m > n$  und  $A$  eine  $m \times n$ -Matrix vom Rang  $n$ . Weiter sei  $\mathbf{y} \in \mathbb{R}^m$  fest. Zeigen Sie:

a) Die Matrix  $A^t \cdot A$  ist positiv definit.

b) Die Funktion  $f(x) := \|A \cdot \mathbf{x} - \mathbf{y}\|^2$  auf  $\mathbb{R}^n$  hat den Gradient

$$\text{grad } f(\mathbf{x}) = 2 \cdot (A^t \cdot A \cdot \mathbf{x} - A^t \cdot \mathbf{y})$$

und die Hesse-Matrix

$$H_f = 2 \cdot A^t \cdot A.$$

c) Im Punkt

$$\mathbf{x}_0 := (A^t \cdot A)^{-1} \cdot A \cdot \mathbf{y}$$

hat die Funktion  $f$  ein eindeutig bestimmtes absolutes Minimum.

**Aufgabe 2.42 (Methode der kleinsten Quadrate)** a) Gegeben seien  $m$  Punkte (Messwerte)

$$(x_1, y_1), \dots, (x_m, y_m) \in \mathbb{R}^2$$

Zeigen Sie: Es gibt eindeutig bestimmte reelle Zahlen  $u$  und  $v$  so, dass die Gerade mit der Gleichung

$$y = ux + v$$

eine minimale Quadratsumme der Abweichungen

$$q(u, v) := \sum_{\mu=1}^m (u \cdot x_\mu + v - y_\mu)^2$$

besitzt. Hinweis: Verwenden Sie Aufgabe 2.41 mit der  $m \times 2$ -Matrix

$$A = \begin{pmatrix} x_1 & 1 \\ \vdots & \vdots \\ x_m & 1 \end{pmatrix}.$$

b) Bestimmen Sie die Gerade aus a) explizit für die vier Punkte

$$(1, 2), (2, 1), (3, 3), (4, 3).$$

## 2.8 Die Differentialoperatoren *grad*, *rot*, *div* und $\Delta$

In Definition 2.5 haben wir für jede differenzierbare Funktion auf  $\mathbb{R}^n$  den Gradienten

$$\text{grad } f(\mathbf{x}) = \nabla f(\mathbf{x}) = \left( \frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n} \right)$$

definiert. Hier wollen wir dem Symbol

$$\nabla = \begin{pmatrix} \frac{\partial}{\partial x_1} \\ \vdots \\ \frac{\partial}{\partial x_n} \end{pmatrix}$$

eine eigene Existenzberechtigung zuerkennen. Es ist eine Art Kombination von Vektorfeld und Ableitungsoperator, eben ein Differentialoperator. Als erstes wollen wir verschiedene Kombinationsmöglichkeiten für diesen Operator  $\nabla$  untersuchen.

1) Ist  $\mathbf{A} = (A_1, \dots, A_n)$  ein Vektorfeld auf dem  $\mathbb{R}^n$ , so kann man das *Skalarprodukt*

$$(\nabla, \mathbf{A}) = \sum_{\nu=1}^n \frac{\partial}{\partial x_\nu} A_\nu$$

bilden. Das Resultat ist eine Funktion

$$\text{div}(\mathbf{A}) := \sum_{\nu=1}^n \frac{\partial A_\nu}{\partial x_\nu}.$$

Sie heißt *Divergenz* des Vektorfeldes  $A$ .

**Beispiel 2.29** Die Divergenz des Ortsvektorfeldes  $\mathbf{A}(\mathbf{x}) = \mathbf{x}$  im  $\mathbb{R}^n$  ist

$$\text{div}(\mathbf{x}) = \sum_{\nu=1}^n \frac{\partial x_\nu}{\partial x_\nu} = n.$$

2) Das *Skalarprodukt* von  $\nabla$  mit sich selbst

$$\Delta := (\nabla, \nabla) = \sum_{\nu=1}^n \frac{\partial^2}{\partial x_\nu^2}$$

ist ein Differentialoperator, der jeder (zweimal differenzierbaren) Funktion  $f$  die Funktion

$$\Delta f = \sum_{\nu=1}^n \frac{\partial^2 f}{\partial x_\nu^2}$$

zuordnet. Dieser Operator  $\Delta$  heißt *Laplace-Operator*. Wir können ihn auch auffassen als

$$\Delta f = \operatorname{div}(\operatorname{grad} f).$$

3) Im  $\mathbb{R}^3$  ist für jedes Vektorfeld  $\mathbf{A} = (A_1, A_2, A_3)$  auch das Kreuzprodukt

$$\nabla \times \mathbf{A} = \begin{pmatrix} \frac{\partial A_3}{\partial x_2} - \frac{\partial A_2}{\partial x_3} \\ \frac{\partial A_1}{\partial x_3} - \frac{\partial A_3}{\partial x_1} \\ \frac{\partial A_2}{\partial x_1} - \frac{\partial A_1}{\partial x_2} \end{pmatrix}$$

definiert. Das Resultat

$$\operatorname{rot}(\mathbf{A}) = \nabla \times \mathbf{A}$$

ist wieder ein Vektorfeld. Es heißt das *Rotationsfeld* des Feldes  $\mathbf{A}$ .

**Beispiel 2.30** *Natürlich hat das Rotationsfeld etwas mit Drehung zu tun. Als Beispiel betrachten wir das Geschwindigkeitsfeld einer Drehung um eine Achse mit Winkelgeschwindigkeit  $\omega$*

$$\mathbf{A}(x_1, x_2, x_3) = \omega \times \mathbf{x} = \begin{pmatrix} \omega_2 x_3 - \omega_3 x_2 \\ \omega_3 x_1 - \omega_1 x_3 \\ \omega_1 x_2 - \omega_2 x_1 \end{pmatrix},$$

$$\operatorname{rot}(\mathbf{A}) = \begin{pmatrix} \frac{\partial}{\partial x_1} \\ \frac{\partial}{\partial x_2} \\ \frac{\partial}{\partial x_3} \end{pmatrix} \times \begin{pmatrix} \omega_2 \cdot x_3 - \omega_3 \cdot x_2 \\ \omega_3 \cdot x_1 - \omega_1 \cdot x_3 \\ \omega_1 \cdot x_2 - \omega_2 \cdot x_1 \end{pmatrix} = \begin{pmatrix} 2\omega_1 \\ 2\omega_2 \\ 2\omega_3 \end{pmatrix} = 2\omega.$$

Aus der Symmetrie der zweiten Ableitung (Satz 2.5) erhält man in Dimension 3 die folgenden fundamentalen Formeln:

**Satz 2.19 (0-Rechenregeln)** *a) Für jede zweimal stetig differenzierbare Funktion  $f(x_1, x_2, x_3)$  gilt*

$$\operatorname{rot}(\operatorname{grad} f) = \mathbf{0}.$$

*b) Für jedes zweimal stetig differenzierbare Vektorfeld  $\mathbf{A} = (A_1, A_2, A_3)$  gilt*

$$\operatorname{div}(\operatorname{rot}(\mathbf{A})) = 0.$$

Wenn man die Rechenregeln

$$\mathbf{A} \times \mathbf{A} = \mathbf{0}, \quad (\mathbf{A}, \mathbf{A} \times \mathbf{B}) = 0$$

für das Kreuz-Produkt kennt, dann kann man sich diese Regeln in der Form

$$(\nabla \times \nabla)f = \mathbf{0}, \quad (\nabla, (\nabla \times f)) = 0$$

leicht merken. Aber man kann sie auch beweisen:

Beweis von Satz 2.19. a) Wir berechnen nur die erste Komponente

$$\operatorname{rot}(\operatorname{grad} f)_1 = \left( \frac{\partial}{\partial x_2} \frac{\partial f}{\partial x_3} - \frac{\partial}{\partial x_3} \frac{\partial f}{\partial x_2} = \frac{\partial^2 f}{\partial x_2 \partial x_3} - \frac{\partial^2 f}{\partial x_3 \partial x_2} \right) = 0.$$

Das gleiche Ergebnis für die beiden anderen Komponenten erhält man durch zyklisches Vertauschen der Indizes.

b) Hier berechnen wir alle auftretenden Summanden:

$$\begin{aligned} \operatorname{div}(\operatorname{rot}(\mathbf{a})) &= \frac{\partial(\operatorname{rot}(\mathbf{A})_1)}{\partial x_1} + \frac{\partial(\operatorname{rot}(\mathbf{A})_2)}{\partial x_2} + \frac{\partial(\operatorname{rot}(\mathbf{A})_3)}{\partial x_3} \\ &= \frac{\partial}{\partial x_1} \left( \frac{\partial A_3}{\partial x_2} - \frac{\partial A_2}{\partial x_3} \right) + \frac{\partial}{\partial x_2} \left( \frac{\partial A_1}{\partial x_3} - \frac{\partial A_3}{\partial x_1} \right) + \frac{\partial}{\partial x_3} \left( \frac{\partial A_2}{\partial x_1} - \frac{\partial A_1}{\partial x_2} \right) \\ &= \frac{\partial^2 A_3}{\partial x_1 \partial x_2} - \frac{\partial^2 A_2}{\partial x_1 \partial x_3} + \frac{\partial^2 A_1}{\partial x_2 \partial x_3} - \frac{\partial^2 A_3}{\partial x_2 \partial x_1} + \frac{\partial^2 A_2}{\partial x_3 \partial x_1} - \frac{\partial^2 A_1}{\partial x_3 \partial x_2}. \end{aligned}$$

Wegen der Symmetrie der zweiten Ableitungen kürzt sich alles weg. □

Die Produktregel der Differentiation führt nach dem Schema

$$\operatorname{grad}(f \cdot g) = g \cdot \operatorname{grad} f + f \cdot \operatorname{grad} g$$

oder

$$\operatorname{div}(f \cdot \mathbf{A}) = (\operatorname{grad} f, \mathbf{A}) + f \cdot \operatorname{div}(\mathbf{A})$$

auf reichhaltiges Formelmateriale. Dem möchte ich hier aber nicht weiter nachgehen, sondern auf Aufgabe 2.47 verweisen. Stattdessen möchte ich mich hier noch etwas um den Laplace-Operator kümmern.

**Definition 2.13** Eine Funktion  $f$  heißt harmonisch, wenn der Laplace-Operator auf sie angewendet

$$\Delta f = \sum_{\nu=1}^n \frac{\partial^2 f}{\partial x_\nu^2} \equiv 0$$

ergibt.

Der Laplace-Operator  $\Delta$  ist in jeder Dimension  $n$  definiert, liefert aber von der Dimension abhängige Ergebnisse. Betrachten wir etwa die Funktion

$$f(x_1, \dots, x_n) := \|\mathbf{x}\|^m = \left( \sum_{\nu=1}^n x_\nu^2 \right)^{m/2}.$$

Dafür finden wir

$$\begin{aligned} \operatorname{grad}(\|\mathbf{x}\|) &= \frac{1}{\|\mathbf{x}\|} \cdot \mathbf{x}, \\ \operatorname{grad}(\|\mathbf{x}\|^m) &= m \|\mathbf{x}\|^{m-1} \cdot \operatorname{grad} \|\mathbf{x}\| \\ &= m \|\mathbf{x}\|^{m-2} \cdot \mathbf{x}, \\ \frac{1}{m} \cdot \Delta(f) &= \operatorname{div}(\|\mathbf{x}\|^{m-2} \cdot \mathbf{x}) \\ &= (\operatorname{grad} \|\mathbf{x}\|^{m-2}, \mathbf{x}) + \|\mathbf{x}\|^{m-2} \cdot \operatorname{div}(\mathbf{x}) \\ &= (m-2) \|\mathbf{x}\|^{m-4} \cdot (\mathbf{x}, \mathbf{x}) + n \cdot \|\mathbf{x}\|^{m-2} \\ &= (n+m-2) \cdot \|\mathbf{x}\|^{m-2}. \end{aligned}$$

Wir haben ausgerechnet: Auf  $\mathbb{R}^n \setminus \{\mathbf{0}\}$  ist die Funktion

$$r^{2-n} = \|\mathbf{x}\|^{2-n}$$

harmonisch. Für  $n = 1$  ist dies die kaum erwähnenswerte Tatsache, dass die zweite Ableitung der Funktion  $|x|$  für  $x \neq 0$  verschwindet, für  $n = 2$  die noch uninteressantere entsprechende Aussage für die konstante Funktion  $\equiv 1$ . Aber etwa für  $n = 3$  stellt sich die Funktion  $1/r$  als harmonisch heraus.

Allerdings ist die Dimension 2 dadurch nicht sehr benachteiligt: Für die Funktion

$$f(x, y) = \ln(x^2 + y^2)$$

berechnen wir

$$\begin{aligned} \Delta(f) &= \frac{\partial}{\partial x} \left( \frac{2x}{x^2 + y^2} \right) + \frac{\partial}{\partial y} \left( \frac{2y}{x^2 + y^2} \right) \\ &= \frac{4}{x^2 + y^2} - \frac{4x^2}{(x^2 + y^2)^2} - \frac{4y^2}{(x^2 + y^2)^2} \\ &= 0, \end{aligned}$$

also ist  $\ln(r) = \frac{1}{2} \ln(x^2 + y^2)$  in Dimension 2 harmonisch.

Äbschließend wollen wir uns noch einmal mit den Formeln

$$\operatorname{rot}(\operatorname{grad}) = 0 \quad \text{und} \quad \operatorname{div}(\operatorname{rot}) = 0$$

aus Satz 2.19 beschäftigen. Wenn man auf die Verwendung des *rot*-Operators verzichtet, kann man die erste der beiden Formeln in beliebiger Dimension  $n$  beweisen:

**Satz 2.20** *Das Vektorfeld  $\mathbf{A}$  sei stetig differenzierbar auf der offenen Menge  $U \subset \mathbb{R}^n$ . Eine notwendige Bedingung dafür, dass  $\mathbf{A} = \operatorname{grad} f$  ein Gradientenfeld ist, ist*

$$\frac{\partial A_\mu}{\partial x_\nu} = \frac{\partial A_\nu}{\partial x_\mu}, \quad \mu, \nu = 1, \dots, n. \quad (\text{Symmetriebedingung}).$$

Der Beweis ist klar. Denn wenn  $\mathbf{A} = \operatorname{grad} f$  gilt, dann ist die Funktion  $f$  zweimal stetig differenzierbar. Aus der Symmetrie ihrer zweiten Ableitungen (Satz 2.5) folgt

$$\frac{\partial A_\mu}{\partial x_\nu} = \frac{\partial^2 f}{\partial x_\nu \partial x_\mu} = \frac{\partial^2 f}{\partial x_\mu \partial x_\nu} = \frac{\partial A_\nu}{\partial x_\mu}.$$

Es ist häufig wichtig zu entscheiden, ob ein gegebenes Vektorfeld

$$\mathbf{A}(\mathbf{x}) = (A_1(x_1, \dots, x_n), \dots, A_n(x_1, \dots, x_n))$$

ein Gradientenfeld ist oder nicht:

$$\mathbf{A} = \operatorname{grad}(f) \quad ??$$

Ist dies der Fall, so heißt die Funktion  $-f$  das *Potential* des Feldes  $\mathbf{A}$ .

**Beispiel 2.31** Die Funktion  $f = x^2 + y^2$  auf  $\mathbb{R}^2$  hat den Gradienten

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

Natürlich ist deswegen die Symmetriebedingung

$$\frac{\partial(2x)}{\partial y} = 0 = \frac{\partial(2y)}{\partial x}$$

erfüllt. Die ist auch für das Feld

$$\mathbf{A} = (y, x)$$

erfüllt ( $f = xy$ ), aber z.B. nicht für

$$\mathbf{A} = (y, -x).$$

Dieses Feld kann also kein Potential besitzen.

In die Symmetriebedingung hat man  $\mu, \nu = 1, \dots, n$  unabhängig voneinander einzusetzen. Natürlich braucht man  $\mu = \nu$  nicht auszuprobieren, und wenn die Symmetriebedingung für  $\mu \neq \nu$  erfüllt ist, dann gilt sie auch nach Vertauschen von  $\mu$  und  $\nu$ . Sie stellt also

$$\frac{1}{2}(n^2 - n) = \frac{1}{2}n \cdot (n - 1)$$

skalare Bedingungen dar. I.A. ist diese Anzahl viel größer als  $n$ , aber in unserem 3-dimensionalen Anschauungsraum ist diese Anzahl genau gleich der Dimension:

$$\frac{1}{2}3 \cdot 2 = 3.$$

Und diese drei Symmetriebedingungen kann man zu einer Vektorgleichung  $\text{rot}(\mathbf{A}) = \mathbf{0}$  zusammenfassen.

Und die einfache Symmetriebedingung ist auch fast *hinreichend* dafür, dass  $\mathbf{A}$  ein Gradientenfeld ist:

**Satz 2.21 (Potential)** Die offene Menge  $U \subset \mathbb{R}^n$  sei konvex. Das Vektorfeld  $\mathbf{A}$  auf  $U$  sei stetig differenzierbar und erfülle die Symmetriebedingung. Dann gibt es eine differenzierbare Funktion  $f : U \rightarrow \mathbb{R}$  mit

$$\mathbf{A} = \text{grad } f.$$

Beweis. Wir fixieren einen Punkt in  $U$  und, um Schreibaufwand zu sparen, nennen wir ihn  $\mathbf{0}$ . (D.h., o.B.d.A. nehmen wir an  $\mathbf{0} \in U$ .) Die Menge  $U$  ist konvex vorausgesetzt, weil dann mit jedem Punkt  $\mathbf{x} \in U$  auch die ganze Verbindungsstrecke zum Nullpunkt  $\mathbf{0}$

$$\{t \cdot \mathbf{x} : t \in \mathbb{R}, 0 \leq t \leq 1\}$$

zu  $U$  gehört. Wenn  $f$  auf  $U$  stetig partiell differenzierbar ist, folgt für  $0 \leq t \leq 1$  mit der Formel für die Richtungsableitung (Beispiel 2.9)

$$\frac{d}{dt}f(t \cdot \mathbf{x}) = (\text{grad } f(t \cdot \mathbf{x}), \mathbf{x}).$$

Wenn wir  $f$  durch  $f - f(\mathbf{0})$  ersetzen, können wir  $f(\mathbf{0}) = 0$  annehmen. Dann ist

$$f(\mathbf{x}) = \int_0^1 \frac{d}{dt} f(t \cdot \mathbf{x}) dt = \int_0^1 (\text{grad } f(t \cdot \mathbf{x}), \mathbf{x}) dt.$$

Wir suchen  $f : U \rightarrow \mathbb{R}$  mit  $\text{grad } f = \mathbf{A}$  und  $f(\mathbf{0}) = 0$ . Wenn  $f$  existiert, muss es durch die Formel

$$f(\mathbf{x}) = \int_0^1 (\mathbf{A}(t \cdot \mathbf{x}), \mathbf{x}) dt = \sum_{\nu=1}^n \int_0^1 A_\nu(t \cdot \mathbf{x}) \cdot x_\nu dt$$

gegeben sein. Umgekehrt definiert diese Formel eine Funktion  $f$  auf  $U$ , weil die  $A_\nu$  stetig sind. Zu zeigen ist also, dass die so definierte Funktion  $f$  differenzierbar ist mit

$$\frac{\partial f}{\partial x_\nu} = A_\nu \quad \text{für } \nu = 1, \dots, n.$$

Es genügt, diesen Nachweis für  $\nu = 1$  zu führen.

Wir rechnen erst mal los und überlegen uns später, warum wir das so dürfen. Als erstes vertauschen wir Differentiation und Integration:

$$\begin{aligned} \frac{\partial f}{\partial x_1}(\mathbf{x}) &= \sum_{\nu=1}^n \int_0^1 \frac{\partial}{\partial x_1} (A_\nu(t \cdot \mathbf{x}) \cdot x_\nu) dt \\ &= \sum_{\nu=1}^n \int_0^1 \frac{\partial A_\nu}{\partial x_1}(t \cdot \mathbf{x}) \cdot t \cdot x_\nu dt + \int_0^1 A_1(t \cdot \mathbf{x}) dt \\ &= \sum_{\nu=1}^n \int_0^1 \frac{\partial A_1}{\partial x_\nu}(t \cdot \mathbf{x}) \cdot t \cdot x_\nu dt + \int_0^1 A_1(t \cdot \mathbf{x}) dt \quad (\text{Symmetrie - Bedingung}) \\ &= \int_0^1 \frac{d}{dt} (A_1(t \cdot \mathbf{x}) \cdot t) dt \\ &= A_1(t \cdot \mathbf{x}) \cdot t \Big|_{t=0}^{t=1} \\ &= A_1(\mathbf{x}). \end{aligned}$$

Wir sind fertig - falls wir wie oben Differentiation und Integration vertauschen dürfen. Das ist aber gerade der Punkt. Das müssen wir noch beweisen. Dazu betrachten wir folgende Situation:

Die Funktion  $f(x_1, \dots, x_n)$  sei stetig auf dem abgeschlossenen Quader

$$Q = \{\mathbf{x} \in \mathbb{R}^n : a_1 \leq x_1 \leq b_1, \dots, a_n \leq x_n \leq b_n\} \subset \mathbb{R}^n.$$

Für

$$\mathbf{x}' = (x_2, \dots, x_n) \in Q' = \{\mathbf{x}' \in \mathbb{R}^{n-1} : a_2 \leq x_2 \leq b_2, \dots, a_n \leq x_n \leq b_n\} \subset \mathbb{R}^{n-1}$$

definieren wir die Funktion

$$F(x_2, \dots, x_n) := \int_{a_1}^{b_1} f(x_1, x_2, \dots, x_n) dx_1.$$

**Satz 2.22** a) (Stetige Abhängigkeit des Integrals von einem Parameter) *Aus der Stetigkeit von  $f$  folgt die Stetigkeit von  $F$ .*

b) (Differentiation unter dem Integral) Die Funktion  $f$  sei zusätzlich stetig partiell differenzierbar nach  $x_2, \dots, x_n$  auf der Menge

$$\{\mathbf{x} \in \mathbb{R}^n : a_1 \leq x_1 \leq b_1, a_2 < x_2 < b_2, \dots, a_n < x_n < b_n\} \subset Q.$$

Dann ist  $F$  partiell nach  $x_2, \dots, x_n$  differenzierbar mit den nach a) stetigen partiellen Ableitungen

$$\frac{\partial F}{\partial x_\nu}(x_2, \dots, x_n) = \int_{a_1}^{b_1} \frac{\partial f}{\partial x_\nu}(x_1, x_2, \dots, x_n) dx_1.$$

Beweis. a) Wir zeigen die  $\epsilon$ - $\delta$ -Stetigkeit von  $F$  in einem festen Punkt  $\mathbf{x}'_0 \in Q'$ . Dazu verwenden wir, dass die stetige Funktion  $f$  auf der kompakten Menge  $Q$  sogar gleichmäßig stetig ist. (Das beweist man ganz ähnlich wie im ersten Semester die entsprechende Aussage in Dimension 1.) Zu jedem  $\epsilon > 0$  gibt es also ein  $\delta(\epsilon) > 0$  mit

$$\mathbf{x}, \mathbf{y} \in Q, \|\mathbf{x} - \mathbf{y}\| < \delta(\epsilon) \Rightarrow |f(\mathbf{x}) - f(\mathbf{y})| < \epsilon.$$

Für  $\mathbf{x}', \mathbf{x}'_0 \in Q'$  mit  $\|\mathbf{x}' - \mathbf{x}'_0\| < \delta(\epsilon)$  folgt daraus für alle  $x_1 \in [a_1, b_1]$

$$|f(x_1, \mathbf{x}') - f(x_1, \mathbf{x}'_0)| < \epsilon.$$

Mit der üblichen Integral-Abschätzung zeigt man damit

$$|F(\mathbf{x}') - F(\mathbf{x}'_0)| \leq \int_{a_1}^{b_1} |f(x_1, \mathbf{x}') - f(x_1, \mathbf{x}'_0)| dx_1 \leq \int_{a_1}^{b_1} \epsilon dx_1 = (b_1 - a_1) \cdot \epsilon.$$

b) Es genügt, den Fall  $n = 2$  und eine stetig nach  $y$  differenzierbare Funktion  $f(x, y)$  zu betrachten. Um die Differenzierbarkeit von

$$F(y) = \int_{a_1}^{b_1} f(x, y) dx$$

zu zeigen, fixieren wir ein  $y_0 \in ]a_2, b_2[$  und ein  $r > 0$  mit

$$a_2 < y_0 - r < y_0 + r < b_2.$$

Auf dem kompakten Streifen  $[a_1, b_1] \times [y_0 - r, y_0 + r]$  ist  $\partial f / \partial y$  stetig und sogar gleichmäßig stetig. Zu jedem  $\epsilon > 0$  gibt es also ein  $\delta \leq r$  mit

$$|y - y_0| < \delta \Rightarrow \left| \frac{\partial f}{\partial y}(x, y) - \frac{\partial f}{\partial y}(x, y_0) \right| < \epsilon \quad \text{für alle } x \in [a_1, b_1].$$

Wir betrachten die Funktion

$$\begin{cases} \frac{f(x, y) - f(x, y_0)}{y - y_0} & \text{für } y \neq y_0 \\ \frac{\partial f}{\partial y}(x, y_0) & \text{für } y = y_0 \end{cases}$$

Mit dem Mittelwertsatz

$$\frac{f(x, y) - f(x, y_0)}{y - y_0} = \frac{\partial f}{\partial y}(x, y_0 + \theta \cdot (y - y_0))$$

folgt, dass diese Funktion stetig ist für alle  $(x, y_0)$  und damit auf dem ganzen Streifen  $[a_1, b_1] \times [y_0 - r, y_0 + r]$ . Nach a) ist deswegen

$$\lim_{y \rightarrow y_0} \frac{F(y) - F(y_0)}{y - y_0} = \lim_{y \rightarrow y_0} \int_{a_1}^{b_1} \frac{f(y) - f(y_0)}{y - y_0} dx = \int_{a_1}^{b_1} \lim_{y \rightarrow y_0} \frac{f(y) - f(y_0)}{y - y_0} dx = \int_{a_1}^{b_1} \frac{\partial f}{\partial y}(x, y_0) dx. \quad \square$$

**Beispiel 2.32 (Gegenbeispiel)** Die Bedingung in Satz 2.22, dass  $U$  konvex ist, garantiert, dass  $U$  keine Löcher hat. Für offene Mengen mit Löchern ist der Satz i.A. falsch. Als Gegenbeispiel betrachten wir die Argumentfunktion  $\varphi$  auf der gelochten Ebene  $\mathbb{R}^2 \setminus \{(0,0)\}$ . Die Polarkoordinatenabbildung

$$x = r \cdot \cos(\varphi), \quad y = r \cdot \sin(\varphi)$$

hat die Funktionalmatrix

$$\frac{\partial(x,y)}{\partial(r,\varphi)} = \begin{pmatrix} \cos(\varphi) & -r\sin(\varphi) \\ \sin(\varphi) & r\cos(\varphi) \end{pmatrix}$$

mit der inversen Matrix

$$\frac{\partial(r,\varphi)}{\partial(x,y)} = \frac{1}{r} \cdot \begin{pmatrix} r\cos(\varphi) & r\sin(\varphi) \\ -\sin(\varphi) & \cos(\varphi) \end{pmatrix}.$$

Daraus folgt

$$\mathbf{A} := \text{grad } \varphi = \frac{1}{r}(-\sin(\varphi), \cos(\varphi)) = \frac{1}{x^2 + y^2}(-y, x).$$

Dieses Feld  $\mathbf{A}$  erfüllt die Symmetrie-Bedingung (weil es lokal Gradient einer Funktion ist), hat aber kein globales Potential, weil die Funktion  $\varphi$  mehrdeutig und somit keine stetige Funktion auf  $\mathbb{R}^2 \setminus \{(0,0)\}$  ist.

Satz 2.22 bedeutet auf konvexen offenen Mengen  $U \subset \mathbb{R}^3$  eine Umkehrung der Formel  $\text{rot}(\text{grad}) = \mathbf{0}$  in dem Sinn

$$\text{rot}(\mathbf{A}) = \mathbf{0} \quad \Rightarrow \quad \mathbf{A} = \text{grad } f.$$

Genauso kann man auch die Formel  $\text{div}(\text{rot}) = 0$  umkehren:

**Satz 2.23 (Vektorpotential)** Die offene Menge  $U \subset \mathbb{R}^3$  sei konvex. Das Vektorfeld  $\mathbf{A}$  auf  $U$  sei stetig differenzierbar mit

$$\text{div}(\mathbf{A}) = 0.$$

Dann gibt es ein differenzierbares Vektorfeld  $\mathbf{B}$  auf  $U$  mit

$$\mathbf{A} = \text{rot}(\mathbf{B}).$$

Beweis. Wir gehen vom Vektorfeld  $\mathbf{A}$  über zur schiefssymmetrischen Matrix

$$(a_{i,j}) := \begin{pmatrix} 0 & A_3 & -A_2 \\ -A_3 & 0 & A_1 \\ A_2 & -A_1 & 0 \end{pmatrix}.$$

Damit definieren wir die Funktionen

$$B_i(\mathbf{x}) := \sum_{j=1}^3 x_j \cdot \int_0^1 a_{j,i}(t \cdot \mathbf{x}) \cdot t dt$$

und das Vektorfeld  $\mathbf{B} = (B_1, B_2, B_3)$ . Dieses Feld tut's. Das rechnen wir nur für die erste Komponente nach. Mit Differentiation unter dem Integral finden wir

$$\begin{aligned}
 \frac{\partial B_3}{\partial x_2} &= \int_0^1 ta_{2,3}(\mathbf{t}\mathbf{x})dt + \sum_{j=1}^3 x_j \int_0^1 \frac{\partial a_{j,3}}{\partial x_2}(\mathbf{t}\mathbf{x})dt \\
 &= \int_0^1 t \left( a_{2,3} + tx_1 \frac{\partial a_{1,3}}{\partial x_2} + tx_2 \frac{\partial a_{2,3}}{\partial x_2} + tx_3 \frac{\partial a_{3,3}}{\partial x_2} \right) dt \\
 &= \int_0^1 t \cdot \left( A_1(\mathbf{t}\mathbf{x}) - tx_1 \frac{\partial A_2}{\partial x_2}(\mathbf{t}\mathbf{x}) + tx_2 \frac{\partial A_1}{\partial x_2}(\mathbf{t}\mathbf{x}) \right) dt \\
 \frac{\partial B_2}{\partial x_3} &= \int_0^1 t \left( a_{3,2} + tx_1 \frac{\partial a_{1,2}}{\partial x_3} + tx_3 \frac{\partial a_{3,2}}{\partial x_3} + tx_2 \frac{\partial a_{2,2}}{\partial x_3} \right) dt \\
 &= \int_0^1 t \cdot \left( -A_1(\mathbf{t}\mathbf{x}) + tx_1 \frac{\partial A_3}{\partial x_3}(\mathbf{t}\mathbf{x}) - tx_3 \frac{\partial A_1}{\partial x_3}(\mathbf{t}\mathbf{x}) \right) dt \\
 \frac{\partial B_3}{\partial x_2} - \frac{\partial B_2}{\partial x_3} &= \int_0^1 t \cdot \left( 2A_1 - tx_1 \underbrace{\left( \frac{\partial A_2}{\partial x_2} + \frac{\partial A_3}{\partial x_3} \right)}_{=-\partial A_1/\partial x_1} + tx_2 \frac{\partial A_1}{\partial x_2} + tx_3 \frac{\partial A_1}{\partial x_3} \right) dt \\
 &= \int_0^1 \left( 2tA_1(\mathbf{t}\mathbf{x}) + t^2 \sum_{j=1}^3 x_j \frac{\partial A_1}{\partial x_j}(\mathbf{t}\mathbf{x}) \right) dt \\
 &= \int_0^1 \frac{d}{dt}(t^2 \cdot A_1(\mathbf{t}\mathbf{x}))dt \\
 &= A_1(\mathbf{x}).
 \end{aligned}$$

Auswendig merken kann sich das so niemand. □

Das Potential eines Vektorfeldes  $\mathbf{A}$  ist nur bis auf eine Konstante bestimmt. Denn

$$\text{grad } f_1 = \text{grad } f_2 = \mathbf{A} \quad \Rightarrow \quad \text{grad}(f_1 - f_2) = \mathbf{0} \quad \Rightarrow \quad f_1 - f_2 = \text{const.}$$

Analog gilt für zwei Vektorpotentiale  $\mathbf{B}_1, \mathbf{B}_2$  eines Feldes  $\mathbf{A}$

$$\text{rot}(\mathbf{B}_1) = \text{rot}(\mathbf{B}_2) = \mathbf{A} \quad \Rightarrow \quad \text{rot}(\mathbf{B}_1 - \mathbf{B}_2) = \mathbf{0}.$$

Auf einer konvexen Menge folgt mit Satz 2.22

$$\mathbf{B}_1 - \mathbf{B}_2 = \text{grad } f.$$

Das Vektorpotential ist nur bis auf ein Gradientenfeld eindeutig bestimmt.

**Aufgabe 2.43** Bestimmen Sie Divergenz und Rotation der Vektorfelder

$$\mathbf{A} = \begin{pmatrix} y \\ z \\ x \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} z \\ x \\ y \end{pmatrix}.$$

**Aufgabe 2.44** Bestimmen Sie  $\text{rot}(\mathbf{a})$  und  $\text{rot}(\text{rot}(\mathbf{a}))$  für das Vektorfeld

$$\mathbf{A} = \begin{pmatrix} yz^2 \\ zx^2 \\ xy^2 \end{pmatrix}.$$

**Aufgabe 2.45** Zeigen Sie für differenzierbare Funktionen  $f, g$  und Vektorfelder  $\mathbf{A}, \mathbf{B}$  auf dem  $\mathbb{R}^3$

- 1)  $\text{grad}(f \cdot g) = g \cdot \text{grad}(f) + f \cdot \text{grad}(g)$
- 2)  $\text{div}(f \cdot \mathbf{A}) = (\text{grad}(f) \cdot \mathbf{A}) + f \cdot \text{div}(\mathbf{A})$
- 3)  $\Delta(f \cdot g) = g \cdot \Delta(f) + 2(\text{grad}(f) \cdot \text{grad}(g)) + f \cdot \Delta(g)$
- 4)  $\text{rot}(f \cdot \mathbf{A}) = \text{grad}(f) \times \mathbf{A} + f \cdot \text{rot}(\mathbf{A})$
- 5)  $\text{rot}(f \cdot \text{grad}(g)) = \text{grad}(f) \times \text{grad}(g)$
- 6)  $\text{div}(\mathbf{A} \times \mathbf{B}) = (\mathbf{B} \cdot \text{rot}(\mathbf{A})) - (\mathbf{A} \cdot \text{rot}(\mathbf{B}))$
- 7)  $\text{div}(\text{grad}(f) \times \text{grad}(g)) = 0$

**Aufgabe 2.46** Es seien  $a \in \mathbb{R}$  und  $\mathbf{x} \neq \mathbf{0}$ .

- a) Bestimmen Sie den Gradient der Funktion  $(x^2 + y^2 + z^2)^a$ .
- b) Bestimmen Sie für  $r := \sqrt{x^2 + y^2 + z^2}$  den Gradient der Funktion  $r^a$ .
- c) Bestimmen Sie Divergenz und Rotation des Vektorfeldes  $\hat{\mathbf{x}} := \frac{1}{r}\mathbf{x}$ .

**Aufgabe 2.47** Zeigen Sie  $\text{rot}(\mathbf{A}) = 0$  für die Felder

$$\text{a) } \mathbf{A} = (yz, xz, xy), \quad \text{b) } \mathbf{A} = (2xz, 0, z^2), \quad \text{c) } \mathbf{A} = (2xy^3, 3x^2y^2, 0),$$

und geben Sie jeweils eine Funktion  $f(x, y, z)$  mit  $\text{grad } f = \mathbf{A}$  an.

**Aufgabe 2.48** Es sei  $f(x, y)$  eine genügend oft differenzierbare Funktion auf  $\mathbb{R}^2$  und

$$f(r, \varphi) = f(x(r, \varphi), y(r, \varphi))$$

dieselbe Funktion in Polarkoordinaten.

$$\text{a) Zeigen Sie: } \Delta f = \frac{1}{r} \frac{\partial}{\partial r} \left( r \frac{\partial f}{\partial r} \right) + \frac{1}{r^2} \frac{\partial^2 f}{\partial \varphi^2}.$$

b) Bestimmen Sie alle harmonischen Funktionen  $f$  auf  $\mathbb{R}^2 \setminus \{\mathbf{0}\}$ , die nur von  $r$  und nicht von  $\varphi$  abhängen.

**Aufgabe 2.49** Es sei  $f : \mathbb{R}^2 \setminus \{\mathbf{0}\} \rightarrow \mathbb{R}$  genügend oft differenzierbar. Zeigen Sie:  $f \cdot \mathbf{x}$  ist genau dann ein Gradientenfeld, wenn  $f = f(r)$  ist.

### 3 Gewöhnliche Differentialgleichungen

Eine Differentialgleichung verknüpft eine unbekannte Funktion  $f$  mit ihren Ableitungen. Im Idealfall kann man sie lösen und die unbekannte Funktion bestimmen. *Gewöhnliche* Differentialgleichungen sind solche, wo es um eine Funktion  $f(x)$  einer Veränderlichen geht. Die einfachste davon wird schon im Gymnasium behandelt:

$$f'(x) = g(x).$$

Um sie zu lösen muss man die Funktion  $g(x)$  integrieren. Daher kommt auch der Sprachgebrauch: Statt zu sagen 'eine Differentialgleichung lösen' benutzt man auch die Formulierung 'eine Differentialgleichung integrieren'.

Wenn die gesuchte Funktion  $f$  nicht von einer, sondern von mehreren Variablen abhängt, nennt man die Differentialgleichung *partiell*. Beispiele dafür sind (ich lasse Konstanten weg)

$$\begin{aligned} \frac{\partial^2 f}{\partial t^2} &= \frac{\partial^2 f}{\partial x^2} && \text{Wellengleichung} \\ \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} &= 0 && \text{Potentialgleichung} \\ \frac{\partial f}{\partial t} &= \frac{\partial^2 f}{\partial x^2} && \text{Wärmeleitungsgleichung.} \end{aligned}$$

Partielle Differentialgleichungen sind viel wichtiger als gewöhnliche, aber auch viel schwerer zu lösen. Doch auch gewöhnliche Differentialgleichungen sind meistens ungewöhnlich schwer zu lösen. So führt deren Behandlung auf die folgenden vier Problemkreise:

Praxis: Zusammenstellung von Differentialgleichungen, die man elementar lösen kann;

Theorie: Existenz und Eindeutigkeit von Lösungen, qualitative Aussagen;

Individualismus: ausgiebige Untersuchung einer einzigen Differentialgleichung, möglicherweise Einführung ihrer Lösungen als neue Funktionen und Beschreibung von deren Eigenschaften;

Numerik: Approximation der Lösungen.

Auch die Bezeichnung der Variablen ist nicht sehr einheitlich. Mathematiker suchen meist eine Funktion  $f(x)$  oder  $y(x)$  von  $x$ , während in den klassischen physikalischen Anwendungen die unabhängige Variable meist die Zeit  $t$  und die abhängige Variable meist der Ort  $x$  ist. Da wird dann also eine Funktion  $x(t)$  gesucht. Bei Funktionen von  $x$  schreibt man die Ableitung  $f'(x)$  oder  $y'(x)$ . Bei Funktionen  $x(t)$  schreibt man die Ableitung  $\dot{x}(t)$ .

Hier eine kurze Liste von Lehrbüchern zu diesem Thema (in alphabetischer Reihenfolge):

- M.Braun: Differentialgleichungen und ihre Anwendungen. Springer Hochschultext 1979
- I.Collatz: Differentialgleichungen. Teubner Studienbücher 1990
- F.Erwe: Gewöhnliche Differentialgleichungen. BI 1961
- O.Forster: Analysis II. Vieweg 1977
- S.Großmann: Mathematischer Einführungskurs für die Physik. Teubner Studienbücher 1991
- K.Jänich: Analysis für Physiker und Ingenieure. Springer 1983
- E.Kamke: Differentialgleichungen. Akad. Verlagsges. Leipzig 1930

Das letztgenannte Buch ist *der* Klassiker, aus dem nicht nur ich abschreibe, sondern auch so mancher Buchautor abgeschrieben hat.

### 3.1 Elementare Lösungsmethoden

So sehr viele elementare Lösungsmethoden gibt es gar nicht. Sie sind - außer für Klausuraufgaben - auch nicht sehr wichtig. Ich stelle sie hier zusammen, um ein Gefühl dafür zu vermitteln, worum es eigentlich geht.

#### 3.1.1 Trennung der Variablen

Hier geht es um Gleichungen der Form

$$y' = a(x) \cdot b(y).$$

Die Variablen kann man dann trennen:

$$\frac{y'}{b(y)} = a(x).$$

Mit Stammfunktionen  $B$  von  $1/b$  und  $A$  von  $a$  kann man diese Gleichung schreiben

$$\frac{d}{dx}B(y) = \frac{d}{dx}A(x)$$

und integrieren

$$B(y) = A(x) + C, \quad C \in \mathbb{R}.$$

Mit einer Umkehrfunktion von  $B$  kann man diese Gleichung nach  $y$  auflösen.

**Beispiel 3.1** *Wir betrachten die Gleichung*

$$y' = \frac{x}{y}.$$

*Trennung der Variablen ergibt*

$$y \cdot y' = x,$$

*und die Integration dieser Gleichung führt auf*

$$\frac{y^2}{2} = \frac{x^2}{2} + C.$$

*Nach  $y$  aufgelöst wird daraus*

$$y = \pm\sqrt{x^2 + C}, \quad C \in \mathbb{R}.$$

Der wichtigste Spezialfall dieser Art von Differentialgleichungen ist die *homogene lineare* Differentialgleichung

$$y' = a(x) \cdot y.$$

Wir trennen die Variablen

$$\frac{y'}{y} = a(x)$$

und integrieren mit einer Stammfunktion  $A$  von  $a$

$$\ln(|y|) = A(x) + C.$$

Nachdem wir hierauf die Exponentialfunktion anwenden erhalten wir

$$|y| = e^C \cdot e^{A(x)}, \quad C \in \mathbb{R}.$$

Jetzt wird es etwas tricky:  $c := e^C$  ist eine beliebige Konstante  $> 0$ . Weil links aber  $|y|$  steht, können wir die Gleichung umschreiben als

$$y = c \cdot e^{A(x)}, \quad c \neq 0.$$

Aber wie das Schicksal so will (nicht immer zu unseren Ungunsten), erhalten wir auch für  $c = 0$  eine Lösung, nämlich  $y \equiv 0$ . Somit haben wir die Lösungen

$$y = c \cdot e^{A(x)}, \quad c \in \mathbb{R}.$$

**Beispiel 3.2** Die einfachste Gleichung dieser Art ist

$$y' = x \cdot y.$$

Wir trennen die Variablen

$$\frac{y'}{y} = x$$

und integrieren

$$\ln(|y|) = \frac{x^2}{2} + C.$$

Nach  $y$  aufgelöst wird daraus

$$y = c \cdot e^{x^2/2}.$$

### 3.1.2 Inhomogene lineare Differentialgleichung

Die inhomogene lineare Differentialgleichung entsteht aus der homogenen linearen durch Addition einer Inhomogenität und sieht so aus:

$$y' = a(x) \cdot y + b(x).$$

Wenn man eine spezielle Lösung  $y_0(x)$  dieser Gleichung kennt, erhält man alle ihre Lösungen, indem man zu  $y_0(x)$  alle Lösungen der zugehörigen homogenen Gleichung  $y' = a(x) \cdot y$  addiert. Es kommt also darauf an, so eine einzige spezielle Lösung  $y_0(x)$  zu finden.

Dazu gibt ein Verfahren mit dem in sich widersprüchlichen Namen 'Variation der Konstanten'. Man beginnt mit einer (nicht-trivialen) Lösung  $y(x)$  der homogenen Gleichung und macht den Ansatz

$$y_0(x) := c(x) \cdot y(x).$$

Dabei ist jetzt natürlich die Funktion  $c(x)$  gesucht. Mit  $y' = a \cdot y$  finden wir

$$y_0'(x) = c'(x) \cdot y(x) + c(x) \cdot y'(x) = c'(x) \cdot y(x) + c(x) \cdot a(x) \cdot y(x).$$

Dies vergleichen wir mit der Bedingung

$$y_0'(x) = a(x) \cdot y_0(x) + b(x) = a(x) \cdot c(x) \cdot y(x) + b(x).$$

Die Differenz beider Gleichungen führt auf

$$c'(x) \cdot y(x) = b(x) \quad \text{bzw.} \quad c'(x) = \frac{b(x)}{y(x)}.$$

Die unbekannte Funktion ermittelt man hieraus durch Integration.

**Beispiel 3.3** *Wir betrachten die inhomogene lineare Differentialgleichung*

$$y' = y + 1.$$

*Die zugehörige homogene Gleichung hat die Lösungen  $c \cdot e^x$  mit  $c \in \mathbb{R}$ . Wir machen also den Ansatz*

$$y_0(x) = c(x) \cdot e^x$$

*und berechnen*

$$y_0'(x) = c'(x) \cdot e^x + c(x) \cdot e^x = c'(x) \cdot e^x + y_0(x).$$

*Vergleich mit der Gleichung  $y_0'(x) = y_0(x) + 1$  liefert*

$$c'(x) \cdot e^x = 1, \quad c'(x) = e^{-x}, \quad c(x) = -e^{-x}.$$

*Damit bekommen wir die spezielle Lösung*

$$y_0(x) = -e^{-x} \cdot e^x = -1.$$

*Die hätten wir auch erraten können. Die allgemeine Lösung der inhomogenen Gleichung ist damit*

$$y(x) = -1 + c \cdot e^x, \quad c \in \mathbb{R}.$$

### 3.1.3 Bernoullische Differentialgleichung

Das ist die hochgradig nichtlineare Gleichung

$$y' = a(x) \cdot y + b(x) \cdot y^k, \quad k \in \mathbb{R}.$$

Man dividiert durch  $y^k$

$$y^{-k} \cdot y' = a(x) \cdot y^{1-k} + b(x)$$

und substituiert

$$z = y^{1-k}, \quad z' = (1-k) \cdot y^{-k} y'.$$

Für  $z$  bekommt man dann die inhomogene lineare Gleichung

$$\frac{1}{1-k} \cdot z' = a(x) \cdot z + b(x) \quad \text{bzw.} \quad z' = (1-k)a(x) \cdot z(x) + (1-k)b(x).$$

Den Fall  $k = 1$  muss man natürlich ausschließen, aber da wäre die ursprüngliche Gleichung eine homogene lineare gewesen.

**Beispiel 3.4** So ziemlich die einfachste Bernoullische Differentialgleichung ist

$$y' = y + y^2.$$

Wir dividieren

$$\frac{y'}{y^2} = \frac{1}{y} + 1$$

und substituieren  $z = 1/y$ . Für  $z$  liefert das die Gleichung

$$-z' = z + 1, \quad z' = -z - 1.$$

Die zugehörige homogene Gleichung hat die allgemeine Lösung  $z(x) = c \cdot e^{-x}$ . Die spezielle Lösung  $z_0(x) = -1$  der inhomogenen Gleichung errät man. Damit haben wir

$$z(x) = -1 + c \cdot e^{-x}, \quad y(x) = \frac{1}{z(x)} = \frac{1}{c \cdot e^{-x} - 1}, \quad c \in \mathbb{R}.$$

### 3.1.4 Homogene Differentialgleichung

Der Name ist etwas irreführend, weil man sie nicht mit der homogenen linearen Gleichung verwechseln darf. Hier bedeutet 'homogen', dass die rechte Seite eine Funktion von  $y/x$  ist. Die Gleichung sieht also so aus:

$$y' = g\left(\frac{y}{x}\right).$$

Man substituiert

$$z = \frac{y}{x}, \quad y = z \cdot x, \quad y' = z' \cdot x + z.$$

Damit wird die Gleichung

$$z' \cdot x + z = g(z), \quad z' = \frac{g(z) - z}{x},$$

eine Gleichung, wo man die Variablen trennen kann.

**Beispiel 3.5** Substituiert man  $z = y/x$  in der Gleichung

$$y' = \frac{1}{2} \left( 1 + \frac{y^2}{x^2} \right),$$

so erhält man

$$x \cdot z' + z = \frac{1}{2}(1 + z^2), \quad z' = \frac{1}{2} \frac{1 - 2z + z^2}{x}, \quad \frac{z'}{(1-z)^2} = \frac{1}{2x}.$$

Nach einer Integration wird daraus

$$\frac{1}{1-z} = \frac{1}{2} \ln|x| + c_1, \quad 1-z = \frac{2}{\ln|x| + c}, \quad z = 1 - \frac{2}{\ln|x| + c}.$$

Die Rücktransformation zeigt

$$y = x - \frac{2x}{\ln|x| + c}, \quad c \in \mathbb{R}.$$

Ich habe bisher die Notation  $y = y(x)$  für die gesuchte Funktion verwendet. Mit einer Differentialgleichung  $y' = f(x, y)$  wird dabei die folgende Vorstellung verbunden: Durch die rechte Seite wird in jedem Punkt  $(x, y)$  eines Teils der Ebene eine Richtung gegeben. Gesucht ist eine Kurve, welche in jedem ihrer Punkte die dort vorgegebene Richtung besitzt.

**Beispiel 3.6** *Durch jeden Punkt  $(x, y)$  der Ebene mit  $x \neq 0 \neq y$  geht genau eine Hyperbel mit der Gleichung  $x \cdot y = \text{const.}$  Gesucht seien die orthogonalen Trajektorien dieser Hyperbelschar, d.h. die Kurven, welche auf allen gegebenen Hyperbeln senkrecht stehen. Eine Hyperbel*

$$x \cdot y = c, \quad \text{bzw.} \quad y = \frac{c}{x}$$

hat die Steigung

$$y' = -\frac{c}{x^2}.$$

Elimination von  $c = x \cdot y$  führt auf das Richtungsfeld

$$y' = -\frac{xy}{x^2} = -\frac{y}{x}.$$

Die orthogonalen Trajektorien zu den gegebenen Hyperbeln haben in jedem Punkt  $(x, y)$  die Richtung

$$-\left(\frac{y}{x}\right)^{-1} = \frac{x}{y}.$$

Sie sind Lösungen der Differentialgleichung

$$y' = \frac{x}{y}.$$

Wir lösen sie durch Trennung der Variablen:

$$yy' = x, \quad \frac{1}{2}y^2 = \frac{1}{2}x^2 + c_1, \quad y^2 = x^2 + c.$$

Hier kann man entweder nach  $y$  auflösen

$$y = \pm\sqrt{x^2 + c}$$

oder die Gleichung als Kegelschnitt-Gleichung umschreiben

$$y^2 - x^2 = c, \quad c \in \mathbb{R}.$$

Die orthogonalen Trajektorien sind achsenparallele Hyperbeln, abhängig vom Vorzeichen der Konstante  $c$  nach oben oder seitwärts geöffnet.

Mindestens genauso wichtig wie geometrische Anwendungen waren für die Entwicklung der Theorie der Differentialgleichungen die physikalischen Anwendungen. Begonnen hat das mit den Bewegungsgleichungen von Newton. Die Gleichungen in der Mechanik enthalten allerdings meist die Beschleunigung, d.h., die zweite Ableitung nach der Zeit. Deswegen ein einfaches Beispiel aus der Elektrizitätslehre.

**Beispiel 3.7** Ein Stromkreislauf enthalte einen Widerstand  $R$  und eine Induktivität  $L$ . Angelegt werde eine Wechselstromspannung

$$U(t) = U_0 \cdot \sin(\omega \cdot t).$$

Dann gilt also

$$U(t) = R \cdot I + L \cdot \frac{dI}{dt}.$$

Gesucht ist die Stromstärke  $I(t)$ . Sie erfüllt die Differentialgleichung

$$\frac{dI}{dt} = -\frac{R}{L}I + \frac{U_0}{L} \cdot \sin(\omega \cdot t).$$

Dies ist eine inhomogene lineare Differentialgleichung. Die zugehörige homogene Gleichung hat die Lösungen

$$I(t) = I_0 \cdot e^{-\frac{R}{L}t}.$$

Wir ermitteln eine spezielle Lösung der inhomogenen Gleichung durch Variation der Konstanten:

$$I_1(t) = c(t) \cdot e^{-\frac{R}{L}t}.$$

Wir finden

$$\frac{dI_1}{dt} = \dot{c} \cdot e^{-\frac{R}{L}t} - \frac{R}{L} \cdot c \cdot e^{-\frac{R}{L}t} = -\frac{R}{L} \cdot I + \frac{U_0}{L} \cdot \sin(\omega \cdot t).$$

Für die gesuchte Funktion  $c(t)$  ergibt sich

$$\dot{c} = \frac{U_0}{L} \cdot e^{\frac{R}{L}t} \cdot \sin(\omega \cdot t), \quad \text{bzw.} \quad c(t) = \frac{U_0}{L} \cdot \int e^{\frac{R}{L}t} \cdot \sin(\omega \cdot t) dt.$$

Zur Abkürzung setzen wir  $\alpha := R/L$  und finden in der Formelsammlung eine Stammfunktion

$$\int e^{\alpha t} \cdot \sin(\omega \cdot t) dt = e^{\alpha t} \left( \frac{\alpha}{\alpha^2 + \omega^2} \sin(\omega \cdot t) - \frac{\omega}{\alpha^2 + \omega^2} \cos(\omega \cdot t) \right).$$

Den Klammerausdruck verschönern wir etwas:

$$\frac{1}{\sqrt{\alpha^2 + \omega^2}} \cdot (\cos(\gamma) \sin(\omega \cdot t) + \sin(\gamma) \cos(\omega \cdot t))$$

mit

$$\cos(\gamma) = \frac{\alpha}{\sqrt{\alpha^2 + \omega^2}}, \quad \sin(\gamma) = \frac{-\omega}{\sqrt{\alpha^2 + \omega^2}}.$$

Dann erhalten wir

$$c(t) = \frac{U_0}{L} \cdot \frac{e^{\frac{R}{L}t}}{\sqrt{\frac{R^2}{L^2} + \omega^2}} \cdot \sin(\omega \cdot t - \gamma)$$

und

$$I_1(t) = \frac{U_0}{\sqrt{R^2 + \omega^2 L^2}} \sin(\omega \cdot t - \gamma).$$

Insgesamt erhalten wir in Abhängigkeit von  $I_0$  die Lösungen

$$I(t) = I_0 \cdot e^{-\frac{R}{L}t} + \frac{U_0}{\sqrt{R^2 + \omega^2 L^2}} \sin(\omega \cdot t - \gamma).$$

**Aufgabe 3.1** Der radioaktive Zerfall einer Substanz wird durch die Differentialgleichung

$$\frac{dN}{dt} = -\lambda N$$

beschrieben. Dabei ist  $N$  die Anzahl der Teilchen und  $1/\lambda$  die mittlere Lebensdauer eines Teilchens. Lösen Sie die Gleichung und geben Sie die Beziehung zwischen  $\lambda$  und der Halbwertszeit an. (Die Halbwertszeit ist die Zeit, nach der die Hälfte der Teilchen zerfallen ist.)

**Aufgabe 3.2** Die Vermehrung der Individuen einer biologischen Population ist proportional zur Anzahl  $N$  der Individuen. Bei ungehemmtem Wachstum wird sie durch die Gleichung

$$\frac{dN}{dt} = a \cdot N$$

beschrieben. Ein Modell für gehemmtes Wachstum, bei dem die Zahl der Individuen die Schranke  $b$  nicht übersteigen kann, ist

$$\frac{dN}{dt} = a \cdot \left(1 - \frac{N}{b}\right) \cdot N.$$

Lösen Sie diese Gleichung.

**Aufgabe 3.3** Ein Körper bewege sich linear mit der Geschwindigkeit  $v = v(t)$  wobei  $v(0) = v_0$ . Die Bewegung werde durch Reibung abgebremst vermöge

$$\dot{v} = -\mu \cdot v^{1+\alpha}, \quad \mu > 0, \alpha \geq -1.$$

Berechnen Sie die Geschwindigkeit  $v(t)$ . (Unterscheiden Sie dabei zwischen  $\alpha = 0$  und  $\alpha \neq 0$ .)

**Aufgabe 3.4** Lösen Sie die folgenden Differentialgleichungen und bestimmen Sie die Existenzintervalle der Lösungen:

a)  $\frac{dx}{dt} = \sqrt{1-x^2} \quad (|x| \leq 1),$

b)  $\frac{dx}{dt} = \frac{1}{x} \sqrt{1-x^2} \quad (0 < |x| < 1),$

c)  $\frac{dx}{dt} = (1+x)(1+t).$

**Aufgabe 3.5** Lösen Sie die folgenden Differentialgleichungen:

a)  $\dot{x} = e^{2t-x},$  b)  $\dot{x} = 2tx^2,$  c)  $\dot{x} = \frac{t}{x}.$

**Aufgabe 3.6** Lösen Sie die folgenden Differentialgleichungen:

a)  $\dot{x} = -4x,$  b)  $\dot{x} = 2tx,$  c)  $\dot{x} = -\frac{2x}{t},$  d)  $\dot{x} = \frac{tx}{1-t^2}.$

**Aufgabe 3.7** Lösen Sie die folgenden Differentialgleichungen:

a)  $\dot{x} = -4x + t^2,$  b)  $\dot{x} = 2tx + t^4,$  c)  $\dot{x} = \frac{tx-1}{1-t^2}.$

### 3.2 Theorie: Existenz und Eindeutigkeit, Systeme

Ich habe mich bisher nicht darum gekümmert zu erklären, was es heißt, eine Differentialgleichung zu lösen. Ich habe nicht einmal genau gesagt, was eine Differentialgleichung ist (Learning by doing). Das wird jetzt nachgeholt.

**Definition 3.1** *Es sei  $U \subset \mathbb{R}^2$  offen und  $f : U \rightarrow \mathbb{R}$  eine reelle Funktion auf  $U$ . Dann heißt*

$$\dot{x} = f(t, x)$$

*eine Differentialgleichung auf  $U$ . Eine Lösung dieser Differentialgleichung auf einem Intervall  $]a, b[ \subset \mathbb{R}$  ist eine differenzierbare Funktion  $x(t)$ ,  $t \in ]a, b[$ , deren Graph  $\{(t, x(t)) : a < t < b\}$  in  $U$  enthalten ist, und die auf ihrem Definitionsintervall die Beziehung*

$$\frac{dx}{dt}(t) = f(t, x(t))$$

*erfüllt.*

*Das Anfangswert-Problem (AWP) zum Anfangswert  $(t_0, x_0) \in U$  besteht darin, eine Lösung  $x(t)$  zu finden, deren Definitionsintervall den Wert  $t_0$  enthält, und die*

$$x(t_0) = x_0$$

*erfüllt.*

Wir machen die Generalvoraussetzung, dass die rechte Seite unserer Differentialgleichung stetig ist. Aber für Eindeutigkeitsausagen genügt das noch nicht.

**Beispiel 3.8** *Wir betrachten die Funktionen  $x(t) = (t - c)^3$ ,  $c \in \mathbb{R}$ . Sie sind auf ganz  $\mathbb{R}$  Lösungen der Differentialgleichung*

$$\dot{x} = 3 \cdot x^{2/3}$$

*mit stetiger rechter Seite. Durch einen Punkt  $(t_0, 0)$  auf der  $t$ -Achse geht aber nicht nur die Funktion  $x(t) = (t - t_0)^3$ , sondern auch die Null-Lösung  $x(t) \equiv 0$ .*

Der Grund für diese Pathologie besteht darin, dass für  $x \rightarrow 0$  die Ableitung der rechten Seite

$$\frac{2}{x^{1/3}}$$

gegen unendlich geht. Die rechte Seite ist keine differenzierbare Funktion von  $x$ .

Wir brauchen eine schärfere Voraussetzung an die Funktion  $f(t, x)$  auf der rechten Seite.

**Definition 3.2** *Die Funktion  $f : U \rightarrow \mathbb{R}$  heißt Lipschitz-stetig mit der Lipschitz-Konstante  $L \in \mathbb{R}$ , wenn für alle  $(t, x_1)$  und  $(t, x_2) \in U$  gilt:*

$$|f(t, x_1) - f(t, x_2)| \leq L \cdot |x_1 - x_2|.$$

*Die Funktion  $f$  genügt auf  $U$  lokal einer Lipschitz-Bedingung, wenn zu jedem  $(t_0, x_0) \in U$  ein Quadrat*

$$V = \{(t, x) \in \mathbb{R}^2 : |t - t_0| < r, |x - x_0| < r\} \subset U$$

*mit  $r > 0$  und eine Lipschitz-Konstante  $L_V$  existieren, so, dass  $f$  auf  $V$  Lipschitz-stetig mit der Lipschitz-Konstante  $L_V$  ist.*

**Beispiel 3.9** Ist die Funktion  $f(t, x)$  auf  $U$  stetig partiell nach  $x$  differenzierbar, dann genügt  $f$  auf  $U$  lokal einer Lipschitz-Bedingung.

*Beweis.* Zu  $(t_0, x_0) \in U$  existiert ein kompaktes Quadrat

$$\bar{V} = \{(t, x) \in \mathbb{R}^2 : |t - t_0| < r, |x - x_0| < r\} \subset U, \quad r > 0.$$

Die stetige Funktion  $|\partial f / \partial x(t, x)|$  ist auf der kompakten Menge  $V$  beschränkt, sei etwa  $|\partial f / \partial x| \leq L$  auf  $\bar{V}$ . Für festes  $t$  mit  $|t - t_0| \leq r$  wenden wir den Mittelwertsatz der Differentialrechnung an auf die Funktion

$$x \mapsto f(t, x), \quad |x - x_0| \leq r.$$

Zu  $x_1$  und  $x_2$  mit  $|x_i - x_0| \leq r$  für  $i = 1, 2$  gibt es deswegen ein  $\xi$  zwischen  $x_1$  und  $x_2$  mit

$$|f(t, x_1) - f(t, x_2)| = \left| \frac{\partial f}{\partial x}(t, \xi) \cdot (x_1 - x_2) \right| \leq L \cdot |x_1 - x_2|. \quad \square$$

Unser Hauptziel in diesem Abschnitt ist der Beweis für folgenden Satz:

**Satz 3.1 (Existenz- und Eindeutigkeitsatz von Picard - Lindelöf)** Die stetige Funktion  $f : U \rightarrow \mathbb{R}$  genüge auf der offenen Menge  $U \subset \mathbb{R}^2$  lokal einer Lipschitz-Bedingung. Dann existiert durch jeden Anfangswert  $(t_0, x_0) \in U$  eine lokale Lösung  $x(t)$  der Differentialgleichung

$$\dot{x} = f(t, x),$$

und diese ist durch  $(t_0, x_0)$  eindeutig bestimmt.

Zunächst präzisieren wir, was *lokale*, eindeutig bestimmte Lösung heißen soll: Es gibt ein  $\epsilon > 0$  und auf den Intervall  $]t_0 - \epsilon, t_0 + \epsilon[$  eine differenzierbare Funktion  $x(t)$  so, dass für alle  $t$  in diesem Intervall gilt

$$(t, x(t)) \in U \quad \text{und} \quad \dot{x}(t) = f(t, x(t)).$$

Und falls  $y(t)$  eine weitere Funktion mit diesen Eigenschaften ist, dann gilt  $x(t) = y(t)$  für alle  $t \in ]t_0 - \epsilon, t_0 + \epsilon[$ .

Beweis des Satzes: Weil  $U$  offen ist, gibt es ein  $r > 0$  so, dass das Quadrat

$$V := \{(t, x) \in \mathbb{R}^2 : |t - t_0| \leq r, |x - x_0| \leq r\}$$

ganz zu  $U$  gehört. Um Schreibarbeit zu sparen, verwende ich auch die Notation

$$I := [t_0 - r, t_0 + r], \quad J := [x_0 - r, x_0 + r], \quad V = I \times J.$$

Weil  $f$  lokal einer Lipschitz-Bedingung genügen soll, können wir außerdem  $r > 0$  so klein wählen, dass für alle  $(t, x_1), (t, x_2) \in V$  gilt

$$|f(t, x_1) - f(t, x_2)| \leq L \cdot |x_1 - x_2| \leq L \cdot r.$$

Die Beweis-Idee besteht darin, die Differentialgleichung in eine Integralgleichung umzuwandeln. Dazu betrachten wir eine stetige Funktion  $x : I \rightarrow J$  und zeigen die Äquivalenz der beiden folgenden Eigenschaften:

- a)  $x$  ist auf  $]t_0 - r, t_0 + r[$  differenzierbar mit  $x(t_0) = x_0$  und  $\dot{x}(t) = f(t, x(t))$  für alle  $t$  mit  $|t - t_0| < r$ .

b)  $x$  genügt auf  $I$  der Integralgleichung

$$x(t) = x_0 + \int_{t_0}^t f(u, x(u)) du.$$

Beweis von a)  $\Rightarrow$  b): Für  $|t - t_0| < r$  ist nach dem HDI

$$x(t) = x(t_0) + \int_{t_0}^t \dot{x}(u) du = x_0 + \int_{t_0}^t f(u, x(u)) du.$$

Weil die linke und die rechte Seite dieser Gleichung stetig in  $t$  sind, gilt die Gleichung auch noch in den Randpunkten  $t_0 \pm r$ .

Beweis von b)  $\Rightarrow$  a): Im offenen Intervall ist das Integral nach seiner oberen Grenze differenzierbar mit Ableitung  $f(t, x(t))$ . Dann ist auch  $x$  differenzierbar mit dieser Ableitung. Und  $x(t_0) = x_0$  ist offensichtlich.

Und die Integralgleichung lösen wir durch Anwendung des Banachschen Fixpunktsatzes auf die *Picard-Abbildung*

$$P : C^0(I, J) \rightarrow C^0(I, \mathbb{R}), \quad P : x(t) \mapsto x_0 + \int_{t_0}^t f(u, x(u)) du.$$

Zuerst untersuchen wir, ob  $P(x)$  wieder in  $C^0(I, J)$  liegt: Dazu setzen wir

$$M := \max_{(t,x) \in V} |f(t, x)|$$

und schätzen ab

$$|P(x)(t) - x_0| = \left| \int_{t_0}^t f(u, x(u)) du \right| \leq \left| \int_{t_0}^t |f(u, x(u))| du \right| \leq M \cdot |t - t_0|.$$

Hier ist  $M \cdot |t - t_0| \leq r$ , falls

$$|t - t_0| \leq r_1 := \min\left\{r, \frac{r}{M}\right\}.$$

Wir müssen also  $I$  ersetzen durch  $I_1 := [t_0 - r_1, t_0 + r_1]$ , aber dann haben wir für die Picard-Abbildung

$$P : C^0(I_1, J) \rightarrow C^0(I_1, J).$$

Jetzt müssen wir noch zeigen, dass  $P$  kontrahierend ist. Wir schätzen ab:

$$\begin{aligned} \|P(x) - P(y)\|_{I_1} &= \max_{t \in I_1} |P(x)(t) - P(y)(t)| \\ &= \max_{t \in I_1} \left| \int_{t_0}^t f(u, x(u)) - f(u, y(u)) du \right| \\ &\leq \max_{t \in I_1} \left| \int_{t_0}^t |f(u, x(u)) - f(u, y(u))| du \right| \\ &\leq \max_{t \in I_1} \left| \int_{t_0}^t L \cdot |x(u) - y(u)| du \right| \\ &\leq |t - t_0| \cdot L \cdot \|x - y\|_{I_1}. \end{aligned}$$

Wir müssen  $r_1$  noch weiter verkleinern zu

$$r_2 := \min\left\{r_1, \frac{1}{2L}\right\}$$

und übergehen zum Intervall  $I_2 := [t_0 - r_2, t_0 + r_2] \subset I_1$ . Dafür haben wir dann bewiesen: Die Abbildung

$$P : C^0(I_2, J) \rightarrow C^0(I_2, J)$$

ist kontrahierend mit Lipschitz-Konstante  $1/2$ . Der Banachsche Fixpunktsatz liefert die Existenz- und die Eindeutigkeitsaussage.  $\square$

**Beispiel 3.10 (Forster II (10.2), Kamke Nr. 32)** *Wir betrachten die homogene lineare Gleichung*

$$\dot{x} = 2t \cdot x.$$

*Mit dem Näherungsverfahren von Picard-Lindelöf wollen wir die Lösung  $x(t)$  des AWP  $x(0) = c$  bestimmen. Es ist also  $x_0 \equiv c$ . Und die Integralformel für die sukzessive Bestimmung der Näherungen  $x_k$  lautet*

$$x_{k+1}(t) = c + \int_0^t 2u \cdot x_k(u) du.$$

*Damit erhalten wir*

$$\begin{aligned} x_1(t) &= c + \int_0^t 2u \cdot c du = c \cdot (1 + t^2) \\ x_2(t) &= c + \int_0^t 2u \cdot c \cdot (1 + u^2) du = c \cdot \left(1 + x^2 + \frac{x^4}{2}\right) \end{aligned}$$

*Und durch Induktion könnten wir leicht zeigen*

$$x_k(t) = c \cdot \left(1 + t^2 + \frac{t^4}{2} + \dots + \frac{t^{2k}}{k!}\right) = c \cdot \sum_{\nu=0}^k \frac{t^{2\nu}}{\nu!}.$$

*Das sind die Partialsummen der Potenzreihe für die Lösung  $x(t) = c \cdot e^{x^2}$ .*

Jetzt noch ein Wort zur 'lokalen Existenz' der Lösung  $x(t)$ . Selbst wenn die rechte Seite auf ganz  $\mathbb{R}^2$  definiert ist, brauchen die Lösungen nicht auf ganz  $\mathbb{R}$  definiert zu sein. Sie können etwa nach  $\infty$  abhauen.

**Beispiel 3.11** *Wir betrachten die Differentialgleichung*

$$\dot{x} = -x^2.$$

*Sie ist nicht linear, aber lokal Lipschitz, weil  $f(t, x) = -x^2$  nach  $x$  stetig partiell differenzierbar ist. Etwa mit Trennung der Variablen findet man die Lösungen*

$$x(t) = \frac{1}{t - c}, \quad c \in \mathbb{R}.$$

*Jede dieser Lösungen ist nur auf dem Intervall  $] -\infty, c[$  oder  $]c, \infty[$  definiert, nicht auf ganz  $\mathbb{R}$ . Und nimmt man noch die Null-Lösung  $x(t) \equiv 0$  hinzu, so hat man durch jeden Punkt  $(t, x) \in \mathbb{R}^2$  eine Lösung angegeben. Nach dem Eindeutigkeitsanteil von Satz 3.1 kann es nicht noch mehr Lösungen geben.*

Es liegt also in der Natur der Sache, dass eine Lösung möglicherweise nur auf einem Intervall definiert ist. Das Wörtchen 'lokal' bei der Existenz-Aussage ist somit unvermeidbar. Anders ist es mit der Eindeutigkeit:

**Satz 3.2 (Globale Eindeutigkeit)** *Mit den Voraussetzungen von Satz 3.1 seien  $x_1, x_2 : ]a, b[ \rightarrow \mathbb{R}$  zwei Lösungen der Differentialgleichung  $\dot{x} = f(t, x)$ . Gilt  $x_1(t_0) = x_2(t_0)$  für ein  $t_0 \in ]a, b[$ , so ist  $x_1(t) = x_2(t)$  für alle  $t \in ]a, b[$ .*

Zwei Lösungen des gleichen Anfangswertproblems stimmen also auf ihrem gemeinsamen Existenzintervall überein.

Beweis des Satzes. Wir zeigen die Identität  $x_1(t) = x_2(t)$  für alle  $t$  mit  $t_0 \leq t < b$ , rechts von  $t_0$ . Links von  $t_0$  geht der Beweis analog.

Es sei

$$T := \{\tau \in [t_0, b[ : x_1(t) = x_2(t) \text{ für alle } t \text{ mit } t_0 \leq t \leq \tau\}.$$

Die Menge  $T$  ist durch  $b$  nach oben beschränkt und wegen  $t_0 \in T$  nicht leer. Also existiert  $t_1 := \sup(T)$ . Wenn  $t_1 < b$  sein sollte, existiert nach Satz 3.1 ein  $r > 0$  derart, dass  $x_1$  und  $x_2$  auch noch auf dem Intervall  $[t_0, t_0 + r]$  übereinstimmen. Das ist ein Widerspruch zur Wahl von  $t_1$ . Also muss  $t_1 = b$  gelten, und der Satz ist bewiesen.  $\square$

**Definition 3.3** *Eine Lösung der Differentialgleichung  $\dot{x} = f(t, x)$  auf einem Intervall  $]a, b[$  heißt maximal, wenn man sie auf kein größeres Intervall als Lösung forsetzen kann. Eine Differentialgleichung zu lösen, das heißt all ihre maximalen Lösungen mit deren Existenzintervallen anzugeben.*

In der Linearen Algebra lernt man mit Systemen linearer Gleichungen umzugehen. Da hat man nicht mehr eine, sondern mehrere Unbekannte. Genauso gibt es Systeme von Differentialgleichungen. Da hat man nicht mehr nur eine, sondern mehrere unbekannte Funktionen. Ein System von Differentialgleichungen erster Ordnung sieht so aus:

$$\begin{aligned} \dot{x}_1 &= f_1(t, x_1, \dots, x_n) \\ \dot{x}_2 &= f_2(t, x_1, \dots, x_n) \\ &\vdots \\ \dot{x}_n &= f_n(t, x_1, \dots, x_n) \end{aligned}$$

Da sind dann also  $n$  differenzierbare Lösungsfunktionen zu suchen. Die Definitionsmenge  $U$  der rechten Seite ist eine Teilmenge des  $\mathbb{R}^{n+1}$ . Eine Lösung des Systems ist ein  $n$ -tupel  $(x_1, \dots, x_n)$  von Funktionen, die auf einem gemeinsamen Intervall  $]a, b[ \subset \mathbb{R}$  definiert sind, und zusammen die Gleichungen des Systems erfüllen. Fass man sie zu einer Vektorwertigen Funktion  $\mathbf{x}(t) = (x_1(t), \dots, x_n(t))$  zusammen, so kann man das System auch in Vektorform

$$\dot{\mathbf{x}} = \mathbf{f}(t, \mathbf{x})$$

schreiben.

**Beispiel 3.12** *So ziemlich das einfachste Beispiel ist das System*

$$\begin{aligned} \dot{x}_1 &= x_2 \\ \dot{x}_2 &= -x_1 \end{aligned}$$

Die Definitionsmenge  $U$  ist der ganze  $\mathbb{R}^3$ . Eine Lösung ist etwa

$$\mathbf{x}(t) = \begin{pmatrix} \sin(t) \\ \cos(t) \end{pmatrix}$$

definiert auf der ganzen reellen Achse.

Ein AWP ist hier folgendes: Vorgegeben wird ein Punkt  $(t_0, \mathbf{a}) \in U$  und gesucht ist eine Vektorwertige Funktion  $\mathbf{x}(t)$  mit  $\mathbf{x}(t_0) = \mathbf{a}$ .

**Beispiel 3.13** Das System in Beispiel 3.12 hat nicht nur die angegebene Lösung

$$\mathbf{x}_1(t) = \begin{pmatrix} \sin(t) \\ \cos(t) \end{pmatrix},$$

sondern auch

$$\mathbf{x}_2(t) = \begin{pmatrix} -\cos(t) \\ \sin(t) \end{pmatrix}.$$

Und auch jede Linearkombination  $\mathbf{x}(t) = c_1\mathbf{x}_1(t) + c_2\mathbf{x}_2(t)$  ist eine Lösung. Um etwa das AWP  $\mathbf{x}(0) = \mathbf{a}$  zu lösen, genügt es  $c_1, c_2 \in \mathbb{R}$  zu finden, so, dass

$$c_1\mathbf{x}_1(0) + c_2\mathbf{x}_2(0) = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$$

ist. Die Koeffizienten  $c_2 = -a_1$  und  $c_1 = a_2$  tun's.

Die bisher für eine Differentialgleichung entwickelte Theorie geht - mit einigen Modifikationen - genauso für Systeme. Ich möchte das nicht ganz ausführen, sondern nur die Modifikationen angeben.

**Definition 3.4** Die Vektorwertige Funktion  $\mathbf{f}(t, \mathbf{x})$  genügt auf  $U \subset \mathbb{R}^{n+1}$  einer Lipschitz-Bedingung mit Lipschitz-Konstante  $L$ , wenn für je zwei Vektoren

$$(t, \mathbf{x}_1), (t, \mathbf{x}_2) \in U$$

gilt

$$\| \mathbf{f}(t, \mathbf{x}_1) - \mathbf{f}(t, \mathbf{x}_2) \| \leq L \cdot \| \mathbf{x}_1 - \mathbf{x}_2 \| .$$

Man braucht hier also nur die Absolut-Striche durch die Vektor-Norm zu ersetzen. Und genau wie in Dimension eins (Beispiel 3.9) beweist man:

**Satz 3.3** Die Komponentenfunktionen  $f_1(t, \mathbf{x}), \dots, f_n(t, \mathbf{x})$  der Vektorwertigen Funktion  $\mathbf{f}(t, \mathbf{x})$  seien auf ihrer Definitionsmenge  $U \subset \mathbb{R}^{n+1}$  stetig partiell differenzierbar nach  $x_1, \dots, x_n$ . Dann genügt  $\mathbf{f}$  auf  $U$  lokal einer Lipschitzbedingung.

Beim Beweis muss man nur den MWS der Differentialrechnung durch die Abschätzung Satz 2.7 ersetzen. Und genau wie in Dimension eins beweist man auch

**Satz 3.4 (Picard-Lindelöf für Systeme)** Die Funktion  $\mathbf{f}(t, \mathbf{x})$  genüge auf ihrem Definitionsbereich  $U \subset \mathbb{R}^{n+1}$  lokal einer Lipschitz-Bedingung. Dann besitzt jedes AWP eine lokale Lösung, und je zwei Lösungen desselben AWP stimmen auf ihrem gemeinsamen Definitionsintervall überein.

Der Beweis mit Picard-Iteration und dem Banachschen Fixpunktsatz geht genau wie in einer Dimension. Überall muss man die skalare Funktion  $f(t, x)$  durch die Vektorwertige Funktion  $\mathbf{f}(t, \mathbf{x})$  ersetzen. Nur an zwei Stellen muss man etwas aufpassen: Bei der Definition des Integrals für Vektorwertige Funktionen und bei der Integralabschätzung für diese Funktionen.

**Definition 3.5 (Komponentenweise Integration)** Die Funktionen  $f_1, \dots, f_n : [a, b] \rightarrow \mathbb{R}$  seien Riemann-integrierbar. Dann definiert man

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

**Satz 3.5 (Integralabschätzung)** Die Abbildung  $\mathbf{f} : [a, b] \rightarrow \mathbb{R}^n$  sei integrierbar im Sinn von Definition 3.5. Dann gilt

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

Beweis. Mit der Integralabschätzung für skalare Funktionen und der Cauchy-Schwarz-Ungleichung folgt für jeden Vektor  $\mathbf{v} \in \mathbb{R}^n$

$$\left| \int_a^b (\mathbf{f}(t), \mathbf{v}) dt \right| \leq \int_a^b |(\mathbf{f}(t), \mathbf{v})| dt \leq \int_a^b \|\mathbf{f}(t)\| \cdot \|\mathbf{v}\| dt = \int_a^b \|\mathbf{f}(t)\| dt \cdot \|\mathbf{v}\|.$$

Für den Integralvektor

$$\mathbf{v} = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} = \begin{pmatrix} \int_a^b f_1(t) dt \\ \vdots \\ \int_a^b f_n(t) dt \end{pmatrix}$$

ist das Skalarprodukt

$$(\mathbf{f}(t), \mathbf{v}) = \sum_{\nu=1}^n f_\nu(t) \cdot \int_a^b f_\nu(t) dt$$

und das Integral darüber

$$\int_a^b (\mathbf{f}(t), \mathbf{v}) dt = \sum_{\nu=1}^n \left( \int_a^b f_\nu(t) dt \right)^2 = \left\| \int_a^b \mathbf{f}(t) dt \right\|^2.$$

Die oben angegebene Ungleichung wird hier

$$\left\| \int_a^b \mathbf{f}(t) dt \right\|^2 \leq \int_a^b \|\mathbf{f}(t)\| dt \cdot \left\| \int_a^b \mathbf{f}(t) dt \right\|.$$

Die Behauptung folgt, wenn wir durch den zweiten Faktor auf der rechten Seite kürzen können. Wenn wir aber nicht kürzen können, ist

$$\left\| \int_a^b \mathbf{f}(t) dt \right\| = 0,$$

und die Behauptung ist trivialerweise richtig. □

Die hier für Systeme von Differentialgleichungen entwickelte Theorie gilt wörtlich auch für Differentialgleichungen höherer Ordnung. Eine Differentialgleichung  $n$ -ter Ordnung hat die Form

$$x^{(n)}(t) = f(t, x, \dot{x}, \dots, x^{(n-1)}).$$

Dabei bezeichnet  $x^{(n)}$  die  $n$ -te Ableitung der Funktion  $x$  nach  $t$ . Die rechte Seite  $f$  ist definiert auf einer offenen Menge  $U \subset \mathbb{R}^{n+1}$ . Wie in Abschnitt 1.4 ist auch diese Differentialgleichung  $n$ -ter Ordnung äquivalent zu einem System aus  $n$  Differentialgleichungen für  $n$  Funktionen. Wir setzen

$$x_1 = x, x_2 = \dot{x}, \dots, x_n = x^{(n-1)}.$$

Dafür erhalten wir das System

$$\begin{aligned} \dot{x}_1 &= x_2, \\ \dot{x}_2 &= x_3, \\ &\vdots \\ \dot{x}_n &= f(t, x_1, \dots, x_n). \end{aligned}$$

Aus dem entsprechenden Satz für Systeme folgt

**Satz 3.6** Die Funktion  $f(t, x, \dot{x}, \dots, x^{(n-1)})$  sei stetig differenzierbar auf der offenen Menge  $U \subset \mathbb{R}^{n+1}$ . Dann gibt es zu jedem Anfangswert  $(t_0, \mathbf{a} = (a_0, \dots, a_{n-1})) \in U$  lokal eine Lösung der Differentialgleichung  $x^{(n)} = f(t, x, \dots, x^{(n-1)})$  mit

$$x(t_0) = a_0, \dot{x}(t_0) = a_1, \dots, x^{(n-1)}(t_0) = a_{n-1}.$$

**Aufgabe 3.8** Lösen Sie das System

$$\begin{aligned} \dot{x}_1 &= x_1^2 + x_1 \cdot x_2 \cdot e^{-t^2} \\ \dot{x}_2 &= 2t \cdot x_2 \end{aligned}$$

und dafür die Anfangswertprobleme  $\mathbf{x}(0) = (1, 1)$ , bzw.  $\mathbf{x}(0) = (1, 0)$ .

**Aufgabe 3.9** a) Lösen Sie das System

$$\begin{aligned} \dot{x}_1 &= 2t \cdot x_2 \\ \dot{x}_2 &= -2t \cdot x_1 \end{aligned}$$

indem Sie  $x_1 \dot{x}_1$  und  $x_2 \dot{x}_2$  vergleichen.

b) Lösen Sie das AWP  $x_1(0) = x_2(0) = 1$ .

**Aufgabe 3.10** Finden Sie alle Lösungen des Systems

$$\begin{aligned} \dot{x}_1 &= \cos(t) \cdot x_1 + \sin(t) \cdot x_2 \\ \dot{x}_2 &= \cos(t) \cdot x_2 \end{aligned}$$

**Aufgabe 3.11** Zeigen Sie dass die Vektorfunktionen

$$\mathbf{x}_1(t) = \begin{pmatrix} \cosh(t) \\ \sinh(t) \end{pmatrix}, \quad \mathbf{x}_2(t) = \begin{pmatrix} \sinh(t) \\ \cosh(t) \end{pmatrix}$$

linear unabhängig sind und geben Sie ein lineares  $2 \times 2$ -Differentialgleichungssystem an, welches  $\mathbf{x}_1(t)$  und  $\mathbf{x}_2(t)$  als Lösungen besitzt.

### 3.3 Lineare Systeme

Hier betrachten wir Systeme von Differentialgleichungen der Form

$$\begin{aligned} \dot{x}_1 &= a_{1,1}(t)x_1 + a_{1,2}(t)x_2 + \dots + a_{1,n}(t)x_n + b_1(t) \\ \dot{x}_2 &= a_{2,1}(t)x_1 + a_{2,2}(t)x_2 + \dots + a_{2,n}(t)x_n + b_2(t) \\ &\vdots \\ \dot{x}_n &= a_{n,1}(t)x_1 + a_{n,2}(t)x_2 + \dots + a_{n,n}(t)x_n + b_n(t) \end{aligned}$$

Fasst man die Koeffizientenfunktionen zu einer  $n \times n$ -Matrix  $A(t) = (a_{\mu,\nu}(t))$  und die Inhomogenitäten zu einem Vektor  $\mathbf{b}(t) = (b_\mu(t))$  zusammen, so kann man es in Vektorform

$$\dot{\mathbf{x}}(t) = A(t) \cdot \mathbf{x}(t) + \mathbf{b}(t)$$

schreiben. Das sieht formal ganz so aus wie eine (inhomogene) lineare Gleichung erster Ordnung. Deswegen geht auch manches ganz genauso wie im Fall  $n = 1$ .

**Satz 3.7 (Struktursatz)** Die Matrix  $A(t)$  und der Vektor  $\mathbf{b}(t)$  seien stetig auf einem Intervall  $I = ]\alpha, \beta[ \subset \mathbb{R}$ .

a) Die Lösungen  $\mathbf{x} : I \rightarrow \mathbb{R}^n$  des homogenen Systems  $\dot{\mathbf{x}} = A \cdot \mathbf{x}$  bilden einen  $\mathbb{R}$ -Vektorraum.

b) Die Lösungen des inhomogenen Systems erhält man, indem man zu einer speziellen Lösung alle Lösungen des homogenen Systems addiert.

Satz 3.7 sagt noch nichts darüber aus, ob Lösungen überhaupt global existieren. Die Eindeutigkeit ist jedoch durch Satz 3.4 geklärt.

**Satz 3.8 (Globale Existenz)** Es seien  $A(t)$  und  $I \subset \mathbb{R}$  wie eben. Weiter sei  $t_0 \in I$  und  $\mathbf{a} \in \mathbb{R}^n$  beliebig. Dann besitzt das AWP

$$\dot{\mathbf{x}} = A \cdot \mathbf{x}, \quad \mathbf{x}(t_0) = \mathbf{a}$$

eine auf dem ganzen Intervall  $I$  definierte (und eindeutig bestimmte) Lösung.

Beweis. Wieder verwenden wir den Banachschen Fixpunktsatz und die Picard-Abbildung

$$P : C^0(I, \mathbb{R}^n) \rightarrow C^0(I, \mathbb{R}^n), \quad \mathbf{x}(t) \mapsto \mathbf{a} + \int_{t_0}^t A(u) \cdot \mathbf{x}(u) du.$$

Weil man  $I$  durch kompakte Teilintervalle ausschöpfen kann, genügt es einen Fixpunkt der Picard-Abbildung in dem Fall zu finden, wo das Intervall  $I$  kompakt ist. Dann ist die stetige Funktion  $\|A(t)\|$  auf diesem Intervall beschränkt, und es gibt eine Konstante  $L \in \mathbb{R}$  mit

$$\|A(t)\| \leq L, \quad \|A(t) \cdot \mathbf{x}(t)\| \leq L \cdot \|\mathbf{x}(t)\|$$

für alle  $t \in I$ . Damit folgt für alle  $\mathbf{x}, \mathbf{y} \in C^0(I, \mathbb{R}^n)$

$$\begin{aligned} \|P(\mathbf{x}) - P(\mathbf{y})\| &= \left\| \int_{t_0}^t A(u) \cdot (\mathbf{x}(u) - \mathbf{y}(u)) du \right\| \leq \\ &\leq \left| \int_{t_0}^t \|A(u) \cdot (\mathbf{x}(u) - \mathbf{y}(u))\| du \right| \leq |t - t_0| \cdot L \cdot \|\mathbf{x} - \mathbf{y}\|_I. \end{aligned}$$

Leider reicht das noch nicht aus, um zu beweisen, dass  $P$  kontrahierend ist. Nur wenn  $|t - t_0| < 1/L$  wäre, hätten wir dieses Ziel erreicht. Pech.

Aber nicht nur Glück, auch Beharrlichkeit führt zum Ziel. Mit  $r := 1/2L$  ist die Abbildung kontrahierend auf dem Teilintervall  $I \cap \{|t - t_0| \leq r\}$ . Eine Lösung existiert also mindestens auf diesem Teilintervall. Und jetzt machen wir weiter, wie im Beweis von Satz 3.2. Wir stückeln Teilintervalle der Länge  $r$  aneinander und schöpfen damit irgend wann das ganze Intervall  $I$  aus.  $\square$

Satz 3.8 kann man auch so ausdrücken:

**Satz 3.9** *Es seien  $A(t)$ ,  $I \subset \mathbb{R}$  und  $t_0 \in I$  wie in Satz 3.8. Weiter sei  $V$  der  $\mathbb{R}$ -Vektorraum der Lösungen  $\mathbf{x}(t)$  des Systems  $\dot{\mathbf{x}}(t) = A(t) \cdot \mathbf{x}(t)$  auf  $I$ . Dann ist der Homomorphismus*

$$V \ni \mathbf{x} \mapsto \mathbf{x}(t_0) \in \mathbb{R}^n$$

*bijektiv. Insbesondere hat der Lösungsraum  $V$  die Dimension  $n$ .*

Das Auffinden aller Lösungen eines homogenen Systems ist also äquivalent mit der Angabe einer Basis für dessen Lösungsraum. Und weil eine solche Basis so wichtig ist, hat sie einen eigenen Namen.

**Definition 3.6** *Ein Lösungsfundamentalsystem (LFS) für das homogene lineare  $n \times n$ -Differentialgleichungssystem ist eine Basis für dessen Lösungsraum.*

**Beispiel 3.14** *Die Vektorfunktionen*

$$\mathbf{x}_1(t) = \begin{pmatrix} \sin(t) \\ \cos(t) \end{pmatrix} \quad \text{und} \quad \mathbf{x}_2(t) = \begin{pmatrix} -\cos(t) \\ \sin(t) \end{pmatrix}$$

*sind ein Lösungsfundamentalsystem für das System*

$$\begin{aligned} \dot{x}_1 &= x_2 \\ \dot{x}_2 &= -x_1 \end{aligned}.$$

Weil der Lösungsraum  $V$  die Dimension  $n$  hat, ist ein  $n$ -tupel  $\mathbf{x}_1(t), \dots, \mathbf{x}_n(t)$  von Lösungen schon dann ein LFS, wenn diese Vektorfunktionen linear unabhängig sind. Hierzu

**Satz 3.10** Ein  $n$ -tupel  $\mathbf{x}_1(t), \dots, \mathbf{x}_n(t)$  von Lösungen des homogenen  $n \times n$ -Systems  $\dot{\mathbf{x}}(t) = A(t) \cdot \mathbf{x}(t)$  auf  $I$  ist genau dann linear unabhängig, wenn für einen einzigen Punkt  $t_0 \in I$  die Vektoren

$$\mathbf{x}_1(t_0), \dots, \mathbf{x}_n(t_0) \in \mathbb{R}^n$$

linear unabhängig sind.

Beweis. a) Seien die Lösungen  $\mathbf{x}_1(t), \dots, \mathbf{x}_n(t)$  linear unabhängig. Um zu zeigen, dass die Vektoren  $\mathbf{x}_1(t_0), \dots, \mathbf{x}_n(t_0)$  linear unabhängig sind, setzen wir den Test

$$c_1 \mathbf{x}_1(t_0) + \dots + c_n \mathbf{x}_n(t_0) = \mathbf{0}, \quad c_1, \dots, c_n \in \mathbb{R},$$

an. Auch  $\mathbf{x} := c_1 \mathbf{x}_1 + \dots + c_n \mathbf{x}_n$  ist eine Lösung. Weil  $\mathbf{x}(t_0) = \mathbf{0}$  gilt, folgt aus dem Eindeutigkeitsatz  $\mathbf{x} \equiv \mathbf{0}$ . Und aus der linearen Unabhängigkeit der Funktionen  $\mathbf{x}_1, \dots, \mathbf{x}_n$  folgt  $c_1 = \dots = c_n = 0$ .

b) Seien die Vektoren  $\mathbf{x}_1(t_0), \dots, \mathbf{x}_n(t_0) \in \mathbb{R}^n$  linear unabhängig. Um die Funktionen  $\mathbf{x}_1(t), \dots, \mathbf{x}_n(t)$  auf lineare Unabhängigkeit zu testen, setzen wir

$$c_1 \mathbf{x}_1 + \dots + c_n \mathbf{x}_n \equiv \mathbf{0}$$

an. Wenn wir in diese Gleichung  $t = t_0$  einsetzen, finden wir  $c_1 = \dots = c_n = 0$ . □

Ein LFS kann man auffassen als eine  $n \times n$ -Matrix

$$X(t) = (\mathbf{x}_1(t), \dots, \mathbf{x}_n(t)) = \begin{pmatrix} x_{1,1}(t) & \dots & x_{n,1}(t) \\ \vdots & & \vdots \\ x_{1,n}(t) & \dots & x_{n,n}(t) \end{pmatrix}$$

von Funktionen. Und wir haben bewiesen

$$\det X(t_0) = 0 \quad \Leftrightarrow \quad \det X(t) = 0 \text{ für alle } t.$$

Mit einem LFS des homogenen Systems  $\dot{\mathbf{x}} = A \cdot \mathbf{x}$  kann man genau wie in einer Dimension Variation der Konstanten machen, um eine spezielle Lösung des inhomogenen Systems  $\dot{\mathbf{x}} = A \cdot \mathbf{x} + \mathbf{b}$  zu finden. Wir setzen an

$$\mathbf{x}_0(t) := X(t) \cdot \mathbf{c}(t).$$

Differentiation ergibt

$$\dot{\mathbf{x}}_0 = \dot{X} \cdot \mathbf{c} + X \cdot \dot{\mathbf{c}} = A \cdot X \cdot \mathbf{c} + X \cdot \dot{\mathbf{c}}.$$

Um

$$\dot{\mathbf{x}}_0 = A \cdot \mathbf{x} + \mathbf{b}$$

zu erfüllen, müssen wir also

$$X \cdot \dot{\mathbf{c}} = \mathbf{b}, \quad \text{bzw.} \quad \dot{\mathbf{c}} = X^{-1} \cdot \mathbf{b}$$

lösen. Das ist schön zu wissen. Von konkreter Durchführung wird aber dringend abgeraten.

Die Theorie der Systeme linearer Differentialgleichungen enthält als Spezialfall die Theorie der linearen Differentialgleichungen höherer Ordnung. Eine solche Differentialgleichung hat, wenn sie homogen ist, die Form

$$x^{(n)} + a_n(t)x^{(n-1)} + \dots + a_2(t)\dot{x} + a_1(t)x = 0.$$

Sie ist äquivalent zum System

$$\begin{aligned} \dot{x}_1 &= && x_2 \\ \dot{x}_2 &= && x_3 \\ &\vdots && \ddots \\ \dot{x}_n &= a_1 x_1 + a_2 x_2 + && + a_n x_n \end{aligned}$$

Satz 3.8 ergibt hier

**Satz 3.11** Die Koeffizientenfunktionen  $a_1, \dots, a_n$  seien stetig auf dem offenen Intervall  $I \subset \mathbb{R}$ . Dann besitzt jedes AWP

$$t_0 \in I, \quad x(t_0) = c_0, \dots, x^{(n-1)}(t_0) = c_{n-1} \text{ mit } c_0, \dots, c_{n-1} \in \mathbb{R}$$

eine eindeutig bestimmte Lösung. Diese existiert auf dem ganzen Intervall  $I$ .

Wieder bilden die Lösungen der Differentialgleichung einen  $\mathbb{R}$ -Vektorraum. Aus der Theorie der Systeme wissen wir sogar: Es gibt Lösungen  $x_1, \dots, x_n$  (schändlicher Missbrauch der Bezeichnung) der Gleichung  $n$ -ter Ordnung so, dass die Vektorwertigen Funktionen

$$\begin{pmatrix} x_1(t) \\ \dot{x}_1(t) \\ \vdots \\ x_1^{(n-1)}(t) \end{pmatrix}, \quad \dots, \quad \begin{pmatrix} x_n(t) \\ \dot{x}_n(t) \\ \vdots \\ x_n^{(n-1)}(t) \end{pmatrix}$$

in jedem Punkt  $t \in I$  linear unabhängig sind. Ein solches  $n$ -tupel von Lösungen bildet eine Basis des Lösungsraums und heißt auch wieder *Lösungsfundamentalsystem*. Das Lösen einer linearen Differentialgleichung  $n$ -ter Ordnung ist äquivalent damit, ein solches LFS anzugeben. Zu einem  $n$ -tupel  $x_1, \dots, x_n$  von Lösungen gehört die  $n \times n$ -Matrix

$$W(t) = \begin{pmatrix} x_1 & x_2 & \dots & x_n \\ \dot{x}_1 & \dot{x}_2 & \dots & \dot{x}_n \\ \vdots & \vdots & & \vdots \\ x_1^{(n-1)} & x_2^{(n-1)} & \dots & x_n^{(n-1)} \end{pmatrix}.$$

Sie heißt *Wronski-Matrix* der Lösungen  $x_1, \dots, x_n$ . Aus Satz 3.10 wissen wir:

$$\det(W(t)) \neq 0 \text{ für alle } t \in I \quad \Leftrightarrow \quad \det(W(t_0)) \neq 0 \text{ für ein } t_0 \in I.$$

Hier gilt sogar:

**Satz 3.12** Für eine Lösungs- $n$ -tupel  $x_1, \dots, x_n$  einer linearen Differentialgleichung  $n$ -ter Ordnung sind äquivalent:

- a) Die Funktionen  $x_1, \dots, x_n$  sind linear unabhängig.
- b) Sie bilden ein LFS.

Beweis. a)  $\Rightarrow$  b): Wenn die Wronski-Matrix nicht überall maximalen Rang hätte, gäbe es ein  $t_0 \in I$  und eine Vektor  $\mathbf{0} \neq \mathbf{c} \in \mathbb{R}^n$  mit

$$W(t_0) \cdot \mathbf{c} = \mathbf{0}.$$

Die Vektor-wertige Funktion  $W(t) \cdot \mathbf{c}$  ist eine Lösung des linearen  $n \times n$ -Systems, das zu unserer Gleichung  $n$ -ter Ordnung gehört. Sie hat die Nullstelle  $t_0$  und muss nach Satz 3.9 identisch verschwinden. Die erste Komponente dieses Vektors ist

$$c_1 x_1(t) + c_2 x_2(t) + \dots + c_n x_n(t) \equiv 0.$$

Wegen  $\mathbf{c} \neq \mathbf{0}$  ist das ist ein Widerspruch zu a).

b)  $\Rightarrow$  a): Seien die Lösungen  $x_1, \dots, x_n$  linear abhängig, es gebe also  $\mathbf{0} \neq (c_1, \dots, c_n) \in \mathbb{R}^n$  mit

$$c_1 x_1(t) + c_2 x_2(t) + \dots + c_n x_n(t) \equiv 0.$$

Durch  $\nu$ -maliges Differenzieren dieser Gleichung folgt für alle  $\nu$

$$c_1 x_1^{(\nu)}(t) + c_2 x_2^{(\nu)}(t) + \dots + c_n x_n^{(\nu)}(t) \equiv 0.$$

Insbesondere ist  $W(t) \cdot \mathbf{c} \equiv \mathbf{0}$  im Widerspruch zu b). □

**Beispiel 3.15** Die Winkelfunktionen  $\cos(\omega \cdot t)$ ,  $\sin(\omega \cdot t)$  sind Lösungen der linearen Differentialgleichung zweiter Ordnung

$$\ddot{x} + \omega^2 x = 0, \quad \omega \neq 0.$$

Die zugehörige Wronski-Matrix

$$W(t) = \begin{pmatrix} \cos(\omega \cdot t) & \sin(\omega \cdot t) \\ -\omega \cdot \sin(\omega \cdot t) & \omega \cdot \cos(\omega \cdot t) \end{pmatrix}$$

hat in jedem Punkt  $t \in \mathbb{R}$  den Rang 2.

Jetzt ist wohl doch der Zeitpunkt gekommen, die schlechte Nachricht nicht länger zu verheimlichen: Anders als im ein-dimensionalen Fall, bzw. bei Gleichungen erster Ordnung, gibt es für  $n > 1$  keine expliziten Lösungsmethoden, jedenfalls dann nicht, wenn die Koeffizienten des Systems, bzw. der Gleichung nicht-konstant sind. Aber genau die Fälle mit nicht-konstanten Koeffizienten sind die wichtigen Fälle. Da kann man nichts anderes machen, als in jedem einzelnen Fall die Lösungen liebevoll zu untersuchen. Meist handelt es sich dabei dann auch um neue unbekannte Funktionen.

**Beispiel 3.16 (Exaktes Pendel)** Die Differentialgleichung des exakten Pendels ist

$$\ddot{x} + \omega^2 \cdot \sin(x) = 0.$$

Um sie zu integrieren multipliziert man mit  $2\dot{x}$

$$2\dot{x}\ddot{x} = -2\omega^2 \sin(x)\dot{x}$$

und integriert einmal

$$(\dot{x})^2 = C + 2\omega^2 \cos(x).$$

Zumindest für  $|C| \leq 2\omega^2$  erhält man daraus

$$\dot{x} = \pm \sqrt{C + 2\omega^2 \cos(x)} = \pm \omega \sqrt{2 \cdot (\cos(x) - \cos(x_0))}, \quad \cos(x_0) := -\frac{C}{\omega^2}.$$

Wir konzentrieren uns auf das negative Vorzeichen und trennen die Variablen:

$$\frac{-dx}{\omega \sqrt{2 \cdot (\cos(x) - \cos(x_0))}} = dt,$$

$$\frac{1}{\omega} \int_x^{x_0} \frac{du}{\sqrt{2 \cdot (\cos(u) - \cos(x_0))}} = t - C$$

mit einer neuen, zweiten Integrationskonstante  $C$ . Jetzt formen wir um

$$\cos(u) = 1 - 2 \cdot \sin^2\left(\frac{u}{2}\right), \quad \frac{1}{2\omega} \int_x^{x_0} \frac{du}{\sqrt{\sin^2\left(\frac{x_0}{2}\right) - \sin^2\left(\frac{u}{2}\right)}} = t - C$$

und substituieren

$$\sin\left(\frac{x}{2}\right) = r, \quad \sin\left(\frac{x_0}{2}\right) = r_0, \quad \sin\left(\frac{u}{2}\right) = \rho.$$

Mit

$$u = 2 \cdot \arcsin(\rho), \quad du = \frac{2 \cdot d\rho}{\sqrt{1 - \rho^2}}$$

erhalten wir schließlich

$$t = C + \frac{1}{\omega} \int_r^{r_0} \frac{d\rho}{\sqrt{(1 - \rho^2)(r_0^2 - \rho^2)}}.$$

Da haben wir jetzt nicht  $x = x(t)$  ausgerechnet, sondern irgend eine Umkehrfunktion davon. Und die ist ein Integral, das nicht mit elementaren Funktionen auszuwerten ist. Es heißt elliptisches Integral. Darüber sind ganze Bücher geschrieben worden.

Als eine Art Bestiarium möchte ich noch ein paar Beispiele von Differentialgleichungen zweiter Ordnung zusammenstellen.

Differentialgleichung

Besselsche	$t\ddot{x} + \dot{x} + (m^2t - \frac{n^2}{t})x = 0$	$m, n \in \mathbb{N}$
Hermiteische	$\ddot{x} - 2t\dot{x} + 2\nu x = 0$	$\nu \in \mathbb{R}$
Hypergeometrische	$t(1-t)\ddot{x} + (\gamma - (1+\alpha+\beta)\dot{x} - \alpha\beta x = 0$	$\alpha, \beta, \gamma \in \mathbb{R}$
von Laguerre	$t\ddot{x} + (1-t)\dot{x} + \nu x = 0$	$\nu \in \mathbb{R}$
von Legendre	$t(1-t)\ddot{x} + (1-2t)\dot{x} + n(n+1)x = 0$	$n \in \mathbb{N}$

Zu beachten ist hier, dass nicht alle diese Differentialgleichungen *explizit*, d.h., nach  $\ddot{x}$  aufgelöst sind. Die Sätze 3.8 und 3.9 gelten nur dort, wo  $\ddot{x}$  ohne Vorfaktor vorkommt. Die Existenz-Intervalle  $I$  sind also für die

Differentialgleichung	$I$
Besselsche, Laguerresche	$] -\infty, 0[, ]0, \infty[$
Hypergeometrische, Legendresche	$] -\infty, 0[, ]0, 1[, ]1, \infty[$

Befriedigend ist die Situation nur für lineare Systeme *mit konstanten Koeffizienten*. In Vektorschreibweise sieht ein solches System so aus:

$$\dot{\mathbf{x}} = A \cdot \mathbf{x}.$$

Dabei ist  $A$  eine *konstante*  $n \times n$ -Matrix. Nach Beispiel 2.8 sind alle Spaltenvektoren der Exponentialmatrix  $e^{tA}$  Lösungen des Systems. Nach Aufgabe 2.1 ist die Matrix  $e^{tA}$  invertierbar. Ihre Spaltenvektoren sind für alle  $t \in \mathbb{R}$  linear unabhängig und bilden ein LFS. Durch Berechnung der Exponentialmatrix  $e^{tA}$  wird das System gelöst! Bemerkenswert ist, dass jede Lösung  $e^{tA} \cdot \mathbf{c}$  für alle Zeiten  $t \in \mathbb{R}$  existiert.

Diese Information ist allerdings nur dann hilfreich, wenn man die Exponentialmatrix explizit ausrechnen kann.

**Beispiel 3.17** *Zum System*

$$\begin{aligned} \dot{x}_1 &= -\omega \cdot x_2 \\ \dot{x}_2 &= \omega \cdot x_1 \end{aligned}$$

gehört die Matrix

$$A = \begin{pmatrix} 0 & -\omega \\ \omega & 0 \end{pmatrix} \quad \text{mit} \quad e^{tA} = \begin{pmatrix} \cos(\omega \cdot t) & -\sin(\omega \cdot t) \\ \sin(\omega \cdot t) & \cos(\omega \cdot t) \end{pmatrix}$$

(Aufgabe 2.3). Mit dieser Matrix ist das System gelöst.

Die Matrix  $A$  aus dem vorhergehenden Beispiel hat das charakteristische Polynom  $\lambda^2 + \omega^2$  mit den beiden komplexen Eigenwerten  $\lambda_{1,2} = \pm i\omega$ . Sie führen auf die Winkelfunktionen. Zu einem reellen Eigenwert  $\lambda_1$  von  $A$  gehört immer ein reeller Eigenvektor  $\mathbf{v}$ . Für  $\mathbf{x}(t) = e^{\lambda_1 \cdot t} \cdot \mathbf{v}$  berechnet man

$$\dot{\mathbf{x}} = \lambda_1 \cdot e^{\lambda_1 \cdot t} \cdot \mathbf{v} = A \cdot \mathbf{x}.$$

Damit hat man eine reelle Lösung. Besitzt die Matrix  $A$  nur reelle, nicht-entartete Eigenwerte, so bekommt man zu den  $n$  Eigenwerten  $n$  linear unabhängige Eigenvektoren  $\mathbf{v}_1, \dots, \mathbf{v}_n$  und ein LFS

$$\mathbf{x}_1(t) = e^{\lambda_1 \cdot t} \cdot \mathbf{v}_1, \dots, \mathbf{x}_n(t) = e^{\lambda_n \cdot t} \cdot \mathbf{v}_n.$$

Die Schwierigkeiten beginnen, wenn Eigenwerte entarten. Hierzu das einfachste Beispiel:

**Beispiel 3.18** *Wir betrachten das System*

$$\begin{aligned} \dot{x}_1 &= \lambda \cdot x_1 + x_2 \\ \dot{x}_2 &= \lambda \cdot x_2 \end{aligned}$$

mit der Koeffizientenmatrix

$$A = \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}.$$

Mit Aufgabe 2.2 kennen wir die Exponentialmatrix

$$e^{tA} = \begin{pmatrix} e^{\lambda \cdot t} & t \cdot e^{\lambda \cdot t} \\ 0 & e^{\lambda \cdot t} \end{pmatrix}.$$

Ihre Spalten bilden ein LFS. Die erste Spalte ist die Exponentialfunktion zum Eigenvektor. Die zweite Spalte zeigt, wie sich die Komplikation auswirkt.

Der für lineare Systeme mit konstanten Koeffizienten skizzierte Formalismus gilt analog für lineare Differentialgleichungen höherer Ordnung mit konstanten Koeffizienten. Betrachten wir etwa die homogene Gleichung  $n$ -ter Ordnung

$$x^{(n)} + a_{n-1}x^{(n-1)} + \dots + a_1\dot{x} + a_0x = 0, \quad a_{n-1}, \dots, a_0 \in \mathbb{R}.$$

Es wäre ungeschickt, diese Gleichung als System umzuschreiben. Mit dem  $e$ -hoch-Ansatz kommt man viel direkter zum Ziel: Wir wenden die linke Seite der Gleichung auf die  $e$ -Funktion

$$x(t) := e^{\lambda t}, \quad \lambda \in \mathbb{R},$$

an. Mit

$$\frac{d^\nu}{dt^\nu} x(t) = \lambda^\nu x(t), \quad \nu = 0, \dots, n,$$

erhalten wir

$$(\lambda^n + a_{n-1}\lambda^{n-1} + \dots + a_1\lambda + a_0) \cdot e^{\lambda t} = 0.$$

Wenn  $\lambda_1$  eine Nullstelle des Polynoms

$$p(\lambda) := \lambda^n + a_{n-1}\lambda^{n-1} + \dots + a_1\lambda + a_0$$

ist, dann ist die zugehörige Exponentialfunktion  $x(t) = e^{\lambda_1 t}$  eine Lösung der Differentialgleichung.

**Definition 3.7** Das Polynom  $p(\lambda)$  vom Grad  $n$ , das zur betrachteten Differentialgleichung  $n$ -ter Ordnung gehört, heißt das charakteristische Polynom der Differentialgleichung.

Zu jeder Nullstelle des charakteristischen Polynoms gehört also eine Exponentialfunktion, welche die Differentialgleichung löst. Und Lösungen zu verschiedenen Nullstellen von  $p$  sind linear unabhängig. Das wollen wir aber nicht mehr nachrechnen. Stattdessen

**Beispiel 3.19** Die Differentialgleichung

$$\ddot{x} - 4x = 0$$

hat das charakteristische Polynom  $\lambda^2 - 4$  mit den beiden Nullstellen  $\pm 2$ . Die beiden Exponentialfunktionen  $e^{2t}$  und  $e^{-2t}$  bilden also ein LFS der Differentialgleichung.

Die Differentialgleichung

$$\ddot{x} + 4x = 0$$

hat das charakteristische Polynom  $\lambda^2 + 4$  mit den beiden komplexen Nullstellen  $\pm 2i$ . Dazu gehören die komplexen Exponentialfunktionen

$$e^{2i \cdot t} = \cos(2t) + i \cdot \sin(2t), \quad e^{-2i \cdot t} = \cos(2t) - i \cdot \sin(2t).$$

Wir sind aber nicht an diesen komplexen Lösungen interessiert, sondern nur an den reellen Lösungen, die man daraus durch Rekombination

$$\frac{1}{2}(e^{2it} + e^{-2it}) = \cos(2t), \quad \frac{1}{2i}(e^{2it} - e^{-2it}) = \sin(2t)$$

erhält.

Wieder einmal ist das Problem vollständig gelöst, wenn das charakteristische Polynom lauter verschiedene Nullstellen besitzt. Schwierigkeiten machen die mehrfachen Nullstellen. Betrachten wir den Extremfall

$$p(\lambda) = (\lambda - c)^m$$

einer einzigen,  $m$ -fachen Nullstellen. Die zugehörige Differentialgleichung

$$\left(\frac{d}{dt} - c\right)^m x = 0$$

multiplizieren wir aber nicht aus, sondern fassen sie auf als iterierte Anwendung des 'Differentialoperators'

$$D := \frac{d}{dt} - c, \quad x \mapsto \dot{x} - c \cdot x.$$

Wegen  $De^{ct} = 0$  ist  $e^{ct}$  eine Lösung. Die kennen wir schon. Aber angesichts Beispiel 3.18 probieren wir mal  $x(t) = t \cdot e^{ct}$  aus:

$$D(t \cdot e^{ct}) = \frac{d}{dt}(t \cdot e^{ct}) - c \cdot t \cdot e^{ct} = e^{ct} + t \cdot ce^{ct} - c \cdot te^{ct} = e^{ct}.$$

Wegen  $m \geq 2$  haben wir

$$\left(\frac{d}{dt} - c\right)^m x = \left(\frac{d}{dt} - c\right)^{m-1} \left(\frac{d}{dt} - c\right)(te^{ct}) = \left(\frac{d}{dt} - c\right)^{m-1} e^{ct} = 0.$$

Also ist  $x(t) = te^{ct}$  eine Lösung der gesamten Gleichung. Ähnlich sieht man, dass die 'Quasi-Polynome'

$$t^{m-1}e^{ct}, t^{m-2}e^{ct}, \dots, te^{ct}, e^{ct}$$

Lösungen, sogar linear unabhängige Lösungen sind.

Und wenn das charakteristische Polynom

$$p(\lambda) = q(\lambda) \cdot (\lambda - c)^m$$

den  $m$ -fachen Linearfaktor  $\lambda - c$  enthält, dann schreiben wir die Differentialgleichung (symbolisch)

$$q\left(\frac{d}{dt}\right) \cdot \left(\frac{d}{dt} - c\right)^m x = 0.$$

Alle eben angegebenen Quasi-Polynome sind auch Lösungen der zusammengesetzten Gleichung. So kann man alle homogenen Differentialgleichungen höherer Ordnung mit konstanten Koeffizienten lösen.

Es bleibt das Problem der inhomogenen Gleichung

$$x^{(n)} + a_{n-1}x^{(n-1)} + \dots + a_1\dot{x} + a_0x = b(t).$$

Wieder kann man sie auf dem Umweg über das zugehörige System und Variation der Konstanten lösen. Wenn aber die Inhomogenität  $b(t)$  einfach gebaut ist, gibt es eine leichtere, direkte Methode:

1) Falls  $b(t) = e^{ct}$  eine Exponentialfunktion ist, stecken wir sie mal in die linke Seite der Gleichung rein und erhalten als Resultat

$$p(c) \cdot e^{ct}.$$

Damit haben wir die spezielle Lösung  $e^{ct}/p(c)$ , allerdings nur wenn  $p(c) \neq 0$  ist. Wenn  $p(c) = 0$  sein sollte (Resonanzfall), müssen wir  $t \cdot e^{ct}$  reinstecken, und uns bei einer mehrfachen Nullstelle  $c$  eventuell mit höheren Polynomen hochhangeln.

2) Falls  $p(t) = \cos(ct)$  oder  $= \sin(ct)$  eine Winkelfunktion ist, stecken wir beide Winkelfunktionen in die Gleichung rein und bekommen als Resultate zwei Linearkombinationen dieser Funktionen. Eine geeignete Linearkombination davon gibt eine spezielle Lösung. Falls nicht ein Resonanzfall vorliegt und wir uns mit Polynomen  $\cdot$  Winkelfunktionen hochhangeln müssen.

3) Falls  $p(t)$  ein Polynom sein sollte, stecken wir ein Polynom gleichen Grades in die Gleichung rein, und schauen, was rauskommt.

4) Und wenn die Inhomogenität eine Linearkombination der Fälle 1)-3) sein sollte, verwenden wir die entsprechende Linearkombination der eben angegebenen Funktionen.

**Beispiel 3.20 (Harmonischer Oszillator)** *Den harmonischen Oszillator kennen wir schon aus dem Gymnasium. Aus dem Physik-Unterricht! Aber realistische Schwingungen sind selten harmonisch. Deswegen ist er eher eine mathematische Abstraktion, als ein realistisches Modell für die Physik. Aber die oben eingeführte Theorie illustriert er sehr schön. In aufsteigender Komplexität betrachtet man folgende Fälle:*

1) *Der ungedämpfte Oszillator hat die Gleichung*

$$\ddot{x} + \omega^2 \cdot x = 0$$

*mit dem charakteristischen Polynom  $p(\lambda) = \lambda^2 + \omega^2$  und dessen Nullstellen  $\pm i \cdot \omega$ . Ein LFS bilden deswegen die Winkelfunktionen  $\cos(\omega \cdot t)$  und  $\sin(\omega \cdot t)$ .*

2) *Der gedämpfte Oszillator hat abhängig von  $\alpha > 0$  die Gleichung*

$$\ddot{x} + 2\alpha \cdot \dot{x} + \omega^2 \cdot x = 0$$

*mit dem charakteristischen Polynom*

$$p(\lambda) = \lambda^2 + 2\alpha \cdot \lambda + \omega^2 = 0$$

*und seinen Nullstellen*

$$\lambda_{1,2} = -\alpha \pm \sqrt{\alpha^2 - \omega^2}.$$

*Für  $\alpha < \omega$  (schwache Dämpfung) sind beide Nullstellen komplex und führen zum LFS*

$$e^{-\alpha \cdot t} \cdot \cos(\sqrt{\omega^2 - \alpha^2} \cdot t), \quad e^{-\alpha \cdot t} \cdot \sin(\sqrt{\omega^2 - \alpha^2} \cdot t).$$

*Für  $\alpha > \omega$  (starke Dämpfung) sind beide Nullstellen reell und negativ. Dazu gehört das LFS*

$$e^{-\alpha + \sqrt{\alpha^2 - \omega^2} \cdot t}, \quad e^{-\alpha - \sqrt{\alpha^2 - \omega^2} \cdot t}.$$

*Bleibt  $\alpha = \omega$  (Grenzfall) mit der doppelten Nullstelle  $\lambda_1 = \lambda_2 = -\alpha$ . Das führt zum LFS*

$$e^{-\alpha \cdot t}, \quad t \cdot e^{-\alpha \cdot t}.$$

3) *Der gedämpfte und mit einer Frequenz  $\rho$  angestoßene Oszillator hat die inhomogene Gleichung*

$$\ddot{x} + 2\alpha \cdot \dot{x} + \omega^2 \cdot x = \sin(\rho \cdot t).$$

Es kommt darauf an, eine spezielle Lösung dieser Gleichung zu finden. Dazu müssen wir die linke Seite der Gleichung auf die beiden Winkelfunktionen  $\sin(\rho \cdot t)$  und  $\cos(\rho \cdot t)$  anwenden mit den Ergebnissen

$$(\omega^2 - \rho^2) \cdot \sin(\rho \cdot t) + 2\alpha\rho \cdot \cos(\rho \cdot t),$$

$$-2\alpha\rho \cdot \sin(\rho \cdot t) + (\omega^2 - \rho^2) \cdot \cos(\rho \cdot t).$$

Als Linearkombination beider Ergebnisse können wir die gegebene rechte Seite  $\sin(\rho \cdot t)$ , ja sogar jede Winkelfunktion mit der Frequenz  $\rho$  erhalten, weil

$$\det \begin{pmatrix} \omega^2 - \rho^2 & 2\alpha\rho \\ -2\alpha\rho & \omega^2 - \rho^2 \end{pmatrix} = (\omega^2 - \rho^2)^2 + 4\alpha^2\rho^2 > 0$$

ist. Nur im ungedämpften Resonanzfall  $\alpha = 0, \rho = \omega$  hat man hier Probleme.

**Aufgabe 3.12** Es seien  $a, b \in \mathbb{R}$ . Lösen Sie das System

$$\begin{aligned} \dot{x}_1 &= ax_1 + bx_2 + (1-a)e^t \\ \dot{x}_2 &= bx_1 + ax_2 - be^t. \end{aligned}$$

**Aufgabe 3.13** Lösen Sie die AWP's

$$\begin{aligned} a) \dot{\mathbf{x}} &= \begin{pmatrix} 1 & -1 \\ -3 & 0 \end{pmatrix}, \quad \mathbf{x}(0) = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \\ b) \text{vec}x &= \begin{pmatrix} -3 & 5 \\ -2 & 4 \end{pmatrix}, \quad \mathbf{x}(0) = \begin{pmatrix} 13 \\ 1 \end{pmatrix}. \end{aligned}$$

**Aufgabe 3.14** Finden Sie ein LFS für

$$\begin{aligned} \dot{x}_1 &= -x_2 + x_3 \\ \dot{x}_2 &= x_1 - x_3 \\ \dot{x}_3 &= -x_1 + x_2 \end{aligned}$$

**Aufgabe 3.15** Lösen Sie das AWP

$$\ddot{x} + 5\dot{x} = t + e^{-5t}, \quad x(0) = \dot{x}(0) = 0.$$

**Aufgabe 3.16** Bestimmen Sie die allgemeine Lösung der Gleichung

$$x^{(3)} + \ddot{x} + 8\dot{x} - 10x = 18e^{-t}.$$

**Aufgabe 3.17** Bestimmen Sie alle Lösungen der Gleichung

- $x^{(4)} + 2x^{(3)} - 8\ddot{x} = e^t,$
- $\ddot{x} + 4\dot{x} - 2x = t,$
- $\ddot{x} + \dot{x} - 2x = t^2,$
- $x^{(3)} + \ddot{x} + \dot{x} + x = 2e^{-t},$
- $\ddot{x} + 16x = -\sin(4t).$

### 3.4 Autonome Systeme, Stabilität

Hier betrachten wir Systeme von Differentialgleichungen, die nicht mehr linear sind. Zum Ausgleich soll die rechte Seite von  $t$  unabhängig sein.

**Definition 3.8** *Ein autonomes System ist ein System von Differentialgleichungen*

$$\begin{aligned}\dot{x}_1 &= f_1(x_1, \dots, x_n) \\ &\vdots \\ \dot{x}_n &= f_n(x_1, \dots, x_n)\end{aligned}\quad \text{bzw.} \quad \dot{\mathbf{x}} = F(\mathbf{x})$$

mit von  $t$  unabhängiger rechter Seite. Eine Lösungskurve ist (genau wie bisher) eine Abbildung  $\mathbf{x} : ]\alpha, \beta[ \rightarrow \mathbb{R}^n$ . Wenn die Funktionen  $f_1, \dots, f_n$  stetig partiell differenzierbar sind, dann gilt wie bisher lokale Existenz und globale Eindeutigkeit (Satz 3.4).

Die rechte Seite des autonomen Systems kann man als eine zeitunabhängige Geschwindigkeitsverteilung deuten. Und Lösungskurven sind die Kurven  $\mathbf{x}(t)$ , die in jedem ihrer Punkte die dort vorgegebene Geschwindigkeit (Richtung und Betrag) haben. Am einfachsten ist das zu verstehen in der Ebene. Nehmen wir also zunächst  $n = 2$  an. Und die Koordinaten nennen wir wieder  $x, y$ . Dann sieht das autonome System so aus:

$$\begin{aligned}\dot{x} &= f(x, y) \\ \dot{y} &= g(x, y)\end{aligned}$$

Das Problem besteht also darin: Ein stationäres Vektorfeld  $\mathbf{v}(x, y) = (f(x, y), g(x, y))$  ist vorgegeben, und Lösungskurven sind solche Kurven,  $(x(t), y(t))$ , die in jedem ihrer Punkte den dort vorgegebenen Vektor als Geschwindigkeitsvektor haben. Um sie zu zeichnen, müsste man das eigentlich im dreidimensionalen  $t - x - y$ -Raum tun. Praktisch lässt man die Zeit-Abhängigkeit meist unter den Tisch fallen, und zeichnet sie als Kurven im zwei-dimensionalen  $x - y$ -Raum.

**Definition 3.9** *Das autonome System heißt Hamilton-System wenn es eine Funktion  $H(x, y)$  (die Hamilton-Funktion) gibt so, dass*

$$f = \frac{\partial H}{\partial y}, \quad g = -\frac{\partial H}{\partial x}.$$

Diese Bedingung können wir auch schreiben

$$(-g, f) = \text{grad } H.$$

Zumindest auf einer konvexen Definitionsmenge ist hierfür die Symmetriebedingung (Sätze 2.20, 2.21)

$$-g_y = f_x \quad \text{bzw.} \quad f_x + g_y = 0$$

notwendig und hinreichend. Hamilton-Systeme sind wichtig, weil man ihre Lösungskurven  $(x(t), y(t))$  ohne Rechnung angeben kann:

**Satz 3.13** *Jede Lösungskurve des Hamilton-Systems ist eine Niveaumenge  $H(x, y) = c$  der Hamiltonfunktion.*

Beweis. Ist  $(x(t), y(t))$  eine Lösungskurve, so gilt

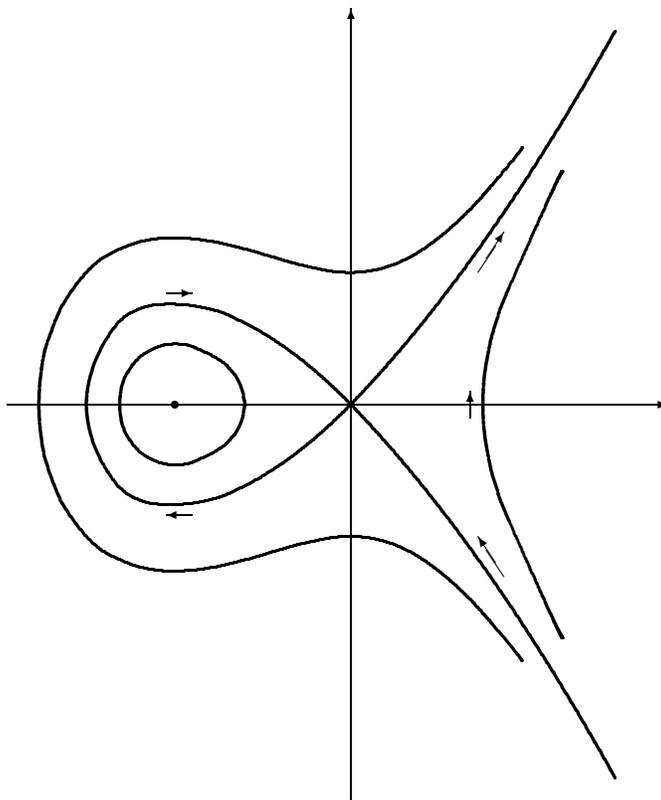
$$\frac{d}{dt}H(x(t), y(t)) = H_x \dot{x} + H_y \dot{y} = -g \cdot f + f \cdot g \equiv 0,$$

also  $H(x(t), y(t)) = \text{const.}$  □

**Beispiel 3.21** *Das autonome System*

$$\begin{aligned}\dot{x} &= 2y \\ \dot{y} &= 2x + 3x^2\end{aligned}$$

hat die Hamiltonfunktion  $H = y^2 - x^2(x + 1)$ . Deren Niveaulinien versuche ich mal zu plotten:



Wie man jede Differentialgleichung höherer Ordnung in ein System erster Ordnung umwandeln kann, so kann man eine zeitunabhängige Differentialgleichung höherer Ordnung

$$x^{(n)} = f(x, \dot{x}, \dots, x^{(n-1)})$$

in ein autonomes System umwandeln.

**Beispiel 3.22 (Exaktes Pendel)** *Die Differentialgleichung des exakten Pendels ist*

$$\ddot{x} + \sin(x) = 0.$$

Mit  $y = \dot{x}$  schreibt sich das zugehörige System als

$$\begin{aligned}\dot{x} &= y \\ \dot{y} &= -\sin(x)\end{aligned}$$

Die Koeffizientenfunktionen sind

$$f(x, y) = y, \quad g(x, y) = -\sin(x).$$

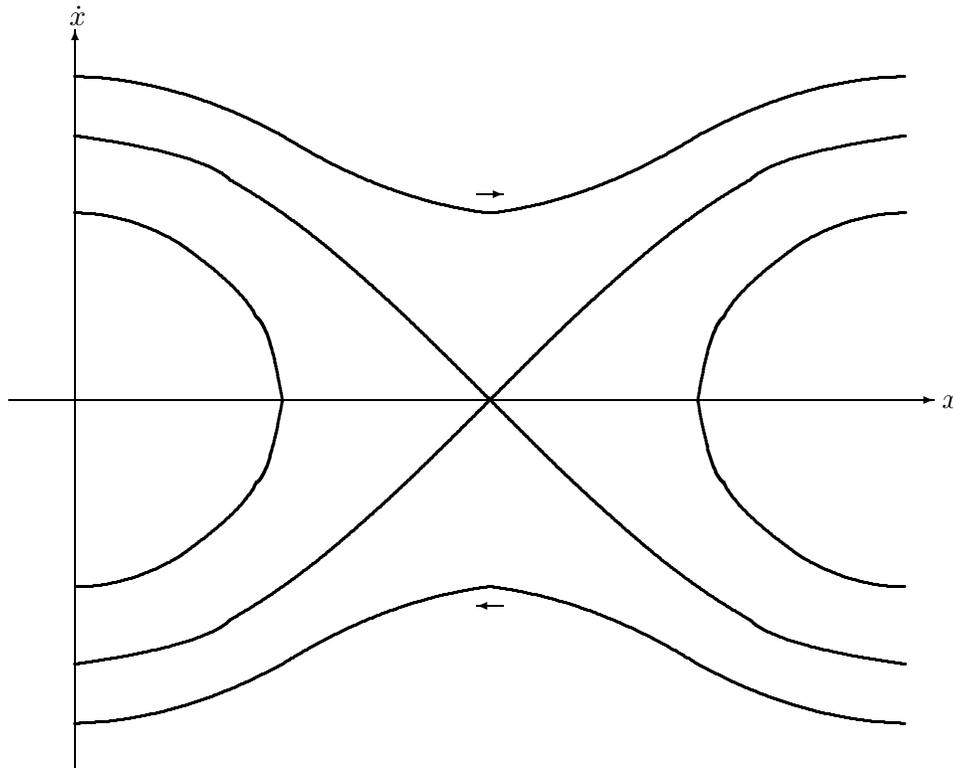
Das System hat die Hamilton-Funktion

$$H(x, y) = \frac{1}{2}y^2 - \cos(x).$$

Deren Niveau-Linien

$$y = \pm\sqrt{2 \cdot (\cos(x) + c)}, \quad -1 \leq c \in \mathbb{R}$$

sehen etwa so aus:



Allgemeiner führt jede Differentialgleichung zweiter Ordnung

$$\ddot{x} + h(x) = 0$$

auf ein System

$$\begin{aligned}\dot{x} &= y \\ \dot{y} &= h(x)\end{aligned}$$

Ein solches System ist hamiltonsch mit der Hamiltonfunktion

$$H(x, y) = \frac{1}{2}y^2 - \int_{x_0}^x h(u)du.$$

Die Hamilton-Funktion  $H(x, y)$  des Systems ist durch ihr Gradientenfeld nur bis auf eine additive Konstante bestimmt. Das drückt sich dadurch aus, dass man in der vorstehenden Formel den Anfangspunkt  $x_0$  der Integration beliebig wählen kann.

**Definition 3.10** *Das autonome System, das zu einer Gleichung höherer Ordnung gehört, hat Lösungskurven der Form*

$$(x(t), \dot{x}(t), \dots, x^{n-1}(t))$$

*Ihre zeitunabhängigen Bildkurven im  $n$ -dimensionalen Raum mit den Koordinaten  $x_1 = x, x_2 = \dot{x}, \dots, x_n = x^{(n-1)}$  heißen das Phasenporträt des Systems. Der zugehörige  $n$ -dimensionale Raum heißt sein Phasenraum.*

Durch jeden Punkt  $\mathbf{x}_0 \in \mathbb{R}^n$  im Definitionsbereich der rechten Seite geht genau eine lokale Lösungskurve. Davon gibt es drei wesentlich verschiedene Sorten:

1) *Stationäre Lösungen:* Ist  $f_1(\mathbf{x}_0) = \dots = f_n(\mathbf{x}_0) = 0$ , so bildet die konstante Kurve  $\mathbf{x}(t) = \mathbf{x}_0$  eine Lösung. Das sieht sehr langweilig aus. Aber praktisch sind diese Lösungen genau die Gleichgewichtszustände des Systems. Und wenn ein System nicht wenigstens näherungsweise im Gleichgewicht ist, dann kann man es bald vergessen.

2) *Periodische Lösungen:* Wenn es zwei Parameter  $t_1, t_2 \in ]\alpha, \beta[$  gibt mit  $\mathbf{x}(t_1) = \mathbf{x}(t_2)$  und  $p := t_2 - t_1 > 0$ , dann ist

$$\mathbf{y}(t) := \mathbf{x}(t - p) : ]\alpha + p, \beta + p[ \rightarrow \mathbb{R}^n$$

eine Lösung mit demselben Anfangswert

$$\mathbf{y}(t_2) = \mathbf{x}(t_1) = \mathbf{x}(t_2)$$

wie  $\mathbf{x}$ . Die Kurve  $\mathbf{y}(t), t_2 \leq t \leq t_2 + p$ , durchläuft dieselbe Bahn, wie  $\mathbf{x}(t), t_1 \leq t \leq t_2$ . Indem wir beide Kurven hintereinander durchlaufen, können wir  $\mathbf{x}$  auf das Intervall  $]t_1, t_2 + p[$  fortsetzen. Das können wir beliebig oft wiederholen, auch nach links, unterhalb von  $t_1$ , und erhalten eine Lösung  $\mathbf{x} : \mathbb{R} \rightarrow \mathbb{R}^n$  mit der Periode  $p$ , d.h.,

$$\mathbf{x}(t + p) = \mathbf{x}(t) \text{ für alle } t \in \mathbb{R}.$$

3) Ist eine Lösungskurve weder stationär, noch periodisch, dann kann  $\mathbf{x}(t_1) = \mathbf{x}(t_2)$  für  $t_1 \neq t_2$  nie passieren. Die Abbildung  $\mathbf{x} : ]\alpha, \beta[ \rightarrow \mathbb{R}^n$  ist injektiv.

**Beispiel 3.23** *In Beispiel 3.21 gibt es zwei stationäre Lösungen  $x = y = 0$  und  $x = -1, y = 0$ . In der Nähe der zweiten stationären Lösung sind alle Lösungen periodisch. Auch in Beispiel 3.22 gibt es zwei stationäre Lösungen  $x = \dot{x} = 0$  und  $x = \pi, \dot{x} = 0$ . Die erste ist die Gleichgewichtslage des unausgelenkten Pendels. Die zweite ist die (ziemlich ungemütliche) Ruhelage, wo das Pendel senkrecht nach oben steht.*

Die folgende Aussage formuliert man kurz so: Jede Lösung verlässt jedes Kompaktum. Ich hätte sie schon in 3.2 beweisen können.

**Satz 3.14** Die rechte Seite  $F$  des autonomen Systems  $\dot{\mathbf{x}} = F(\mathbf{x})$  sei stetig differenzierbar auf der offenen Menge  $U \subset \mathbb{R}^n$ . Der Anfangswert  $\mathbf{x}(t_0)$  der Lösung  $\mathbf{x}(t)$  sei enthalten in der kompakten Menge  $K \subset U$ . Dann gilt:

Entweder ist  $x(t)$  als Lösung fortsetzbar auf ein Intervall  $[t_0, t_1[$  und es gibt darin Parameter  $t$  mit  $\mathbf{x}(t) \notin K$ . (Die Lösung verlässt das Kompaktum  $K$ .)

Oder die Lösung ist auf  $[t_0, \infty[$  fortsetzbar und für alle Zeiten  $t \geq t_0$  ist  $\mathbf{x}(t) \in K$ .

Beweis. Die Funktion  $F$  genügt lokal einer Lipschitz-Bedingung. Weil wir  $K$  durch endlich viele offene Mengen überdecken können, können wir für alle  $\mathbf{x} \in K$  die gleiche Lipschitz-Konstante  $L$  nehmen. Aus dem Beweis von Satz 3.1 folgt dann: Es gibt ein  $r > 0$  so, dass für jeden Anfangswert  $\mathbf{x}(t) \in K$  eine Lösung  $\mathbf{x} : [t, t + r[ \rightarrow \mathbb{R}^n$  existiert. Nehmen wir an, die gegebene Lösung  $\mathbf{x}(t)$  verlasse das Kompaktum  $K$  nicht. Ihr maximales Existenz-Intervall sei  $[t_0, t_1[$ . Dann existiert sie also für  $t_1 - \frac{r}{2}$  und dann auch für  $t_1 - \frac{r}{2} + r = t_1 + \frac{r}{2}$ . Für  $t_1 < \infty$  wäre das ein Widerspruch.  $\square$

Eine stationäre Lösung heißt auch *Gleichgewichtslage* des Systems. Ist das System genau in einer solchen Lage, tut sich nie mehr was. Aber praktisch interessiert, wie sich das System in der Nähe einer Gleichgewichtslage verhält. Es kann dort mehr oder weniger stabil sein.

**Definition 3.11** Es sei  $\mathbf{x}_0$  mit  $F(\mathbf{x}_0) = \mathbf{0}$  ein stationärer Punkt des autonomen Systems

$$\dot{\mathbf{x}} = F(\mathbf{x}).$$

Dieser Punkt heißt

- stabil, wenn zu jedem  $\epsilon > 0$  ein  $\delta > 0$  existiert so, dass
  - a) für alle  $\mathbf{x}_1$  mit  $\|\mathbf{x}_1 - \mathbf{x}_0\| < \delta$  die Lösungskurve  $\mathbf{x}(t)$  zum Anfangswert  $\mathbf{x}(0) = \mathbf{x}_1$  für alle Zeiten  $t > 0$  existiert, und
  - b)  $\|\mathbf{x}(t) - \mathbf{x}_0\| < \epsilon$  für alle  $t > 0$ .

Eine Lösungskurve, welche nahe genug bei  $\mathbf{x}_0$  anfängt, bleibt für alle Zeiten in der Nähe von  $\mathbf{x}_0$ .

- asymptotisch stabil, wenn er stabil ist, und wenn  $\delta > 0$  so gewählt werden kann, dass für  $\|\mathbf{x}_1 - \mathbf{x}_0\| < \delta$  gilt: Die Lösungskurve  $\mathbf{x}(t)$  mit Anfangswert  $\mathbf{x}_1$  erfüllt

$$\lim_{t \rightarrow \infty} \mathbf{x}(t) = \mathbf{x}_0.$$

Eine Lösungskurve, welche nahe genug bei  $\mathbf{x}_0$  losgeht, geht im Grenzwert gegen diesen Fixpunkt.

- instabil, wenn er nicht stabil ist. D.h., es gibt ein  $\epsilon > 0$  und zu jedem  $\delta > 0$  gibt es Anfangswerte  $\mathbf{x}_1$  mit  $\|\mathbf{x}_1 - \mathbf{x}_0\| < \delta$ , so dass die Lösungskurve  $\mathbf{x}(t)$  mit Anfangswert  $\mathbf{x}_1$  Punkte  $\mathbf{x}(t)$  enthält, für welche  $\|\mathbf{x}(t) - \mathbf{x}_0\| > \epsilon$ . Es gibt Lösungen, welche beliebig nahe bei  $\mathbf{x}_0$  beginnen aber nicht ganz in der  $\epsilon$ -Kugel um  $\mathbf{x}_0$  liegen.

**Beispiel 3.24** In den Beispielen 3.21 und 3.22 ist je eine Gleichgewichtslage stabil und die andere instabil.

Als erstes überlegen wir uns, was diese Stabilitätsbegriffe für ein *lineares* autonomes System

$$\dot{\mathbf{x}} = A \cdot \mathbf{x}$$

bedeuten. In Abhängigkeit von den (reellen oder komplexen) Eigenwerten der Matrix  $A$  können wir hier ja alle Lösungen hinschreiben.

Ist  $\lambda$  ein reeller Eigenwert der Matrix  $A$  mit Eigenvektor  $\mathbf{v}$ , so ist  $\mathbf{x}(t) = e^{\lambda t} \cdot \mathbf{v}$  eine Lösung. Wenn  $\lambda = \alpha + i\beta$  echt komplex ist, haben wir Lösungen mit

$$e^{\lambda t} = e^{\alpha t} \cdot \cos(\beta \cdot t), \quad \text{bzw.} \quad e^{\lambda t} = e^{\alpha t} \cdot \sin(\beta \cdot t).$$

Abhängig vom Vorzeichen des Realteils  $Re(\lambda)$  finden wir:

$$Re(\lambda) < 0 \quad \Rightarrow \quad \lim_{t \rightarrow \infty} \mathbf{x}(t) = \mathbf{0}, \quad Re(\lambda) > 0 \quad \Rightarrow \quad \lim_{t \rightarrow \infty} \|\mathbf{x}(t)\| = \infty.$$

An diesem Verhalten ändert sich auch dann nichts, wenn wir bei einem entarteten Eigenwert noch Lösungen haben, die Polynom-Faktoren enthalten.

Es bleibt der Fall eines rein imaginären Eigenwertes  $\lambda = \beta \cdot i$ . Zunächst gehören dazu zwei reelle Lösungen

$$\mathbf{x}(t) = \cos(\beta \cdot t) \cdot \mathbf{u} - \sin(\beta \cdot t) \cdot \mathbf{v} \quad \text{und} \quad \mathbf{x}(t) = \cos(\beta \cdot t) \cdot \mathbf{v} + \sin(\beta \cdot t) \cdot \mathbf{u}.$$

Diese Lösungen bewegen sich auf Ellipsen, sind stabil, aber nicht asymptotisch stabil. Jetzt ändert sich die Situation wesentlich, wenn der Eigenwert  $\beta i$  entartet ist.

Ist seine geometrische Vielfachheit  $k$  gleich seiner algebraischen, so gehören zu  $\beta i$  und  $-\beta i$  insgesamt  $2k$  Eigenvektoren und  $2k$  reelle stabile Lösungen auf Ellipsenbahnen. Ist die geometrische Vielfachheit aber echt kleiner als die algebraische, dann gibt es zusätzlich Lösungen, die Polynom-Faktoren enthalten und für  $t \rightarrow \infty$  nach unendlich gehen.

Insgesamt haben wir bewiesen:

**Satz 3.15** *Der Nullpunkt ist ein stationärer Punkt des autonomen Systems  $\dot{\mathbf{x}} = A \cdot \mathbf{x}$ . Er ist*

<i>asymptotisch stabil</i>	$\Leftrightarrow$	<i>alle Eigenwerte von <math>A</math> haben Realteile <math>&lt; 0</math>,</i>
<i>stabil</i>	$\Leftrightarrow$	<i>alle Eigenwerte von <math>A</math> haben Realteile <math>\leq 0</math></i>
		<i>und für alle imaginären Eigenwerte ist</i>
		<i>geometrische Vielfachheit = der algebraischen</i>
<i>instabil</i>	$\Leftrightarrow$	<i><math>A</math> hat einen Eigenwert mit Realteil <math>&gt; 0</math> oder</i>
		<i>einen imaginären Eigenwert</i>
		<i>mit geometrischer Vielfachheit <math>&lt;</math> der algebraischen</i>

**Beispiel 3.25** *Betrachten wir etwa den gedämpften harmonischen Oszillator mit der Gleichung*

$$\ddot{x} + \alpha \cdot \dot{x} + x = 0$$

*und dem stationären Punkt  $x = \dot{x} = 0$ . Die Eigenwerte sind*

$$\lambda_{1,2} = -\alpha \pm \sqrt{\alpha^2 - 1}.$$

*Für  $\alpha = 0$  ist der stationäre Punkt stabil, für  $\alpha > 0$  asymptotisch stabil.*

Jetzt betrachten wir wieder ein nichtlineares autonomes System mit Gleichgewichtslage  $\mathbf{x}_0$

$$\dot{\mathbf{x}} = F(\mathbf{x}), \quad F(\mathbf{x}_0) = \mathbf{0}.$$

Wir setzen voraus, dass die rechte Seite  $F(\mathbf{x})$  differenzierbar ist. Dann ist das System also

$$\dot{\mathbf{x}} = F'(\mathbf{x}_0) \cdot (\mathbf{x} - \mathbf{x}_0) + \Phi(\mathbf{x} - \mathbf{x}_0) \quad \text{mit} \quad \lim_{\mathbf{x} \rightarrow \mathbf{x}_0} \frac{\Phi(\mathbf{x} - \mathbf{x}_0)}{\|\mathbf{x} - \mathbf{x}_0\|} = \mathbf{0}.$$

Nachdem wir  $\mathbf{x}_0$  in den Ursprung schieben, also  $\mathbf{x}$  durch

$$\mathbf{x} + \mathbf{x}_0 \quad \text{mit} \quad \frac{d}{dt}(\mathbf{x} + \mathbf{x}_0) = \dot{\mathbf{x}}$$

ersetzen, können wir o.B.d.A. das System als

$$\dot{\mathbf{x}} = F'(\mathbf{0}) \cdot \mathbf{x} + \Phi(\mathbf{x})$$

annehmen. Die Funktionalmatrix kürzen wir ab

$$A := F'(\mathbf{0}).$$

**Definition 3.12** *Das lineare System mit konstanten Koeffizienten*

$$\dot{\mathbf{x}} = A \cdot \mathbf{x}$$

heißt das in der Gleichgewichtslage linearisierte System.

Ziel dieses Abschnitts ist der Beweis des folgenden Satzes

**Satz 3.16** *Ist das in einer Gleichgewichtslage linearisierte System asymptotisch stabil, so gilt dies auch für das ursprüngliche nichtlineare System.*

Vor dem eigentlichen Beweis zwei Hilfsaussagen.

**Satz 3.17 (Hilfssatz, Variation der Konstanten)** *Es sei  $\mathbf{x}(t)$ ,  $0 \leq t < b$ , eine Lösung des Systems*

$$\dot{\mathbf{x}} = A \cdot \mathbf{x} + \Phi(\mathbf{x}).$$

Dann ist

$$\mathbf{x}(t) = e^{t \cdot A} \cdot \mathbf{x}(0) + \int_0^t e^{(t-s) \cdot A} \cdot \Phi(\mathbf{x}(s)) ds.$$

Beweis. Wir definieren  $\mathbf{c}(t) := e^{-t \cdot A} \cdot \mathbf{x}(t)$  und haben dann

$$\mathbf{x}(t) = e^{t \cdot A} \cdot \mathbf{c}(t), \quad \mathbf{c}(0) = \mathbf{x}(0).$$

Durch Differenzieren und Vergleich mit dem System finden wir

$$\begin{aligned} \dot{\mathbf{x}}(t) &= A \cdot e^{t \cdot A} \cdot \mathbf{c}(t) + e^{t \cdot A} \cdot \dot{\mathbf{c}}(t) = A \cdot \mathbf{x}(t) + e^{t \cdot A} \cdot \dot{\mathbf{c}}(t) \\ &= A \cdot \mathbf{x}(t) + \Phi(\mathbf{x}(t)) \\ e^{t \cdot A} \cdot \dot{\mathbf{c}}(t) &= \Phi(\mathbf{x}(t)) \\ \dot{\mathbf{c}}(t) &= e^{-t \cdot A} \cdot \Phi(\mathbf{x}(t)) \end{aligned}$$

Die letzte Gleichung integrieren wir (rein formal)

$$\mathbf{c}(t) = \mathbf{c}(0) + \int_0^t e^{-s \cdot A} \cdot \Phi(\mathbf{x}(s)) ds.$$

Multiplikation dieser Gleichung von links mit  $e^{t \cdot A}$  liefert die Behauptung. □

**Satz 3.18 (Lemma von Gronwall)** *Es seien  $f, g : [a, b] \rightarrow \mathbb{R}$  zwei stetige Funktionen mit  $f \geq 0, g \geq 0$ . Es gebe ein  $c \geq 0$  so, dass für  $a \leq t < b$  die Ungleichung*

$$f(t) \leq c + \int_a^t f(s) \cdot g(s) ds$$

*gilt. Dann gilt für diese  $t$  auch*

$$f(t) \leq c \cdot \exp\left(\int_a^t g(s) ds\right).$$

Beweis. Es genügt den Fall  $c > 0$  zu behandeln. Denn falls  $c = 0$  sein sollte gilt die Voraussetzung und dann auch die Behauptung automatisch für alle  $c > 0$ . Es folgt  $f \equiv 0$ .

Sei also  $c > 0$ . Dann ist die Funktion auf der rechten Seite der Voraussetzung

$$h(t) := c + \int_a^t f(s) \cdot g(s) ds > 0.$$

Diese Funktion ist differenzierbar und für ihre Ableitung gilt

$$h'(t) = f(t) \cdot g(t) \leq h(t) \cdot g(t), \quad \frac{h'(t)}{h(t)} \leq g(t).$$

Durch Integration erhalten wir

$$\ln(h(t)) - \ln(h(a)) = \ln(h(t)) - \ln(c) = \int_a^t \frac{h'(s)}{h(s)} ds \leq \int_a^t g(s) ds,$$

$$\ln(h(t)) \leq \ln(c) + \int_a^t g(s) ds.$$

Anwendung der Exponentialfunktion liefert

$$h(t) \leq c \cdot \exp\left(\int_a^t g(s) ds\right)$$

und damit die Behauptung. □

**Satz 3.19 (Matrix-Abschätzung)** *Die Matrix  $A$  habe lauter Eigenwerte mit negativem Realteil, etwa*

$$\operatorname{Re}(\lambda_i) < -\Lambda \text{ für alle } i, \quad \Lambda > 0.$$

*Dann gibt es eine Konstante  $c$  mit*

$$\|e^{tA}\| \leq c \cdot e^{-\Lambda t} \text{ für alle } t \geq 0.$$

Beweis. Es sei  $A = T^{-1} \cdot J \cdot T$  mit einer Matrix  $J$  in Jordan-Normalform. Wegen

$$\|e^{tA}\| = \|T^{-1} \cdot e^{tJ} \cdot T\| \leq \|T\| \cdot \|T^{-1}\| \cdot \|e^{tJ}\|$$

genügt es die Ungleichung für die Matrix  $J$  zu beweisen. Wegen

$$\left\| \begin{pmatrix} J_1 & & \\ & \ddots & \\ & & J_k \end{pmatrix} \right\| \leq \sum_1^k \|J_i\|$$

brauchen wir uns sogar nur um einen einzigen Jordanblock kümmern. Den schreiben wir

$$J = \lambda_i \cdot \mathbb{1} + N$$

mit der Nebendiagonalmatrix  $N$ . Für diesen haben wir

$$e^{t \cdot \lambda_i \cdot \mathbb{1} + t \cdot N} = e^{t \cdot \lambda_i} \cdot e^{t \cdot N}.$$

Für große  $k$  ist  $N^k = 0$ , deswegen sind alle Einträge in  $e^{t \cdot N}$  Polynome in  $t$ . Dann ist auch  $\|e^{t \cdot N}\|$  ein Polynom  $p(t)$ . Nun ist

$$\|e^{t \cdot \Lambda} \| e^{t \cdot J} \| = e^{t \cdot (\lambda_i + \Lambda)} \cdot p(t).$$

Wegen  $\operatorname{Re}(\lambda_i + \Lambda) < 0$  geht für  $t \rightarrow \infty$  diese Funktion gegen 0. Insbesondere ist sie für alle  $t \geq 0$  beschränkt  $< c$ .  $\square$

Beweis von Satz 3.16. Nachdem wir den stationären Punkt in den Ursprung verschoben haben wir das System

$$\dot{\mathbf{x}} = A \cdot \mathbf{x} + \Phi(\mathbf{x}).$$

Nach Voraussetzung hat die Matrix lauter reelle, negative Eigenwerte  $\lambda_i$ , etwa  $\lambda_i < -\Lambda$  mit  $\Lambda > 0$ . Wir wählen eine Konstante  $c$  wie in Satz 3.19. Weiter wählen wir einen Radius  $r > 0$  so klein, dass

- 1) die Kugel  $K := \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\| \leq r\}$  ganz im Definitionsbereich  $U$  der rechten Seite unserer Differentialgleichung, d.h., im Definitionsbereich von  $\Phi$  liegt, und
- 2) für  $\|\mathbf{x}\| \leq r$  gilt

$$\|\Phi(\mathbf{x})\| < \frac{\Lambda}{2c} \|\mathbf{x}\|.$$

Schließlich wählen wir  $\epsilon := r/c$  und wollen für dieses  $\epsilon$  die Eigenschaft aus Definition 3.15 nachweisen. Sei also  $\mathbf{x}(t)$  eine Lösung mit  $\|\mathbf{x}(0)\| < \epsilon$ . Für alle  $t \geq 0$  im Definitionsintervall dieser Lösung, für welche  $\mathbf{x}(t) \in K$  ist, gilt nach Satz 3.17

$$\mathbf{x}(t) = e^{t \cdot A} \mathbf{x}(0) + \int_0^t e^{(t-s) \cdot A} \cdot \Phi(\mathbf{x}(s)) ds.$$

Mit Satz 3.19 erhalten wir daraus die Abschätzung

$$\begin{aligned} \|\mathbf{x}(t)\| &\leq c \cdot e^{-t \cdot \Lambda} \cdot \|\mathbf{x}(0)\| + \int_0^t c \cdot e^{(s-t) \cdot \Lambda} \cdot \|\Phi(\mathbf{x}(s))\| ds \\ &\leq c \cdot e^{-t \cdot \Lambda} \cdot \|\mathbf{x}(0)\| + \int_0^t c \cdot e^{(s-t) \cdot \Lambda} \cdot \frac{\Lambda}{2c} \cdot \|\mathbf{x}(s)\| ds \\ &= e^{-t \cdot \Lambda} \cdot \left( c \cdot \|\mathbf{x}(0)\| + \int_0^t e^{s \cdot \Lambda} \|\mathbf{x}(s)\| \cdot \frac{\Lambda}{2} ds \right). \end{aligned}$$

Jetzt setzen wir

$$f(t) := e^{t \cdot \Lambda} \cdot \|\mathbf{x}(t)\|, \quad g(t) := \frac{\Lambda}{2}$$

und haben

$$f(t) \leq c \cdot \|\mathbf{x}(0)\| + \int_0^t f(s) \cdot g(s) ds.$$

Mit Satz 3.18 (Gronwall) folgt daraus

$$f(t) \leq c \cdot \|\mathbf{x}(0)\| \cdot \exp\left(\int_0^t \frac{\Lambda}{2} ds\right) \leq c \cdot \epsilon \cdot e^{t \cdot \frac{\Lambda}{2}} \leq r \cdot e^{t \cdot \frac{\Lambda}{2}},$$

$$\|\mathbf{x}(t)\| \leq r \cdot e^{-t \cdot \frac{\Lambda}{2}}.$$

Deswegen kann für kein  $t$  im Definitionsbereich der Lösung  $\mathbf{x}(t)$  diese Lösung tatsächlich an den Rand von  $K$  gelangen. Sie bleibt immer in der kompakten Kugel  $K$ . Nach Satz 3.14 existiert die Lösung für alle Zeiten  $t \geq 0$ . Und aus der bewiesenen Ungleichung folgt

$$\lim_{t \rightarrow \infty} \mathbf{x}(t) = \mathbf{0}. \quad \square$$

In den anderen Fällen (stabil, instabil) kann man vom Stabilitätsverhalten der Linearisierung nicht auf das des nicht-linearen Problems schließen. Eine Aussage kann man allerdings machen:

**Satz 3.20** *Das autonome System  $\dot{\mathbf{x}} = F(\mathbf{x})$  habe einen stationären Punkt  $\mathbf{x}_0$ , in dem die Funktionalmatrix  $A = F'(\mathbf{x}_0)$  nur Eigenwerte mit echt positivem Realteil besitzt. Dann ist dieser Fixpunkt instabil.*

Beweis. Um Schreibarbeit zu sparen nehmen wir wieder o.B.d.A. an, dass  $\mathbf{x}_0 = \mathbf{0}$ . Unser System ist dann

$$\dot{\mathbf{x}} = F(\mathbf{x}) = A \cdot \mathbf{x} + \Phi(\mathbf{x}).$$

Zu jeder Lösungskurve  $\mathbf{x}(t)$  dieses Systems gehört via Zeit-Umkehr eine Kurve  $\mathbf{y}(t) := \mathbf{x}(-t)$ . Diese Kurve erfüllt

$$\dot{\mathbf{y}}(t) = -\dot{\mathbf{x}}(-t) = -F(\mathbf{x}(-t)) = -F(\mathbf{y}) = -A \cdot \mathbf{y} - \Phi(\mathbf{y}).$$

Damit ist  $\mathbf{y}$  Lösungskurve eines autonomen Systems, welches die Voraussetzung von Satz 3.16 erfüllt. Also ist der Nullpunkt ein asymptotisch stabiler stationärer Punkt für dieses System. Es gibt eine Lösungskurve mit

$$\|\mathbf{y}(0)\| =: 2\epsilon > 0 \quad \text{und} \quad \lim_{t \rightarrow \infty} \mathbf{y}(t) = \mathbf{0}.$$

Für die Lösung  $\mathbf{x}(t) = \mathbf{y}(-t)$  des ursprünglichen Systems gilt entsprechend

$$\|\mathbf{x}(0)\| =: 2\epsilon > 0 \quad \text{und} \quad \lim_{t \rightarrow -\infty} \mathbf{x}(t) = \mathbf{0}.$$

Zu jedem  $\delta > 0$  gibt es also einen Parameter  $t_0 < 0$  mit  $\|\mathbf{x}(t_0)\| < \delta$  und  $\|\mathbf{x}(0)\| > \epsilon$ . Es gibt Lösungskurven, die beliebig nahe bei  $\mathbf{0}$  beginnen, aber irgendwann außerhalb der  $\epsilon$ -Kugel um  $\mathbf{0}$  verlaufen. □

**Aufgabe 3.18** *Finden Sie für das System*

$$\begin{aligned} \dot{x} &= -y^3 \\ \dot{y} &= x^3 \end{aligned}$$

*eine Hamiltonfunktion und folgern Sie daraus: Der Nullpunkt ist ein stabiler, aber nicht asymptotisch stabiler stationärer Punkt.*

**Aufgabe 3.19** *Finden Sie alle stationären Punkte des Systems*

$$\begin{aligned} \dot{x} &= -x + y \\ \dot{y} &= x + y^2 \end{aligned}$$

*und zeigen Sie, dass einer davon asymptotisch stabil ist.*

**Aufgabe 3.20** Finden Sie alle stationären Punkte des Systems

$$\begin{aligned}\dot{x} &= -x - y^2 \\ \dot{y} &= x^2 - y\end{aligned}$$

und zeigen Sie, dass diese asymptotisch stabil sind.

**Aufgabe 3.21** Finden Sie alle stationären Punkte des Systems

$$\begin{aligned}\dot{x} &= 2x + 2xy \\ \dot{y} &= -x^2 + y^2 + 2y\end{aligned}$$

und bestimmen Sie deren Stabilitätstyp.

**Aufgabe 3.22** Finden Sie alle stationären Punkte des Systems

$$\begin{aligned}\dot{x} &= 2xy - 4y - 8 \\ \dot{y} &= 4y^2 - x^2\end{aligned}$$

und bestimmen Sie deren Stabilitätstyp.

**Aufgabe 3.23** Die van der Pol-Gleichung lautet

$$\ddot{x} + a \cdot (x^2 - 1) \cdot \dot{x} + x = 0.$$

Ihr einziger stationärer Punkt ist der Nullpunkt. Bestimmen Sie in Abhängigkeit von  $a \in \mathbb{R}$  seinen Stabilitätstyp.

## Mathematik II für Physiker

SS 08

Probeklausur

**Aufgabe 1:** Entscheiden Sie, welche der folgenden drei Matrizen unitär, hermitesch, bzw. normal sind:

$$A = \frac{1}{13} \begin{pmatrix} 12 & 5i \\ 5i & 12 \end{pmatrix}, \quad B = \frac{1}{13} \begin{pmatrix} 12 & 5i \\ -5i & 12 \end{pmatrix}, \quad C = \frac{1}{13} \begin{pmatrix} 12 & 5i \\ 5 & 12 \end{pmatrix}.$$

**Aufgabe 2a)** In Abhängigkeit von  $t \in \mathbb{R}$  sei die Matrix

$$A_t = \begin{pmatrix} t & 2+t \\ 2+t & t \end{pmatrix}$$

gegeben. Entscheiden Sie, für welche  $t$  diese Matrix positiv, bzw. negativ definit, positiv bzw. negativ semidefinit oder indefinit ist.

b) Bestimmen Sie in Abhängigkeit von  $t \in \mathbb{R}$  den affinen Typ des Kegelschnitts mit der Gleichung

$$t \cdot x^2 + (4 + 2t) \cdot xy + t \cdot y^2 = 1 + 4 \cdot t.$$

**Aufgabe 3:** Bestimmen Sie die absoluten Extrema der Funktion  $f(x, y) = x^3 + y^3$  auf der Menge

$$\{(x, y) \in \mathbb{R}^2 : x^4 + y^4 \leq 1\}.$$

**Aufgabe 4a)** Entscheiden Sie, ob das Vektorfeld

$$A = \begin{pmatrix} x \cdot (x^2 + y^2) \\ y \cdot (x^2 + y^2) \end{pmatrix}$$

auf  $\mathbb{R}^2$  ein Gradientenfeld ist.

b) Entscheiden Sie, ob das Vektorfeld

$$B = \begin{pmatrix} x \\ y \\ -2z \end{pmatrix}$$

auf  $\mathbb{R}^3$  ein Rotationsfeld ist.

**Aufgabe 5:** In welchen Punkten  $(x, y) \in \mathbb{R}^2$  ist die Abbildung

$$F : \mathbb{R}^2 \rightarrow \mathbb{R}^2, \quad \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} e^x \cdot \sin(y) \\ e^y \cdot \sin(x) \end{pmatrix}$$

lokal umkehrbar? Berechnen Sie die Funktionalmatrix der Umkehrabbildung  $F^{-1}$  im Punkt  $F(\pi, 0)$ .

**Aufgabe 6:** Lösen Sie das Anfangswertproblem

$$\dot{x} = x^2 - 1, \quad x(0) = 0.$$

**Aufgabe 7:** Finden Sie alle Lösungen der Differentialgleichung

$$\ddot{x} + \dot{x} + x = \sin(t).$$