

1 Topologie von metrischen Räumen

1.1 Metrische Räume

Definition 1.1. Ein *metrischer Raum* ist ein Paar (X, d) bestehend aus einer Menge X und einer Abbildung

$$d: X \times X \rightarrow \mathbb{R},$$

so daß für alle $x, y, z \in X$ gilt

- $d(x, y) \geq 0$ mit $d(x, y) = 0 \Leftrightarrow x = y$,
- $d(x, y) = d(y, x)$ und
- $d(x, z) \leq d(x, y) + d(y, z)$ (Dreiecksungleichung). (BILD)

Beispiel 1.2. 1. $X = \mathbb{R}$ oder $X = \mathbb{C}$ mit $d(x, y) = |x - y|$.

2. $X = \mathbb{R}^n$ mit euklidischer Abstandsfunktion

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

(Beweis später.)

3. (Berliner U-Bahn Netz)

- $X = \{\text{U-Bahnhöfe in Berlin}\}$
- $d(x, y) =$ Minimale Anzahl von Stationen auf Weg von x nach y .

(Die analoge Definition für das Tübinger Busnetz führt nicht auf eine Metrik im obigen Sinne. Warum?)

4. (Diskrete Metrik)

- X ist beliebige Menge
- $d(x, y) = \begin{cases} 0 & x = y \\ 1 & x \neq y \end{cases}$

Definition 1.3. Ist (X, d) metrischer Raum und $A \subset X$, so nennt man die Einschränkung d_A von d auf A die *induzierte Metrik*.

Bemerkung 1.4. (A, d_A) ist wieder ein metrischer Raum.

Definition 1.5. Eine Teilmenge $A \subset X$ eines metrischen Raumes (X, d) nennt man *beschränkt*, wenn es M gibt mit $d(x, y) \leq M$ für alle $x, y \in A$. Der *Durchmesser* von A ist

$$\text{diam}(A) = \begin{cases} \sup\{d(x, y) \mid x, y \in A\} & \text{falls } A \text{ beschränkt} \\ \infty & \text{sonst} \end{cases}$$

1.2 Topologie eines metrischen Raumes

Definition 1.6. Sei (X, d) ein metrischer Raum.

- Für $\epsilon > 0$, $a \in X$ bezeichne

$$U_\epsilon(a) := \{x \in X \mid d(x, a) < \epsilon\}$$

die *offene ϵ -Umgebung* von a oder *offene ϵ -Kugel* um a .

- $O \subset X$ heißt *offen*, wenn es für jedes $a \in O$ ein $\epsilon > 0$ mit $U_\epsilon(a) \subset O$ gibt.

Bemerkung 1.7. Jede offene ϵ -Umgebung $U_\epsilon(a)$ eines Punktes $a \in X$ in einem metrischen Raum (X, d) ist eine offene Menge.

Beweis. Für $x \in U_\epsilon(a)$ ist $d(x, a) < \epsilon$. Wegen der Dreiecksungleichung liegt damit $U_{\epsilon'}(x)$ mit $\epsilon' = \epsilon - d(x, a)$ ganz in $U_\epsilon(a)$, denn $y \in U_{\epsilon'}(x)$ impliziert

$$d(a, y) \leq d(a, x) + d(x, y) < d(a, x) + \epsilon' = d(a, x) + (\epsilon - d(a, x)) = \epsilon$$

und somit $y \in U_\epsilon(a)$. (BILD)

□

Das System aller offenen Mengen eines metrischen Raumes ist eine Topologie im Sinne der folgenden Definition:

Definition 1.8. Eine Menge \mathcal{T} von Teilmengen einer Menge X heißt *Topologie*, falls

- i) $\emptyset, X \in \mathcal{T}$,
- ii) $O_1, O_2 \in \mathcal{T} \Rightarrow O_1 \cap O_2 \in \mathcal{T}$ (d.h. \mathcal{T} ist stabil unter endlichen Durchschnitten) und
- iii) $O_i \in \mathcal{T}$ für $i \in I \Rightarrow \bigcup_{i \in I} O_i \in \mathcal{T}$ (d.h. \mathcal{T} ist stabil unter beliebigen Vereinigungen).

(X, \mathcal{T}) heißt dann ein topologischer Raum.

Konvention: Wir betrachten in dieser Vorlesung “offiziell” nur metrische Räume. Viele Konstruktionen funktionieren aber auch für allgemeine topologische Räume. Wenn dies ohne Mehraufwand möglich ist bevorzugen wir im Folgenden allgemeingültige Formulierungen. Konstruktionen, die nur in metrischen Räumen funktionieren, werden explizit gekennzeichnet.

(Der Rest dieses Abschnitts überträgt sich wörtlich auf den Fall allgemeiner topologischer Räume.)

Definition 1.9. Sei (X, d) ein metrischer Raum.

- $A \subset X$ heißt *abgeschlossen*, falls $X \setminus A$ offen ist.

- $U \subset X$ heißt *Umgebung* von $a \in X$, falls es eine offene Menge $O \subset X$ gibt mit $a \in O \subset U$.

(Jede offene ϵ -Umgebung $U_\epsilon(a)$ eines Punktes $a \in X$ in einem metrischen Raum (X, d) ist eine Umgebung von a , da sie offen ist, vgl. Bemerkung 1.7).

Bemerkung 1.10. Wir haben hier basierend auf dem Begriff “offene Menge” erklärt, was “abgeschlossene Mengen” und “Umgebungen” sind. Man könnte auch umgekehrt vorgehen und offene Mengen über den Begriff abgeschlossene Menge bzw. Umgebung charakterisieren:

- $O \subset X$ ist offen $\Leftrightarrow X \setminus O$ ist abgeschlossen.
- $O \subset X$ ist offen $\Leftrightarrow O$ ist Umgebung aller Punkte $a \in O$.

(Entsprechend kann man eine Topologie auf einer Menge definieren, indem man ein konsistentes System abgeschlossener Mengen oder Umgebungen vorgibt. Zum Beispiel ist eine Menge \mathcal{A} von Teilmenge einer Menge X genau dann ein System abgeschlossener Mengen bezüglich einer Topologie auf X , wenn \mathcal{A} stabil unter endlichen Vereinigungen und beliebigen Durchschnitten ist sowie $\emptyset, X \in \mathcal{A}$ gilt.)

Beispiel 1.11. 1. Bezüglich der euklidischen Abstandsfunktion auf \mathbb{R}^2 ist die offene ϵ -Umgebung $U_\epsilon(a)$ eines Punktes $a \in \mathbb{R}^2$ eine offene Kreisscheibe mit Mittelpunkt a und Radius ϵ . Der Rand der Scheibe gehört nicht dazu. (BILD)

2. Für $a, b \in X = \mathbb{R}$, $a < b$ mit Standardmetrik $d(x, y) = |x - y|$ ist

- $]a, b[$ offen,
- $[a, b]$ abgeschlossen,
- $[a, b[$ sowie $]a, b]$ weder offen noch abgeschlossen,
- $]a, \infty[$ offen und
- $[a, \infty[$ abgeschlossen.

3. In einer Menge (X, d) mit diskreter Metrik ist jede Teilmenge offen. (Die Topologie, bezüglich der jede Teilmenge offen ist, bezeichnet man auch als *diskrete Topologie*.)

Beweis. Da $U_{1/2}(a)$ (wie immer) offen ist und hier $U_{1/2}(a) = \{a\}$ gilt, ist jede Teilmenge offen (denn jedes $O \in X$ ist $O = \bigcup_{a \in O} U_{1/2}(a)$). \square

Definition 1.12. Sei (X, d) ein metrischer Raum, $Y \subset X$.

- Das *Innere* bzw. der *offene Kern* $\overset{\circ}{Y}$ von Y ist

$$\overset{\circ}{Y} = \bigcup_{O \text{ offen}, O \subset Y} O,$$

also die größte offene Menge, die in Y enthalten ist.

- Die *abgeschlossene Hülle* \bar{Y} von Y ist

$$\bar{Y} = \bigcap_{A \text{ abgeschlossen}, Y \subset A} A,$$

also die kleinste abgeschlossene Menge, in der Y enthalten ist.

- Der Rand ∂Y von Y ist

$$\partial Y = \bar{Y} \setminus \overset{\circ}{Y}.$$

Bemerkung 1.13. Es gilt:

- $\overset{\circ}{Y} = \{y \in Y \mid Y \text{ ist Umgebung von } y\}$,
- $\bar{Y} = X \setminus \overbrace{(X \setminus Y)}^{\circ}$ und somit $\bar{Y} = \{x \in X \mid \text{jede Umgebung von } x \text{ schneidet } Y\}$,
- $\partial Y = \bar{Y} \setminus \overset{\circ}{Y} = (X \setminus \overbrace{(X \setminus Y)}^{\circ}) \setminus \overset{\circ}{Y} = X \setminus (\overbrace{(X \setminus Y)}^{\circ} \cup \overset{\circ}{Y})$, womit ∂Y abgeschlossen ist und gilt $\partial Y = \{x \in X \mid \text{jede Umgebung von } x \text{ schneidet } Y \text{ und } X \setminus Y\}$.

Beispiel 1.14. 1. Sei $X = \mathbb{R}^2$ mit euklidischem Abstand und

$$Y = \{(x, y) \mid x^2 + y^2 \leq 1 \text{ und } x \geq 0 \text{ oder } x^2 + y^2 < 1 \text{ und } x < 0\}.$$

Dann ist (BILD)

- $\overset{\circ}{Y} = \{(x, y) \mid x^2 + y^2 < 1\}$,
 - $\bar{Y} = \{(x, y) \mid x^2 + y^2 \leq 1\}$ und
 - $\partial Y = \{(x, y) \mid x^2 + y^2 = 1\}$.
2. Die Begriffe offen und abgeschlossen sowie die Operationen $Y \rightsquigarrow \overset{\circ}{Y}$, \bar{Y} und ∂Y haben nur relativ zum umgebenden Raum X eine Bedeutung! Sei

$$Y = \{(x, y, 0) \in \mathbb{R}^3 \mid x^2 + y^2 < 1\}.$$

- Betrachtet man Y als Teilmenge von $X = \mathbb{R}^3$, dann ist
 - $\overset{\circ}{Y} = \emptyset$ und
 - $\bar{Y} = \partial Y = \{(x, y, 0) \in X \mid x^2 + y^2 \leq 1\}$.
 - Betrachtet man Y als Teilmenge von $X = \{(x, y, 0) \mid (x, y) \in \mathbb{R}^2\}$ ausgestattet mit der von \mathbb{R}^3 induzierten Metrik, so ist
 - $\overset{\circ}{Y} = Y$,
 - $\bar{Y} = \{(x, y, 0) \in X \mid x^2 + y^2 \leq 1\}$ und
 - $\partial Y = \{(x, y, 0) \in X \mid x^2 + y^2 = 1\}$.
3. Sei $X = \mathbb{R}$ mit Standardmetrik und $Y = \mathbb{Q}$. Dann ist $\overset{\circ}{Y} = \emptyset$, $\bar{Y} = X$ und $\partial Y = \mathbb{R}$.

Definition 1.15. Eine Teilmenge Y eines metrischen Raumes (X, d) heißt *dicht*, falls $\bar{Y} = X$.

1.3 Konvergenz von Folgen

Definition 1.16. Eine Folge $(x_n)_{n \in \mathbb{N}}$ in einem metrischen Raum (X, d) *konvergiert* gegen $a \in X$, wenn jede Umgebung von a fast alle (d.h. alle bis auf endlich viele) Folgenglieder enthält.

Schreibweise: a heißt dann *Limes* oder *Grenzwert* von $(x_n)_{n \in \mathbb{N}}$. Für Konvergenz von $(x_n)_{n \in \mathbb{N}}$ gegen $a \in X$ schreibt man auch

$$a = \lim_{n \rightarrow \infty} x_n \quad \text{oder} \quad x_n \xrightarrow[n \rightarrow \infty]{} a.$$

Eine nicht konvergente Folge heißt *divergent*.

Bemerkung 1.17. Die folgenden Aussagen sind äquivalente Charakterisierungen der Konvergenz von $(x_n)_{n \in \mathbb{N}}$ gegen $a \in X$:

- i) für jede Umgebung U von a gibt es ein N , so daß für alle $n \geq N$ gilt $x_n \in U$,
- ii) für alle $\epsilon > 0$ gibt es ein N , so daß für alle $n \geq N$ gilt $d(a, x_n) < \epsilon$,
- iii) $\lim_{n \rightarrow \infty} d(a, x_n) = 0$, d.h. der Abstand $d(a, x_n)$ ist eine (reelle!) Nullfolge (wie in Analysis I).

Insbesondere stimmt für $X = \mathbb{R}$ mit der Standardmetrik diese Definition der Konvergenz von Folgen mit der aus Analysis I überein.

Die Definition der Konvergenz von Folgen überträgt sich wörtlich auf allgemeine topologische Räume. Anders als im allgemeinen Fall ist für metrische Räume der Grenzwert von Folgen, sofern er existiert, auch eindeutig. Das ist eine direkte Konsequenz der im folgenden Satz bewiesenen Hausdorff-Eigenschaft:

Satz 1.18. *Metrische Räume haben die Hausdorff-Eigenschaft, das heißt für $x, y \in X$ mit $x \neq y$ gibt es Umgebungen U, V von x bzw. y mit $U \cap V = \emptyset$.*

Beweis. Für $x, y \in X$ mit $x \neq y$ setze $\epsilon = \frac{d(x, y)}{2} > 0$. Dann sind die offenen ϵ -Umgebungen $U = U_\epsilon(x)$ und $V = U_\epsilon(y)$ um x bzw. y disjunkt, denn gäbe es $z \in U \cap V$, so wäre

$$d(x, y) \leq d(x, z) + d(z, y) < \epsilon + \epsilon = d(x, y)$$

(was ein Widerspruch ist). (BILD) □

Einen topologischen Raum mit Hausdorff-Eigenschaft bezeichnet man auch als *Hausdorff-Raum*. Die Hausdorff-Eigenschaft ist eine sogenannte Trennungseigenschaft: in einem Hausdorff-Raum kann man je zwei verschiedene Punkte trennen, indem man disjunkte Umgebungen der Punkte angibt.)

Beispiel 1.19. Ein Beispiel für einen topologischen Raum, der die Hausdorff-Eigenschaft nicht erfüllt, ist eine Menge X mit mindestens zwei Elementen und der Topologie $\mathcal{T} = \{\emptyset, X\}$. Bezüglich dieser Topologie hat jedes Element nur eine Umgebung, nämlich X selbst. Insbesondere konvergiert in X also jede Folge gegen jeden Punkt von X .

Satz 1.20. Eine Folge $(x_k)_{k \in \mathbb{N}}$ in \mathbb{R}^n konvergiert bezüglich der euklidischen Metrik gegen $a \in \mathbb{R}^n \Leftrightarrow$ die Komponentenfolgen konvergieren gegen die Komponenten von a .

Beweis. Sei

$$x_k = \begin{pmatrix} x_{1k} \\ \vdots \\ x_{nk} \end{pmatrix} \quad \text{und} \quad \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix},$$

dann ist $x_k \xrightarrow[k \rightarrow \infty]{} a$ äquivalent zu $\lim_{k \rightarrow \infty} \sqrt{(x_{1k} - a_1)^2 + \dots + (x_{nk} - a_n)^2} = 0$ und somit zu $\lim_{k \rightarrow \infty} (x_{jk} - a_j) = 0$ für $j = 1, \dots, n$. \square

Satz 1.21. Für eine Teilmenge $A \subset X$ eines metrischen Raumes (X, d) sind äquivalent

- i) A ist abgeschlossen
- ii) für jede konvergente Folge $(x_n)_{n \in \mathbb{N}}$ in X mit $x_n \in A$ liegt auch der Grenzwert in A .

(ii) \Rightarrow i) gilt in allgemeinen topologischen Räumen nicht.)

Beweis. i) \Rightarrow ii) Sei $(x_n)_{n \in \mathbb{N}}$ eine konvergente Folge in X mit $x_n \in A$. Jede Umgebung U des Grenzwertes x enthält fast alle x_n , hat also nicht-leeren Durchschnitt mit A . Damit gilt $x \in \bar{A}$ und, weil A abgeschlossen und somit $\bar{A} = A$, auch $x \in A$.

ii) \Rightarrow i) Sei $x \in \bar{A}$. Wähle $x_n \in U_{1/n}(x) \cap A$. Die Folge $(x_n)_{n \in \mathbb{N}}$ konvergiert dann gegen x . Also ist $x \in A$. Damit gilt $\bar{A} = A$ und A ist abgeschlossen. \square

Beispiel 1.22. Sind $A_1, \dots, A_n \subset \mathbb{R}$ abgeschlossen, so ist $A_1 \times \dots \times A_n \subset \mathbb{R}^n$ abgeschlossen. (Folgt aus den beiden vorhergehenden Sätzen.)

(Der Rest dieses Abschnitts überträgt sich nicht auf allgemeine topologische Räume.)

Definition 1.23. • Eine Folge $(x_n)_{n \in \mathbb{N}}$ in einem metrischen Raum (X, d) heißt *Cauchy-Folge*, falls es für jedes $\epsilon > 0$ ein N gibt mit $d(x_m, x_n) < \epsilon$ für alle $m, n \geq N$.

- (X, d) heißt *vollständig*, wenn jede Cauchy-Folge konvergiert.

Jede konvergente Folge ist eine Cauchy-Folge.

Beispiel 1.24. \mathbb{R}^n ist vollständig. (Für $n = 1$ wurde das in Analysis I gezeigt. Der allgemeine Fall folgt aus Satz 1.20, da die Komponentenfolgen einer Cauchy-Folge in \mathbb{R}^n Cauchy-Folgen in \mathbb{R} sind.)

Satz 1.25. Sei (X, d) vollständiger metrischer Raum, $A \subset X$. Dann ist A mit der induzierten Metrik d_A vollständig $\Leftrightarrow A \subset X$ ist abgeschlossen.

Beweis. “ \Rightarrow ” Wegen Satz 1.21 reicht es zu zeigen, daß eine in X konvergente Folge, deren Glieder in A liegen, einen Grenzwert in A hat. Dies folgt daraus, daß jede konvergente Folge Cauchy ist und A vollständig ist.

“ \Leftarrow ” Sei $(x_n)_{n \in \mathbb{N}}$ eine Cauchy-Folge in A . Da X vollständig ist, konvergiert die Folge in X . Wegen Satz 1.21 liegt der Grenzwert in A . \square

Beispiel 1.26. Jede abgeschlossene Teilmenge des \mathbb{R}^n ist vollständig. (Folgt aus dem vorhergehenden Beispiel und Satz.)

1.4 Stetige Abbildungen

Definition 1.27. Eine Abbildung $f: X \rightarrow Y$ zwischen metrischen Räumen ist *stetig in* $x \in X$, wenn es für jede Umgebung V von $f(x)$ eine Umgebung U von x gibt mit $f(U) \subset V$. Sie ist *stetig*, wenn sie in jedem Punkt stetig ist.

(Äquivalent ist f stetig in x , wenn das Urbild jeder Umgebung von $f(x)$ eine Umgebung von x ist. Diese Umformulierung der Definition ist dichter am zweiten Punkt der folgenden Bemerkung; die obige Definition 1.27 von Stetigkeit ist ähnlicher der Definition aus Analysis I, vgl. den ersten Punkt der folgenden Bemerkung.)

Bemerkung 1.28. • f ist stetig in $x \Leftrightarrow$ für alle $\epsilon > 0$ gibt es $\delta > 0$ mit

$$f(U_\delta(x)) \subset U_\epsilon(f(x))$$

bzw.

$$d_X(x, x') < \delta \Rightarrow d_Y(f(x), f(x')) < \epsilon \quad \text{für alle } x' \in X$$

(wobei d_X und d_Y die Metriken auf X und Y bezeichnen).

- f ist stetig \Leftrightarrow für alle $O \subset Y$ offen ist $f^{-1}(O)$ offen in $X \Leftrightarrow$ für alle $A \subset Y$ abgeschlossen ist $f^{-1}(A)$ abgeschlossen in X (Hausaufgabe Blatt 2)

Satz 1.29. (Folgenkriterium für Stetigkeit) Eine Abbildung $f: X \rightarrow Y$ zwischen metrischen Räumen ist genau dann stetig in $x \in X$, wenn für jede Folge $(x_n)_{n \in \mathbb{N}}$ in X gilt

$$x_n \xrightarrow[n \rightarrow \infty]{} x \quad \Longrightarrow \quad f(x_n) \xrightarrow[n \rightarrow \infty]{} f(x).$$

Beweis. “ \Rightarrow ” Sei $(x_n)_{n \in \mathbb{N}}$ eine Folge mit $x_n \xrightarrow[n \rightarrow \infty]{} x$. Da f stetig ist, gibt es zu jeder Umgebung V von $f(x)$ eine Umgebung U mit $f(U) \subset V$. Fast alle Folgenglieder von $(x_n)_{n \in \mathbb{N}}$ liegen in U . Damit liegen fast alle Folgenglieder von $(f(x_n))_{n \in \mathbb{N}}$ in V .

“ \Leftarrow ” Angenommen f ist nicht stetig in x . Dann gibt es eine Umgebung V von $f(x)$ ohne “passende” Umgebung U von x . Insbesondere gibt es zu jedem n ein $x_n \in U_{1/n}(x)$ mit $f(x_n) \notin V$. Das ist ein Widerspruch, denn $(x_n)_{n \in \mathbb{N}}$ konvergiert gegen x aber $(f(x_n))_{n \in \mathbb{N}}$ konvergiert nicht gegen $f(x)$. \square

(In allgemeinen topologischen Räumen gilt nur "=>".)

Beispiel 1.30. 1. Die Abbildungen

- $\mathbb{R}^2 \rightarrow \mathbb{R}, (x, y) \mapsto x + y$
- $\mathbb{R}^2 \rightarrow \mathbb{R}, (x, y) \mapsto x \cdot y$
- $\{(x, y) \in \mathbb{R}^2 \mid y \neq 0\} \rightarrow \mathbb{R}, (x, y) \mapsto x/y$

sind stetig. (Mit den Sätzen 1.29 und 1.20 folgt das direkt aus den Regeln für Grenzwerte von Summen, Produkten bzw. Quotienten zweier konvergenter Folgen aus Analysis I.)

2. Mit den gleichen Argumenten sind auch

- $\mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}^n, (x, y) \mapsto x + y$
- $\mathbb{R} \times \mathbb{R}^n \rightarrow \mathbb{R}^n, (\lambda, x) \mapsto \lambda \cdot x$
- $\mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}, (x, y) \mapsto \sum_{i=1}^n x_i y_i$

stetig.

3. Genauso folgt Stetigkeit von Abbildungen, die polynomial in den Koordinaten sind, z.B.

$$\det: M_{n \times n}(\mathbb{R}) \rightarrow \mathbb{R} \quad (a_{ij}) \mapsto \sum_{\sigma} \operatorname{sgn}(\sigma) a_{1\sigma(1)} \cdots a_{n\sigma(n)}.$$

4. Die Abbildungen

- $f = pr_i: \mathbb{R}^n \rightarrow \mathbb{R} \quad (x_1, \dots, x_m) \mapsto x_i$ und
- $g: \mathbb{R} \rightarrow \mathbb{R}^n, x \mapsto x \cdot e_i = (0, \dots, x, \dots, 0)$

sind stetig (auch eine direkte Konsequenz der Sätze 1.29 und 1.20).

Lemma 1.31. *Ist $f: X_1 \rightarrow X_2$ stetig in p und $g: X_2 \rightarrow X_3$ stetig in $f(p)$, so ist $g \circ f$ stetig in p .*

Beweis. Sei U_3 eine Umgebung von $(g \circ f)(p)$. Dann gibt es eine Umgebung U_2 von $f(p)$ mit $g(U_2) \subset U_3$ und eine Umgebung U_1 von p mit $f(U_1) \subset U_2$. Also ist $(g \circ f)(U_1) \subset U_3$. (BILD) □

Korollar 1.32. *Eine Abbildung $f: X \rightarrow \mathbb{R}^n$ von einem metrischen Raum X nach \mathbb{R}^n ist genau dann stetig, wenn die Komponentenfunktionen $f_i = pr_i \circ f$ stetig sind.*

Beweis. “ \Rightarrow ” Folgt aus dem vorhergehenden Lemma, da pr_i stetig ist (Teil 4 des vorhergehenden Beispiels).

“ \Leftarrow ” Sind die f_i stetig, so ist wegen Teil 2. und 4. des vorhergehenden Beispiels auch

$$f = \begin{pmatrix} f_1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \dots + \begin{pmatrix} 0 \\ \vdots \\ 0 \\ f_n \end{pmatrix}$$

stetig.

□

Wichtig! (Das vorhergehende Korollar und das folgende Beispiel zeigen:)

- Es gilt: $f: X \rightarrow \mathbb{R}^n$ stetig \Leftrightarrow alle Komponentenfunktionen sind stetig
- Aber: $f: \mathbb{R}^m \rightarrow Y$ stetig $\not\Rightarrow x_i \mapsto f(x_1, \dots, x_i, \dots, x_m)$, x_j , $j \neq i$ fest, stetig

Beispiel 1.33. (Für Stetigkeit in Null reicht es nicht aus, daß f entlang der Achsen stetig ist.) Wir betrachten $f: \mathbb{R}^2 \rightarrow \mathbb{R}$ definiert durch

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

Dann sind $x \mapsto f(x, 0)$ und $y \mapsto f(0, y)$ stetig, d.h. f ist “stetig entlang der Koordinatenachsen”. Entlang aller anderen Geraden durch Null ist f jedoch unstetig: betrachtet man f entlang der Geraden $y = \lambda x$ mit Steigung $\lambda \in \mathbb{R}$, so gilt $f(x, \lambda x) = \frac{\lambda}{1+\lambda^2}$ falls $x \neq 0$. Für $\lambda \neq 0$ ist also der Grenzwert von $f(1/k, \lambda/k)$ für $k \rightarrow \infty$ nicht gleich $f(0, 0) = 0$. Da aber $(1/k, \lambda/k)$ für $k \rightarrow \infty$ gegen $(0, 0)$ konvergiert, kann f in $(0, 0)$ nicht stetig sein.

Beispiel 1.34. (Für Stetigkeit in Null reicht es nicht aus, daß f in Richtung aller Geraden durch Null stetig ist.) Sei $f: \mathbb{R}^2 \rightarrow \mathbb{R}$ definiert durch (BILD)

$$f(x, y) = \begin{cases} 1 & y = x^2, x \neq 0 \\ 0 & \text{sonst.} \end{cases}$$

Für alle $\lambda \in \mathbb{R}$ ist $x \mapsto f(x, \lambda x)$ stetig in $x = 0$. Auch $y \mapsto f(0, y)$ ist stetig. Aber $f(1/k, 1/k^2) = 1$ für alle $k \in \mathbb{N}$. Also konvergiert $(1/k, 1/k^2)$ für $k \rightarrow \infty$ gegen $(0, 0)$, aber $f(1/k, 1/k^2)$ konvergiert nicht gegen $f(0, 0) = 0$, womit f in $(0, 0)$ unstetig ist.

Erinnerung (Analysis I)

In den folgenden beiden Abschnitten werden wir unter anderem diese drei wichtigen Tatsachen aus der Analysis I verallgemeinern:

1. (Zwischenwertsatz) Ist $f: I \rightarrow \mathbb{R}$ stetig auf einem Intervall I , so ist auch $f(I)$ ein Intervall, d.h. für $a, b \in I$ werden alle Werte zwischen $f(a)$ und $f(b)$ angenommen.
2. (Satz vom Maximum/Minimum) Ist $f: [a, b] \rightarrow \mathbb{R}$ stetig auf einem kompakten Intervall $[a, b]$, so nimmt f Maximum und Minimum an.
3. Eine stetig Funktion $f: [a, b] \rightarrow \mathbb{R}$ auf einem kompakten Intervall $[a, b]$ ist gleichmäßig stetig.

1.5 Zusammenhängende Räume

Angenommen $f: X \rightarrow \mathbb{R}$ ist eine stetige Abbildung von einem metrischen Raum X nach \mathbb{R} , für die “der Zwischenwertsatz nicht gilt”, d.h. es gibt Punkte p und $q \in X$ für die ein Wert $R \in \mathbb{R}$ mit

$$f(p) < R < f(q)$$

nicht angenommen wird, es also kein $r \in X$ gibt mit $R = f(r)$. Dann sind $U = f^{-1}(]-\infty, R])$ und $V = f^{-1}(]R, \infty[)$ zwei nicht-leere offene Teilmengen mit

$$U \cup V = X \quad \text{und} \quad U \cap V = \emptyset.$$

Definition 1.35. Ein metrischer Raum (X, d) heißt *zusammenhängend*, falls für alle offenen Teilmengen $U, V \subset X$

$$U \cup V = X \quad \text{und} \quad U \cap V = \emptyset \quad \Rightarrow \quad U = \emptyset \quad \text{oder} \quad V = \emptyset.$$

Bemerkung 1.36. Ein Raum X ist genau dann zusammenhängend, wenn X und \emptyset die einzigen Teilmengen sind, die sowohl offen als auch abgeschlossen sind.

Der folgende Satz verallgemeinert den Zwischenwertsatzes aus Analysis I

Satz 1.37. *Ist (X, d) zusammenhängend und $f: X \rightarrow \mathbb{R}$ stetig, so gibt es für alle $p, q \in X$ und $R \in \mathbb{R}$ mit $f(p) < R < f(q)$ ein $r \in X$, so daß $R = f(r)$.*

Beweis. Der Satz folgt direkt aus den Betrachtungen am Anfang des Abschnitts (vor Definition 1.35). \square

Definition 1.38. Eine Teilmenge $B \subset X$ eines metrischen Raumes (X, d) ist *zusammenhängend*, wenn (B, d_B) mit der induzierten Metrik zusammenhängend ist, d.h. wenn für alle in X offenen Mengen $U, V \subset X$ gilt

$$B \subset U \cup V \quad \text{und} \quad U \cap V \cap B = \emptyset \quad \Rightarrow \quad U \cap B = \emptyset \quad \text{oder} \quad V \cap B = \emptyset.$$

Beispiel 1.39. Eine Teilmenge $I \subset \mathbb{R}$ ist genau dann zusammenhängend, wenn sie ein Intervall ist (d.h. wenn für alle $p, q \in I$ und $r \in \mathbb{R}$ gilt $p < r < q \Rightarrow r \in I$).

Beweis. “ \Rightarrow ” Annahme I ist kein Intervall. Dann gibt es $p < r < q$ mit $p, q \in I$ und $r \notin I$. Damit sind $U =]-\infty, r[$ und $V =]r, \infty[$ offene Teilmengen von \mathbb{R} mit $I \subset U \cup V$ und $U \cap V \cap I = \emptyset$. Da $U \cap I$ und $V \cap I$ nicht leer sind (denn sie enthalten p bzw. q), ist I nicht zusammenhängend.

“ \Leftarrow ” Annahme I ist Intervall aber nicht zusammenhängend, d.h. es gibt offene Mengen $U, V \subset \mathbb{R}$ in \mathbb{R} mit $I \subset U \cup V$ sowie $U \cap V \cap I = \emptyset$, aber $U \cap I \neq \emptyset$ und $V \cap I \neq \emptyset$.

Sei $p \in U \cap I$ und $q \in V \cap I$. ObdA sei $p < q$ (sonst vertauschen wir U und V). Wir definieren

$$r = \sup\{t \in \mathbb{R} \mid [p, t] \subset U \cap I\}.$$

Nach Konstruktion ist $p < r < q$ und $r \in I$.

Es gilt entweder $r \in U$ oder $r \in V$. Im ersten Fall gibt es, da U offen und $[p, q] \subset I$ ist, ein $\epsilon > 0$ mit $[p, r + \epsilon] \subset U \cap I$, was der Definition von r widerspricht. Im zweiten Fall gibt es, da V offen und $[p, q] \subset I$ ist, ein $\epsilon > 0$ mit $r - \epsilon \in V \cap I$. Da $V \cap I$ und $U \cap I$ disjunkt sind, widerspricht auch das der Definition von r . Die Annahme I sei nicht zusammenhängend führt also auf einen Widerspruch. \square

Mit diesem Beispiel folgt der klassische Zwischenwertsatz (wie in Analysis I) aus Satz 1.37 und Satz 1.37 folgt aus dem folgenden Satz.

Satz 1.40. Ist $f: X \rightarrow Y$ stetig und X zusammenhängend, so ist auch $f(X)$ zusammenhängend.

Beweis. Seien $U, V \subset Y$ offene Mengen mit $f(X) \subset U \cup V$ und $f(X) \cap U \cap V = \emptyset$. Dann sind die offenen Mengen $f^{-1}(U)$ und $f^{-1}(V)$ eine disjunkte Zerlegung von X . Da X zusammenhängend ist, muss also eine der beiden Mengen leer sein und somit $f(X) \cap U = \emptyset$ oder $f(X) \cap V = \emptyset$ gelten. \square

Satz 1.41. Für Teilmengen $A \subset B \subset \bar{A} \subset X$ eines metrischen Raumes gilt

$$A \text{ zusammenhängend} \quad \Rightarrow \quad B \text{ zusammenhängend.}$$

Beweis. Seien $O_1, O_2 \subset X$ offen und $B \subset O_1 \cup O_2$ sowie $B \cap O_1 \cap O_2 = \emptyset$. Da A zusammenhängend ist, gilt dann $O_1 \cap A = \emptyset$ oder $O_2 \cap A = \emptyset$. ObdA sei $O_1 \cap A = \emptyset$. Dann muss auch $B \cap O_1 = \emptyset$ gelten, denn gäbe es $b \in B \cap O_1$, so müßte wegen $b \in \bar{A}$ auch $O_1 \cap A$ nicht leer sein. \square

Definition 1.42. Ein metrischer Raum ist *wegzusammenhängend*, wenn es für alle $p, q \in X$ eine stetige Abbildung $c: [0, 1] \rightarrow X$ gibt mit $c(0) = p$ und $c(1) = q$. (BILD)

Übungsaufgaben (Blatt 2):

- wegzusammenhängend \Rightarrow zusammenhängend
- $\{(x, y) \in \mathbb{R}^2 \mid \text{entweder } x = 0 \text{ oder } y = \sin(1/x)\}$ ist zusammenhängend aber nicht wegzusammenhängend. (Man kann zeigen, daß offene Mengen in \mathbb{R}^n genau dann wegzusammenhängend sind wenn sie zusammenhängend sind.)

1.6 Kompaktheit

Definition 1.43. Sei X ein metrischer Raum.

- Eine *offene Überdeckung* von X ist eine Familie $(U_i)_{i \in I}$ von offenen Mengen mit

$$X = \bigcup_{i \in I} U_i.$$

- X heißt *kompakt*, wenn jede offene Überdeckung $(U_i)_{i \in I}$ eine endliche Teilüberdeckung hat, d.h. es eine endliche Menge $I' \subset I$ gibt mit $X = \bigcup_{i \in I'} U_i$.

(Für allgemeine topologische Räume fordert man in der Definition von Kompaktheit meist noch, daß X Hausdorff ist.)

Achtung! Kompaktheit ist nicht gleichbedeutend dazu, daß es eine endliche Überdeckung gibt (denn die gibt es immer).

Bemerkung 1.44. Eine Teilmenge $A \subset X$ eines metrischen Raumes (X, d) ist kompakt, wenn (A, d_A) mit der induzierten Metrik d_A kompakt ist. Das ist genau dann der Fall, wenn für jede Familie $(U_i)_{i \in I}$ von offenen Mengen in X mit $A \subset \bigcup_{i \in I} U_i$ schon für ein endliches $I' \subset I$ gilt $A \subset \bigcup_{i \in I'} U_i$.

Beispiel 1.45. Ist $(x_n)_{n \in \mathbb{N}}$ eine gegen x konvergente Folge. Dann ist

$$A = \{x_n \mid n \in \mathbb{N}\} \cup \{x\}$$

kompakt.

Beweis. Sei $(U_i)_{i \in I}$ eine Familie offener Menge in X , die A überdeckt. Dann gibt es $i_0 \in I$ mit $x \in U_{i_0}$. Wegen der Konvergenz der Folge liegen fast alle x_n in U_{i_0} . Die endlich vielen Folgenglieder außerhalb von U_{i_0} kann man durch endlich viele Mengen in der Familie überdecken. \square

1.6.1 Satz von Heine–Borel

Der Satz von Heine–Borel charakterisiert kompakte Teilmengen des euklidischen Raumes. Der Beweis ist aufgeteilt in ein Beispiel und zwei vorbereitende Sätze, die beide auch in einem allgemeineren Rahmen gelten.

Beispiel 1.46. Seien $I_k = [a_k, b_k]$, $k = 1, \dots, n$ abgeschlossene und beschränkte Intervalle in \mathbb{R} . Dann ist $Q = I_1 \times \dots \times I_n \subset \mathbb{R}^n$ kompakt.

Beweis. Sei $(U_i)_{i \in I}$ eine Familie offener Menge in \mathbb{R}^n die Q überdeckt. Wie nehmen an, es gibt keine endliche Teilüberdeckung. Durch Halbierung der Intervalle zerlegen wir den Quader in 2^n kleinere Quader. Von diesen besitzt mindestens einer keine endliche Teilüberdeckung. Durch sukzessives Halbieren der Intervalle kann man eine Folge $(Q_k)_{k \in \mathbb{N}}$ von Quadern konstruieren mit

- $Q_{k+1} \subset Q_k$ und $\text{diam}(Q_k) = \frac{1}{2^k} \text{diam}(Q)$
- Q_k hat keine endliche Teilüberdeckung.

Nach dem Intervallschachtelungsprinzip (angewandt auf die Seiten der Quader) gibt es genau ein $x \in \mathbb{R}^n$ mit $\{x\} = \bigcap_k Q_k$. Nun gibt es i_0 und ϵ mit $U_{\epsilon}(x) \subset U_{i_0}$. Weil die Durchmesser der Quader beliebig klein werden, gilt dann aber $Q_k \subset U_{i_0}$ für k groß genug. Widerspruch. \square

Satz 1.47. Sei $A \subset X$ eine kompakte Teilmenge eines metrischen Raumes. Dann ist A abgeschlossen und beschränkt.

Beweis. A ist beschränkt: OBdA sei $A \neq \emptyset$ und $x \in A$. Dann ist

$$U_n = U_n(x) = \{y \in X \mid d(x, y) < n\}, \quad n \in \mathbb{N}$$

eine offene Überdeckung von A . Da es eine endliche Teilüberdeckung gibt, ist $A \subset U_N$ für ein N . Damit ist wegen der Dreiecksungleichung $\text{diam}(X) \leq 2N$. (BILD)

A ist abgeschlossen: sein $x \in X \setminus A$. Dann gibt es wegen der Hausdorff-Eigenschaft für jeden Punkt $p \in A$ disjunkte offene Mengen U_p und V_p mit $p \in U_p$ und $x \in V_p$. Da die $(U_p)_{p \in A}$ eine offene Überdeckung von A bilden, gibt es endlich viele Punkte p_1, \dots, p_n mit $X \subset U_{p_1} \cup \dots \cup U_{p_n}$. Damit ist $V_{p_1} \cap \dots \cap V_{p_n}$ eine offene Umgebung von x , die A nicht schneidet. Also ist $X \setminus A$ offen. \square

(Abgeschlossenheit von kompakten Teilmengen gilt mit dem gleichen Argument für allgemeine Hausdorff-Räume.)

Satz 1.48. *Jede abgeschlossene Teilmenge eines kompakten Raumes ist kompakt.*

Beweis. Sei $A \subset X$ eine abgeschlossene Teilmenge eines kompakten Raumes und $(U_i)_{i \in I}$ eine Überdeckung von A durch offene Teilmengen von X . Dann ist $(X \setminus A, U_i)_{i \in I}$ eine offene Überdeckung von X und es gibt i_1, \dots, i_n , so daß X von $X \setminus A \cup U_{i_1} \cup \dots \cup U_{i_n}$ überdeckt wird. Insbesondere wird A von $U_{i_1} \cup \dots \cup U_{i_n}$ überdeckt. \square

Satz 1.49. (Heine–Borel) $A \subset \mathbb{R}^n$ ist genau dann kompakt, wenn A abgeschlossen und beschränkt ist.

Beweis. “ \Rightarrow ” gilt mit Satz 1.47. “ \Leftarrow ” Da A beschränkt ist, gibt es wegen Beispiel 1.46 einen kompakten Quader Q mit $A \subset Q$. Da A abgeschlossen ist, ist A somit wegen Satz 1.48 auch kompakt. \square

Beispiel 1.50. (Es gibt metrische Räume, in denen die Heine–Borel–Eigenschaft nicht gilt.) Sei X eine Menge und $B(X, \mathbb{R}) = \{f: X \rightarrow \mathbb{R} \mid f \text{ beschränkt}\}$ mit Metrik

$$d(f, g) = \sup_{x \in X} |f(x) - g(x)|.$$

Dann ist $A = \{\delta_x \mid x \in X\}$ mit

$$\delta_x(y) = \begin{cases} 1 & y = x \\ 0 & y \neq x \end{cases}$$

abgeschlossen (da das Komplement offen ist) und beschränkt. Wenn X keine endliche Menge ist, ist A nicht kompakt, denn die offene Überdeckung $(U_x)_{x \in X}$ mit $U_x = U_{1/2}(x)$ besitzt dann keine endliche Teilüberdeckung.

(Im Abschnitt über normierte Vektorräume werden wir zeigen, daß die Heine–Borel–Eigenschaft genau für endlichdimensionale Vektorräume gilt.)

1.6.2 Satz von Bolzano–Weierstraß

Definition 1.51. Ein metrischer Raum X heißt *folgenkompakt*, wenn jede Folge in X eine konvergente Teilfolge hat.

Satz 1.52. (Bolzano–Weierstraß) *Ein metrischer Raum X ist genau dann kompakt, wenn er folgenkompakt ist.*

(In allgemeinen topologischen Räumen gilt weder die Implikation “kompakt \Rightarrow folgenkompakt” noch “folgenkompakt \Rightarrow kompakt” .)

Beweis. “ \Rightarrow ” Sei $(x_k)_{k \in \mathbb{N}}$ eine Folge, die keine konvergente Teilfolge besitzt. Dann gibt es für jedes $p \in X$ eine Umgebung U_p , die nur endliche viele Folgenglieder enthält. (Sonst gäbe es $p \in X$ mit der Eigenschaft, daß jede Umgebung von $p \in X$ unendlich viele Folgenglieder enthält. Insbesondere gäbe es dann eine Teilfolge x_{n_k} mit $x_{n_k} \in U_{1/k}(p)$ für alle $k \in \mathbb{N}$. Diese Teilfolge würde dann gegen p konvergieren.) Die offene Überdeckung $(U_p)_{p \in X}$ hat eine endliche Teilüberdeckung U_{p_1}, \dots, U_{p_n} . Damit kann auch $X = U_{p_1} \cup \dots \cup U_{p_n}$ nur endliche viele Folgenglieder enthalten. Widerspruch.

“ \Leftarrow ” Folgt aus Lemma 1.53 und Lemma 1.54 unten. □

Lemma 1.53. *Sei X ein folgenkompakter metrischer Raum. Dann gibt es zu jeder offenen Überdeckung $(U_i)_{i \in I}$ ein $\delta > 0$, so daß jede offene δ -Umgebung $U_\delta(x)$, $x \in X$ in einem U_i liegt. (δ wird als Lebesgue–Zahl der Überdeckung bezeichnet.)*

Beweis. Angenommen, es gibt kein δ . Dann gibt es für jedes $n \in \mathbb{N}$ ein x_n , so daß $U_{1/n}(x_n)$ in keinem U_i liegt. Sei x_{n_k} eine gegen x konvergente Teilfolge. Dann gibt es ein $\epsilon > 0$ mit $U_\epsilon(x) \subset U_{i_0}$ für ein $i_0 \in I$. Das widerspricht unserer Annahme, denn $U_{1/n_k}(x_{n_k}) \subset U_\epsilon(x)$ für große k . □

Lemma 1.54. *Sei X ein folgenkompakter metrischer Raum. Dann gibt es zu jedem $\delta > 0$ eine endliche Überdeckung von X durch offene δ -Umgebungen.*

Beweis. Angenommen, für δ gibt es keine endliche Überdeckung. Dann kann man eine Folge $(x_n)_{n \in \mathbb{N}}$ konstruieren mit $d(x_m, x_n) \geq \delta$ für alle $m \neq n$. Diese Folge hat keine konvergente Teilfolge, da keine Teilfolge eine Cauchy–Folge ist. □

Korollar 1.55. *Jeder kompakte metrische Raum ist vollständig.*

Beweis. Sei $(x_n)_{n \in \mathbb{N}}$ eine Cauchy-Folge. Dann gibt es nach Bolzano-Weierstraß eine konvergente Teilfolge. Eine Cauchy-Folge mit konvergenter Teilfolge ist aber selbst konvergent. \square

1.6.3 Stetige Abbildungen auf kompakten Räumen

Satz 1.56. *Ist $f: X \rightarrow Y$ eine stetige Abbildung zwischen metrischen Räumen und X kompakt, so ist auch $f(X) \subset Y$ kompakt.*

Beweis. Sei $(U_i)_{i \in I}$ eine Überdeckung von $f(X)$ durch in Y offene Mengen. Dann ist $(f^{-1}(U_i))_{i \in I}$ eine offene Überdeckung von X . Diese besitzt eine endliche Teilüberdeckung $f^{-1}(U_{i_1}), \dots, f^{-1}(U_{i_n})$. Damit ist U_{i_1}, \dots, U_{i_n} eine endliche Überdeckung von $f(X)$. \square

Satz 1.57. *(Existenz von Maximum und Minimum) Sei $f: X \rightarrow \mathbb{R}$ eine stetige Funktion auf einem kompakten metrischen Raum X . Dann nimmt f ihr Maximum und Minimum an.*

Beweis. Nach dem vorhergehenden Satz ist $f(X) \subset \mathbb{R}$ kompakt, also abgeschlossen und beschränkt (Satz 1.49 von Heine-Borel). Damit enthält die Menge $f(X)$ ihr Supremum und Infimum: sei zum Beispiel $(x_n)_{n \in \mathbb{N}}$ eine Folge in X , für die $(f(x_n))_{n \in \mathbb{N}}$ gegen das Supremum konvergiert. Da $f(X)$ abgeschlossen ist, liegt auch der Grenzwert der Folge in $f(X)$ (Satz 1.21). \square

(Die beiden vorhergehenden Sätze gelten für allgemeine topologische Räume, der folgende benötigt metrische Räume.)

Definition 1.58. Eine Abbildung $f: X \rightarrow Y$ zwischen metrischen Räumen heißt *gleichmäßig stetig*, wenn es für jedes $\epsilon > 0$ ein $\delta > 0$ gibt mit

$$d(x, y) < \delta \quad \Rightarrow \quad d(f(x), f(y)) < \epsilon$$

für alle $x, y \in X$.

(Erinnerung: f ist stetig, wenn es für jedes x und $\epsilon > 0$ ein $\delta > 0$ gibt mit $d(x, y) < \delta \Rightarrow d(f(x), f(y)) < \epsilon$ für alle $y \in X$.

Im Gegensatz zur gleichmäßigen Stetigkeit kann δ hier von x abhängen.)

Satz 1.59. *Sei $f: X \rightarrow Y$ eine stetige Abbildung zwischen metrischen Räumen. Ist X kompakt, so ist f gleichmäßig stetig.*

Beweis. Sei $\epsilon > 0$. Für $x \in X$ gibt es $\delta_x > 0$ mit

$$f(U_{\delta_x}(x)) \subset U_{\epsilon/2}(f(x)).$$

Es gibt x_1, \dots, x_n , so daß X von $U_{\delta_{x_1}/2}(x_1), \dots, U_{\delta_{x_n}/2}(x_n)$ überdeckt wird. Für $\delta = \min\{\delta_{x_1}/2, \dots, \delta_{x_n}/2\}$ gilt dann

$$d(y, y') < \delta \quad \Rightarrow \quad d(f(y), f(y')) < \epsilon$$

für alle $y, y' \in X$, denn jedes y liegt in einem $U_{\delta_{x_i}/2}(x_i)$ und mit der Dreiecksungleichung ist mit $d(y, y') < \delta$ auch $y' \in U_{\delta_{x_i}}(x_i)$ (BILD), so daß

$$d(f(y), f(y')) \leq d(f(y), f(x_i)) + d(f(x_i), f(y')) \leq \epsilon/2 + \epsilon/2 = \epsilon.$$

□

(Den Beweis kann man etwas vereinfachen, indem man Lemma 1.53 über Lebesgue-Zahlen benutzt.)

Zwischenbetrachtung

Eine Abbildung $f: X \rightarrow Y$ ist genau dann stetig, wenn

- O offen $\Rightarrow f^{-1}(O)$ ist offen,

bzw. wenn

- A abgeschlossen $\Rightarrow f^{-1}(A)$ ist abgeschlossen.

Für eine stetige Abbildung $f: X \rightarrow Y$ gilt:

- X zusammenhängend $\Rightarrow f(X)$ zusammenhängend und
- X kompakt $\Rightarrow f(X)$ kompakt.

(Das sind die Sätze 1.40 bzw. 1.56, die man als Verallgemeinerungen des Zwischenwertsatzes bzw. des Satzes vom Maximum und Minimum ansehen kann.)

Für einen metrischen Raum X gilt

- X ist zusammenhängend \Leftrightarrow der Zwischenwertsatz gilt für alle stetigen $f: X \rightarrow \mathbb{R}$ und
- X ist kompakt \Leftrightarrow jede stetige Abbildung $f: X \rightarrow \mathbb{R}$ nimmt ihr Minimum und Maximum an.

(Im Fall von zusammenhängend ist “ \Rightarrow ” genau Satz 1.37 und “ \Leftarrow ” gilt, da es auf einem nicht-zusammenhängenden Raum nicht-leere offene Mengen U, V gibt mit $U \cup V = X$ sowie $U \cap V = \emptyset$ und damit

$$f: X \rightarrow \mathbb{R} \quad \begin{cases} 0 & x \in U \\ 1 & x \in V \end{cases}$$

eine stetige Funktion ist, die keinen Wert zwischen 0 und 1 annimmt.

Im Fall von kompakt ist “ \Rightarrow ” genau Satz 1.57. Der Teil “ \Leftarrow ” ist z.B. Übungsaufgabe 85 in C.C. Pugh, *Real mathematical analysis*, Springer 2002.)

(Alle Aussagen bis auf “ \Leftarrow ” im Fall von Kompaktheit gelten auch für allgemeine topologische Räume.)

1.7 Normierte Vektorräume

Definition 1.60. Ein *normierter Vektorraum* $(V, \|\cdot\|)$ ist ein Vektorraum über $\mathbb{K} = \mathbb{R}$ oder \mathbb{C} und eine Abbildung $\|\cdot\|: V \rightarrow \mathbb{R}$, genannt *Norm*, mit den Eigenschaften

- $\|x\| \geq 0$ und $\|x\| = 0 \Leftrightarrow x = 0$,
- $\|\lambda x\| = |\lambda| \|x\|$, und
- $\|x + y\| \leq \|x\| + \|y\|$ (Dreiecksungleichung)

für alle $x, y \in V$ und $\lambda \in \mathbb{K}$.

Bemerkung 1.61. Ist $(V, \|\cdot\|)$ ein normierter Vektorraum, so definiert

$$d(x, y) = \|x - y\|$$

eine Metrik auf V . Wir werden im Folgenden normierte Vektorräume immer als metrische Räume mit dieser Abstandsfunktion auffassen. Man überlegt sich leicht, daß eine Metrik d auf einem Vektorraum V genau dann von einer Norm induziert wird, wenn

- Translationen, also Abbildungen der Form $x \mapsto x+t$ mit $t \in V$, den Abstand erhalten und wenn
- Streckungen, also Abbildungen der Form $x \mapsto \lambda x$ mit $\lambda \in \mathbb{K}$, den Abstand um den Faktor $|\lambda|$ strecken.

Definition 1.62. Ein *Skalarprodukt* $\langle \cdot, \cdot \rangle$ auf einem Vektorraum über $\mathbb{K} = \mathbb{R}$ (oder \mathbb{C}) ist eine Abbildung $\langle \cdot, \cdot \rangle: V \times V \rightarrow \mathbb{K}$ mit

- $\langle \cdot, \cdot \rangle$ ist bilinear (über \mathbb{C} *sesquilinear*, d.h. anti-linear im ersten Eingang und linear im zweiten Eingang),
- $\langle \cdot, \cdot \rangle$ ist symmetrisch (über \mathbb{C} *hermitsch*, d.h. $\langle x, y \rangle = \overline{\langle y, x \rangle}$ für alle $x, y \in V$), und
- $\langle \cdot, \cdot \rangle$ ist *positiv definit*, d.h. $\langle x, x \rangle \geq 0$ für alle $x \in V$ mit Gleichheit genau für $x = 0$.

Bemerkung 1.63. Ein Vektorraum V mit Skalarprodukt wird durch

$$\|x\| = \sqrt{\langle x, x \rangle}$$

zu einem normierten Vektorraum. Die ersten beiden Normeigenschaften sind offensichtlich erfüllt. Die Dreiecksungleichung folgt aus der Cauchy–Schwarzschen Ungleichung (siehe Satz 1.64):

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

(wobei Cauchy–Schwarz genau bei der Abschätzung “ \leq ” eingeht).

Satz 1.64. (*Cauchy–Schwarz Ungleichung*) In einem Vektorraum V mit Skalarprodukt gilt

$$|\langle x, y \rangle| \leq \|x\| \|y\|$$

für alle $x, y \in V$ mit Gleichheit genau für linear abhängige Vektoren.

Beweis. OBdA können wir annehmen, daß x nicht der Nullvektor ist und $\|x\| = 1$ gilt. Sind nun x und y linear abhängig, so ist $y = \lambda x$ und $|\langle x, y \rangle| = |\lambda| \|x\|^2 = \|\lambda x\| \|x\| = \|y\| \|x\|$. Sind x und y linear unabhängig, so liefert das Gram–Schmidt Verfahren eine Orthonormalbasis e_1, e_2 des erzeugten Unterraumes für die $x = e_1$ und $y = y_1 e_1 + y_2 e_2$ mit $y_2 \neq 0$ gilt. Damit ist

$$|\langle x, y \rangle| = |y_1| < \sqrt{|y_1|^2 + |y_2|^2} = \|y\| \|x\|.$$

□

Bemerkung 1.65. Eine Norm auf einem reellen Vektorraum wird genau dann von einem Skalarprodukt induziert, wenn die *Parallelogrammungleichung*

$$2(\|x\|^2 + \|y\|^2) = \|x + y\|^2 + \|x - y\|^2$$

für alle $x, y \in V$ gilt (BILD). Dann ist

$$\langle x, y \rangle = \frac{1}{4}(\|x + y\|^2 - \|x - y\|^2).$$

1.7.1 Normen auf \mathbb{R}^n und \mathbb{C}^n (und Einschub über Ungleichungen)

Auf \mathbb{R}^n bzw. \mathbb{C}^n induziert das Standard–Skalarprodukt

$$\langle x, y \rangle = \sum_{i=1}^n x_i y_i \quad \text{bzw.} \quad \langle x, y \rangle = \sum_{i=1}^n \bar{x}_i y_i$$

die Norm

$$\|x\| = \sqrt{\sum_{i=1}^n x_i^2} \quad \text{bzw.} \quad \|x\| = \sqrt{\sum_{i=1}^n |x_i|^2}$$

und damit die (in Beispiel 1.2 eingeführte) euklidische Abstandsfunktion

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

Es gibt noch viele andere Möglichkeiten, Normen auf \mathbb{R}^n bzw. \mathbb{C}^n zu definieren, zum Beispiel:

Definition 1.66. Für $x \in \mathbb{R}^n$ bzw. \mathbb{C}^n und $p \in [1, \infty[$ definieren wir die p -Norm

$$\|x\|_p = \sqrt[p]{|x_1|^p + \dots + |x_n|^p}.$$

Weiter ist die *Maximumsnorm* definiert durch

$$\|x\|_\infty = \max\{|x_i| \mid i = 1, \dots, n\}.$$

Die ersten beiden Normeigenschaften sind jeweils offensichtlich erfüllt, genauso wie die Dreiecksungleichung für $\|\cdot\|_1$ und $\|\cdot\|_\infty$. Die Dreiecksungleichung für $\|x\|_p$, $p \in]1, \infty[$ wird auch Minkowski Ungleichung genannt und im folgenden Einschub hergeleitet (siehe Satz 1.75).

(BILD Einheitsbälle in \mathbb{R}^2 bezüglich $\|\cdot\|_1$, $\|\cdot\|_2$ und $\|\cdot\|_\infty$.)

Bemerkung 1.67. Wir werden sehen (in Satz 1.94), daß alle Normen auf \mathbb{R}^n bzw. \mathbb{C}^n "äquivalent" sind, also die gleiche Topologie beschreiben und auch dieselben Cauchy-Folgen haben. (Das liegt daran, daß jeder offene Ball um 0 bezüglich einer Norm einen offenen Ball bezüglich jeder beliebigen anderen Norm enthält. Zum Beispiel gilt für $p \in [1, \infty[$

$$\|x\|_\infty \leq \|x\|_p \leq \sqrt[p]{n} \|x\|_\infty.$$

(BILD für $p=2$)

Einschub über Ungleichungen

Satz 1.68. Ist $f: I \rightarrow \mathbb{R}$ eine zweimal differenzierbare Funktion mit $f'' \geq 0$, so gilt für alle $x, y \in I$ und $\lambda \in [0, 1]$

$$f((1 - \lambda)x + \lambda y) \leq (1 - \lambda)f(x) + \lambda f(y). \quad (*)$$

Beweis. OBdA nehmen wir an, daß $x < y$ und $\lambda \in]0, 1[$, womit $z = (1 - \lambda)x + \lambda y$ ungleich x und y ist. Mit dem Mittelwertsatz gibt es $\tilde{x} \in]x, z[$ und $\tilde{y} \in]z, y[$ mit

$$\frac{f(z) - f(x)}{z - x} = f'(\tilde{x}) \leq f'(\tilde{y}) = \frac{f(y) - f(z)}{y - z},$$

wobei die Ungleichung \leq gilt, da f' wegen $f'' \geq 0$ monoton wachsend ist. Setzt man in diese Ungleichung die Definition von z ein, so erhält man

$$\frac{f((1 - \lambda)x + \lambda y) - f(x)}{\lambda(y - x)} \leq \frac{f(y) - f((1 - \lambda)x + \lambda y)}{(1 - \lambda)(y - x)},$$

und somit (*). □

Bemerkung 1.69. • Eine Funktion, für die (*) gilt, bezeichnet man als *konvex*.

- Konvexität bedeutet genau, daß der Graph der Funktion “unter” jeder Sekante liegt (BILD), denn für $\lambda \in [0, 1]$ durchläuft $(1-\lambda)x + \lambda y = x + \lambda(y-x)$ genau das Intervall zwischen x und y und

$$\begin{pmatrix} x \\ f(x) \end{pmatrix} + \lambda \begin{pmatrix} y-x \\ f(y)-f(x) \end{pmatrix} = \begin{pmatrix} (1-\lambda)x + \lambda y \\ (1-\lambda)f(x) + \lambda f(y) \end{pmatrix}$$

parametrisiert das entsprechende Sekantenstück.

- Ist f konvex, so gilt (per Induktion) für alle $x_1, \dots, x_n \in I$ und $\lambda_1, \dots, \lambda_n \geq 0$ mit $\lambda_1 + \dots + \lambda_n = 1$ die *Jensensche Ungleichung*

$$f(\lambda_1 x_1 + \dots + \lambda_n x_n) \leq \lambda_1 f(x_1) + \dots + \lambda_n f(x_n).$$

Beispiel 1.70. Die Funktion $f(x) = -\log(x)$ auf $\mathbb{R}_{>0}$ ist konvex, da

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

Damit gilt für alle $x_1, \dots, x_n \in \mathbb{R}_{>0}$ und $\lambda_1, \dots, \lambda_n \geq 0$ mit $\lambda_1 + \dots + \lambda_n = 1$

$$-\log(\lambda_1 x_1 + \dots + \lambda_n x_n) \leq -\lambda_1 \log(x_1) - \dots - \lambda_n \log(x_n)$$

und somit

$$\lambda_1 \log(x_1) + \dots + \lambda_n \log(x_n) \leq \log(\lambda_1 x_1 + \dots + \lambda_n x_n).$$

Nach Anwenden der (streng monotonen) Exponentialfunktion erhält man

$$x_1^{\lambda_1} \cdots x_n^{\lambda_n} \leq \lambda_1 x_1 + \dots + \lambda_n x_n$$

(wobei wir benutzt haben, daß nach Definition $x^\lambda = \exp(\lambda \log(x))$).

Satz 1.71. (*Ungleichung vom gewichteten arithmetischen und geometrischen Mittel*) Sind $x_1, \dots, x_n \geq 0$, $w_1, \dots, w_n > 0$ und $w = w_1 + \dots + w_n$, so gilt

$$\sqrt[w]{x_1^{w_1} \cdots x_n^{w_n}} \leq \frac{w_1 x_1 + \dots + w_n x_n}{w}.$$

Beweis. Falls ein $x_i = 0$, so ist $x_i^{w_i} = 0$ und die Ungleichung gilt. Sind alle $x_i > 0$, so folgt die Ungleichung mit $\lambda_i = w_i/w$ aus der vorhergehenden Ungleichung. \square

Beispiel 1.72. Im Spezialfall $n = 2$ mit Gewichten $w_1 = w_2 = 1$ erhält man, daß für alle $a, b \geq 0$ gilt

$$\sqrt{a \cdot b} \leq \frac{a+b}{2}.$$

(Diesen Fall kann man leicht direkt mittels der binomischen Formeln verifizieren.)

Satz 1.73. (*Youngsche Ungleichung*) Für $a, b \geq 0$ und $p, q > 1$ mit $\frac{1}{p} + \frac{1}{q} = 1$ gilt

$$a \cdot b \leq \frac{a^p}{p} + \frac{b^q}{q}.$$

Beweis. OBdA seien $a, b > 0$ (im Fall $a = 0$ oder $b = 0$ gilt die Ungleichung klarerweise). Dann folgt die Ungleichung direkt mit $n = 2$, $x_1 = a^p$, $x_2 = b^q$ sowie $\lambda_1 = 1/p$ und $\lambda_2 = 1/q$ aus der letzten Ungleichung von Beispiel 1.70. \square

Satz 1.74. (*Höldersche Ungleichung*) Für $x, y \in \mathbb{R}^n$ oder \mathbb{C}^n und $p, q \in [1, \infty]$ mit $\frac{1}{p} + \frac{1}{q} = 1$ gilt

$$\sum_{i=1}^n |x_i y_i| \leq \|x\|_p \|y\|_q.$$

(Für $p = q = 2$ liefert das die Cauchy–Schwarz Ungleichung, also Satz 1.64.)

Beweis. Die Fälle $p = 1$ und $q = \infty$ sowie $p = \infty$ und $q = 1$ sind offensichtlich. Nehmen wir daher $p, q > 1$ an. Weiter seien oBdA $x, y \neq 0$. Wir setzen

$$a_i = \frac{|x_i|}{\|x\|_p} \quad \text{und} \quad b_i = \frac{|y_i|}{\|y\|_q}.$$

Dann gilt $\sum_{i=1}^n a_i^p = \sum_{i=1}^n b_i^q = 1$ und n -fache Anwendung der Youngschen Ungleichung liefert

$$\frac{\sum_{i=1}^n |x_i y_i|}{\|x\|_p \|y\|_q} = \sum_{i=1}^n a_i b_i \leq \sum_{i=1}^n \frac{a_i^p}{p} + \frac{b_i^q}{q} = \frac{1}{p} + \frac{1}{q} = 1.$$

\square

Satz 1.75. (*Minkowski Ungleichung*) Für $x, y \in \mathbb{R}^n$ oder \mathbb{C}^n und $p \in [1, \infty[$ gilt

$$\sqrt[p]{\sum_{i=1}^n |x_i + y_i|^p} \leq \sqrt[p]{\sum_{i=1}^n |x_i|^p} + \sqrt[p]{\sum_{i=1}^n |y_i|^p}$$

Beweis. Der Fall $p = 1$ ist offensichtlich. Sei $p > 1$ und $q \in]1, \infty[$ die eindeutige Lösung von $\frac{1}{p} + \frac{1}{q} = 1$. Dann ist

$$\begin{aligned} \sum_{i=1}^n |x_i + y_i|^p &= \sum_{i=1}^n |x_i + y_i| |x_i + y_i|^{p-1} \leq \\ &\leq \sum_{i=1}^n |x_i| |x_i + y_i|^{p-1} + \sum_{i=1}^n |y_i| |x_i + y_i|^{p-1} \leq \\ &\stackrel{\text{Hölder}}{\leq} (\|x\|_p + \|y\|_p) \left(\sum_{i=1}^n |x_i + y_i|^{q(p-1)} \right)^{1/q}. \end{aligned}$$

Wegen $q(p-1) = p$ zeigt das die Ungleichung. \square

1.7.2 Vollständigkeit

Definition 1.76. Ein *Banachraum* ist ein vollständiger normierter Vektorraum. Ein *Hilbertraum* ist ein Banachraum, dessen Norm von einem Skalarprodukt induziert wird.

Beispiel 1.77. • Jeder endlichdimensionale euklidische Vektorraum ist ein Hilbertraum (weil isometrisch zu \mathbb{R}^n mit Standardskalarprodukt, der nach Beispiel 1.24 vollständig ist). Analog ist jeder endlichdimensionale unitäre Vektorraum ein komplexer Hilbertraum (weil bezüglich der Standardmetriken \mathbb{C}^n isometrisch zu \mathbb{R}^{2n} ist).

- \mathbb{R}^n bzw. \mathbb{C}^n wird mit jeder der Normen $\|\cdot\|_p$, $p \in [1, \infty]$ zu einem Banachraum, denn wie im Fall der euklidischen Norm gilt
 - eine Folge konvergiert genau dann bzgl. $\|\cdot\|_p$, wenn sie komponentenweise konvergiert, und
 - eine Folge ist genau dann eine Cauchy-Folge bzgl. $\|\cdot\|_p$, wenn die Komponentenfolgen Cauchy-Folgen sind.

(Beide Tatsachen sieht man, völlig analog zu Satz 1.20 und Beispiel 1.24, leicht direkt. Sie folgen auch aus der noch zu beweisenden und schon in Bemerkung 1.67 angekündigten Tatsache, daß alle Normen auf endlichdimensionalen Vektorräumen äquivalent sind, siehe Satz 1.94.)

1.7.3 Beispiele für unendlichdimensionale normierte Vektorräume

Definition 1.78. Sei $\mathbb{K} = \mathbb{R}$ oder \mathbb{C} .

- Für eine beliebige Menge X betrachten wir den Vektorraum

$$B(X, \mathbb{K}) = \{f: X \rightarrow \mathbb{K} \mid f \text{ beschränkt}\}$$

der beschränkten Funktionen auf X mit der sogenannten *Supremumsnorm*

$$\|f\|_\infty = \sup_{x \in X} \{|f(x)|\}.$$

- Für ein metrischer Raum X betrachten wir den Untervektorraum

$$BC(X, \mathbb{K}) = \{f: X \rightarrow \mathbb{K} \mid f \text{ beschränkt und stetig}\}.$$

- Für einen kompakten metrischer Raum X ist

$$C(X, \mathbb{K}) = \{f: X \rightarrow \mathbb{K} \mid f \text{ stetig}\} = BC(X, \mathbb{K}).$$

Beispiel 1.79. Ist $X = \{1, \dots, n\}$, so kann man $B(X, \mathbb{K})$ mit \mathbb{K}^n identifizieren. Unter dieser Identifizierung wird die Supremumsnorm zur Maximumsnorm (aus Definition 1.66).

Definition 1.80. Konvergenz bezüglich der Supremumsnorm bezeichnet man auch als *gleichmäßige Konvergenz*.

Bemerkung 1.81. Gleichmäßige Konvergenz

$$f_n \xrightarrow{\|\cdot\|_\infty} f$$

einer Funktionenfolge bedeutet anschaulich, daß in jedem “ ϵ -Schlauch” (BILD) um f fast alle f_n liegen. Gleichmäßige Konvergenz impliziert punktweise Konvergenz $f_n(x) \rightarrow f(x)$ für alle $x \in X$.

Satz 1.82. Sei X eine Menge.

- a) $(B(X, \mathbb{K}), \|\cdot\|_\infty)$ ist ein Banachraum.
- b) Ist X ein metrischer Raum, so ist $BC(X, \mathbb{K})$ bezüglich $\|\cdot\|_\infty$ ein abgeschlossener Unterraum von $B(X, \mathbb{K})$ und damit selbst ein Banachraum.
- c) Für einen kompakten Raum X ist $C(X, \mathbb{K})$ mit $\|\cdot\|_\infty$ ein Banachraum.

Beweis. a) Ist $(f_n)_n \in \mathbb{N}$ eine Cauchy-Folge in $B(X, \mathbb{K})$, so ist $(f_n(x))_n \in \mathbb{N}$ für alle x eine Cauchy-Folge, denn

$$|f_n(x) - f_m(x)| \leq \|f_n - f_m\|_\infty.$$

Wir definieren $f(x) = \lim_{n \rightarrow \infty} f_n(x)$. Zu jedem $\epsilon > 0$ gibt es N mit $\|f_n - f_m\|_\infty < \epsilon/2$ für $n, m \geq N$. Zu jedem x gibt es $m \geq N$ mit $|f(x) - f_m(x)| < \epsilon/2$, so daß

$$|f(x) - f_n(x)| \leq |f(x) - f_m(x)| + |f_m(x) - f_n(x)| < \epsilon/2 + \epsilon/2 = \epsilon$$

für alle $n \geq N$. Damit ist $f \in B(X, \mathbb{K})$ und $f_n \rightarrow f$ bezüglich $\|\cdot\|_\infty$.

b) Es reicht zu zeigen, daß für eine in $B(X, \mathbb{K})$ konvergente Folge $(f_n)_{n \in \mathbb{N}}$ aus der Stetigkeit der f_n die Stetigkeit des Grenzwertes f folgt (Satz 1.21). Sei also $x \in X$ und $\epsilon > 0$. Dann gibt es n mit $\|f - f_n\|_\infty < \epsilon/3$ und $\delta > 0$ mit $|f_n(y) - f_n(x)| \leq \epsilon/3$ für $y \in U_\delta(x)$. Insbesondere ist damit für $y \in U_\delta(x)$

$$|f(y) - f(x)| \leq |f(y) - f_n(y)| + |f_n(y) - f_n(x)| + |f_n(x) - f(x)| \leq \epsilon.$$

Als abgeschlossene Teilmenge eines vollständigen Raumes ist $BC(X, \mathbb{K})$ insbesondere selbst vollständig (Satz 1.25). \square

Teil b) des Satzes beinhaltet die folgende wichtige Aussage (die man für Funktionen $f: [a, b] \rightarrow \mathbb{R}$ auch in Analysis I Büchern findet):

Korollar 1.83. Ein gleichmäßiger Limes stetiger Funktionen ist stetig.

Bemerkung 1.84. Die Resultate dieses Abschnitts gelten ohne Änderung auch, wenn man \mathbb{K} in $B(X, \mathbb{K})$ durch einen beliebigen Banachraum ersetzt.

1.7.4 Stetigkeit linearer Operatoren

Definition 1.85. Eine Abbildung $f: X \rightarrow Y$ zwischen metrischen Räumen heißt *Lipschitz stetig*, wenn es $C > 0$ gibt mit

$$d(f(x), f(y)) \leq C \cdot d(x, y) \quad \text{für alle } x, y \in X.$$

Lipschitz stetige Abbildungen sind gleichmäßig stetig.

Beispiel 1.86. Die Norm $\|\cdot\|: V \rightarrow \mathbb{R}$ eines normierten Vektorraumes $(V, \|\cdot\|)$ ist Lipschitz stetig, denn

$$\left| \|u\| - \|v\| \right| \leq \|u - v\| \quad \text{für alle } u, v \in V.$$

Beweis. Mit der Dreiecksungleichung gilt für alle $u, v \in V$

$$\|u\| \leq \|u - v\| + \|v\| \quad \text{und} \quad \|v\| \leq \|v - u\| + \|u\|.$$

□

Definition 1.87. Ein linearer Operator $F: V \rightarrow W$ zwischen normierten Räumen heißt *beschränkt*, wenn es C gibt mit

$$\|F(v)\| \leq C\|v\| \quad \text{für alle } v \in V.$$

Bemerkung 1.88. Ein linearer Operator ist genau dann beschränkt, wenn das Bild der (abgeschlossenen) Einheitskugel in V eine beschränkte Menge in W ist, denn aufgrund der Linearität von F und der Homogenität von $\|\cdot\|$ gilt

$$\|F(v)\| \leq C\|v\| \quad \forall v \in V \quad \iff \quad \|F(v)\| \leq C \quad \forall v \in V, \|v\| \leq 1.$$

(“ \Rightarrow ” ist hier offensichtlich und “ \Leftarrow ” gilt, weil für $v \neq 0$ auch $\|F(v/\|v\|)\| \leq C$.)

Es sei angemerkt, daß das Bild $F(V)$ des ganzen Vektorraumes V unter $F \neq 0$ nie eine beschränkte Menge ist! (Ist $F(v) \neq 0$, so kann nämlich $F(\lambda v) = \lambda F(v)$ mit $\lambda \in \mathbb{K}$ beliebig große Norm haben.)

Satz 1.89. Eine lineare Abbildung $F: V \rightarrow W$ zwischen normierten Vektorräumen ist genau dann stetig, wenn sie beschränkt ist.

Beweis. “ \Rightarrow ” Ist F stetig, so ist F insbesondere stetig in 0 und es gibt δ mit

$$\|v\| \leq \delta \quad \Rightarrow \quad \|F(v)\| \leq 1.$$

Damit gilt für alle $v \neq 0$

$$\left\| F\left(\delta \frac{v}{\|v\|}\right) \right\| \leq 1 \quad \text{und mit } C = \frac{1}{\delta} \text{ gilt} \quad \|F(v)\| \leq C\|v\|$$

(wobei wir im letzten Schritt die Linearität von F sowie die Homogenität von $\|\cdot\|$ ausgenutzt haben).

“ \Leftarrow ” Ist F beschränkt, so gilt für alle $u, v \in V$

$$\|F(u) - F(v)\| = \|F(u - v)\| \leq C\|u - v\|,$$

so daß F Lipschitz stetig und somit stetig ist. □

Bemerkung 1.90. Für eine lineare Abbildung F gilt also: F stetig $\Rightarrow F$ stetig in 0 $\Rightarrow F$ beschränkt $\Rightarrow F$ Lipschitz stetig $\Rightarrow F$ gleichmäßig stetig $\Rightarrow F$ stetig.

Satz 1.91. Ist V ein endlichdimensionaler normierter Vektorraum, so ist jede lineare Abbildung $F: V \rightarrow W$ in einen normierten Vektorraum W stetig.

Beweis. a) Sei zuerst $V = \mathbb{K}^n$ mit euklidischer Norm $\|x\|_2 = \sqrt{|x_1|^2 + \dots + |x_n|^2}$. Dann gilt für jede \mathbb{K} -lineare Abbildung $F: V \rightarrow W$

$$\|F(x)\| = \left\| F\left(\sum_{i=1}^n x_i e_i\right) \right\| \leq \sum_{i=1}^n \|F(e_i)\| |x_i| \leq \underbrace{\sum_{i=1}^n \|F(e_i)\|}_{C:=} \|x\|_\infty \leq C \|x\|_2$$

(wobei e_1, \dots, e_n die Standardbasis bezeichnet und wir $\|\cdot\|_\infty \leq \|\cdot\|_2$ benutzt haben).

b) Ist F nun eine bijektive lineare Abbildung auf $V = \mathbb{K}^n$ mit $\|\cdot\|_2$, so ist auch F^{-1} stetig: die Einheitssphäre (in \mathbb{R}^n bzw. $\mathbb{C}^n = \mathbb{R}^{2n}$)

$$S = \{x \in \mathbb{K}^n \mid |x_1|^2 + \dots + |x_n|^2 = 1\} = \{x \in V \mid \|x\|_2 = 1\}$$

ist dann kompakt, da abgeschlossen und beschränkt (Satz 1.49 von Heine–Borel). Also nimmt die, wegen a) stetige, Abbildung $\|F\|$ auf S ihr Minimum an (Satz 1.57). Da F bijektiv ist, muß dieses größer als 0 sein und es gibt $C' > 0$ mit

$$C' \leq \|F(x)\| \quad \text{für alle } x \in V, \|x\| = 1.$$

Für $x \neq 0$ gilt somit

$$C' \leq \left\| F\left(\frac{x}{\|x\|}\right) \right\| \quad \text{und damit auch} \quad C' \|x\| \leq \|F(x)\|.$$

Für alle $y \in W$ gilt also, indem man $x = F^{-1}(y)$ setzt, daß

$$\|F^{-1}(y)\| \leq 1/C' \|y\|.$$

c) Sei nun V ein beliebiger endlichdimensionaler normierter Vektorraum und $F: V \rightarrow W$ eine lineare Abbildung. Wir wählen einen Isomorphismus $\Phi: \mathbb{K}^n \rightarrow V$. Stattdessen wir wie oben \mathbb{K}^n mit der euklidischen Norm aus, so ist wegen a) die Abbildung $\tilde{F} = F \circ \Phi$ stetig. Da wegen b) auch Φ^{-1} stetig ist, muß insbesondere auch $F = \tilde{F} \circ \Phi^{-1}$ stetig sein. (DIAGRAMM) \square

Definition 1.92. Zwei Normen $\|\cdot\|$ und $\|\cdot\|'$ auf einem Vektorraum V heißen äquivalent, wenn es $C, C' > 0$ gibt mit

$$C \|x\| \leq \|x\|' \leq C' \|x\| \quad \text{für alle } x \in V.$$

Bemerkung 1.93. Man überzeugt sich leicht, daß äquivalente Normen (wie schon in Bemerkung 1.67 erwähnt) dieselbe Topologie erzeugen und dieselben Cauchy-Folgen haben.

Satz 1.94. *Je zwei Normen auf einem endlichdimensionalen Vektorraum sind äquivalent. Insbesondere ist jeder endlichdimensionale normierte Vektorraum ein Banachraum.*

Beweis. Wegen Satz 1.91 ist die Abbildung $id: (V, \|\cdot\|) \rightarrow (V, \|\cdot\|')$ stetig, so daß es C' gibt mit

$$\|x\|' \leq C'\|x\|$$

für alle $x \in V$. Falls $V \neq \{0\}$ ist $C > 0$. Da auch $id: (V, \|\cdot\|') \rightarrow (V, \|\cdot\|)$ stetig ist, gibt es $C > 0$ mit

$$\|x\| \leq \frac{1}{C}\|x\|'$$

für alle $x \in V$ und die beiden Normen sind äquivalent.

Die Aussage über Vollständigkeit gilt, da jeder endlichdimensionale Vektorraum isomorph zu \mathbb{R}^n bzw. $\mathbb{C}^n = \mathbb{R}^{2n}$ ist (welche bezüglich der euklidischen Norm und der p -Normen vollständig sind, siehe Beispiel 1.24 und Beispiel 1.77). \square

Definition 1.95. Für einen stetigen linearen Operator $F: V \rightarrow W$ zwischen normierten Vektorräumen V und W ist die *Operatornorm* definiert als

$$\|F\| = \sup_{v \neq 0} \frac{\|F(v)\|}{\|v\|} = \sup_{v, \|v\|=1} \|F(v)\|.$$

Bemerkung 1.96. • $\|F\|$ ist wohldefiniert wegen Satz 1.89.

- $\|F\|$ ist die kleinstmögliche Schranke C , so daß für alle $v \in V$ gilt

$$\|F(v)\| \leq C\|v\|.$$

- Die Operatornorm ist eine Norm auf dem Vektorraum

$$L(V, W) = \{F: V \rightarrow W \mid F \text{ ist stetig und linear}\}.$$

Satz 1.97. *Ist W ein Banachraum, so ist auch $L(V, W)$ ein Banachraum.*

Beweis. Sei $(A_n)_{n \in \mathbb{N}}$ eine Cauchy-Folge in $L(V, W)$. Für $v \in V$ ist dann $A_n(v)$ eine Cauchy-Folge in W , denn

$$\|A_n(v) - A_m(v)\| \leq \|A_n - A_m\| \|v\|.$$

Setzt man $A(v) = \lim_{n \rightarrow \infty} A_n(v)$, so ist $A: V \rightarrow W$ linear.

Bleibt also zu zeigen, daß A beschränkt ist und A_n bezüglich der Operatornorm gegen A konvergiert. Für $\epsilon > 0$ gibt es N , so daß für alle $n, m \geq N$ gilt $\|A_n - A_m\| < \epsilon$ und somit

$$\|A_n(v) - A_m(v)\| \leq \|A_n - A_m\| \|v\| < \epsilon \|v\|.$$

Geht man zum Grenzwert $m \rightarrow \infty$ über, so erhält man

$$\|A_n(v) - A(v)\| \leq \epsilon \|v\| \quad (*)$$

für alle $n \geq N$. Wegen der Dreiecksungleichung gilt damit

$$\|A(v)\| \leq \|A_n(v)\| + \|A(v) - A_n(v)\| \leq (\|A_n\| + \epsilon) \|v\|,$$

so daß A beschränkt ist. Wegen (*) ist $\|A_n - A\| \leq \epsilon$ für alle $n \geq N$. □

Satz 1.98. *Ist $D \subset V$ ein dichter Untervektorraum eines normierten Vektorraumes V , so hat jede stetige lineare Abbildung $F: D \rightarrow W$ in einen Banachraum W eine eindeutige stetige lineare Fortsetzung $\tilde{F}: V \rightarrow W$ mit $\tilde{F}|_D = F$.*

Beweis. Sei $v \in V$ und $(v_n)_{n \in \mathbb{N}}$ eine Folge in D mit $v_n \rightarrow v$. Wegen

$$\|F(v_n) - F(v_m)\| \leq \|F\| \|v_n - v_m\|$$

ist $(F(v_n))_{n \in \mathbb{N}}$ eine Cauchy-Folge in W . Wir definieren

$$\tilde{F}(v) = \lim_{n \rightarrow \infty} F(v_n).$$

(Zur Übung zeige man, daß $\tilde{F}(v)$ unabhängig von der Wahl der Folge und $\tilde{F}: V \rightarrow W$ linear ist.) Da $\|\cdot\|$ stetig ist, folgt aus $\|F(v_n)\| \leq \|F\| \|v_n\|$ auch

$$\|\tilde{F}(v)\| \leq \|F\| \|v\| \quad (*)$$

für alle $v \in V$, womit \tilde{F} stetig ist. \square

Bemerkung 1.99. Die Gleichung $(*)$ zeigt

$$\|\tilde{F}\| = \sup_{0 \neq v \in V} \frac{\|\tilde{F}(v)\|}{\|v\|} \leq \|F\|.$$

Andererseits gilt

$$\|\tilde{F}\| = \sup_{0 \neq v \in V} \frac{\|\tilde{F}(v)\|}{\|v\|} \geq \sup_{0 \neq v \in D} \frac{\|F(v)\|}{\|v\|} = \|F\|$$

und somit

$$\|\tilde{F}\| = \|F\|.$$

Beispiel 1.100. (Regelintegral) Sei D der Vektorraum der reellwertigen Treppenfunktionen auf einem kompakten Intervall $[a, b]$ und

$$F: D \rightarrow \mathbb{R} \quad f \mapsto \int_a^b f(x) dx$$

der Operator, der jeder Treppenfunktion ihr Integral zuordnet. Dann ist F stetig bezüglich der Supremumsnorm $\|\cdot\|_\infty$, denn

$$|F(f)| \leq |b - a| \|f\|_\infty.$$

Der Abschluß $V = \bar{D}$ des Raumes der Treppenfunktionen, aufgefaßt als Teilmenge des Banachraumes $(B([a, b], \mathbb{R}), \|\cdot\|_\infty)$ der beschränkten Funktionen auf $[a, b]$, ist der Vektorraum der Regelfunktionen auf $[a, b]$. Nach Satz 1.98 kann man F eindeutig fortsetzen zu einem stetigen linearen Operator $\tilde{F}: V \rightarrow \mathbb{R}$. Diese Fortsetzung bezeichnet man als Regelintegral.

Satz 1.101. *Ein normierter Vektorraum $(V, \|\cdot\|)$ ist genau dann endlichdimensional, wenn die abgeschlossene Einheitskugel $\{v \in V \mid \|v\| \leq 1\}$ kompakt ist.*

Beweis. “ \Rightarrow ” Ist V endlichdimensional, so gibt es einen Isomorphismus

$$\Phi: \mathbb{K}^n \rightarrow V.$$

Nach Satz 1.91 sind Φ und Φ^{-1} stetig. Die Einheitskugel $\{v \in V \mid \|v\| \leq 1\}$ in V ist abgeschlossen und beschränkt, so daß wegen der Stetigkeit von Φ und Φ^{-1} auch

$$\Phi^{-1}(\{v \in V \mid \|v\| \leq 1\})$$

abgeschlossen und beschränkt in \mathbb{K}^n ist. Mit dem Satz von Heine–Borel (Satz 1.49) ist also $\Phi^{-1}(\{v \in V \mid \|v\| \leq 1\})$ kompakt. Wegen der Stetigkeit von Φ ist damit auch $\{v \in V \mid \|v\| \leq 1\}$ kompakt (Satz 1.56).

“ \Leftarrow ” Ist die Einheitskugel $\{v \in V \mid \|v\| \leq 1\}$ kompakt, so hat sie eine endliche Überdeckung durch offene Kugeln

$$B_{1/2}(w_1), \dots, B_{1/2}(w_m)$$

von Radius $1/2$. Wir definieren

$$W = \text{Span}\{w_1, \dots, w_m\}.$$

Für $v \in \{v \in V \mid \|v\| \leq 1\}$ gibt es $v_1 \in W$ mit $\|v - v_1\| < 1/2$. Es gilt: gibt es $v_i \in W$ mit

$$\|v - v_i\| < 1/2^i,$$

so gibt es $v_{i+1} \in W$ mit

$$\|v - v_{i+1}\| < 1/2^{i+1}$$

(denn dann ist $\|2^i(v - v_i)\| < 1$, womit $\|2^i(v - v_i) - w_j\| < 1/2$ für ein $j = 1, \dots, m$ gilt, so daß man $v_{i+1} = v_i + \frac{1}{2^i}w_j$ setzen kann).

Damit haben wir eine Folge $(v_i)_{i \in \mathbb{N}}$ in W mit $v_i \rightarrow v$. Da W endlichdimensional ist, ist W ein abgeschlossener Unterraum (denn wegen Satz 1.94 ist W mit der Einschränkung der Norm auf V vollständig) und es gilt $v \in W$. Da v ein beliebiger Vektor in V mit $\|v\| \leq 1$ war, gilt $V = W$ und V ist endlichdimensional. \square

2 Differentialrechnung

In diesem Kapitel entwickeln wir die Differentialrechnung in endlichdimensionalen reellen Vektorräumen. Wir diskutieren dabei sowohl eine invariante Formulierung für Abbildungen zwischen abstrakten reellen Vektorräumen, als auch eine Formulierung in Koordinaten. In beiden Fällen verzichten wir in der Regel auf die konkrete Wahl von Normen, da nach Satz 1.94 im Endlichdimensionalen alle Normen äquivalent sind. (Die meisten koordinatenunabhängigen Resultate lassen sich mit geringen Änderungen auf den Fall allgemeiner reeller oder komplexer Banachräumen übertragen.)

2.1 Kurven

Für Abbildungen $f:]a, b[\subset \mathbb{R} \rightarrow W$ von einem Intervalle $]a, b[$ in einen endlichdimensionalen Vektorraum W kann man Differenzierbarkeit genauso definieren wie in Analysis I.

Definition 2.1. Eine Abbildung $f:]a, b[\subset \mathbb{R} \rightarrow W$ von einem Intervall $]a, b[$ in einen endlichdimensionalen Vektorraum W heißt differenzierbar in $x_0 \in]a, b[$, wenn der Grenzwert

$$f'(x_0) = \lim_{x \rightarrow x_0} \frac{f(x) - f(x_0)}{x - x_0}$$

existiert.

Bemerkung 2.2. • In dieser Definition benutzen wir die “natürliche” Topologie auf W , die von einer (und damit jeder) Norm erzeugt wird (Satz 1.94). (Alternativ könnte man auch fordern, daß W mit einer Norm ausgestattet ist. Jedoch würde die Definition im Endlichdimensionalen nach Satz 1.94 nicht von der Wahl der Norm abhängen.)

- Wie in Analysis I ist $f'(x_0)$ (nach Definition) genau dann Grenzwert von $\frac{f(x)-f(x_0)}{x-x_0}$ für $x \rightarrow x_0$, wenn man die Funktion $x \mapsto \frac{f(x)-f(x_0)}{x-x_0}$ stetig nach x_0 fortsetzen kann, indem man sie dort $f'(x_0)$ setzt. Aufgrund des Folgenkriteriums für Stetigkeit (Satz 1.29) ist Differenzierbarkeit von f in x_0 äquivalent dazu, daß für jede Folge $(x_n)_{n \in \mathbb{N}}$ mit Grenzwert x_0 und $x_n \in]a, b[\setminus \{x_0\}$ gilt

$$f'(x_0) = \lim_{n \rightarrow \infty} \frac{f(x_n) - f(x_0)}{x_n - x_0}.$$

- Für $W = \mathbb{R}^n$ ist (nach Satz 1.20 oder Korollar 1.32) eine Funktion

$$f = \begin{pmatrix} f_1 \\ \vdots \\ f_n \end{pmatrix}$$

genau dann differenzierbar in x_0 , wenn die Komponentenfunktionen dies sind, und

$$f'(x_0) = \begin{pmatrix} f'_1(x_0) \\ \vdots \\ f'_n(x_0) \end{pmatrix}.$$

Im Hinblick auf die Formulierung des Differenzierbarkeitsbegriffs für Abbildungen auf höherdimensionalen Vektorräumen ist es nützlich, Differenzierbarkeit wie folgt als “lineare Approximierbarkeit” zu verstehen und

$$x \mapsto f(x_0) + f'(x_0)(x - x_0)$$

als “bestmögliche” (affin) lineare Approximation von f in x_0 zu charakterisieren.

Satz 2.3. *Eine Abbildung $f:]a, b[\subset \mathbb{R} \rightarrow W$ von einem Intervall $]a, b[$ in einen endlichdimensionalen Vektorraum W ist differenzierbar in $x_0 \in]a, b[$, wenn es $a \in W$ gibt, so daß*

$$f(x) = f(x_0) + a(x - x_0) + o(x - x_0)$$

mit Restterm $o(x - x_0)$, der für $x \rightarrow x_0$ so schnell fällt, daß

$$\lim_{x \rightarrow x_0} \frac{o(x - x_0)}{|x - x_0|} = 0.$$

Dann ist a gleich $f'(x_0)$.

(Der Beweis geht wie in der Analysis I und sei hier als Übungsaufgabe gelassen.)

Definition 2.4. Eine *parametrisierte Kurve* in einem endlichdimensionalen reellen Vektorraum W ist eine stetige Abbildung $\gamma: I \rightarrow W$ von einem Intervall $I \subset \mathbb{R}$ nach W . Eine *Kurve* ist eine Äquivalenzklasse von parametrisierten Kurven unter $\tilde{\gamma} \sim \gamma : \Leftrightarrow$ es gibt eine Umparametrisierung $\varphi: I' \rightarrow I$ mit $\tilde{\gamma} = \gamma \circ \varphi$ (wobei φ stetig und bijektiv ist und eine stetige Inverse hat).

Bezeichnung: Eine stetige bijektive Abbildung mit stetiger Inverser bezeichnet man als *Homöomorphismus*.

Bemerkung 2.5. Eine Kurve kann man visualisieren, indem man die Menge ihrer Bildpunkte

$$\gamma(I) = \{\gamma(t) \mid t \in I\} \subset W$$

zeichnet. Diese Menge ändert sich nicht, wenn man die Kurve umparametrisiert. Will man die Parametrisierung visualisieren, so kann man die t -Werte entlang von $\gamma(I)$ markieren. Physikalisch kann man eine parametrisierte Kurve $\gamma(t)$ als Trajektorie eines Teilchens im Raum W interpretieren.

Beispiel 2.6. • Eine *Gerade* in W kann man affin linear parametrisieren durch die “Punkt–Richtungs–Form”

$$\gamma(t) = p + tv, \quad t \in \mathbb{R}.$$

Dabei ist p der Punkt auf der Gerade, der dem Parameter $t = 0$ entspricht und $v \neq 0$ beschreibt die “Richtung” der Gerade (BILD). Schränkt man $\gamma(t)$ auf $[0, \infty[$ bzw. $[0, 1]$ ein, so erhält man einen *Strahl* bzw. eine *Strecke*.

- Jeden *Kreis* mit Mittelpunkt $p \in W$ und Radius $r > 0$ in einem euklidischen Vektorraum W kann man parametrisieren durch

$$\gamma(t) = p + r \cos(t)v + r \sin(t)w,$$

wobei v, w eine Orthonormalbasis eines zwei-dimensionalen Unterraums von W ist.

Den Kreis

$$\{(x, y) \in \mathbb{R}^2 \mid (x-p_1)^2 + (y-p_2)^2 = r^2\} = \left\{ \begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2 \mid d\left(\begin{pmatrix} x \\ y \end{pmatrix}, \begin{pmatrix} p_1 \\ p_2 \end{pmatrix}\right) = r \right\}$$

in $\mathbb{R}^2 = \mathbb{C}$ mit Standardskalarprodukt kann man parametrisieren durch

$$\gamma(t) = \begin{pmatrix} p_1 \\ p_2 \end{pmatrix} + r \begin{pmatrix} \cos(t) \\ \sin(t) \end{pmatrix} = p + re^{it}.$$

- Der Graph einer stetigen Abbildung $f: I \rightarrow \mathbb{R}$ ist das Bild der parametrisierten Kurve

$$\gamma: I \rightarrow \mathbb{R}^2 \quad x \mapsto \begin{pmatrix} x \\ f(x) \end{pmatrix}.$$

Bemerkung 2.7. In der Definition 2.1 haben wir Differenzierbarkeit nur für Abbildungen auf offenen Intervallen $]a, b[$ definiert. Wie in Analysis I kann man dieselbe Definition aber übertragen auf den Fall abgeschlossener (oder halboffener) Intervalle.

Der Grund für die Beschränkung auf den Fall offener Definitionsbereiche ist, daß wir uns (wie allgemein üblich) in der Definition von Differenzierbarkeit für Abbildungen auf höherdimensionalen Räumen (Definition 2.10) auf den Fall offener Definitionsbereiche beschränken. Den Fall allgemeiner Definitionsbereiche behandelt man dann so, daß man die Existenz einer differenzierbaren Fortsetzung auf eine offene Umgebung fordert.

Für Abbildungen $\gamma: I \rightarrow W$ mit abgeschlossenem (oder halboffenem) Definitionsbereich sieht man leicht (vgl. Blatt 6, Aufgabe 4), daß die Definition der Differenzierbarkeit in den Randpunkten wie in Analysis I äquivalent zur Definition über die differenzierbare Fortsetzung auf eine offene Umgebung ist.

Interpretation von Differenzierbarkeit einer parametrisierten Kurve: Ist $\gamma: I \rightarrow W$ eine stetige parametrisierte Kurve, $p = \gamma(t_0)$ ein Punkt im Bild von γ und $v \in W$ beliebig. Dann gilt

$$\gamma(t) = p + (t - t_0)v + r(t)$$

mit $r(t) \rightarrow 0$ für $t \rightarrow t_0$. Aber nur wenn γ in t_0 differenzierbar ist und $v = \gamma'(t_0)$ gilt, ist $\frac{r(t)}{t-t_0} \rightarrow 0$ für $t \rightarrow t_0$. (BILD)

- Ist $v = \gamma'(t_0) \neq 0$, so bezeichnet man die Gerade $p + tv$ als *Tangente* an die Kurve.
- Physikalisch kann man die Ableitung als *Geschwindigkeit* eines Teilchens ansehen, das sich zur Zeit t am Punkt $\gamma(t)$ befindet.

Bezeichnung: Zwei differenzierbare parametrisierte Kurven sieht man als *äquivalent* an, wenn es eine Umparametrisierung gibt, die differenzierbar ist und eine differenzierbare Inverse hat. Eine derartige Umparametrisierung nennt man einen *Diffeomorphismus*.

Beispiel 2.8. Die ersten beiden Beispiele für parametrisierte Kurven in Beispiel 2.6 sind differenzierbar. Die Parametrisierung

$$\gamma(x) = \begin{pmatrix} x \\ f(x) \end{pmatrix}$$

des Graphen von f ist genau dann differenzierbar, wenn $f: I \rightarrow \mathbb{R}$ differenzierbar ist. Dann gilt

$$\gamma'(x) = \begin{pmatrix} 1 \\ f'(x) \end{pmatrix}.$$

Bemerkung 2.9. • Man stellt sich das Bild einer Kurve gerne als “1–dimensionales Objekt” in W vor. Wenn man stetige Kurven betrachtet, ist diese Vorstellung so nicht korrekt, wie das Beispiel sogenannter *Peano–Kurven*, also stetiger Kurve $\gamma: [0, 1] \rightarrow \mathbb{R}^2$ mit Bild $[0, 1] \times [0, 1]$, zeigt (siehe z.B. K. Königsberger, Analysis 1).

- Betrachtet man parametrisierte Kurve $\gamma: I \rightarrow W$, die differenzierbar sind, kann so etwas nicht passieren. (Der *Satz von Sard* impliziert dann, daß das Bild der Kurve “klein”, nämlich eine sogenannte Nullmenge, ist, vgl. T. Bröcker, Analysis II.)

(Wir werden sehen, daß man in der Nähe von Punkten mit $\gamma'(t_0) \neq 0$ eine parametrisierte Kurve lokal so umparametrisieren kann, daß sie den Graph einer Funktion auf einem Intervall parametrisiert.)

2.2 Differenzierbarkeit in mehreren Veränderlichen

Definition 2.10. Seien V und W endlichdimensionale reelle Vektorräume und $f: U \subset V \rightarrow W$ eine Abbildung auf einer offenen Teilmenge $U \subset V$.

- f ist *differenzierbar* in $p \in U$, wenn es eine lineare Abbildung $A: V \rightarrow W$ gibt, so daß

$$f(x) = f(p) + A(x - p) + o(x - p)$$

mit Restterm $o(x - p)$, der für $x \rightarrow p$ schnell fällt im Sinne von

$$\lim_{x \rightarrow p} \frac{o(x - p)}{\|x - p\|} = 0.$$

A heißt dann die *Ableitung* von f in p und wird mit $Df|_p$ bezeichnet.

- f heißt *differenzierbar*, wenn f in jedem Punkt $p \in U$ differenzierbar ist.

Bemerkung 2.11. Die Definition ist unabhängig von der Wahl von Normen, da im Endlichdimensionalen alle Normen äquivalent sind (Satz 1.94). Da lineare Abbildungen zwischen endlichdimensionalen Vektorräumen stetig sind (Satz 1.91), impliziert Differenzierbarkeit in einem Punkt p insbesondere auch Stetigkeit in p .

Lemma 2.12. *Ist $f: U \subset V \rightarrow W$ differenzierbar in p , so gilt für alle $v \in V$*

$$Df|_p(v) = \lim_{t \rightarrow 0} \frac{f(p + tv) - f(p)}{t} = \frac{d}{dt}|_{t=0} f(p + tv).$$

Beweis. Nach Definition der Differenzierbarkeit (Definition 2.10) gilt

$$\frac{f(p + tv) - f(p)}{t} = Df|_p(v) + \frac{o(tv)}{t}.$$

Mit $\lim_{t \rightarrow 0} \frac{o(tv)}{t} = \lim_{t \rightarrow 0} \left(\frac{o(tv)}{\|tv\|} \frac{\|tv\|}{t} \right) = 0$ folgt die Behauptung. \square

Definition 2.13. Seien V und W endlichdimensionale reelle Vektorräume und $f: U \subset V \rightarrow W$ eine Abbildung auf einer offenen Teilmenge $U \subset V$.

- Die *Richtungsableitung* von f im Punkt $p \in U$ in Richtung $v \in V$ ist

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

- Ist $V = \mathbb{R}^n$, so bezeichnet man die Richtungsableitung $\partial_{e_i} f(p)$ in Richtung des i -ten Basisvektors e_i als die i -te *partielle Ableitung* von f in p und schreibt dafür auch $\frac{\partial f}{\partial x_i}(p)$. Für $p = (x_1, \dots, x_n)$ ist also

$$\frac{\partial f}{\partial x_i}(p) = \lim_{h \rightarrow 0} \frac{f(x_1, \dots, x_i + h, \dots, x_n) - f(x_1, \dots, x_i, \dots, x_n)}{h}.$$

Bemerkung 2.14. • Ist $f: U \subset V \rightarrow W$ in p differenzierbar, so besagt Lemma 2.12, daß alle Richtungsableitungen existieren und es gilt

$$\partial_v f(p) = Df|_p(v).$$

Insbesondere kann man die Ableitung von f bestimmen, indem man für alle v die Ableitung $\frac{d}{dt} \Big|_{t=0} f(p + tv)$ berechnet. Ist $V = \mathbb{R}^n$, so ist $Df|_p$ bestimmt durch die partiellen Ableitungsvektoren $\frac{\partial f}{\partial x_1}(p), \dots, \frac{\partial f}{\partial x_n}(p)$.

- Eine notwendige Bedingung für Differenzierbarkeit von f in einem Punkt p ist, daß alle Richtungsableitungen $\partial_v f(p)$ existieren. Weiter muß $\partial_v f(p)$ linear von v abhängen. Ist dies der Fall, muß zum Nachweis der Differenzierbarkeit jedoch immer noch überprüft werden, daß die lineare Abbildung, die man aus den Richtungsableitungen erhält, wirklich die Bedingung aus Definition 2.10 erfüllt. Das ist nicht automatisch der Fall, wie das folgende Beispiel zeigt.

Beispiel 2.15. (Existenz aller Richtungsableitungen impliziert nicht Differenzierbarkeit) Die Funktion $f: \mathbb{R}^2 \rightarrow \mathbb{R}$ definiert durch (BILD)

$$f(x, y) = \begin{cases} 1 & y = x^2, x \neq 0 \\ 0 & \text{sonst.} \end{cases}$$

aus Beispiel 1.34 hat im Punkt $p = (0, 0)$ Richtungsableitung $\partial_v f(p) = 0$ in alle Richtungen. Die Funktion ist aber in p nicht differenzierbar, da sie dort nicht einmal stetig ist.

Im Fall einer Abbildung $f: U \subset V \rightarrow W$ zwischen einer offenen Menge $U \subset V = \mathbb{R}^n$ und $W = \mathbb{R}^m$ kann man jede lineare Approximation $Df|_p: V \rightarrow W$ durch eine $m \times n$ -Matrix darstellen, denn eine Abbildung $A: \mathbb{R}^n \rightarrow \mathbb{R}^m$ ist genau dann linear, wenn sie von der Form

$$\begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} \mapsto \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \cdot \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} = \begin{pmatrix} a_{11}x_1 + \cdots + a_{1n}x_n \\ \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n \end{pmatrix}$$

ist für eine $m \times n$ -Matrix (a_{ij}) . Die Spalten der Matrix sind dabei genau die Bilder der Basisvektoren e_1, \dots, e_n unter der linearen Abbildung A .

Die Spalten der Matrix, die die Ableitung $Df|_p$ einer Abbildung $f: U \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$ im Punkt p repräsentiert, sind (nach Lemma 2.12 und Definition 2.13) genau die Richtungsableitungen in Richtung der Vektoren e_1, \dots, e_n der Standardbasis, also die partiellen Ableitungen $\frac{\partial f}{\partial x_1}(p), \dots, \frac{\partial f}{\partial x_n}(p)$. Die partiellen Ableitungen kann man berechnen durch Ableiten der Kurven die man erhält, wenn man f einschränkt $h \mapsto f(p + he_i)$ auf Geraden $\{p + he_i \mid h \in \mathbb{R}\}$ parallel zu den Koordinatenachsen.

Da die Ableitung einer Kurven in \mathbb{R}^m gegeben ist durch die Ableitung der Komponentenfunktionen (Bemerkung 2.2), sind die Zeilen der Ableitung genau gegeben durch die Ableitungen der Komponentenfunktionen. Damit ist die Ableitung von

$$f = \begin{pmatrix} f_1 \\ \vdots \\ f_m \end{pmatrix} \quad \text{von der Form} \quad Df|_p = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(p) & \cdots & \frac{\partial f_1}{\partial x_n}(p) \\ \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1}(p) & \cdots & \frac{\partial f_m}{\partial x_n}(p) \end{pmatrix}. \quad (*)$$

Definition 2.16. Die Matrix $(*)$, die die Ableitung von $f: U \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$ in einem Punkt $p \in U$ repräsentiert, bezeichnet man als *Jacobimatrix* von f in p .

Beispiel 2.17. (Konstante Abbildungen) Ist $f: V \rightarrow W$ konstant, so gilt für alle $p, x \in V$

$$f(x) = f(p),$$

so daß der lineare Term und der Rest in Definition 2.10 gleich Null sind. Also ist f in allen Punkte $p \in V$ differenzierbar mit $Df|_p = 0$.

Beispiel 2.18. (Lineare Abbildungen) Ist $f: V \rightarrow W$ eine lineare Abbildung zwischen V und W , so ist f in jedem Punkt $p \in V$ differenzierbar und die Ableitung $Df|_p$ ist gleich f , denn aufgrund der Linearität ist

$$f(x) = f(p) + f(x - p),$$

so daß der Restterm in der Definition 2.10 identisch verschwindet und die lineare Approximation A in p gleich f selbst ist. (Da Differenzierbarkeit definiert ist

als “lineare Approximierbarkeit”, ist dies wenig erstaunlich.) Die Ableitung einer linearen Abbildung $f: V \rightarrow W$ ist also die konstante Abbildung

$$Df: V \rightarrow L(V, W) \quad p \in V \mapsto Df|_p = f.$$

Insbesondere ist für eine lineare Abbildung $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ die Jacobimatrix

$$Df|_p = (a_{ij})$$

gleich der darstellenden $m \times n$ -Matrix (a_{ij}) . Die j -te partielle Ableitung $\frac{\partial f_i}{\partial x_j}(p)$ der i -ten Komponentenfunktion $f_i(x_1, \dots, x_n) = a_{i1}x_1 + \dots + a_{in}x_n$ ist mit anderen Worten dann für alle $p \in \mathbb{R}^n$ gleich a_{ij} .

Beispiel 2.19. (Multilineare Abbildungen) Eine k -lineare Abbildung

$$f: V_1 \times \dots \times V_k \rightarrow W$$

ist differenzierbar und die Ableitung in einem Punkt $p = (p_1, \dots, p_k)$ in Richtung $v = (v_1, \dots, v_k)$ ist gleich

$$Df|_p(v) = f(v_1, p_2, \dots, p_k) + f(p_1, v_2, p_3, \dots, p_k) + \dots + f(p_1, \dots, p_{k-1}, v_k).$$

In dem Beweis werden wir benutzen, daß es für eine k -lineare Abbildung f von endlichdimensionalen Vektorräumen V_1, \dots, V_k in einen endlichdimensionalen Vektorraum W für beliebige Normen auf den Vektorräumen eine Konstante C gibt mit

$$\|f(v_1, \dots, v_k)\| \leq C \|v_1\| \cdot \|v_2\| \cdots \|v_k\| \quad (*)$$

für alle $v_1 \in V_1, \dots, v_k \in V_k$.

Beweis. 1.) (Beweis der Ungleichung $(*)$) OBdA seien $V_1 = \mathbb{R}^{n_1}, \dots, V_k = \mathbb{R}^{n_k}$ mit Maximumsnormen (Satz 1.94). Dann gilt (analog zum Beweis von Satz 1.91) für alle $v_1 \in V_1, \dots, v_k \in V_k$

$$\begin{aligned} \|f(v_1, \dots, v_k)\| &= \left\| f\left(\sum_{i_1} e_{i_1} v_{i_1 1}, \dots, \sum_{i_k} e_{i_k} v_{i_k k}\right) \right\| \leq \\ &\leq \sum_{i_1, \dots, i_k} \|f(e_{i_1}, \dots, e_{i_k})\| |v_{i_1 1}| \cdots |v_{i_k k}| \leq \underbrace{\sum_{i_1, \dots, i_k} \|f(e_{i_1}, \dots, e_{i_k})\|}_{C:=} \|v_1\|_\infty \cdots \|v_k\|_\infty. \end{aligned}$$

2.) (Differenzierbarkeit) Für $p = (p_1, \dots, p_k)$ und $v = (v_1, \dots, v_k)$ gilt aufgrund der Multilinearität

$$\begin{aligned} f(p_1 + v_1, p_2 + v_2, \dots, p_k + v_k) &= f(p_1, p_2, \dots, p_k) + \\ &+ f(v_1, p_2, \dots, p_k) + f(p_1, v_2, p_3, \dots, p_k) + \dots + f(p_1, \dots, p_{k-1}, v_k) + \\ &+ R(p_1, \dots, p_k, v_1, \dots, v_k), \end{aligned}$$

wobei der Rest $R(p_1, \dots, p_k, v_1, \dots, v_k)$ aus $2^k - k - 1$ Summanden besteht, in die jeweils mindestens in 2 der Eingänge v_i 's eingesetzt werden. Aufgrund von (*) ist dieser Rest schnell genug fallend. Zum Beispiel ist im bilinearen Fall, also dem Fall $k = 2$, einfach

$$f(p_1 + v_1, p_2 + v_2) = f(p_1, p_2) + f(p_1, v_2) + f(v_1, p_2) + R(p_1, p_2, v_1, v_2)$$

mit $R(p_1, p_2, v_1, v_2) = f(v_1, v_2)$. Damit gilt, wenn man $V_1 \times V_2$ mit der Summe der Normen auf den Faktoren, also der Norm $\|(v_1, v_2)\| = \|v_1\| + \|v_2\|$, ausstattet, wegen (*)

$$\frac{\|R(p_1, p_2, v_1, v_2)\|}{\|(v_1, v_2)\|} \leq C \frac{\|v_1\| \|v_2\|}{\|v_1\| + \|v_2\|} \leq C \frac{\|v_1\|}{\|v_1\| + \|v_2\|} \|v_2\| \leq C \|v_2\| \xrightarrow{v_1, v_2 \rightarrow 0} 0,$$

so daß

$$Df_{|(p_1, p_2)}(v_1, v_2) = f(p_1, v_2) + f(v_1, p_2).$$

□

Zwischenbemerkung: Wir haben gesehen, daß man Ableitungen von differenzierbaren Abbildungen zwischen endlichdimensionalen reellen Vektorräumen berechnen kann, indem man Richtungsableitungen bzw. partielle Ableitungen ausrechnet. Damit ist die Berechnung von Ableitungen auf den Fall von Kurven, und damit im wesentlichen auf Analysis I zurückgeführt.

Unabhängig davon muß jedoch die Differenzierbarkeit einer Abbildung überprüft werden (vgl. Beispiel 2.15). In der Praxis kann man die oft relativ schnell mit den Regeln des folgenden Abschnitts sehen.

2.3 Regeln zur Berechnung von Ableitungen

In diesem Abschnitt diskutieren wir Regeln, mit denen man Ableitungen berechnen und vor allem auch die Differenzierbarkeit von Abbildungen überprüfen kann.

Satz 2.20. *Seien $f, g: U \subset V \rightarrow W$ Abbildungen von einer offenen Menge $U \subset V$ nach W , wobei V und W endlichdimensionale reelle Vektorräume sind. Sind f und g in $p \in U$ differenzierbar, so ist für alle $\lambda, \mu \in \mathbb{R}$ die Abbildung $\lambda f + \mu g$ in p differenzierbar und es gilt*

$$D(\lambda f + \mu g)|_p = \lambda Df|_p + \mu Dg|_p.$$

Der Beweis dieser und der folgenden Aussage ist einfach und sei dem Leser überlassen.

Satz 2.21. Eine Abbildung $f = (f_1, \dots, f_k): U \subset V \rightarrow W = W_1 \times \dots \times W_k$ ist genau dann differenzierbar im Punkt $p \in U$, wenn alle Komponentenabbildungen $f_i: U \subset V \rightarrow W_i$, $i = 1, \dots, k$ in p differenzierbar sind. Dann gilt

$$Df|_p = \begin{pmatrix} Df_{1|p} \\ \vdots \\ Df_{k|p} \end{pmatrix}.$$

2.3.1 Kettenregel

Satz 2.22. (Kettenregel) Seien V_1, V_2, V_3 endlichdimensionale reelle Vektorräume sowie $f: U_1 \subset V_1 \rightarrow V_2$ und $g: U_2 \subset V_2 \rightarrow V_3$ Abbildungen auf offenen Mengen $U_i \subset V_i$, $i = 1, 2$, so daß die Verkettung $g \circ f$ definiert ist. Ist f differenzierbar in $p \in U_1$ und g differenzierbar in $f(p)$, so ist $g \circ f$ differenzierbar in p und die Ableitung der Verkettung ist gleich der Verkettung der Ableitungen, also

$$D(g \circ f)|_p = Dg|_{f(p)} \circ Df|_p.$$

Beweis. Wegen der Differenzierbarkeit von f in p und g in $q = f(p)$ gilt

$$f(x) = f(p) + Df|_p(x - p) + \|x - p\|\psi_1(x) \quad \text{mit} \quad \psi_1(x) \xrightarrow{x \rightarrow p} 0 \quad (1)$$

und

$$g(y) = g(q) + Dg|_q(y - q) + \|y - q\|\psi_2(y) \quad \text{mit} \quad \psi_2(y) \xrightarrow{y \rightarrow q} 0. \quad (2)$$

Setzt man $y = f(x)$, so erhält man aus (2)

$$g(f(x)) = g(f(p)) + Dg|_{f(p)}(f(x) - f(p)) + \|f(x) - f(p)\|\psi_2(f(x)).$$

Setzt man für $f(x) - f(p)$ die Gleichung (1) ein, so erhält man

$$\begin{aligned} g(f(x)) &= g(f(p)) + (Dg|_{f(p)} \circ Df|_p)(x - p) + \\ &\quad Dg|_{f(p)}(\|x - p\|\psi_1(x)) + \|Df|_p(x - p) + \|x - p\|\psi_1(x)\| \psi_2(f(x)). \end{aligned}$$

Das zeigt die Behauptung, denn die Terme in der zweiten Zeile sind alle schnell genug fallend:

$$\frac{\|Dg|_{f(p)}(\|x - p\|\psi_1(x))\|}{\|x - p\|} \leq \|Dg|_{f(p)}\| \|\psi_1(x)\| \xrightarrow{x \rightarrow p} 0,$$

$$\begin{aligned} \frac{\|Df|_p(x - p) + \|x - p\|\psi_1(x)\| \|\psi_2(f(x))\|}{\|x - p\|} &\leq \\ &\leq \|Df|_p\| \|\psi_2(f(x))\| + \|\psi_1(x)\| \|\psi_2(f(x))\| \xrightarrow{x \rightarrow p} 0. \end{aligned}$$

□

Bemerkung 2.23. (Kettenregel in Koordinaten) Ist die Verkettung der Abbildungen $f: U_1 \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$ und $g: U_2 \subset \mathbb{R}^m \rightarrow \mathbb{R}^l$ definiert und sind f in p sowie g in $f(p)$ differenzierbar, so kann man die Kettenregel in Matrizenform

$$\begin{pmatrix} \frac{\partial(g \circ f)_1}{\partial x_1}(p) & \cdots & \frac{\partial(g \circ f)_1}{\partial x_n}(p) \\ \vdots & & \vdots \\ \frac{\partial(g \circ f)_l}{\partial x_1}(p) & \cdots & \frac{\partial(g \circ f)_l}{\partial x_n}(p) \end{pmatrix} = \begin{pmatrix} \frac{\partial g_1}{\partial y_1}(f(p)) & \cdots & \frac{\partial g_1}{\partial y_m}(f(p)) \\ \vdots & & \vdots \\ \frac{\partial g_l}{\partial y_1}(f(p)) & \cdots & \frac{\partial g_l}{\partial y_m}(f(p)) \end{pmatrix} \cdot \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(p) & \cdots & \frac{\partial f_1}{\partial x_n}(p) \\ \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1}(p) & \cdots & \frac{\partial f_m}{\partial x_n}(p) \end{pmatrix}$$

schreiben, wobei (x_1, \dots, x_n) bzw. (y_1, \dots, y_m) die Koordinaten auf \mathbb{R}^n und \mathbb{R}^m bezeichnen und $f_i = f_i(x_1, \dots, x_n)$, $i = 1, \dots, m$ sowie $g_j = g_j(y_1, \dots, y_m)$, $j = 1, \dots, l$ die Komponentenfunktionen von f bzw. g sind. Schreibt man das Matrizenprodukt aus, so ergibt die Kettenregel

$$\frac{\partial(g \circ f)_i}{\partial x_j}(p) = \sum_{k=1}^m \frac{\partial g_i}{\partial y_k}(f(p)) \frac{\partial f_k}{\partial x_j}(p).$$

Beispiel 2.24. Die Funktion

$$h: U \subset \mathbb{R}^2 \rightarrow \mathbb{R}, \quad (x, y) \mapsto \sqrt{1 - x^2 - y^2}$$

auf $U = \{x^2 + y^2 < 1\} \subset \mathbb{R}^2$ ist differenzierbar, denn die linearen Abbildungen $(x, y) \mapsto x$ bzw. y sind differenzierbar, die Abbildungen $x \mapsto x^2$ bzw. $y \mapsto y^2$ sind differenzierbar (nach Analysis I), so daß mit der Kettenregel $(x, y) \mapsto x^2$ bzw. y^2 differenzierbar sind. Damit ist auch $(x, y) \mapsto 1 - x^2 - y^2$ differenzierbar, so daß h , wegen der Differenzierbarkeit von $\sqrt{\cdot}: \mathbb{R}_{>0} \rightarrow \mathbb{R}$ (vgl. Analysis I), mit der Kettenregel differenzierbar ist.

Die Berechnung der Ableitungen erfolgt am einfachsten durch Bestimmung der partiellen Ableitungen $\frac{\partial h}{\partial x}$ und $\frac{\partial h}{\partial y}$, welche ergibt

$$Dh = \begin{pmatrix} \frac{\partial h}{\partial x} & \frac{\partial h}{\partial y} \end{pmatrix} = \begin{pmatrix} \frac{-x}{\sqrt{1-x^2-y^2}} & \frac{-y}{\sqrt{1-x^2-y^2}} \end{pmatrix}.$$

Der Graph $\{(x, y, z) \in \mathbb{R}^3 \mid x^2 + y^2 < 1, z = h(x, y)\}$ von h ist die Einheitshalb-sphäre im durch $z > 0$ gegebenen Halbraum in \mathbb{R}^3 . Die Abbildung $g: U \rightarrow \mathbb{R}^3$, $(x, y) \mapsto (x, y, h(x, y))$ ist eine Parametrisierung dieser Halbsphäre. Die partiellen Ableitungen

$$\frac{\partial g}{\partial x} = \begin{pmatrix} 1 \\ 0 \\ \frac{-x}{\sqrt{1-x^2-y^2}} \end{pmatrix} \quad \text{und} \quad \frac{\partial g}{\partial y} = \begin{pmatrix} 0 \\ 1 \\ \frac{-y}{\sqrt{1-x^2-y^2}} \end{pmatrix}$$

sind Tangentialvektoren an die Koordinatenkurven $x \mapsto g(x, y)$ bzw. $y \mapsto g(x, y)$.
(BILD)

Beispiel 2.25. Die Abbildung

$$f: \mathbb{R}^2 \mapsto \mathbb{R}^3 \quad (\vartheta, h) \mapsto \begin{pmatrix} \cos(\vartheta) \\ \sin(\vartheta) \\ h \end{pmatrix}$$

ist differenzierbar, da die Projektionen $(\vartheta, h) \mapsto \vartheta$ bzw. h linear und somit differenzierbar sind und die Funktionen $\vartheta \mapsto \cos(\vartheta)$ und $\vartheta \mapsto \sin(\vartheta)$ (nach Analysis I) differenzierbar sind, und damit nach der Kettenregel die Komponentenfunktionen $(\vartheta, h) \mapsto \cos(\vartheta)$ und $(\vartheta, h) \mapsto \sin(\vartheta)$ differenzierbar sind. Die Ableitung ist

$$Df = \left(\frac{\partial f}{\partial \vartheta} \quad \frac{\partial f}{\partial h} \right) = \begin{pmatrix} -\sin(\vartheta) & 0 \\ \cos(\vartheta) & 0 \\ 0 & 1 \end{pmatrix}.$$

Das Bild von f ist ein Zylinder. Die partiellen Ableitungsvektoren $\frac{\partial f}{\partial \vartheta}$ und $\frac{\partial f}{\partial h}$ sind Tangentialvektoren an die Koordinatenkurven $\vartheta \mapsto f(\vartheta, h)$ und $h \mapsto f(\vartheta, h)$. (BILD)

2.3.2 Produktregel

Der folgende Satz ist eine Ableitungsregel für Produkte aller Art, z.B.

- Produkte mit Skalaren $\mathbb{R} \times V \rightarrow V$,
- Skalarprodukte $\langle \cdot, \cdot \rangle: V \times V \rightarrow \mathbb{R}$,
- das Kreuzprodukt $\times: \mathbb{R}^3 \times \mathbb{R}^3 \rightarrow \mathbb{R}^3$,
- Matrizenprodukte,
- Verkettungen von Linearen Abbildungen $L(V_2, V_3) \times L(V_1, V_2) \rightarrow L(V_1, V_3)$,
- Einsetzungsabbildungen $L(V, W) \times V \rightarrow W$,
- Determinanten $\det: \underbrace{V \times \dots \times V}_{\dim(V)} \rightarrow \mathbb{R}, \dots$

Satz 2.26. (*Produktregel*) Seien $\alpha: V_1 \times \dots \times V_k \rightarrow W$ eine k -lineare Abbildung und $f_i: U \subset V \rightarrow V_i$, $i = 1, \dots, k$ differenzierbar in $p \in U$, wobei V, V_1, \dots, V_k sowie W endlichdimensionale reelle Vektorräume sind und $U \subset V$ offen ist. Dann ist $f = \alpha(f_1, \dots, f_k): U \subset V \rightarrow W$ differenzierbar in p und

$$\begin{aligned} Df|_p(v) &= \alpha(Df_{1|p}(v), f_2(p), \dots, f_k(p)) + \\ &\quad + \alpha(f_1(p), Df_{2|p}(v), f_3(p), \dots, f_k(p)) + \\ &\quad \dots \\ &\quad + \alpha(f_1(p), \dots, f_{k-1}(p), Df_{k|p}(v)). \end{aligned}$$

Beweis. Folgt direkt aus der Kettenregel (Satz 2.22), weil

$$\alpha: V_1 \times \cdots \times V_k \rightarrow W$$

differenzierbar ist (Beispiel 2.19) und

$$(f_1, \dots, f_k): U \subset V \rightarrow V_1 \times \cdots \times V_k$$

in p differenzierbar ist (Satz 2.21). □

Beispiel 2.27. Ist $\gamma: I \subset \mathbb{R} \rightarrow V$ eine differenzierbare Kurve in einem euklidischen Vektorraum. Dann ist $f(t) = \langle \gamma(t), \gamma(t) \rangle$ differenzierbar und

$$f'(t) = 2\langle \gamma(t), \gamma'(t) \rangle.$$

Insbesondere steht für eine Kurve mit Werten in der Einheitssphäre $\{v \in V \mid \langle v, v \rangle = 1\}$ die Ableitung (“Geschwindigkeit”) $\gamma'(t)$ immer senkrecht auf $\gamma(t)$.

Beispiel 2.28. Die Funktion

$$f: \mathbb{R}^2 \rightarrow \mathbb{R} \quad (x, y) \mapsto 1 + 3x + 4y + 5xy^2$$

ist differenzierbar, denn sie ist die Summe aus einer konstanten Funktion, der linearen Funktion $(x, y) \mapsto 3x + 4y$ und der Funktion $(x, y) \mapsto xy^2$, welche nach der Produktregel (Satz 2.26) differenzierbar ist (mit Produkt $\alpha: \mathbb{R} \times \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$, $(x_1, x_2, x_3) \mapsto x_1 \cdot x_2 \cdot x_3$).

Die Ableitung von f kann man durch Berechnung der partiellen Ableitungen bestimmen. Es gilt

$$Df = \left(\frac{\partial f}{\partial x} \quad \frac{\partial f}{\partial y} \right) = (3 + 5y^2 \quad 4 + 10xy).$$

Die Ableitung in den Punkten $p = (0, 0)$ und $p = (1, 2)$ ist damit zum Beispiel

$$Df|_{p=(0,0)} = (3 \quad 4) \quad \text{bzw.} \quad Df|_{p=(1,2)} = (23 \quad 24).$$

Beispiel 2.29. Die euklidische Norm $\|x\| = \sqrt{\langle x, x \rangle}$ auf einem euklidischen Vektorraum V ist auf $V \setminus \{0\}$ differenzierbar (und falls $\dim(V) > 0$ in 0 nicht differenzierbar, siehe Bem 2.30). Die Ableitung von $f: V \setminus \{0\} \rightarrow \mathbb{R}$, $x \mapsto \sqrt{\langle x, x \rangle}$ in $p \neq 0$ ist nach Ketten- und Produktregel

$$Df|_p(v) = \frac{1}{2\sqrt{\langle p, p \rangle}} (\langle p, v \rangle + \langle v, p \rangle) = \frac{\langle p, v \rangle}{\|p\|}.$$

(Die Kettenregel ist in $p = 0$ nicht anwendbar, da die Wurzel in 0 nicht differenzierbar ist.)

Bemerkung 2.30. Eine Norm auf einem Vektorraum $V \neq \{0\}$ ist im Nullpunkt nicht differenzierbar, da die Richtungsableitungen nicht existieren, denn für $v \in V$ gilt

$$\lim_{t \searrow 0} \frac{\|tv\| - \|0\|}{t} = \|v\|$$

und

$$\lim_{t \nearrow 0} \frac{\|tv\| - \|0\|}{t} = -\|v\|,$$

so daß für $v \neq 0$ die Richtungsableitung von $\| \cdot \|$ im Punkt 0 in Richtung v nicht existiert.

2.3.3 Ein hinreichendes Kriterium für Differenzierbarkeit

Wir haben gesehen, daß man Ableitungen zwar berechnen kann, indem man Richtungsableitungen bzw. partielle Ableitungen bestimmt (Lemma 2.12), deren Existenz jedoch nicht die Differenzierbarkeit impliziert (Beispiel 2.15). Die folgenden Kriterien zeigen, daß die

- Existenz der Richtungsableitungen bzw. partielle Ableitungen in einer Umgebung von p
- zusammen mit deren Stetigkeit in p

die Differenzierbarkeit in p impliziert.

Satz 2.31. *Seien V und W endlichdimensionale reelle Vektorräume und $f: U \subset V \rightarrow W$ eine Abbildung auf einer offenen Menge $U \subset V$. Existieren alle Richtungsableitungen $\partial_v f(q)$, $q \in U$, $v \in V$ und ist für alle $v \in V$ die Abbildung*

$$q \mapsto \partial_v f(q)$$

stetig in $p \in U$, so ist f in p differenzierbar.

Der Satz folgt (weil eine Basis von V eine Identifikation $V = \mathbb{R}^n$ definiert) aus dem folgenden Satz, welcher auch zeigt, daß es reicht, die Existenz der Richtungsableitungen $\partial_v f(q)$ und Stetigkeit von $q \mapsto \partial_v f(q)$ für alle Vektoren $v = v_i$ einer Basis v_1, \dots, v_n von V zu fordern

Satz 2.32. *Ist $f: U \subset \mathbb{R}^n \rightarrow W$ eine Abbildung von einer offenen Menge $U \subset \mathbb{R}^n$ in einen endlichdimensionalen reellen Vektorraum W . Existieren alle partiellen Ableitungen $\frac{\partial f}{\partial x_i}(q)$, $q \in U$, $i = 1, \dots, n$ und sind diese in p stetig, so ist f in p differenzierbar.*

Beweis. Ist f in p differenzierbar, so muß aufgrund von Lemma 2.12 gelten, daß $Df|_p = A$ mit

$$A \left(\sum_{i=1}^n v_i e_i \right) = \sum_{i=1}^n v_i \frac{\partial f}{\partial x_i}(p).$$

Schreibt man

$$\begin{aligned} f(x) - f(p) &= f(x_1, \dots, x_n) - f(p_1, \dots, p_n) = \\ &= f(x_1, \dots, x_n) - f(p_1, x_2, \dots, x_n) + \\ &\quad + f(p_1, x_2, \dots, x_n) - f(p_1, p_2, x_3, \dots, x_n) \\ &\quad \vdots \\ &\quad + f(p_1, \dots, p_{n-1}, x_n) - f(p_1, \dots, p_n), \end{aligned}$$

so ergibt zeilenweise Anwendung des Mittelwertsatzes aus Analysis I

$$\begin{aligned} f(x) - f(p) &= \frac{\partial f}{\partial x_1}(\xi_1, x_2, \dots, x_n)(x_1 - p_1) + \\ &+ \frac{\partial f}{\partial x_2}(p_1, \xi_2, x_3, \dots, x_n)(x_2 - p_2) + \\ &\quad \vdots \\ &+ \frac{\partial f}{\partial x_n}(p_1, \dots, p_{n-1}, \xi_n)(x_n - p_n), \end{aligned}$$

wobei jedes ξ_i eine passende Stelle aus dem Intervall $[x_i, p_i]$ bzw. $[p_i, x_i]$ ist. Damit ist für $x \neq p$

$$\begin{aligned} \frac{f(x) - f(p) - A(x - p)}{\|x - p\|} &= \\ &= \sum_{i=1}^n \frac{x_i - p_i}{\|x - p\|} \left(\frac{\partial f}{\partial x_i}(p_1, \dots, p_{i-1}, \xi_i, x_{i+1}, \dots, x_n) - \frac{\partial f}{\partial x_i}(p_1, \dots, p_n) \right). \end{aligned}$$

Das zeigt die Behauptung, da die rechte Seite aufgrund der Stetigkeit der partiellen Ableitungen in p (und der Beschränktheit der $\frac{x_i - p_i}{\|x - p\|}$) für $x \rightarrow p$ gegen Null geht. \square

Beispiel 2.33. In den obigen Beispielen 2.24 und 2.28 hätte man die Differenzierbarkeit auch über Satz 2.32 beweisen können, indem man die partiellen Ableitungen berechnet und deren Stetigkeit nachweist. (Den Nachweis der Stetigkeit würde man jedoch in der Regel mit Argumenten führen, die denen ähnlich sind, welche wir oben zum Nachweis der Differenzierbarkeit angegeben haben.)

2.3.4 Der Gradient

In einem endlichdimensionalen euklidischen Vektorraum $(V, \langle \cdot, \cdot \rangle)$ kann man jedem Vektor $v \in V$ über

$$\alpha = \langle v, _ \rangle$$

eine eindeutige Linearform $\alpha \in V^*$ zuordnen und umgekehrt. Da die Ableitung $Df|_p$ einer in p differenzierbaren Funktion $f: U \subset V \rightarrow \mathbb{R}$ auf einer offenen Menge $U \subset V$ ein Element des Dualraums $L(V, \mathbb{R}) = V^*$ ist, kann man daraus durch obigen Isomorphismus zwischen V und V^* einen Vektor machen.

Definition 2.34. Der *Gradient* einer in $p \in U$ differenzierbaren Funktion $f: U \subset V \rightarrow \mathbb{R}$ auf einer offenen Teilmenge U eines endlichdimensionalen euklidischen Vektorraums $(V, \langle \cdot, \cdot \rangle)$ ist der Vektor $\text{grad}(f)(p)$ mit

$$\langle \text{grad}(f)(p), v \rangle = Df|_p(v)$$

für alle $v \in V$.

Bemerkung 2.35. • $\langle \text{grad}(f)(p), v \rangle$ ist gleich der Richtungsableitung $\partial_v f(p)$.

- Für $V = \mathbb{R}^n$ mit Standardskalarprodukt ist

$$\text{grad}(f)(p) = \begin{pmatrix} \frac{\partial f}{\partial x_1}(p) \\ \vdots \\ \frac{\partial f}{\partial x_n}(p) \end{pmatrix}.$$

Satz 2.36. *Der Gradientenvektor zeigt in Richtung des stärksten Anstiegs der Funktion. Der Betrag des Gradienten ist gleich dem Maximum der Richtungsableitungen.*

Beweis. Für $\text{grad}(f)(p) \neq 0$ gilt nach Cauchy–Schwarz (Satz 1.64) für alle Vektoren $v \in V$ von Länge 1

$$|\partial_v f(p)| = |\langle \text{grad}(f)(p), v \rangle| \leq \|\text{grad}(f)(p)\|$$

mit Gleichheit für $v = \pm \frac{\text{grad}(f)(p)}{\|\text{grad}(f)(p)\|}$. Das Maximum von $\partial_v f(p)$ wird erreicht für $v = + \frac{\text{grad}(f)(p)}{\|\text{grad}(f)(p)\|}$ und ist gleich $\|\text{grad}(f)(p)\|$. \square

Bemerkung 2.37. Physiker betrachten oft den Vektor $-\text{grad}(f)(p)$, der in Richtung des “steilsten Abstiegs” zeigt.

Beispiel 2.38. In Beispiel 2.29 haben wir gesehen, daß für jeden endlichdimensionalen euklidischen Vektorraum die Funktion $f: V \rightarrow \mathbb{R}$, $x \mapsto \|x\| = \sqrt{\langle x, x \rangle}$ auf $V \setminus \{0\}$ differenzierbar ist und ihre Ableitung berechnet. Die dort errechnete Formel ergibt

$$\text{grad}(f)(p) = \frac{p}{\|p\|}$$

für alle $p \in V \setminus \{0\}$. Man überlege sich, daß man diese Formel direkt (“ohne Rechnen”) aus dem vorhergehenden Satz folgern kann, wenn man berücksichtigt, daß der Wert der Funktion f in p genau der Abstand zum Nullpunkt ist.

Der Vektor $\text{grad}(f)(p)$ steht in jedem Punkt $p \in V \setminus \{0\}$ senkrecht auf die Niveauflächen von f (BILD). Wir werden sehen (Beispiel 4.13), daß dies ein Phänomen ist, das für beliebige differenzierbare Funktionen gilt.

2.4 Höhere Ableitungen

Seien V , W endlichdimensionale reelle Vektorräume und $U \subset V$ eine offenen Menge. Das Differential einer differenzierbaren Abbildung $f: U \rightarrow W$ ist eine Abbildung

$$Df: U \rightarrow L(V, W)$$

von U in den Vektorraum $L(V, W)$ der linearen Abbildungen von V nach W . Ist diese Abbildung selbst differenzierbar, so ist deren Ableitung eine Abbildung

$$D^2f: U \rightarrow L(V, L(V, W)).$$

Den Vektorraum $L(V, L(V, W))$ identifizieren wir dabei mit dem Vektorraum $L^2(V, W)$ der bilinearen Abbildungen von V nach W , indem wir folgendes allgemeines Prinzip anwenden.

Notation: Für $k \geq 2$ identifizieren wir den Vektorraum

$$L^k(V, W) = \{\alpha: \underbrace{V \times \dots \times V}_k \rightarrow W \mid \alpha \text{ ist } k\text{-linear}\}$$

über den Isomorphismus

$$j_k: L(V, L^{k-1}(V, W)) \rightarrow L^k(V, W)$$

definiert durch

$$j_k(\Phi)(v_1, \dots, v_k) = \Phi(v_1)(v_2, \dots, v_n) \quad \text{für } v_1, \dots, v_n \in V$$

mit dem Vektorraum $L(V, L^{k-1}(V, W))$.

Zum Beispiel kann man ein Skalarprodukt $\langle \cdot, \cdot \rangle \in L^2(V, \mathbb{R})$ auf einem endlichdimensionalen reellen Vektorraum interpretieren als die Abbildung $L(V, L(V, \mathbb{R})) = L(V, V^*)$, die einem Vektor $v \in V$ die Linearform $\alpha = \langle v, _ \rangle \in V^*$ zuordnet (vgl. Abschnitt 2.3.4).

Definition 2.39. Rekursiv definiert man für $k \geq 2$: eine Abbildung

$$f: U \rightarrow W$$

ist k -mal differenzierbar in $p \in U$, wenn es eine offene Umgebung $\tilde{U} \subset U$ von $p \in \tilde{U}$ gibt, so daß $f|_{\tilde{U}}$ $(k-1)$ -mal differenzierbar ist und die $(k-1)$ -te Ableitung

$$D^{k-1}f: \tilde{U} \rightarrow L^{(k-1)}(V, W)$$

im Punkt p differenzierbar ist. Die k -te Ableitung $D^k f|_p \in L^k(V, W)$ in p ist die Ableitung von $D^{k-1}f$ in p , also

$$D^k f|_p(v_1, \dots, v_k) = j_k(D(D^{k-1}f)|_p)(v_1, \dots, v_k) = D(D^{k-1}f)|_p(v_1)(v_2, \dots, v_k)$$

(wobei wir im letzten Schritt die obige Identifizierung angewandt haben).

Lemma 2.40. Ist f in p k -mal differenzierbar, so gilt für $v_1, \dots, v_k \in V$

$$D^k f|_p(v_1, \dots, v_k) = \partial_{v_1}(\partial_{v_2} \dots (\partial_{v_k} f))(p)$$

(insbesondere existiert die rechte Seite).

Beweis. (Per Induktion) Für $k = 1$ ist die Aussage gleich Lemma 2.12. Um zu zeigen, daß die Aussage für $k-1$ die Aussage für k impliziert, benutzen wir die lineare Abbildung (den "Einsetzungshomomorphismus")

$$g: L^{k-1}(V, W) \rightarrow W \quad \eta \mapsto \eta(v_2, \dots, v_k)$$

definiert durch $v_2, \dots, v_k \in V$. Da g linear ist, ist die Ableitung von g konstant und $Dg|_\eta = g$. Damit impliziert die Kettenregel angewandt auf $g \circ D^{k-1}f$, daß

$$D(g \circ D^{k-1}f)|_p(v_1) = g \circ D(D^{k-1}f)|_p(v_1)$$

für alle $v_1 \in V$. Die rechte Seite ist per definitionem gleich $D^k f|_p(v_1, \dots, v_k)$. Die linke Seite ist nach Lemma 2.12 (und Induktionsannahme) gleich

$$\partial_{v_1}(g \circ D^{k-1}f)(p) = \partial_{v_1}(\partial_{v_2} \dots (\partial_{v_k} f))(p).$$

□

Bemerkung 2.41. Will man $D^2 f|_p(v_1, v_2)$ berechnen, so kann man also anstatt

$$Df: U \rightarrow L(V, W)$$

abzuleiten auch erst v_2 einsetzen und dann in Richtung v_1 ableiten (und analog für höhere Ableitungen).

Per Induktion erhält man aus Satz 2.32:

Satz 2.42. *Besitzt $f: U \subset V \rightarrow W$ beliebige Richtungsableitungen (oder partielle Ableitungen, falls $V = \mathbb{R}^n$) bis zur Ordnung k und sind diese stetig in $p \in U$, so ist f k -mal differenzierbar in p .*

Satz 2.43. (H.A. Schwarz) *Ist $f: U \subset V \rightarrow W$ in $p \in U$ zweimal differenzierbar, so gilt für $v_1, v_2 \in V$*

$$D^2 f|_p(v_1, v_2) = D^2 f|_p(v_2, v_1).$$

Der Satz folgt aus dem folgenden Lemma:

Lemma 2.44. *Ist $f: U \subset V \rightarrow W$ in $p \in U$ zweimal differenzierbar, so gilt für $v_1, v_2 \in V$*

$$D^2 f|_p(v_1, v_2) = \lim_{t \rightarrow 0} \frac{f(p + tv_1 + tv_2) - f(p + tv_1) - f(p + tv_2) + f(p)}{t^2}.$$

Beweis. OBdA sein $W = \mathbb{R}$ (sonst wähle man eine Basis von W und argumentiere komponentenweise). Weiter seien v_1, v_2 oBdA so klein, daß $p + tv_1 + sv_2 \in U$ für alle $t, s \in [0, 1]$. Wir betrachten

$$\Delta(t, v_1, v_2) = f(p + tv_1 + tv_2) - f(p + tv_1) - f(p + tv_2) + f(p).$$

Dann gilt $\Delta(t, v_1, v_2) = F(1) - F(0)$, wobei

$$F(s) = f(p + tv_1 + stv_2) - f(p + stv_2).$$

Der Mittelwertsatz aus Analysis I impliziert, daß es für jedes t ein $\tau = \tau_t \in [0, 1]$ gibt mit

$$\Delta(t, v_1, v_2) = F'(\tau) = Df_{p+tv_1+\tau tv_2}(tv_2) - Df_{p+\tau tv_2}(tv_2)$$

(das t im Index von $\tau = \tau_t$ lassen wir zur Vereinfachung der Notation weg). Die zweimalige Differenzierbarkeit in p ergibt, daß für kleine $w \in V$

$$Df_{p+w} = Df|_p + D^2 f|_p(w) + \underbrace{\|w\|\psi(p+w)}_{R(p+w):=}$$

mit $\psi(p+w) \rightarrow 0$ für $w \rightarrow 0$. Damit ist

$$\begin{aligned} \frac{\Delta(t, v_1, v_2)}{t^2} &= \frac{Df_{p+tv_1+\tau tv_2}(tv_2) - Df_{p+\tau tv_2}(tv_2)}{t^2} = \\ &= \frac{(Df|_p + D^2 f|_p(tv_1 + \tau tv_2) + R(p + tv_1 + \tau tv_2))(tv_2)}{t^2} \\ &\quad - \frac{(Df|_p + D^2 f|_p(\tau tv_2) + R(p + \tau tv_2))(tv_2)}{t^2} \\ &= D^2 f|_p(v_1, v_2) + \frac{R(p + tv_1 + \tau tv_2)}{t}(v_2) - \frac{R(p + \tau tv_2)}{t}(v_2) \end{aligned}$$

Damit folgt die Behauptung, da die beiden R -Terme für $t \rightarrow 0$ gegen Null gehen.
Zum Beispiel ist

$$\frac{R(p + tv_1 + \tau tv_2)}{|t|} = \frac{\|tv_1 + \tau tv_2\| \psi(p + tv_1 + \tau tv_2)}{|t|} = \|v_1 + \tau v_2\| \psi(p + tv_1 + \tau tv_2),$$

was wegen der Beschränktheit von $\|v_1 + \tau v_2\| \leq \|v_1\| + |\tau| \|v_2\| \leq \|v_1\| + \|v_2\|$ und
Aufgrund der Eigenschaften von ψ gegen Null geht. \square

Korollar 2.45. Ist $f: U \subset V \rightarrow W$ in $p \in U$ k -mal differenzierbar, so ist die k -lineare Abbildung

$$D^k f|_p: \underbrace{V \times \dots \times V}_k \rightarrow W$$

symmetrisch.

Beweis. Für $k = 2$ ist das der Satz von Schwarz (Satz 2.43). Dieser sagt wegen Lemma 2.40 aus, daß in jedem Punkt q , in dem eine Abbildung g zweimal differenzierbar ist, die Richtungsableitungen ∂_{v_l} und $\partial_{v_{l+1}}$, $l \in \{1, \dots, k-1\}$ "kommutieren", also

$$\partial_{v_l}(\partial_{v_{l+1}}g)(q) = \partial_{v_{l+1}}(\partial_{v_l}g)(q).$$

Das impliziert die Behauptung im allgemeinen Fall, da (wegen Lemma 2.40)

$$D^k f|_p(v_1, \dots, v_k) = \partial_{v_1}(\partial_{v_2} \dots (\partial_{v_k} f))(p).$$

□

Definition 2.46. Eine Abbildung $f: U \subset V \rightarrow W$ ist k -mal stetig differenzierbar, wenn sie k -mal differenzierbar ist und $D^k f: U \rightarrow L^k(V, W)$ stetig ist.

$$C^k(U, W) = \{f: U \rightarrow W \mid f \text{ ist } k\text{-mal stetig differenzierbar}\}$$

Bemerkung 2.47. Wegen Satz 2.42 ist f genau dann k -mal stetig differenzierbar, wenn für alle $v_1, \dots, v_l \in V$ mit $l \leq k$ und alle $p \in U$ die Richtungsableitung

$$\partial_{v_1}(\partial_{v_2}(\dots \partial_{v_l} f))(p)$$

existiert und stetig von p abhängt.

2.4.1 Höhere partielle Ableitungen

Definition 2.48. Für $f: U \subset \mathbb{R}^n \rightarrow W$ und $i_1, \dots, i_k \in \{1, \dots, n\}$ schreiben wir

$$\frac{\partial^k f}{\partial x_{i_1} \dots \partial x_{i_k}}(p) = \partial_{e_{i_1}}(\partial_{e_{i_2}}(\dots \partial_{e_{i_k}} f))(p)$$

für die höheren Richtungsableitungen in Richtung der Standardbasisvektoren.

Bemerkung 2.49. • Ist $f: U \subset \mathbb{R}^n \rightarrow W$ k -mal differenzierbar in $p \in U$, so gilt

$$D^k f|_p(v_1, \dots, v_k) = \sum_{i_1, \dots, i_k=1}^n \frac{\partial^k f}{\partial x_{i_1} \dots \partial x_{i_k}}(p) \cdot v_{i_1 1} \dots v_{i_k k},$$

$$\text{wobei } v_j = \begin{pmatrix} v_{1j} \\ \vdots \\ v_{nj} \end{pmatrix} = \sum_{i_j=1}^n e_{i_j} v_{i_j j}, \quad j = 1, \dots, n.$$

- Ist $f: U \subset \mathbb{R}^n \rightarrow W$ 2-mal differenzierbar in $p \in U$, so gilt mit dem Satz von Schwarz (Satz 2.43)

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(p) = \frac{\partial^2 f}{\partial x_j \partial x_i}(p)$$

für alle $i, j \in \{1, \dots, n\}$. Falls f k -mal differenzierbar ist, kommt es analog (wegen Korollar 2.45) bei der Berechnung einer höheren partiellen Ableitung

$$\frac{\partial^k f}{\partial x_{i_1} \cdots \partial x_{i_k}}(p)$$

nicht auf die Reihenfolge der i_1, \dots, i_k an.

- Eine Abbildung $f: U \subset \mathbb{R}^n \rightarrow W$ ist genau dann k -mal stetig differenzierbar, wenn alle partiellen Ableitungen bis Ordnung k existieren und stetig sind (wobei die nicht-triviale Implikation hier wegen Satz 2.42 gilt).

2.5 Mittelwertsatz und Satz von Taylor

Satz 2.50. (Mittelwertsatz) Sei $f: U \subset V \rightarrow \mathbb{R}$ eine reellwertige differenzierbare Funktion auf einer offenen Teilmenge $U \subset V$ eines endlichdimensionalen reellen Vektorraumes. Liegt für zwei Punkte $p, q \in U$ die Strecke

$$\overline{pq} = \{(1 - \lambda)p + \lambda q = p + \lambda(q - p) \mid \lambda \in [0, 1]\}$$

ganz in U , so gibt es $\tau \in]0, 1[$ mit

$$f(q) - f(p) = Df_{p+\tau(q-p)}(q - p).$$

Beweis. Nach dem Mittelwertsatz aus Analysis I gibt es $\tau \in]0, 1[$, so daß für $h: [0, 1] \rightarrow \mathbb{R}$, $t \mapsto f(p + t(q - p))$ gilt

$$h(1) - h(0) = h'(\tau).$$

Mit der Kettenregel ergibt sich daraus die Aussage des Satzes. □

Beispiel 2.51. (Der Mittelwertsatz in dieser Form gilt nicht für Abbildungen in höherdimensionale Räume.) Für

$$f: \mathbb{R} \rightarrow \mathbb{C} = \mathbb{R}^2 \quad t \mapsto e^{2\pi i t} = \begin{pmatrix} \cos(2\pi t) \\ \sin(2\pi t) \end{pmatrix}$$

gilt $0 = f(1) - f(0) \neq f'(\tau)$ für alle $\tau \in]0, 1[$, denn

$$f'(t) = 2\pi i e^{2\pi i t} = 2\pi \begin{pmatrix} -\sin(2\pi t) \\ \cos(2\pi t) \end{pmatrix} \neq 0$$

für alle $t \in \mathbb{R}$.

Satz 2.52. (Schranksatz) Seien V, W endlichdimensionale reelle normierte Vektorräume und $f: U \subset V \rightarrow W$ eine differenzierbare Abbildung auf einer offenen Teilmenge $U \subset V$. Liegt für zwei Punkte $p, q \in U$ die Strecke

$$\overline{pq} = \{(1 - \lambda)p + \lambda q = p + \lambda(q - p) \mid \lambda \in [0, 1]\}$$

ganz in U , so gilt

$$\|f(q) - f(p)\| \leq \sup_{x \in \overline{pq}} \|Df_x\| \|q - p\|.$$

Beweis. OBdA sei $f(q) \neq f(p)$. Der Einfachheit halber nehmen wir an, daß die Norm auf W von einem Skalarprodukt induziert wird (den allgemeinen Fall findet man z.B. als Satz 145 im Skript “Analysis 2” von Dirk Ferus; alternativ dazu skizzieren wir am Ende des Beweises, wie man den hier gegebenen Beweis an den allgemeinen Fall anpassen kann). Nach dem Mittelwertsatz aus Analysis I gibt es für die Funktion

$$g: [0, 1] \rightarrow \mathbb{R} \quad t \mapsto \langle f(p + t(q - p)), v \rangle \quad \text{mit} \quad v = \frac{f(q) - f(p)}{\|f(q) - f(p)\|} \quad (*)$$

ein $\tau \in]0, 1[$, so daß

$$g(1) - g(0) = g'(\tau).$$

Nun ist

$$g(1) - g(0) = \langle f(q), v \rangle - \langle f(p), v \rangle = \|f(q) - f(p)\|$$

und

$$g'(\tau) = \langle Df_{p+\tau(q-p)}(q - p), v \rangle,$$

so daß mit Cauchy–Schwarz (Satz 1.64) und $\|v\| = 1$ gilt

$$|g'(\tau)| \leq \|Df_{p+\tau(q-p)}\| \|q - p\| \|v\| \leq \sup_{x \in \overline{pq}} \|Df_x\| \|q - p\|.$$

(Im Fall einer allgemeinen Norm auf W wählt man eine Linearform $\alpha \in W^*$ mit $\|\alpha\| = 1$ und $\alpha(f(q) - f(p)) = \|f(q) - f(p)\|$ und ersetzt g in $(*)$ durch $g(t) = \alpha(f(p + t(q - p)))$; der Rest des Beweises geht dann mit offensichtlichen Modifikationen durch. Die Existenz von α folgt aus dem Satz von Hahn–Banach, den man z.B. auf Seite 69 der “Real and functional analysis” (3. Ausgabe) von Serge Lang findet. Im endlichdimensionalen Fall kann man im zweiten Teil des Beweises dort das Argument mit dem Lemma von Zorn durch eine einfach Induktion über die Dimension ersetzen.) \square

Korollar 2.53. Seien V, W endlichdimensionale reelle Vektorräume und $f: U \subset V \rightarrow W$ eine differenzierbare Funktion auf einer offenen Teilmenge $U \subset V$. Ist U zusammenhängend und $Df|_p = 0$ für alle $p \in U$, so ist f konstant.

Beweis. Für jeden Punkt $p \in U$ gibt es ein $\epsilon > 0$, so daß der ϵ -Ball $B_\epsilon(p)$ um p ganz in U liegt. Nach dem Schrankensatz (Satz 2.52) ist f konstant auf $B_\epsilon(p)$.

Sei nun $p_0 \in U$ ein beliebiger Punkt und $V = \{p \in U \mid f(p) = f(p_0)\}$. Dann ist $V \subset U$ abgeschlossen, da f stetig ist. Andererseits ist V offen, da f auf einer ϵ -Umgebung um jeden Punkt konstant ist. Also ist $V = U$, da U zusammenhängend ist. \square

Der folgende Satz zeigt, daß jede stetig differenzierbare Abbildung auf einer kompakten Menge Lipschitz stetig ist (wobei man unter einer stetig differenzierbaren Abbildung auf einer kompakten Menge K eine Abbildung versteht, die man stetig differenzierbar fortsetzen kann auf eine offene Umgebung U von $K \subset U$).

Satz 2.54. Seien V, W endlichdimensionale reelle Vektorräume, $U \subset V$ eine offene Teilmenge und $f: U \subset V \rightarrow W$ eine stetig differenzierbare Abbildung. Für jede kompakte Menge $K \subset U$ gibt es $L > 0$ mit

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

Beweis. Gäbe es kein derartiges L , so könnte man Folgen $(x_n)_{n \in \mathbb{N}}$ und $(y_n)_{n \in \mathbb{N}}$ in K finden mit

$$\|f(x_n) - f(y_n)\| > n\|x_n - y_n\|.$$

Da K kompakt ist, können wir oBdA annehmen, daß $x_n \rightarrow x$ für $n \rightarrow \infty$ (Satz 1.52 von Bolzano–Weierstraß). Für $r > 0$, so daß $\overline{B_r(x)} \subset U$, gilt dann für $y, y' \in \overline{B_r(x)}$ und $L = \sup_{y \in \overline{B_r(x)}} \|Df|_y\| < \infty$ (nach Satz 1.57) mit dem Schrankensatz (Satz 2.52)

$$\|f(y) - f(y')\| \leq L\|y - y'\|.$$

Ist M eine obere Schranke für $\|f\|$ auf K (Satz 1.57), so gilt

$$\|x_n - y_n\| < \frac{1}{n}\|f(x_n) - f(y_n)\| \leq \frac{2M}{n}.$$

Also gibt es ein N , so daß $x_n, y_n \in \overline{B_r(x)}$ für alle $n \geq N$. Das ist ein Widerspruch, denn dann gilt für alle $n \geq N$

$$n\|x_n - y_n\| < \|f(x_n) - f(y_n)\| \leq L\|x_n - y_n\|$$

und somit $n < L$. \square

2.5.1 Satz von Taylor und Lokale Extremwerte

Definition 2.55. Sind V, W endlichdimensionale reelle Vektorräume und ist $f: U \subset V \rightarrow W$ eine Abbildung auf einer offenen Menge $U \subset V$, die n -mal differenzierbar im Punkt $p \in U$ ist, so nennt man

$$x \in V \mapsto \sum_{k=0}^n \frac{1}{k!} D^k f|_p(\underbrace{x-p, \dots, x-p}_{k \text{ mal}}) = f(p) + Df|_p(x-p) + \frac{1}{2} D^2 f|_p(x-p, x-p) + \frac{1}{3!} D^3 f|_p(x-p, x-p, x-p) + \dots + \frac{1}{n!} D^n f|_p(\underbrace{x-p, \dots, x-p}_{n \text{ mal}})$$

die *Taylorapproximation n . Ordnung von f im Punkt p* .

Die Taylorapproximation n . Ordnung in p ist die polynomiale Abbildung von Ordnung n , die f in p "am besten approximiert" im folgenden Sinne.

Satz 2.56. (*Satz von Taylor*) Sind V, W endlichdimensionale reelle Vektorräume und ist $f: U \subset V \rightarrow W$ eine n -mal differenzierbare Abbildung auf einer offenen Menge $U \subset V$. Dann gilt für $p, x \in U$

$$f(x) = f(p) + Df|_p(x-p) + \frac{1}{2} D^2 f|_p(x-p, x-p) + \dots + \frac{1}{n!} D^n f|_p(\underbrace{x-p, \dots, x-p}_{n \text{ mal}}) + R(x)$$

mit $\frac{R(x)}{\|x-p\|^n} \rightarrow 0$ für $x \rightarrow p$.

Beweis. (Per Induktion nach n .) Der Fall $n = 1$ gilt nach Definition der Differenzierbarkeit. Wir zeigen die Aussage für n unter der Induktionsannahme, daß die Aussage für $n - 1$ gezeigt ist.

Dazu berechnen wir die Ableitung nach x des Restglieds

$$R(x) = f(x) - \sum_{k=0}^n \frac{1}{k!} D^k f|_p(\underbrace{x-p, \dots, x-p}_{k \text{ mal}})$$

der Taylorapproximation n . Ordnung von f im Punkt p mittels Produktregel:

$$DR|_x(v) = Df|_x(v) - \sum_{k=1}^n \frac{1}{k!} k D^k f|_p(\underbrace{x-p, \dots, x-p, v}_{(k-1) \text{ mal}})$$

für alle $v \in V$. Nach Definition der höheren Ableitungen ist

$$DR|_x(v) = Df|_x(v) - \sum_{k=1}^n \frac{1}{(k-1)!} D^{k-1}(Df)|_p(\underbrace{x-p, \dots, x-p}_{(k-1) \text{ mal}})(v)$$

und somit

$$DR_{|x} = Df_{|x} - \sum_{k=0}^{n-1} \frac{1}{k!} D^k(Df)_{|p} \underbrace{(x-p, \dots, x-p)}_{k \text{ mal}}.$$

Damit ist $DR_{|x}$ gleich dem Restglied der Taylorapproximation $(n-1)$. Ordnung von Df_x im Punkt p , so daß es per Induktionsannahme für $\epsilon > 0$ ein $\delta > 0$ gibt mit

$$\|DR_{|x}\| \leq \epsilon \|x-p\|^{n-1}$$

für $x \in B_\delta(p) \subset U$. Wegen $R(p) = 0$ folgt damit für $x \in B_\delta(p)$ aus dem Schrankeinsatz (Satz 2.52)

$$\|R(x)\| = \|R(x) - R(p)\| \leq \left(\sup_{y \in B_\delta(p)} \|DR_y\| \right) \|x-p\| \leq \epsilon \|x-p\|^n.$$

□

Satz 2.57. (Lagrange-Form des Restglieds) Für eine $(n+1)$ -mal differenzierbare reellwertige Funktion $f: U \subset V \rightarrow \mathbb{R}$ gibt es für p, x mit

$$\overline{px} = \{(1-\lambda)p + \lambda x \mid \lambda \in [0, 1]\} \subset U$$

ein $q \in \overline{px}$, so daß

$$R(x) = \frac{1}{(n+1)!} D^{n+1} f_{|q} \underbrace{(x-p, \dots, x-p)}_{n+1 \text{ mal}}.$$

Beweis. Folgt mit der Kettenregel direkt aus dem Satz von Taylor mit Lagrange-Form des Restglieds aus Analysis I (siehe Dirk Ferus, *Analysis I* für eine Version, die ohne Stetigkeit der $(n+1)$ -ten Ableitung auskommt) angewandt auf

$$g:]-\epsilon, 1+\epsilon[\rightarrow \mathbb{R} \quad t \mapsto f(p+tv)$$

mit $v = x-p$.

□

Definition 2.58. Wie nennen eine symmetrische k -lineare Form

$$\alpha \in L^k(V, \mathbb{R}) = \left\{ \alpha: \underbrace{V \times \dots \times V}_{k \text{ mal}} \rightarrow \mathbb{R} \mid k\text{-linear} \right\}$$

auf einem endlichdimensionalen Vektorraum V

- *positiv definit* $\Leftrightarrow \alpha(v, \dots, v) > 0$ für alle $v \in V \setminus \{0\}$,
- *positiv semidefinit* $\Leftrightarrow \alpha(v, \dots, v) \geq 0$ für alle $v \in V \setminus \{0\}$,
- *negativ definit* $\Leftrightarrow \alpha(v, \dots, v) < 0$ für alle $v \in V \setminus \{0\}$,

- *negativ semidefinit* $\Leftrightarrow \alpha(v, \dots, v) \leq 0$ für alle $v \in V \setminus \{0\}$,
- *indefinit* \Leftrightarrow es gibt sowohl $v \in V \setminus \{0\}$ mit $\alpha(v, \dots, v) > 0$ also auch mit $\alpha(v, \dots, v) < 0$.

Das folgende Lemma zeigt, daß jede nicht-triviale symmetrische k -lineare Form α von genau einem der obigen Typen ist:

Lemma 2.59. *Ist $\alpha \in L^k(V, \mathbb{R})$ eine symmetrische k -lineare Form, so gilt*

i) $\alpha(v, \dots, v) = 0$ für alle $v \in V \Rightarrow \alpha = 0$ und

ii) $\alpha \neq 0$ und k ungerade $\Rightarrow \alpha$ indefinit.

Beweis. *i)* (Per Induktion) Für $k = 1$ gilt die Behauptung. Sei also die Behauptung für $k - 1$ gezeigt. Ist $\alpha \in L^k(V, \mathbb{R})$ und gilt $\alpha(v, \dots, v) = 0$ für alle $v \in V$, dann gilt für alle $v, w_1 \in V$ und $t \in \mathbb{R}$

$$0 = \alpha(v + tw_1, \dots, v + tw_1) = \sum_{i=1}^k \binom{k}{i} t^i \alpha(\underbrace{w_1, \dots, w_1}_i, \underbrace{v, \dots, v}_{k-i}).$$

Insbesondere ist $\alpha(w_1, \underbrace{v, \dots, v}_{k-1}) = 0$ und per Induktionsannahme ist

$$\alpha(w_1, \dots, w_k) = 0$$

für alle w_1, \dots, w_k .

ii) ist $l(v, \dots, v) \neq 0$, so ist $l(-v, \dots, -v) = -l(v, \dots, v)$. □

Definition 2.60. Eine reellwertige Funktion $f: U \subset V \rightarrow \mathbb{R}$ auf einer offenen Teilmenge $U \subset V$ eines endlichdimensionalen reellen Vektorraums hat in $p \in U$ ein *lokales Maximum (Minimum)*, falls es ein $\epsilon > 0$ gibt mit

$$f(p) \geq f(x) \quad (f(p) \leq f(x)) \quad \text{für alle } x \in U \text{ mit } \|x - p\| < \epsilon.$$

Sie hat ein *isoliertes lokales Maximum (Minimum)*, falls es ein $\epsilon > 0$ gibt mit

$$f(p) > f(x) \quad (f(p) < f(x)) \quad \text{für alle } x \in U \text{ mit } 0 < \|x - p\| < \epsilon.$$

Satz 2.61. (*Notwendiges Bedingung für lokale Extrema*) Ist eine reellwertige Funktion $f: U \subset V \rightarrow \mathbb{R}$ auf einer offenen Teilmenge $U \subset V$ eines endlichdimensionalen reellen Vektorraums im Punkt $p \in U$ differenzierbar und hat dort ein lokales Maximum oder Minimum, so ist

$$Df|_p = 0.$$

Beweis. Wir betrachten den Fall eines lokalen Minimums. Für alle $v \in V$ ist dann

$$Df|_p(v) = \lim_{t \rightarrow 0} \frac{f(p + tv) - f(p)}{t} = 0,$$

denn der linksseitige Grenzwert ist ≤ 0 und der rechtsseite Grenzwert ist ≥ 0 . \square

Einen Punkt $p \in U$ mit $Df|_p = 0$ nennt man auch einen *kritischen Punkt* der Funktion f .

Satz 2.62. (*Notwendiges und hinreichendes Kriterium für lokale Extrema*) Ist für eine k -mal differenzierbare reellwertige Funktion $f: U \subset V \rightarrow \mathbb{R}$ auf einer offenen Teilmenge $U \subset V$ eines endlichdimensionalen reellen Vektorraums im Punkt $p \in U$

$$Df|_p = 0, \dots, D^{k-1}f|_p = 0 \quad \text{und} \quad D^k f|_p \neq 0,$$

so gilt

- i) ist $D^k f|_p$ positiv definit, so hat f in p ein isoliertes lokales Minimum,
- ii) ist $D^k f|_p$ negativ definit, so hat f in p ein isoliertes lokales Maximum, und
- iii) ist $D^k f|_p$ indefinit, so hat f in p kein lokales Extremum. (Das ist zum Beispiel immer der Fall, wenn k ungerade ist.)

Wir beweisen den Satz nach Beispiel 2.65 unten.

Bemerkung 2.63. • Teil iii) von Satz 2.62 beinhaltet im Spezialfall $k = 1$ das obige notwendige Kriterium (Satz 2.61).

- Gilt in Teil iii) des Satzes $k > 1$, so spricht man von einem *Sattelpunkt*. (Allgemein ist ein *Sattelpunkt* ein kritischer Punkt, der kein lokales Extremum ist; geometrisch bedeutet dies, daß der Graph der Funktion in p eine horizontale Tangentialebene (siehe auch Beispiel 4.13) besitzt und die Funktion in jeder Umgebung von p Werte auf beiden Seiten der Tangentialebene annimmt.)
- Der Satz macht keine Aussage über semidefinite Punkte (über die man, wie das folgende Beispiel 2.65 zeigt, auch nicht so einfach allgemeingültige Aussagen treffen kann).
- Der Satz wird am häufigsten im Spezialfall $k = 2$ angewandt (und oft auch nur für diesen Fall formuliert), also für den Fall, daß

$$Df|_p = 0 \quad \text{und} \quad D^2f|_p \neq 0.$$

Definition 2.64. Ist eine reellwertige Funktion $f: U \subset \mathbb{R}^n \rightarrow \mathbb{R}$ auf einer offenen Menge $U \subset \mathbb{R}^n$ zweimal differenzierbar, so ist ihre *Hessesche Matrix* in $p \in U$

$$Hess(f)(p) = \left(\frac{\partial^2 f}{\partial x_i \partial x_j}(p) \right)_{i,j=1,\dots,n} = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1}(p) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(p) \\ \vdots & & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(p) & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_n}(p) \end{pmatrix}.$$

Die Hessesche Matrix ist die Koordinatendarstellung der zweiten Ableitung, für $u, v \in \mathbb{R}^n$ gilt

$$D^2f|_p(u, v) = \sum_{i,j=1}^n \frac{\partial^2 f}{\partial x_i \partial x_j}(p) u_i v_j = \langle u, Hessf(p)v \rangle = \langle Hessf(p)u, v \rangle$$

und damit symmetrisch wegen des Satzes von Schwarz (Satz 2.43). Die Hessesche Matrix ist die Ableitung des Gradienten von f .

Beispiel 2.65. Im Folgenden diskutieren wir zweimal differenzierbare Funktionen $f: \mathbb{R}^2 \rightarrow \mathbb{R}$, für die in $p = (0, 0)$ gilt

$$Df|_p = 0 \quad \text{und} \quad D^2f|_p \neq 0.$$

- Für $f(x, y) = x^2 + y^2$ ist $Hess(f)(0, 0) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$ und $D^2f|_{(0,0)}$ ist positiv definit. Wie man leicht direkt sieht, ist $p = (0, 0)$ ein isoliertes lokales Minimum (was auch aus Teil i) von Satz 2.62 folgt).

- Für

$$- f_1(x, y) = x^2$$

$$- f_2(x, y) = x^2 + y^4$$

$$- f_3(x, y) = x^2 - y^4$$

ist jeweils $Hess(f_i)(0, 0) = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}$, $i = 1, \dots, 3$ und die entsprechende zweite Ableitung ist positiv semidefinit. Wie man leicht sieht, ist $p = (0, 0)$

- ein lokales Minimum von f_1 ,
- ein isoliertes lokales Minimum von f_2 und
- kein lokales Extremum, sondern ein Sattelpunkt, von f_3 .

(Der Satz 2.62 macht über keinen der Fälle eine Aussage.)

- Die Funktionen $-f$ bzw. $-f_1$, $-f_2$ und $-f_3$ sind Beispiele mit negativ definiten bzw. semidefiniten zweiter Ableitung in $p = (0, 0)$.

- Für $g(x, y) = x^2 - y^2$ ist $Hess(g)(0, 0) = \begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix}$ und $D^2g|_{(0,0)}$ ist indefinit.

Wie man leicht direkt sieht, ist $p = (0, 0)$ kein lokales Extremum, sondern ein Sattelpunkt (was auch aus Teil iii) von Satz 2.62 folgt).

Beweis von Satz 2.62. Nach dem Satz von Taylor (Satz 2.56) gilt für $x \in U$

$$f(x) = f(p) + \frac{1}{k!} D^k f|_p(x-p, \dots, x-p) + R(x)$$

mit $\frac{R(x)}{\|x-p\|^k} \rightarrow 0$ für $x \rightarrow p$. Mit anderen Worten gibt es für $\epsilon > 0$ ein $\delta > 0$ mit

$$-\epsilon \|x-p\|^k \leq R(x) \leq \epsilon \|x-p\|^k \quad \text{für alle } x \in B_\delta(p). \quad (*)$$

Andererseits nimmt, da $S = \{v \in V \mid \|v\| = 1\}$ wegen des Satzes von Heine-Borel (Satz 1.49) kompakt ist, die Funktion $v \in S \mapsto \frac{1}{k!} D^k f|_p(v, \dots, v)$ auf S ihr Minimum m und Maximum M an, so daß gilt

$$m \|v\|^k \leq \frac{1}{k!} D^k f|_p(v, \dots, v) \leq M \|v\|^k \quad \text{für alle } v \in V.$$

i) ist $D^k f|_p$ positiv definit, so ist $m > 0$. Wählen wir $\delta > 0$ oben zu $\epsilon = \frac{m}{2}$, so gilt für $x \in B_\delta(p)$

$$\begin{aligned} f(x) - f(p) &= \frac{1}{k!} D^k f|_p(x-p, \dots, x-p) + R(x) \geq \\ &\geq m \|x-p\|^k - \epsilon \|x-p\|^k = \frac{m}{2} \|x-p\|^k. \end{aligned}$$

ii) folgt, indem man i) auf $-f$ anwendet.

iii) ist $D^k f|_p$ indefinit, so ist $m < 0 < M$. Weiter gibt es Strahlen $l = v_0 \mathbb{R}_{>0}$ und $L = w_0 \mathbb{R}_{>0} \subset V$ (wobei $v_0, w_0 \in S$ Punkte sind, an denen m bzw. M angenommen wird), für die

$$\frac{1}{k!} D^k f|_p(v, \dots, v) = m \|v\|^k \quad \text{für alle } v \in l \text{ und}$$

$$\frac{1}{k!} D^k f|_p(w, \dots, w) = M \|w\|^k \quad \text{für alle } w \in L.$$

Wählen wir also $v \in l$ klein genug, so daß $(*)$ für $x = p + v$ mit $\epsilon = \frac{|m|}{2}$ gilt, so folgt

$$f(p+v) - f(p) \leq m \|v\|^k + \epsilon \|v\|^k = -\frac{|m|}{2} \|v\|^k.$$

Und wählen wir $w \in L$ klein genug, so daß $(*)$ für $x = p + w$ mit $\epsilon = \frac{M}{2}$ gilt, so folgt

$$f(p+w) - f(p) \geq M \|w\|^k - \epsilon \|w\|^k = \frac{M}{2} \|w\|^k.$$

(Die Bemerkung über ungerades k folgt aus Teil ii) von Lemma 2.59.) □

2.5.2 Analyse der Hesseschen Matrix

Will man Satz 2.62 im Spezialfall $k = 2$, also dem Fall, daß

$$Df|_p = 0 \quad \text{und} \quad D^2f|_p \neq 0.$$

anwenden, sind folgende Kriterien nützlich. Sei $f: U \subset \mathbb{R}^n \rightarrow \mathbb{R}$ auf einer offenen Menge $U \subset \mathbb{R}^n$ zweimal differenzierbar. Dann ist die darstellende Matrix

$$A = \text{Hess}(f)(p) = \left(\frac{\partial^2 f}{\partial x_i \partial x_j}(p) \right)_{i,j=1,\dots,n} = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1}(p) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(p) \\ \vdots & & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(p) & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_n}(p) \end{pmatrix}$$

(Definition 2.5.1) der symmetrischen Bilinearform $D^2f|_p$, also die Matrix A , so daß $D^2f|_p(u, v) = \langle u, Av \rangle$ für alle $u, v \in \mathbb{R}^n$, symmetrisch und somit, wie aus der linearen Algebra bekannt, diagonalisierbar über \mathbb{R} .

Satz 2.66. *Ist A eine reelle symmetrische $n \times n$ -Matrix und $\alpha(u, v) = \langle u, Av \rangle$, $u, v \in \mathbb{R}^n$ die dazugehörige symmetrische Bilinearform, so gilt*

- α ist positiv definit \Leftrightarrow alle Eigenwerte von A sind > 0 ,
- α ist positiv semidefinit \Leftrightarrow alle Eigenwerte von A sind ≥ 0 ,
- α ist negativ definit \Leftrightarrow alle Eigenwerte von A sind < 0 ,
- α ist negativ semidefinit \Leftrightarrow alle Eigenwerte von A sind ≤ 0 und
- α ist indefinit \Leftrightarrow A hat positive und negative Eigenwerte.

Beweis. Ist v_1, \dots, v_n eine Orthonormalbasis von Eigenvektoren zu den Eigenwerten $\lambda_1, \dots, \lambda_n \in \mathbb{R}$, so gilt für $v = \sum_{i=1}^n x_i v_i \in \mathbb{R}^n$

$$\alpha(v, v) = \langle v, Av \rangle = \sum_{i=1}^n \lambda_i x_i^2.$$

□

Beispiel 2.67. Ist A eine symmetrische 2×2 -Matrix, so ist die Determinante $\det(A) = \lambda_1 \lambda_2$ gleich dem Produkt der beiden Eigenwerte λ_1, λ_2 und für die dazugehörige symmetrische Bilinearform $\alpha(u, v) = \langle u, Av \rangle$ gilt

- α ist (positiv oder negativ) definit $\Leftrightarrow \det(A) > 0$,
- α ist (positiv oder negativ) semidefinit, aber nicht definit $\Leftrightarrow \det(A) = 0$ und

- α ist indefinit $\Leftrightarrow \det(A) < 0$.

Ohne Beweis (vgl. Anhang zu Dirk Ferus, *Analysis II*) geben wir das folgende Kriterium an:

Satz 2.68. (*Hauptminorenkriterium*) Alle Eigenwerte einer symmetrischen reellen $n \times n$ -Matrix $A = (a_{ij})_{i,j=1,\dots,n}$ sind positiv $\Leftrightarrow \det(A_k) > 0$ für alle $k = 1, \dots, n$ und $A_k = (a_{ij})_{i,j=1,\dots,k}$.

Beispiel 2.69. Wir beweisen das Hauptminorenkriterium für eine symmetrische 2×2 -Matrix

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

Die zu A gehörige symmetrische Bilinearform ist genau dann positiv definit, wenn

$$\left\langle \begin{pmatrix} 1 \\ 0 \end{pmatrix}, A \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\rangle = a > 0 \quad \text{und} \quad \left\langle \begin{pmatrix} x \\ 1 \end{pmatrix}, A \begin{pmatrix} x \\ 1 \end{pmatrix} \right\rangle = ax^2 + 2bx + c > 0$$

für alle $x \in \mathbb{R}$. Letzteres ist äquivalent dazu, daß $c > 0$ und $ax^2 + 2bx + c$ keine reellen Nullstellen hat, was wegen der "pq-Formel" (und $a > 0$) äquivalent zu $ac - b^2 = \det(A) > 0$ ist.

Mit dem vorhergehenden Beispiel erhält man aus Satz 2.62 das folgende Kriterium:

Satz 2.70. Ist $f: U \subset \mathbb{R}^2 \rightarrow \mathbb{R}$ zweimal differenzierbar auf einer offenen Menge $U \subset \mathbb{R}^2$. Dann gilt

- hat f in p ein lokales Extremum, so ist $\frac{\partial f}{\partial x}(p) = \frac{\partial f}{\partial y}(p) = 0$,
- gilt $\frac{\partial f}{\partial x}(p) = \frac{\partial f}{\partial y}(p) = 0$ und

- $\frac{\partial^2 f}{\partial x \partial x}(p) \frac{\partial^2 f}{\partial y \partial y}(p) - \left(\frac{\partial^2 f}{\partial x \partial y}(p) \right)^2 > 0$ und $\frac{\partial^2 f}{\partial x \partial x}(p) > 0 \Rightarrow f$ hat in p ein isoliertes lokales Minimum,
- $\frac{\partial^2 f}{\partial x \partial x}(p) \frac{\partial^2 f}{\partial y \partial y}(p) - \left(\frac{\partial^2 f}{\partial x \partial y}(p) \right)^2 > 0$ und $\frac{\partial^2 f}{\partial x \partial x}(p) < 0 \Rightarrow f$ hat in p ein isoliertes lokales Maximum und
- $\frac{\partial^2 f}{\partial x \partial x}(p) \frac{\partial^2 f}{\partial y \partial y}(p) - \left(\frac{\partial^2 f}{\partial x \partial y}(p) \right)^2 < 0 \Rightarrow f$ hat in p kein Extremum.

2.6 Satz über implizite Funktionen und Satz über die Umkehrabbildung

Vorbetrachtung: Im vorhergehenden Abschnitt haben wir kritische Punkte von differenzierbaren Funktionen $f: U \subset V \rightarrow \mathbb{R}$ auf einer offenen Menge $U \subset V$ in

einem endlichdimensionalen euklidischen Vektorraum V analysiert, also Punkte $p \in U$, in denen gilt $Df|_p = 0$.

Im Folgenden geht es um *reguläre* Punkte. Das sind Punkte $p \in U$, in denen $Df|_p \neq 0$ gilt. Der Satz über implizite Funktionen besagt, daß für eine stetig differenzierbare Funktion $f: U \subset V \rightarrow \mathbb{R}$ die Lösungsmenge der Gleichung

$$f(x) = y$$

mit $y = f(p)$ (also die *Niveauhypersfläche* von f bestehend aus den Punkten $x \in U$, an denen der Funktionswert gleich dem an der Stelle p ist) in einer Umgebung von p und in geeigneten Koordinaten als Graph einer stetig differenzierbaren Funktion von $n - 1$ Variablen geschrieben werden kann, wobei $n = \dim(V)$. Insbesondere sind für eine Funktion von $n = 2$ Variablen die *Niveaulinien* dann Kurven, die man lokal als Graph einer reellwertigen Funktion von einer Variablen schreiben kann.

Beispiel 2.71. Sei $f(x, y) = x^2 + y^2$. Dann ist $\frac{\partial f}{\partial x}(x, y) = 2x$ und $\frac{\partial f}{\partial y}(x, y) = 2y$, so daß $\frac{\partial f}{\partial x}$ und $\frac{\partial f}{\partial y}$ entlang der y -Achse bzw. der x -Achse verschwinden. Der einzige kritische Punkt ist $(x, y) = (0, 0)$ (an dem ein isoliertes Minimum von f vorliegt). Die Niveaumenge zu $z = 0 = f(0, 0)$ ist ein Punkt, die Niveaumengen $f^{-1}(z)$ für $z > 0$ sind Kreise (von Radius $r = \sqrt{z}$, BILD!).

In der Umgebung eines regulären Punktes $(x_0, y_0) \neq 0$ kann man die Gleichung

$$x^2 + y^2 = f(x_0, y_0) = z$$

immer nach x oder y auflösen und Lösungen schreiben als

$$x = \pm \sqrt{z - y^2} \quad \text{bzw.} \quad y = \pm \sqrt{z - x^2},$$

wobei die Funktionen $y \mapsto \sqrt{z - y^2}$ und $x \mapsto \sqrt{z - x^2}$ nur für $y^2 < z$ bzw. $x^2 < z$ stetig differenzierbar sind, also weg von den Punkten, wo x und damit $\frac{\partial f}{\partial x}$ bzw. y und damit $\frac{\partial f}{\partial y}$ gleich Null werden.

Der Satz über implizierte Funktionen besagt, daß man für eine beliebige stetig differenzierbare Funktion $f(x, y)$ in einem Punkt (x_0, y_0) mit $\frac{\partial f}{\partial x}(x_0, y_0) \neq 0$ die Gleichung

$$f(x, y) = z$$

mit $z = f(x_0, y_0)$ lokal um (x_0, y_0) "auflösen kann nach x " und die Lösungsmenge durch eine Funktion $x = g(y)$ schreiben kann (im obigen Beispiel durch eine der beiden Funktionen $g(y) = \pm \sqrt{z - y^2}$). Analog kann man an einem Punkt (x_0, y_0) mit $\frac{\partial f}{\partial y}(x_0, y_0) \neq 0$ die Gleichung

$$f(x, y) = z$$

lokal "nach y auflösen" und die Lösungsmenge durch eine Funktion $y = g(x)$ schreiben (im obigen Beispiel durch eine der beiden Funktionen $g(x) = \pm \sqrt{z - x^2}$).

Satz 2.72. (Satz über implizite Funktionen) Seien V_1, V_2 und W endlichdimensionale reelle Vektorräume und $f: U \subset V_1 \times V_2 \rightarrow W$ eine C^k -Abbildung auf einer offenen Menge $U \subset V_1 \times V_2$. Ist in einem Punkt $(p, q) \in U$ die lineare Abbildung

$$D_2f|_{(p,q)}: V_2 \rightarrow W \quad \text{invertierbar,}$$

wobei $D_2f|_{(p,q)}(v_2) = Df|_{(p,q)}(0, v_2)$, und gilt $f(p, q) = 0$, so gibt es offene Mengen $U_i \subset V_i$, $i = 1, 2$ mit $(p, q) \in U_1 \times U_2 \subset U$ und eine C^k -Abbildung $g: U_1 \rightarrow U_2$ mit der Eigenschaft, daß für $(x, y) \in U_1 \times U_2$ gilt

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

Insbesondere ist die lineare Abbildung $D_2f|_{(x,g(x))}: V_2 \rightarrow W$ für alle $x \in U_1$ invertierbar und

$$Dg|_x = -(D_2f|_{(x,g(x))})^{-1} \circ D_1f|_{(x,g(x))},$$

wobei $D_1f|_{(p,q)}(v_1) = Df|_{(p,q)}(v_1, 0)$.

Wir zeigen später (in Abschnitt 4), daß der Satz über implizite Funktionen aus dem (folgenden) Satz über die Umkehrabbildung (Satz 2.75) folgt.

Bemerkung 2.73. Daß (*) für alle $(x, y) \in U_1 \times U_2$ gilt, ist gleichbedeutend dazu, daß die Gleichung $f(x, y) = 0$ für jedes $x \in U_1$ genau eine Lösung $y \in U_2$ hat. Die Abbildung $g: U_1 \rightarrow U_2$ ordnet jedem $x \in U_1$ genau diese eindeutige Lösung $y \in U_2$ zu. (Man sagt g sei "implizit definiert", da man erst eine Gleichung lösen muß um g zu bestimmen.)

Beispiel 2.74. Sei $f: \mathbb{R}^2 \rightarrow \mathbb{R}$, $(x, y) \mapsto x - y^2$. Die Nullstellenmenge

$$M = \{(x, y) \in \mathbb{R}^2 \mid f(x, y) = 0\}$$

ist dann eine Parabel (BILD!). Man kann den Teil der Nullstellenmenge in der Halbebene $y > 0$ darstellen als Graph

$$M \cap \{y > 0\} = \{(x, g(x)) \mid x \in \mathbb{R}_{>0}\}$$

der C^∞ -Funktion $g: \mathbb{R}_{>0} \rightarrow \mathbb{R}$, $x \mapsto \sqrt{x}$. Der Graph von $\tilde{g}: \mathbb{R}_{>0} \rightarrow \mathbb{R}$, $x \mapsto -\sqrt{x}$ ist analog gleich dem Teil der Nullstellenmenge in der Halbebene $y < 0$, also

$$M \cap \{y < 0\} = \{(x, \tilde{g}(x)) \mid x \in \mathbb{R}_{>0}\}.$$

(Der Punkt $(0, 0) \in M$ hat keine Umgebung, in welcher M der Graph einer Funktion von x ist, da es in jeder Umgebung für kleine $x > 0$ zwei mögliche y -Werte gibt für die gilt $(x, y) \in M$, nämlich $y = +\sqrt{x}$ und $y = -\sqrt{x}$.)

Der Satz über implizite Funktionen bestätigt genau das beobachtete Verhalten, den er besagt, daß für alle Punkte

$$(x, y) \in M \quad \text{mit} \quad \frac{\partial f}{\partial y}(x, y) = -2y \neq 0$$

die Nullstellenmenge M lokal als Graph einer Funktion $y = g(x)$ geschrieben werden kann. Die Formel für die Ableitung von g ergibt für $g(x) = \sqrt{x}$ (also den oberen Zweig der Parabel)

$$Dg|_x = - \left(\frac{\partial f}{\partial y}(x, y) \right)^{-1} \frac{\partial f}{\partial x}(x, y) = \frac{1}{2\sqrt{x}}.$$

(Im Punkt $(0, 0) \in M$, in dem M nicht lokal als Graph einer Funktion $y = g(x)$ geschrieben werden kann, gilt $\frac{\partial f}{\partial y}(0, 0) = 0$. Da $\frac{\partial f}{\partial x}(x, y) = 1$ für alle (x, y) impliziert der Satz über implizite Funktionen auch, daß man die Nullstellenmenge überall lokal als Graph einer Funktion $x = g(y)$ schreiben kann. Diese Tatsache ist wenig überraschend, denn $M = \{(x, y) \in \mathbb{R}^2 \mid x = y^2\}$ ist sogar global der Graph der Funktion $x = g(y) = y^2$.)

Satz 2.75. (Satz über die Umkehrabbildung) Seien V, W endlichdimensionale reelle Vektorräume und sei $f: U \subset V \rightarrow W$ eine C^k -Abbildung auf einer offenen Menge $U \subset V$. Ist für $p \in U$ die lineare Abbildung

$$Df|_p: V \rightarrow W$$

invertierbar, so gibt es offene Mengen $O \subset U$ und $\tilde{O} \subset W$ mit $p \in O$, so daß die Einschränkung $f|_O$ von f auf O eine bijektive Abbildung

$$f|_O: O \rightarrow \tilde{O}$$

ist und $f|_O^{-1}: \tilde{O} \rightarrow O$ auch C^k ist. Für $p \in O$ und $q = f(p) \in \tilde{O}$ gilt dann

$$D(f^{-1})|_q = (Df|_p)^{-1}.$$

Der Satz über die Umkehrabbildung wird am Ende der Vorlesung (in Abschnitt 4) gezeigt.

Beispiel 2.76. Eine C^k -Funktion $f: I \subset \mathbb{R} \rightarrow \mathbb{R}$ hat in einem Punkt $t \in I$ mit $f'(t) \neq 0$ lokal eine Inverse, die wieder eine C^k -Funktion ist. Für die Ableitung der lokalen Umkehrfunktion in $y = f(x)$ gilt dann

$$(f^{-1})'(y) = \frac{1}{f'(x)}.$$

Zum Beispiel hat $f(t) = t^2$ überall außer im Punkt $t = 0$ eine lokale Inverse, denn $f'(t) = 2t \neq 0$ für $t \neq 0$. (Für $t > 0$ ist die Inverse $f^{-1}(s) = \sqrt{s}$ und für $t < 0$ ist entsprechend $f^{-1}(s) = -\sqrt{s}$. BILD!)

Definition 2.77. Seien $U \subset V$ und $\tilde{U} \subset W$ offen Teilmengen von endlichdimensionalen reellen Vektorräumen V bzw. W . Eine bijektive C^k -Abbildung

$$f: U \rightarrow \tilde{U},$$

deren Inverse f^{-1} ebenfalls C^k ist, bezeichnet man als C^k -Diffeomorphismus.

Der Satz über die Umkehrabbildung besagt, daß eine C^k -Abbildung $f: U \subset V \rightarrow W$ um jedem Punkt $p \in U$, in dem $Df|_p$ invertierbar ist, lokal – also eingeschränkt auf passende Umgebungen – ein Diffeomorphismus ist. Eine wichtige Konsequenz aus dem Satz über die Umkehrabbildung ist:

Korollar 2.78. *Eine C^k -Abbildung $f: U \subset V \rightarrow W$, für die $Df|_p$ in jedem Punkt $p \in U$ invertierbar ist, ist offen (d.h. bildet offene Menge wieder auf eine offene Menge ab). Ist f außerdem noch injektiv, so ist f eine C^k -Diffeomorphismus*

$$f: U \rightarrow f(U).$$

Beispiel 2.79. Diffeomorphismen zwischen offenen Teilmengen des \mathbb{R}^n kann man sich vorstellen als nicht-lineare Koordinatenwechsel.

- 1.) (Polarkoordinaten auf $U = \mathbb{R}^2 \setminus \{(x, 0) \mid x \leq 0\}$) Die bijektive C^∞ -Abbildung

$$\Phi: \mathbb{R}_{>0} \times]-\pi, \pi[\rightarrow U \subset \mathbb{R}^2 \quad \begin{pmatrix} r \\ \varphi \end{pmatrix} \mapsto \begin{pmatrix} r \cos(\varphi) \\ r \sin(\varphi) \end{pmatrix}$$

(vgl. Blatt 7, Aufgabe 3) ist wegen Korollar 2.78 ein C^∞ -Diffeomorphismus, denn

$$D\Phi|_{(r,\varphi)} = \begin{pmatrix} \cos(\varphi) & -r \sin(\varphi) \\ \sin(\varphi) & r \cos(\varphi) \end{pmatrix}$$

ist in jedem Punkt invertierbar ($\det(D\Phi|_{(r,\varphi)}) = r \neq 0$).

Daß die Umkehrabbildung unendlich oft differenzierbar ist kann man hier auch (ohne den Satz über die Umkehrabbildung zu benutzen) der leicht zu bestimmenden expliziten Form der Umkehrabbildung ansehen, es ist nämlich

$$\Phi^{-1}(x, y) = \begin{cases} \left(\sqrt{x^2 + y^2}, \arccos\left(\frac{x}{\sqrt{x^2 + y^2}}\right) \right) & \text{für } y > 0 \\ \left(\sqrt{x^2 + y^2}, \arctan\left(\frac{y}{x}\right) \right) & \text{für } x > 0 \\ \left(\sqrt{x^2 + y^2}, -\arccos\left(\frac{x}{\sqrt{x^2 + y^2}}\right) \right) & \text{für } y < 0. \end{cases}$$

- 2.) (Zylinderkoordinaten auf $U = \mathbb{R}^3 \setminus \{(x, 0, z) \mid x \leq 0, z \in \mathbb{R}\}$) Die bijektive C^∞ -Abbildung

$$\Phi: \mathbb{R}_{>0} \times]-\pi, \pi[\times \mathbb{R} \rightarrow U \subset \mathbb{R}^3 \quad \begin{pmatrix} r \\ \varphi \\ h \end{pmatrix} \mapsto \begin{pmatrix} r \cos(\varphi) \\ r \sin(\varphi) \\ h \end{pmatrix}$$

ist wegen Korollar 2.78 ein C^∞ -Diffeomorphismus, denn

$$D\Phi|_{(r,\varphi,h)} = \begin{pmatrix} \cos(\varphi) & -r \sin(\varphi) & 0 \\ \sin(\varphi) & r \cos(\varphi) & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

ist in jedem Punkt invertierbar (auch hier gilt $\det(D\Phi|_{(r,\varphi,h)}) = r \neq 0$).

- 3.) (Sphärische Koordinaten auf $U = \mathbb{R}^3 \setminus \{(x, 0, z) \mid x \leq 0, z \in \mathbb{R}\}$) Die bijektive C^∞ -Abbildung

$$\Phi: \mathbb{R}_{>0} \times]-\pi, \pi[\times]-\frac{\pi}{2}, \frac{\pi}{2}[\rightarrow U \subset \mathbb{R}^3 \quad \begin{pmatrix} r \\ \varphi \\ \vartheta \end{pmatrix} \mapsto \begin{pmatrix} r \cos(\varphi) \cos(\vartheta) \\ r \sin(\varphi) \cos(\vartheta) \\ r \sin(\vartheta) \end{pmatrix}$$

ist wegen Korollar 2.78 ein C^∞ -Diffeomorphismus, denn

$$D\Phi|_{(r,\varphi,h)} = \begin{pmatrix} \cos(\varphi) \cos(\vartheta) & -r \sin(\varphi) \cos(\vartheta) & -r \cos(\varphi) \sin(\vartheta) \\ \sin(\varphi) \cos(\vartheta) & r \cos(\varphi) \cos(\vartheta) & -r \sin(\varphi) \sin(\vartheta) \\ \sin(\vartheta) & 0 & r \cos(\vartheta) \end{pmatrix}$$

ist in jedem Punkt invertierbar ($\det(D\Phi|_{(r,\varphi)}) = r^2 \cos(\vartheta) \neq 0$).

3 Gewöhnliche Differentialgleichungen (Kurzeinführung)

Definition 3.1. Eine (explizite) *gewöhnliche Differentialgleichung erster Ordnung* ist eine Gleichung der Form

$$\frac{dx}{dt}(t) = f(t, x(t)), \quad (*)$$

wobei $f: I \times U \rightarrow \mathbb{R}^n$ eine stetige Funktion ist auf dem Produkt eines Intervalls $I =]a, b[$ mit einer offenen Menge $U \subset \mathbb{R}^n$ ist. Eine *Lösung* oder *Integralkurve* ist eine differenzierbare Funktion $x: I' \rightarrow U$, die (*) für alle $t \in I' \subset I$ erfüllt. (Etwas allgemeiner werden wir später, in Abschnitt 3.2, den Fall betrachten, daß f definiert ist auf einer offenen Menge $U \subset \mathbb{R} \times \mathbb{R}^n$. Als Lösungen kommen dann Abbildungen $x: I \rightarrow \mathbb{R}^n$ in Frage, die $(t, x(t)) \in U$ für alle $t \in I$ erfüllen.)

Wir schreiben im Folgenden $\dot{x}(t)$ für $\frac{dx(t)}{dt}$.

Bemerkung 3.2. Für eine Lösung $x(t)$ von (*) setzen wir erstmal nur Differenzierbarkeit voraus (damit wir $x(t)$ überhaupt in (*) einsetzen können). Da aber f als stetig angenommen ist, ist jede Lösung automatisch C^1 . Per Induktion sieht man, daß allgemein für eine Lösung $x(t)$ einer Gleichung (*) gilt

$$f \in C^k \implies x \in C^{k+1}.$$

Beispiel 3.3. (Exponentielles Wachstum) In vielen Wachstums- und Zerfallsprozessen (zum Beispiel Verzinsung, radioaktiver Zerfall...) ist die Änderung $\dot{x}(t)$ einer Größe $x(t)$ in guter Näherung proportional zur Größe selbst. Das kann man modellieren durch die Differentialgleichung

$$\dot{x}(t) = \lambda x(t) \quad \text{mit } \lambda \in \mathbb{R}.$$

(Hier ist $f(t, x) = \lambda x$.) Die Funktionen $x(t) = e^{\lambda t} x_0$ mit $x_0 \in \mathbb{R}$ sind Lösungen dieser Gleichung zum ‘‘Anfangswert’’ $x(0) = x_0$.

Alle Lösungen dieser Differentialgleichung sind von der angegebenen Form, denn ist $x(t)$ eine beliebige Lösung, so hat $\tilde{x}(t) = e^{-\lambda t} x(t)$ verschwindende Ableitung und ist damit konstant. (Insbesondere ist die Exponentialfunktion $x(t) = e^t$ die eindeutige differenzierbare Funktion mit $\dot{x}(t) = x(t)$ und $x(0) = 1$.)

Bemerkung 3.4. • Eine Differentialgleichung, für die $f(t, x)$ (wie im vorhergehenden Beispiel 3.3) nicht explizit von t abhängt, also $f(t, x) = f(x)$ ist, bezeichnet man als *autonom*. Eine Differentialgleichung, bei der $f(t, x)$ von t abhängt, kann man umschreiben in eine autonome Gleichung

$$\dot{X}(t) = F(X(t)),$$

wobei $X(t) = \begin{pmatrix} x(t) \\ t \end{pmatrix}$ und $F: \tilde{U} = I \times U \rightarrow \mathbb{R}^{n+1}$ mit $F(X) = \begin{pmatrix} F(x) \\ 1 \end{pmatrix}$.

- Eine Differentialgleichung N -ter Ordnung

$$x^{(N)}(t) = f(t, x(t), \dot{x}(t), \dots, x^{(N-1)}(t))$$

kann man umschreiben in eine Gleichung erster Ordnung für

$$X(t) = (x(t), \dot{x}(t), \dots, x^{(N-1)}(t))$$

auf $\tilde{U} = I \times U \times \underbrace{\mathbb{R}^n \times \dots \times \mathbb{R}^n}_{N-1}$.

Zum Beispiel kann man die 2. Ordnung Gleichung (die sogenannte Schwingungsgleichung)

$$\ddot{x}(t) = -x(t)$$

auf \mathbb{R} (hier ist $f(t, x, \dot{x}) = -x$) umschreiben in die Gleichung

$$\dot{X}(t) = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} X(t)$$

für $X(t) = (x(t), \dot{x}(t))$.

Aufgrund dieser Bemerkung reicht es, Eindeutigkeits- und Existenzaussagen für autonome Differentialgleichungen erster Ordnung zu beweisen. Eine derartige Gleichung $\dot{x}(t) = f(x(t))$ mit $f: U \subset \mathbb{R}^n \rightarrow \mathbb{R}^n$ kann man sich vorstellen als ein *Vektorfeld* auf $U \subset \mathbb{R}^n$, d.h. eine Abbildung, die jedem Punkt $p \in U$ einen Vektor $f(p) \in \mathbb{R}^n$ zuordnet (BILD!). Eine Integralkurve $x(t)$ ist dann eine Kurve in U , deren Ableitung $\dot{x}(t)$ in jedem Punkt $p = x(t)$ genau gleich dem entsprechenden Vektor $f(p)$ ist.

Satz 3.5. (*Existenz- und Eindeutigkeitsatz*) Für stetiges $f: U \subset \mathbb{R}^n \rightarrow \mathbb{R}^n$ hat für jedes $x_0 \in U$ das Anfangswertproblem (AWP)

$$\dot{x}(t) = f(x(t)) \quad x(0) = x_0$$

eine Lösung $x:]a, b[\rightarrow U$ mit $0 \in]a, b[$. Ist f stetig differenzierbar (oder auch nur lokal Lipschitz stetig), hat das AWP lokal eine eindeutige Lösung, d.h. sind x und \tilde{x} Lösungen des AWP auf I bzw. \tilde{I} mit $x(0) = \tilde{x}(0)$, so gilt

$$x_{I \cap \tilde{I}} = \tilde{x}_{I \cap \tilde{I}}.$$

Für lokal Lipschitz stetiges f werden wir den Satz in Abschnitt 3.2 (Satz 3.27) beweisen. (Der Existenzsatz für stetiges f wird oft als *Satz von Peano* bezeichnet und normalerweise in der *Numerik* Vorlesung bewiesen.)

Beispiel 3.6. • Das AWP $\dot{x}(t) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, $x(0) = \begin{pmatrix} x_0 \\ y_0 \end{pmatrix}$ hat die eindeutige

$$\text{Lösung } x(t) = \begin{pmatrix} x_0 + t \\ y_0 \end{pmatrix}.$$

- Das AWP $\dot{x}(t) = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} x(t)$, $x(0) = \begin{pmatrix} x_0 \\ y_0 \end{pmatrix}$ hat die eindeutige Lösung $x(t) = \begin{pmatrix} \cos(t) & -\sin(t) \\ \sin(t) & \cos(t) \end{pmatrix} \begin{pmatrix} x_0 \\ y_0 \end{pmatrix}$.
- Die der Schwingungsgleichung entsprechende Gleichung erster Ordnung (Bemerkung 3.4) führt auf das AWP $\dot{x}(t) = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} x(t)$, $x(0) = \begin{pmatrix} x_0 \\ y_0 \end{pmatrix}$, dessen eindeutige Lösung $x(t) = \begin{pmatrix} \cos(t) & \sin(t) \\ -\sin(t) & \cos(t) \end{pmatrix} \begin{pmatrix} x_0 \\ y_0 \end{pmatrix}$ ist.
- Auf die Differentialgleichung

$$\dot{x}(t) = \sqrt[3]{x^2(t)} = x^{2/3}(t)$$

läßt sich nur der Existenzteil des Satzes anwenden. Für das AWP $x(0) = 0$ gilt keine Eindeutigkeit der Lösung: neben $x(t) = 0$ ist auch für jedes $b > 0$ die Funktion

$$x(t) = \begin{cases} 0 & t \leq b \\ \frac{1}{27}(t-b)^3 & t > b \end{cases}$$

eine Lösung des AWP (BILD!).

Bemerkung 3.7. Für eine zeitabhängige Gleichung von Ordnung N erhält man aus Satz 3.5 durch die Umformungen in Bemerkung 3.4 für alle $x_0 \in U$ und $x_1, \dots, x_{N-1} \in \mathbb{R}^n$ die (eindeutige) lokale Lösbarkeit des AWP

$$x^{(N)}(t) = f(t, x(t), \dot{x}(t), \dots, x^{(N-1)}(t))$$

$$x(t_0) = x_0, \quad \dot{x}(t_0) = x_1, \quad \dots \quad x^{(N-1)}(t_0) = x_{N-1}.$$

Beispiel 3.8. $x(t) = \sin(t)$ ist die eindeutige Lösung des AWP

$$\ddot{x}(t) = -x(t) \quad x(0) = 0, \dot{x}(0) = 1$$

und $x(t) = \cos(t)$ ist die eindeutige Lösung des AWP

$$\ddot{x}(t) = -x(t) \quad x(0) = 1, \dot{x}(0) = 0.$$

(Diese beiden Lösungen entsprechen den Lösungen zu

$$x_0 = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad \text{bzw.} \quad x_0 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

im dritten Teil von Beispiel 3.6.)

Das explizite Lösen von gewöhnlichen Differentialgleichungen ist (ähnlich wie die explizite Integration von Funktionen) oft selbst dann schwierig, wenn Lösungen existieren, die durch elementare Funktionen ausgedrückt werden können. Eine umfangreiche Sammlung von Lösungsmethoden und expliziten Lösungen findet man in den Büchern des Tübinger Mathematikers *Erich Kamke*. (Findet man dort keinen Lösungsansatz, ist es wahrscheinlich ratsam, sich mit entsprechenden numerischen Verfahren zu beschäftigen.)

Beispiel 3.9. (Trennung der Variablen) Nicht-konstante Lösungen einer Differentialgleichung vom Typ

$$\dot{x}(t) = h(t)g(x(t))$$

mit $h \in C^0$ und $g \in C^1$ findet man durch Integration von

$$\frac{1}{g(x(t))} \frac{dx}{dt} = h(t)$$

(wobei das Teilen durch $g(x(t))$ hier problemlos ist, denn für eine nicht-konstante Lösung kann $g(x(t))$ nie Null werden). Integriert man von t_0 bis t gilt mit der Substitutionsregel

$$\int_{x(t_0)}^{x(t)} \frac{1}{g(x)} dx = \int_{t_0}^t \frac{\dot{x}(\tau)}{g(x(\tau))} d\tau = \int_{t_0}^t h(\tau) d\tau,$$

so daß man eine implizite Darstellung der Lösung $x(t)$ erhält.

Für die Gleichung $\dot{x} = \lambda x$ in Beispiel 3.3 ergibt diese Methode zum Beispiel

$$\log(x(t)/x(t_0)) = \lambda(t - t_0),$$

woraus man durch Exponenzieren die oben angegebene Lösung erhält.

3.1 Lineare Differentialgleichungen

Definition 3.10. Eine *lineare Differentialgleichung erster Ordnung* ist eine Differentialgleichung der Form

$$\dot{x}(t) = A(t)x(t) + b(t) \quad (*)$$

mit stetigen Abbildungen $A: I \rightarrow \text{Mat}_{n \times n}(\mathbb{R})$ und $b: I \rightarrow \mathbb{R}^n$ auf $I =]a, b[$. Ist $b = 0$, so nennt man die Gleichung *homogen*, sonst ist sie *inhomogen*.

Am Ende des Kapitels (in Abschnitt 3.2.3) zeigen wir folgenden Satz:

Satz 3.11. Eine lineare Differentialgleichung erster Ordnung $(*)$ hat für jedes $x_0 \in \mathbb{R}^n$ und $t_0 \in I$ eine eindeutige Lösung $x: I \rightarrow \mathbb{R}^n$ mit Anfangswert

$$x(t_0) = x_0.$$

Bemerkung 3.12. Der Unterschied zum Satz 3.5 für allgemeine Differentialgleichungen ist, daß hier die Existenz der Lösungen auf dem ganzen Intervall garantiert ist.

Für nicht-lineare Differentialgleichungen gibt es allgemein keine "globale" Existenz. Zum Beispiel hat das Anfangswertproblem

$$\dot{x}(t) = 1 + x^2(t) \quad x(0) = 0$$

die nur auf dem Intervall $] -\frac{\pi}{2}, \frac{\pi}{2}[$ existierende Lösung $x(t) = \tan(t)$ (wie man durch Einsetzen oder mit Trennung der Variablen (Beispiel 3.9) verifizieren kann).

3.1.1 Homogene lineare Differentialgleichungen

Korollar 3.13. Die Menge

$$L = \{x \in C^1(I, \mathbb{R}^n) \mid \dot{x}(t) = A(t)x(t) \text{ für alle } t \in I\}$$

der Lösungen einer homogenen linearen Gleichung $\dot{x}(t) = A(t)x(t)$ mit stetigem $A: I \rightarrow \text{Mat}_{n \times n}(\mathbb{R})$, $I =]a, b[$ ist ein n -dimensionaler Vektorraum. Für jedes $t_0 \in I$ ist die Auswertungsabbildung

$$x \in L \mapsto x(t_0) \in \mathbb{R}^n$$

ein Isomorphismus. Insbesondere gilt für $x \in L$

$$x(t_0) = 0 \quad \text{für ein } t_0 \in I \quad \implies \quad x(t) = 0 \quad \text{für alle } t \in I.$$

Beweis. Satz 3.11 impliziert direkt, daß

$$x \in L \mapsto x(t_0) \in \mathbb{R}^n$$

ein Isomorphismus ist. Insbesondere hat damit nur der Nullvektor in L , also die konstante Lösung $x(t) = 0$, eine Nullstelle in t_0 . Da die Aussage für alle $t_0 \in I$ gilt, haben nicht-triviale Lösungen keine Nullstellen. \square

Korollar 3.14. Wählt man eine Basis $x_1, \dots, x_n \in L$ des Vektorraumes L der Lösungen einer homogenen linearen Gleichung $\dot{x}(t) = A(t)x(t)$ mit stetigem $A: I \rightarrow \text{Mat}_{n \times n}(\mathbb{R})$, $I =]a, b[$, so ist $\Phi(t) = (x_1(t), \dots, x_n(t))$ eine matrixwertige Lösung $\Phi: I \rightarrow M_{n \times n}(\mathbb{R})$ der Gleichung, d.h.

$$\dot{\Phi}(t) = A(t)\Phi(t). \quad (*)$$

Für eine matrixwertige Lösung $\Phi: I \rightarrow M_{n \times n}(\mathbb{R})$ von $(*)$ gilt allgemein

$$\Phi(t_0) \text{ ist invertierbar für ein } t_0 \in I \implies \Phi(t) \text{ ist invertierbar für alle } t \in I.$$

Dies ist genau dann der Fall, wenn die Spalten $x_1(t), \dots, x_n(t)$ von $\Phi(t)$ eine Basis von L bilden.

Beweis. Ist Φ eine matrixwertige Lösung von $(*)$ und ist $\Phi(t_1)$ nicht invertierbar für ein $t_1 \in I$. Dann gibt es einen Vektor $x_1 \in \mathbb{R}^n \setminus \{0\}$ mit $\Phi(t_1)x_1 = 0$. Insbesondere hat dann die Lösung $x(t) = \Phi(t)x_1$ von $\dot{x}(t) = A(t)x(t)$ eine Nullstelle in t_1 , womit für alle $t \in I$ gilt $x(t) = 0$ (Korollar 3.13). Also ist $\Phi(t)$ für alle $t \in I$ nicht invertierbar.

Für eine matrixwertige Lösung Φ von $(*)$ sind die Spalten $x_1(t), \dots, x_n(t)$ von $\Phi(t)$ Lösungen von $\dot{x}(t) = A(t)x(t)$. Weil $x \in L \mapsto x(t_0) \in \mathbb{R}^n$ in Korollar 3.13 ein Vektorraum-Isomorphismus ist, ist $\Phi(t)$ genau dann für ein (und damit für alle) $t \in I$ invertierbar, wenn x_1, \dots, x_n eine Basis von L ist. \square

Definition 3.15. Ein *Fundamentalsystem* einer homogenen linearen Differentialgleichung $\dot{x}(t) = A(t)x(t)$ mit stetigem $A: I \rightarrow \text{Mat}_{n \times n}(\mathbb{R})$, $I =]a, b[$ ist eine matrixwertige Lösung $\Phi: I \rightarrow \text{Gl}(n, \mathbb{R})$ von $\dot{\Phi}(t) = A(t)\Phi(t)$.

Korollar 3.16. Eine homogenen linearen Differentialgleichung $\dot{x}(t) = A(t)x(t)$ hat ein Fundamentalsystem Φ . Dies ist eindeutig bis auf Rechtsmultiplikation mit einer konstanten Matrix aus $\text{Gl}(n, \mathbb{R})$. Jede Lösung $x(t)$ der Differentialgleichung ist von der Form $x(t) = \Phi(t)y$ für einen Vektor $y \in \mathbb{R}^n$.

Beweis. Die Aussage folgt daraus, daß die Spalten eines Fundamentalsystems genau eine Basis des Raums L der Lösungen der Differentialgleichung sind (Korollar 3.14). Zwei Fundamentalsysteme Φ und $\tilde{\Phi}$ unterscheiden sich damit $\tilde{\Phi} = \Phi g$ durch eine Basiswechselmatrix $g \in \text{Gl}(n, \mathbb{R})$. Und jede Lösung ist von der Form $x(t) = \Phi(t)y$, da sie eine Linearkombination der Basisvektoren ist. \square

Beispiel 3.17. (Zwei Fälle, in denen man ein Fundamentalsystem explizit angeben kann.)

- Die Differentialgleichung $\dot{x}(t) = a(t)x(t)$ mit $a: I \rightarrow \mathbb{R}$ auf $I =]a, b[$ hat (wie man zum Beispiel mit Trennung der Variablen, also Beispiel 3.9, findet) die Lösungen

$$x(t) = e^{\int_{t_0}^t a(\tau) d\tau} x_0.$$

(Für $x_0 \neq 0$ hat die Lösung keine Nullstelle und ist eine Basis des 1-dimensionalen Raumes aller Lösungen, also ein Fundamentalsystem.) Für konstantes a erhalten wir dabei insbesondere wieder das exponentielle Wachstum (Bsp. 3.3).

- Eine lineare Differentialgleichung “mit konstanten Koeffizienten”

$$\dot{x}(t) = Ax(t)$$

mit $A \in M_{n \times n}(\mathbb{R})$ hat ein Fundamentalsystem

$$\Phi(t) = e^{tA},$$

wobei $e^X = \sum_{k=1}^{\infty} \frac{1}{k!} X^k$ die Matrixexponentialfunktion ist (einen Beweis der Konvergenz dieser Reihe und der Tatsache, daß diese ein Fundamentalsystem ergibt, kann man im Prinzip wie in Analysis I führen, vgl. dazu den Beweis von Lemma 4.1, Teil a). In Beispiel 3.32 geben wir aber noch einen anderen Beweis).

Zur Berechnung der Matrixexponentialfunktion benutzt man oft die Jordansche Normalform (oder, falls es komplexe Eigenwerte gibt, eine entsprechende reelle Normalform,...). Sein als einfaches Beispiel $A = C(\lambda \text{Id} + N)C^{-1}$ mit $\lambda \in \mathbb{R}$ und nilpotenter Matrix N (also $N^n = 0$) sowie $C \in Gl(n, \mathbb{R})$, so ist

$$\begin{aligned} \Phi(t) &= e^{tA} = C e^{t(\lambda \text{Id} + N)} C^{-1} = \\ &= C \left(e^{\lambda t} (\text{Id} + tN + \frac{1}{2} t^2 N^2 + \dots + \frac{1}{n!} t^n N^n) \right) C^{-1}, \end{aligned}$$

da für kommutierende Matrizen X, Y gilt $e^{X+Y} = e^X e^Y$. (Gibt es mehrere Jordanblöcke, so kann man mit den einzelnen Blöcken analog verfahren.)

3.1.2 Inhomogene lineare Differentialgleichungen

Satz 3.18. *Die Lösungsmenge einer inhomogenen linearen Differentialgleichung*

$$\dot{y}(t) = A(t)y(t) + b(t)$$

mit stetigen Abbildungen $A: I \rightarrow \text{Mat}_{n \times n}(\mathbb{R})$ und $b: I \rightarrow \mathbb{R}^n$ auf $I =]a, b[$ ist eine n -dimensionaler affiner Raum, dessen Vektorraum L der Lösungsraum der dazugehörigen homogenen Gleichung $\dot{x}(t) = A(t)x(t)$ ist. Mit anderen Worten: gegeben eine beliebige Lösung y_p (oft genannt “partikuläre Lösung”) der inhomogenen Gleichung, so ist jede Lösung der inhomogenen Gleichung von der Form

$$y(t) = x(t) + y_p(t) \quad (\text{“homogene Lösung + partikuläre Lösung”}),$$

wobei $x(t)$ eine Lösung der homogenen Gleichung ist.

Eine Lösung der inhomogenen Gleichung kann man mit dem “Variation der Konstanten”-Ansatz

$$y(t) = \Phi(t)c(t)$$

finden, wobei Φ ein Fundamentalsystem der homogenen Gleichung ist und für $c: I \rightarrow \mathbb{R}^n$ durch Integration bestimmt werden kann aus der Gleichung

$$\dot{c}(t) = \Phi(t)^{-1}b(t).$$

Beweis. Der erste Teil des Satzes ist eine direkte Konsequenz von Satz 3.11, der die Existenz einer Lösung $y_p: I \rightarrow \mathbb{R}^n$ der inhomogenen Gleichung garantiert, und Korollar 3.13 über den Lösungsraum der homogenen Gleichung. Denn genau wie in der linearen Algebra gilt:

- ist y eine weitere Lösung der inhomogenen Gleichung, so ist $y - y_p$ eine Lösung der homogenen Gleichung und
- ist x eine Lösung der homogenen Gleichung, so ist $y_p + x$ eine Lösung der inhomogenen Gleichung.

Da $\Phi(t)$ für jedes $t \in I$ invertierbar ist, kann man jede Funktion $y: I \rightarrow \mathbb{R}^n$ schreiben als $y(t) = \Phi(t)c(t)$. Eine Lösung der inhomogenen Gleichung ist y genau dann, wenn

$$\dot{\Phi}c + \Phi\dot{c} = A\Phi c + b,$$

was wegen $\dot{\Phi} = A\Phi$ äquivalent ist zu

$$\Phi\dot{c} = b \quad \iff \quad \dot{c} = \Phi^{-1}b.$$

□

3.1.3 Skalare lineare Differentialgleichungen höherer Ordnung

Eine skalare lineare Differentialgleichung n -ter Ordnung

$$y^{(n)}(t) + a_1(t)y^{(n-1)}(t) + \dots + a_{n-1}(t)\dot{y}(t) + a_n(t)y(t) = b(t) \quad (*)$$

mit stetigen Funktionen $a_1, \dots, a_n, b: I \rightarrow \mathbb{R}$ auf $I =]a, b[$ kann man wie in Bemerkung 3.4 umschreiben in eine lineare Differentialgleichung erster Ordnung

$$\dot{Y}(t) = \begin{pmatrix} 0 & 1 & 0 & \dots \\ 0 & 0 & 1 & \dots \\ \vdots & \vdots & & 1 \\ -a_n(t) & -a_{n-1}(t) & \dots & -a_1(t) \end{pmatrix} Y(t) + \begin{pmatrix} 0 \\ \vdots \\ 0 \\ b(t) \end{pmatrix},$$

wobei $Y(t) = (y(t), \dot{y}(t), \dots, y^{(n-1)}(t))$. Aus Satz 3.18 erhält man damit sofort:

Satz 3.19. *Der Lösungsraum einer skalaren linearen Differentialgleichung n -ter Ordnung (*) ist ein n -dimensionaler affiner Raum, dessen Elemente alle von der Form*

$$y(t) = y_p(t) + x(t)$$

sind, wobei y_p eine beliebige (“partikuläre”) Lösung der inhomogenen Gleichung und x eine Element des n -dimensionalen Vektorraumes der Lösungen der dazugehörigen homogenen Gleichung (also der Gleichung mit $b(t) = 0$) ist.

Eine Lösung der inhomogenen Gleichung erhält man (“durch Variation der Konstanten”) mit dem Ansatz

$$y(t) = x_1(t)c_1(t) + \dots + x_n(t)c_n(t),$$

wobei x_1, \dots, x_n eine Basis des Raumes der Lösungen der homogenen Gleichung ist und für die c_1, \dots, c_n gilt

$$\begin{pmatrix} x_1(t) & \dots & x_n(t) \\ \dot{x}_1(t) & \dots & \dot{x}_n(t) \\ \vdots & & \vdots \\ x_1^{(n-1)}(t) & \dots & x_n^{(n-1)}(t) \end{pmatrix} \begin{pmatrix} \dot{c}_1(t) \\ \dot{c}_2(t) \\ \vdots \\ \dot{c}_n(t) \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ b(t) \end{pmatrix}.$$

Bemerkung 3.20. Eine Basis des Raumes der Lösungen einer homogenen linearen Differentialgleichung n -ter Ordnung mit konstanten Koeffizienten

$$x^{(n)}(t) + a_1 x^{(n-1)}(t) + \dots + a_{n-1} \dot{x}(t) + a_n x(t) = 0$$

$a_1, \dots, a_n \in \mathbb{R}$ findet man mit dem Ansatz $x(t) = e^{\lambda t}$. Eine derartige Funktion $x(t)$ ist genau dann Lösung der Gleichung, wenn

$$\lambda^n + a_1 \lambda^{n-1} + \dots + a_{n-1} \lambda + a_n = 0.$$

Der Ansatz liefert n linear unabhängige Lösungen, wenn man mit komplexen bzw. vielfachen Nullstellen dieses “charakteristischen Polynoms” wie folgt verfährt:

- Für eine komplexe Nullstelle $\lambda = \alpha + i\omega$ kann man das Paar komplexer Lösungen $x(t) = e^{\lambda t}$, $\bar{x}(t) = e^{\bar{\lambda} t}$ ersetzen durch die Real- und Imaginärteile

$$x(t) = e^{\alpha t} \cos(\omega t) \quad \text{und} \quad \bar{x}(t) = e^{\alpha t} \sin(\omega t).$$

- Ist λ ein Nullstellen von Ordnung $k + 1$, so nimmt man

$$x(t) = e^{\lambda t}, t e^{\lambda t}, \dots, t^k e^{\lambda t}$$

(bzw. bei Paaren komplexer Nullstellen wieder die Real- und Imaginärteile dieser Lösungen).

(Details zu diesem Verfahren findet man zum Beispiel in Königsberger *Analysis I* oder in Ferus *Analysis II*.)

Beispiel 3.21. (Schwingungsgleichung mit Dämpfung) Die homogene lineare Gleichung zweiter Ordnung mit konstanten Koeffizienten

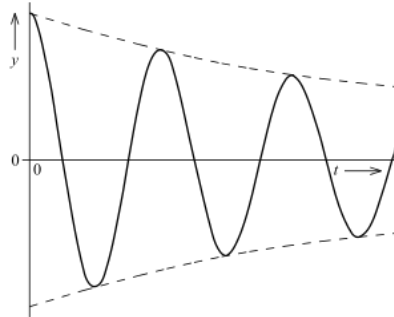
$$\ddot{x}(t) + 2d\dot{x}(t) + \omega^2 x(t) = 0$$

mit $d \geq 0$ und $\omega > 0$ beschreibt gedämpfte Schwingungen. Der Exponentialansatz $x(t) = e^{\lambda t}$ führt auf die charakteristische Gleichung

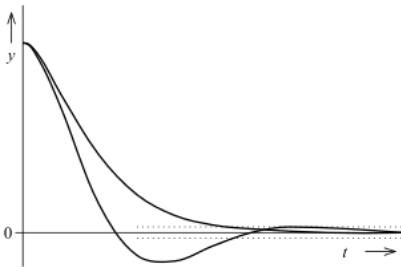
$$\lambda^2 + 2d\lambda + \omega^2 = 0$$

und somit $\lambda = -d \pm \sqrt{d^2 - \omega^2}$.

- Ist die ‘‘Dämpfungskonstante’’ $d = 0$, so erhält man *ungedämpfte* Schwingungen beschrieben durch Linearkombinationen von $\cos(\omega t)$ und $\sin(\omega t)$ (vgl. auch Beispiel 3.8).
- Für $0 < d < \omega$ erhält man *schwach gedämpfte* Schwingungen beschrieben durch Linearkombinationen der beiden Funktionen $e^{-dt} \cos(\omega_0 t)$ und $e^{-dt} \sin(\omega_0 t)$ mit $\omega_0 = \sqrt{\omega^2 - d^2}$.



- Für $d = \omega$ erhält man den sogenannten *aperiodischen Grenzfall* beschrieben durch Linearkombinationen von e^{-dt} und te^{-dt} .



- Für $d > \omega$ erhält man *stark gedämpfte* Schwingungen (den sogenannten ‘‘Kriechfall’’) beschrieben durch Linearkombinationen der beiden Funktionen $e^{-\lambda_1 t}$ und $e^{-\lambda_2 t}$, wobei $\lambda_1 = d + \sqrt{d^2 - \omega^2}$ und $\lambda_2 = d - \sqrt{d^2 - \omega^2} > 0$.

Beispiel 3.22. (Resonanzkatastrophe) Die Differentialgleichung

$$\ddot{y}(t) + y(t) = \cos(t)$$

hat die (partikuläre) Lösung $y(t) = \frac{1}{2}t \sin(t)$.

3.2 Beweis des Existenz- und Eindeutigkeitsatzes

Die Idee des folgenden Beweises ist, das Anfangswertproblem

$$\dot{x}(t) = f(t, x(t)) \quad x(t_0) = x_0$$

umzuschreiben in die (nach dem Hauptsatz der Differential- und Integralrechnung) äquivalente Integralgleichung

$$x(t) = x(t_0) + \int_{t_0}^t f(\tau, x(\tau)) d\tau$$

und diese mit einem auf Picard zurückgehenden Iterationsverfahren zu lösen.

Bemerkung 3.23. Für eine stückweise stetige Funktion h von einer Veränderlichen mit Werten in \mathbb{R}^n kann man das Integral komponentenweise definieren. Dabei gilt für jede Norm

$$\left\| \int_a^b h(\tau) d\tau \right\| \leq \int_a^b \|h(\tau)\| d\tau$$

(für Treppenfunktionen verifiziert man das direkt mittels der Definition des Integrals; der allgemeine Fall folgt daraus via Approximation durch Treppenfunktionen).

3.2.1 Banachscher Fixpunktsatz

Satz 3.24. (*Banachscher Fixpunktsatz*) Sei X ein vollständiger metrischer Raum und $\Phi: X \rightarrow X$ eine Kontraktion, d.h. eine Abbildung, für die es $0 < K < 1$ gibt mit

$$d(\Phi(x), \Phi(y)) \leq Kd(x, y) \quad \text{für alle } x, y \in X.$$

Dann gibt es einen eindeutigen Punkt x^* mit $\Phi(x^*) = x^*$. Für ein beliebiges $x_0 \in X$ ist

$$x^* = \lim_{n \rightarrow \infty} \Phi^n(x_0) = \lim_{n \rightarrow \infty} \underbrace{\Phi \circ \dots \circ \Phi}_{n \text{ mal}}(x_0).$$

Beweis. (Eindeutigkeit) Seien x_1^* und x_2^* Fixpunkte, dann ist

$$d(x_1^*, x_2^*) = d(\Phi(x_1^*), \Phi(x_2^*)) \leq Kd(x_1^*, x_2^*),$$

so daß wegen $K < 1$ gilt $d(x_1^*, x_2^*) = 0$ und somit $x_1^* = x_2^*$.

(Existenz) Für $x_0 \in X$ definieren wir $x_n = \Phi^n(x_0)$. Dann gilt für $n > 1$

$$d(x_{n+1}, x_n) \leq Kd(x_n, x_{n-1}) \leq \dots \leq K^n d(x_1, x_0) = K^n d(\Phi(x_0), x_0).$$

Damit gilt auch für $n \geq 0, l \geq 1$

$$\begin{aligned} d(x_{n+l}, x_n) &\leq d(x_{n+l}, x_{n+l-1}) + \dots + d(x_{n+1}, x_n) \leq \\ &\leq K^n \underbrace{(K^{l-1} + \dots + 1)}_{\frac{1-K^l}{1-K}} d(\Phi(x_0), x_0). \end{aligned}$$

Da $K < 1$, ist

$$d(x_{n+l}, x_n) \leq K^n \frac{1}{1-K} d(\Phi(x_0), x_0)$$

und x_n ist eine Cauchy-Folge. Geht man zum Limes $x^* = \lim_{n \rightarrow \infty} x_n$ über, wird $x_{n+1} = \Phi(x_n)$ zu $x^* = \Phi(x^*)$. \square

3.2.2 Satz von Picard–Lindelöf

Sei $U \subset \mathbb{R} \times \mathbb{R}^n$ offen und $f: U \rightarrow \mathbb{R}^n$ stetig. Wir beweisen im Folgenden für $(t_0, x_0) \in U$ die Eindeutigkeit und lokale Existenz von Lösungen des Anfangswertproblems

$$\dot{x}(t) = f(t, x(t)) \quad x(t_0) = x_0$$

unter der Annahme, daß f lokal Lipschitz stetig in x -Richtung ist.

Definition 3.25. Die Funktion $f(t, x)$ ist *lokal Lipschitz stetig in x -Richtung*, wenn es für alle $(t_0, x_0) \in U$ eine Umgebung V und $L > 0$ existiert mit

$$\|f(t, x) - f(t, x')\| \leq L\|x - x'\| \quad \text{für alle } (t, x), (t, x') \in V,$$

Bemerkung 3.26. Ist $f(t, x)$ stetig und stetig differenzierbar in x -Richtung, so ist f lokal Lipschitz stetig in x -Richtung. (Folgt direkt aus dem Schrankensatz 2.52 zusammen mit Satz 1.57).

Satz 3.27. (Satz von Picard–Lindelöf) *Ist $f(t, x)$ stetig und lokal Lipschitz stetig in x -Richtung, dann gibt es für alle $(t_0, x_0) \in U$ ein $\epsilon > 0$, so daß das Anfangswertproblem*

$$\dot{x}(t) = f(t, x(t)) \quad x(t_0) = x_0$$

für alle $0 < \epsilon' \leq \epsilon$ eine eindeutige Lösung $x:]t_0 - \epsilon', t_0 + \epsilon'[\rightarrow \mathbb{R}^n$ hat.

Beweis. Durch Verkleinern von U können wir annehmen, daß es C und L gibt mit

$$\|f(t, x)\| \leq C \quad \text{und} \quad \|f(t, x) - f(t, x')\| \leq L\|x - x'\|$$

für alle $(t, x), (t, x') \in U$. Für $(t_0, x_0) \in U$ wählen wir $\epsilon, \delta > 0$ mit

$$J \times \overline{B_\delta(x_0)} \subset U \quad \text{für} \quad J =]t_0 - \epsilon, t_0 + \epsilon[$$

und $\epsilon < \frac{\delta}{C}$, $\epsilon < \frac{1}{L}$. Dann ist

$$X = \{x: J \rightarrow \overline{B_\delta(x_0)} \mid x \text{ ist stetig und } x(t_0) = x_0\}$$

als abgeschlossene Teilmenge des vollständigen normierten Vektorraumes $BC(J, \mathbb{R}^n)$ (Satz 1.82) selbst ein vollständiger metrischer Raum. Wir betrachten nun den Operator

$$\Phi: X \rightarrow X \quad x \mapsto \left(t \mapsto x_0 + \int_{t_0}^t f(\tau, x(\tau)) d\tau \right).$$

Dieser Operator ist wohldefiniert, denn $\Phi(x)$ ist wieder stetig (nach dem Hauptsatz der Differential- und Integralrechnung), es gilt $\Phi(x)(t_0) = x_0$ und für alle $t \in J$ gilt $\Phi(x)(t) \in \overline{B_\delta(x_0)}$, denn

$$\|\Phi(x)(t) - x_0\| = \left\| \int_{t_0}^t f(\tau, x(\tau)) d\tau \right\| \leq \int_{t_0}^t \|f(\tau, x(\tau))\| d\tau \leq C\epsilon < \delta.$$

Weiter ist Φ eine Kontraktion, denn

$$\|\Phi(x) - \Phi(y)\| \leq \epsilon \sup_{t \in J} \|f(t, x(t)) - f(t, y(t))\| \leq \underbrace{\epsilon L}_{=:K} \|x - y\|,$$

wobei $K < 1$ nach Wahl von ϵ gilt.

Nach dem Banachschen Fixpunktsatz (Satz 3.24) hat Φ einen eindeutigen Fixpunkt und es gibt eine eindeutige Lösung des Anfangswertproblems

$$\dot{x}(t) = f(t, x(t)) \quad x(t_0) = x_0$$

mit $x: J \rightarrow \overline{B_\delta(x_0)}$ (denn mit dem Hauptsatz der Differential- und Integralrechnung ist jede stetige Lösung der Integralgleichung $\Phi(x^*) = x^*$ differenzierbar und erfüllt die Differentialgleichung).

Es bleibt zu zeigen, daß eine mögliche weitere Lösung des Anfangswertproblems $\tilde{x}: J \rightarrow \mathbb{R}^n$ auch $\tilde{x}(t) \in \overline{B_\delta(x_0)}$ erfüllt (und somit gleich x ist): sei also $t' \in J$, so ist (da \tilde{x} nach dem Hauptsatz der Differential- und Integralrechnung die Integralgleichung $\tilde{x}(t) = x_0 + \int_{t_0}^t f(\tau, \tilde{x}(\tau)) d\tau$ erfüllt)

$$\|\tilde{x}(t') - x_0\| \leq \left\| \int_{t_0}^{t'} f(\tau, \tilde{x}(\tau)) d\tau \right\| \leq \int_{t_0}^{t'} \|f(\tau, \tilde{x}(\tau))\| d\tau \leq \epsilon C < \delta.$$

Der Beweis funktioniert ohne Änderung auch für alle kleineren Wahlen von ϵ , also für alle $\epsilon' \leq \epsilon$. \square

Bemerkung 3.28. Der Beweis des Satzes von Picard–Lindelöf liefert (mit minimaler und offensichtlicher Änderung) auch die eindeutige Existenz von Lösungen für halboffene Intervalle $]t_0 - \epsilon', t_0]$ bzw. $[t_0, t_0 + \epsilon'[,$ Damit kann man eine Lösung $x: I \rightarrow \mathbb{R}^n$ des Anfangswertproblems

$$\dot{x}(t) = f(t, x(t)) \quad x(t_0) = x_0,$$

die auf einem abgeschlossenen oder halboffenen Intervall I definiert ist, in jedem Randpunkt $t_1 \in I$ für kleine $\epsilon > 0$ eindeutig fortsetzen auf ein um ϵ verlängertes offenes Intervall. (Sei zum Beispiel x definiert auf $I =]a, t_1]$ und $x_1 = x(t_1)$, so gibt es ein $\epsilon > 0$ mit $\epsilon < t_1 - a$, so daß das Anfangswertproblem

$$\dot{\tilde{x}}(t) = f(t, \tilde{x}(t)) \quad \tilde{x}(t_1) = x_1$$

auf $]t_1 - \epsilon, t_1 + \epsilon[$ eine eindeutige Lösung \tilde{x} besitzt, die auch auf $]t_1 - \epsilon, t_1]$ eindeutig ist. Damit gilt $x_{]t_1 - \epsilon, t_1]} = \tilde{x}_{]t_1 - \epsilon, t_1]}$ und man kann die Lösung x fortsetzen.)

Korollar 3.29. *Ist $f(t, x)$ stetig und lokal Lipschitz stetig in x -Richtung, und sind $x: I \rightarrow \mathbb{R}^n$ und $\tilde{x}: \tilde{I} \rightarrow \mathbb{R}^n$ zwei Lösungen des Anfangswertproblems*

$$\dot{x}(t) = f(t, x(t)) \quad x(t_0) = x_0 \quad (*)$$

für $(t_0, x_0) \in U$, so ist $x_{I \cap \tilde{I}} = \tilde{x}_{I \cap \tilde{I}}$.

Beweis. Sei $A = \{t \in I \cap \tilde{I} \mid x(t) = \tilde{x}(t)\}$. Dann ist $A \neq \emptyset$ (da $t_0 \in A$). Als Nullstellenmenge einer stetigen Funktion ist A abgeschlossen in $I \cap \tilde{I}$. Weiter ist A mit dem Satz von Picard–Lindelöf (Satz 3.27) offen in $I \cap \tilde{I}$ (denn ist $t_1 \in A$, so stimmen x und \tilde{x} auf einem Intervall um t_1 überein). Da das Intervall $I \cap \tilde{I}$ zusammenhängend ist, gilt $A = I \cap \tilde{I}$. \square

Bemerkung 3.30. Mit Korollar 3.29 kann man, indem man alle Lösungen “vereinigt”, eine *maximale Lösung* $x: I_{max} \rightarrow \mathbb{R}^n$ des Anfangswertproblems $(*)$ definieren, so daß für jede Lösung $\tilde{x}: \tilde{I} \rightarrow \mathbb{R}^n$ von $(*)$ gilt $\tilde{I} \subset I_{max}$ (und damit auch $\tilde{x} = x|_{\tilde{I}}$).

3.2.3 Beweis der globalen Existenz für lineare Differentialgleichungen

In diesem Abschnitt zeigen wir Satz 3.11 über die globale Existenz einer eindeutigen Lösung $x: I \rightarrow \mathbb{R}^n$ des Anfangswertproblems für eine lineare Differentialgleichung erster Ordnung

$$\dot{x}(t) = A(t)x(t) + b(t) \quad x(t_0) = x_0$$

mit stetigen $A: I \rightarrow \text{Mat}_{n \times n}(\mathbb{R})$ und $b: I \rightarrow \mathbb{R}^n$ und $t_0 \in I$, $x_0 \in \mathbb{R}^n$.

Lemma 3.31. *Für eine lineare Differentialgleichung ist $f(t, x) = A(t)x + b(t)$ lokal Lipschitz stetig in x -Richtung.*

Beweis. Da f stetig differenzierbar in x -Richtung ist, folgt die Aussage aus Bemerkung 3.26.

Man kann sie aber auch leicht direkt einsehen, denn für jedes (t_0, x_0) gilt für $(t, x), (t, x')$ mit $t \in K \subset I$ in einer kompakten Umgebung $K \subset I$ von t_0

$$\|f(t, x) - f(t, x')\| = \|A(t)(x - x')\| \leq \|A(t)\| \|x - x'\| \leq L \|x - x'\|,$$

wobei $L = \sup_{t \in K} \|A(t)\|$. □

Damit folgt die lokale und globale Eindeutigkeit von Lösungen direkt aus dem Satz von Picard–Lindelöf (Satz 3.27) und dessen Korollar 3.29.

Beweis von Satz 3.11. Wegen der vorhergehenden Anmerkung zur lokalen und globalen Eindeutigkeit reicht es, die globale Existenz einer Lösung des Anfangswertproblems zu zeigen. Wir werden diese für jedes kompakte Teilintervall $K \subset I$ mit $t_0 \in K$ beweisen. Da das offene Intervall I gleich der Vereinigung aller kompakten Teilintervalle ist, liefert das (wegen der Eindeutigkeit von Lösungen auf dem Durchschnitt von Intervallen) eine Lösung des Anfangswertproblems auf ganz I (vgl. Bemerkung 3.30).

Sei also $K \subset I$ ein kompaktes Teilintervall mit $t_0 \in K$ und

$$L = \sup_{t \in K} \|A(t)\|.$$

Im vollständigen normierten Vektorraum $C^0(K, \mathbb{R}^n)$ betrachten wir die rekursiv definierte Folge $x_n = \Phi^n(x_0)$, wobei x_0 die konstante Abbildung mit Wert x_0 ist und

$$\Phi: C^0(K, \mathbb{R}^n) \rightarrow C^0(K, \mathbb{R}^n) \quad x \mapsto \left(t \mapsto x_0 + \int_{t_0}^t (A(\tau)x_n(\tau) + b_n(\tau)) d\tau \right).$$

Für $n \geq 1$ gilt dann

$$\|x_{n+1}(t) - x_n(t)\| = \left\| \int_{t_0}^t A(\tau)(x_n(\tau) - x_{n-1}(\tau)) d\tau \right\| \leq L \int_{t_0}^t \|x_n(\tau) - x_{n-1}(\tau)\| d\tau$$

und somit per Induktion

$$\|x_{n+1}(t) - x_n(t)\| \leq C \frac{L^n |t - t_0|^n}{n!},$$

wobei $C = \sup_{t \in K} \|x_1(t) - x_0\|$. Damit konvergiert x_n auf K gleichmäßig gegen ein $x \in C^0(K, \mathbb{R}^n)$ (denn die Teleskopsummandarstellung $x_{n+1} = x_0 + \sum_{k=0}^n (x_{k+1} - x_k)$ führt für jedes $t \in K$ auf eine Reihe, die majoriert wird durch ein Vielfaches der Exponentialreihe $e^{L|t-t_0|} = \sum_k \frac{L^k |t-t_0|^k}{k!}$). Der Grenzwert x ist ein Fixpunkt von Φ und somit Lösung des Anfangswertproblems. □

Beispiel 3.32. Das Picardsche Iterationsverfahren (aus dem vorhergehenden Beweis und dem Beweis des Satzes von Picard–Lindelöf (Satz 3.27)) liefert für das Anfangswertproblem

$$\dot{x}(t) = Ax(t) \quad x(0) = x_0$$

mit konstantem $A \in M(n \times n)$ die Folge

$$\begin{aligned} x_1(t) &= x_0 + Atx_0, \\ x_2(t) &= x_0 + Atx_0 + A^2 \frac{t^2}{2} x_0 \end{aligned}$$

und per Induktion

$$x_n(t) = \sum_{k=0}^n \frac{A^k t^k}{k!} x_0.$$

Damit erhalten wir in diesem Spezialfall aus dem vorhergehenden Beweis auch einen Beweis, daß die Matrixexponentialreihe e^{tA} für alle $t \in \mathbb{R}$ konvergiert und ein Fundamentalsystem der Gleichung $\dot{x}(t) = Ax(t)$ liefert.

4 Mehr zum Satz über die Umkehrabbildung und zum Satz über implizite Funktionen...

Wir beweisen zuerst den Satz über die Umkehrabbildung (Satz 2.75) und danach den Satz über implizite Funktionen (Satz 2.72). Danach diskutieren wir Anwendungen der beiden Sätze.

4.1 Die Beweise

Beweis des Satzes über die Umkehrabbildung (Satz 2.75). OBdA nehmen wir an, daß $p = 0$ und $f(p) = 0$ sowie $Df|_{p=0} = \text{Id}$ (das kann man durch Translation im Bild- und Urbildbereich sowie Verkettung mit $(Df|_p)^{-1}$ erreichen). Für die Abbildung

$$g(x) = x - f(x)$$

gilt $Dg|_x = \text{Id} - Df|_x$ und somit $Dg|_0 = 0$, so daß, weil f und somit g stetig differenzierbar ist, ein $r > 0$ existiert mit $\overline{B_r(0)} = \{x \in V \mid \|x\| \leq r\} \subset U$ und

$$\|Dg|_x\| \leq \frac{1}{2} \quad \text{für alle } x \in \overline{B_r(0)}.$$

Da $\overline{B_r(0)}$ konvex ist, gilt mit dem Schrankensatz (Satz 2.52)

$$\|g(x_1) - g(x_2)\| \leq \frac{1}{2} \|x_1 - x_2\| \quad \text{für alle } x_1, x_2 \in \overline{B_r(0)}.$$

Wegen $g(0) = 0$ ist die Einschränkung von g auf $\overline{B_r(0)}$ eine Abbildung

$$g: \overline{B_r(0)} \rightarrow \overline{B_{r/2}(0)}.$$

Für $y \in \overline{B_{r/2}(0)}$ betrachten wir die Abbildung

$$\varphi_y(x) = y + x - f(x) = y + g(x).$$

Für deren Einschränkung auf $\overline{B_r(0)}$ gilt

$$\varphi_y: \overline{B_r(0)} \rightarrow \overline{B_r(0)}$$

und

$$\|\varphi_y(x_1) - \varphi_y(x_2)\| = \|g(x_1) - g(x_2)\| \leq \frac{1}{2}\|x_1 - x_2\|.$$

Nach dem Banachschen Fixpunktsatz (Satz 3.24) gibt es damit für jedes $y \in \overline{B_{r/2}(0)}$ einen eindeutigen Punkt $x \in \overline{B_r(0)}$ mit

$$\varphi_y(x) = x \quad \iff \quad y = f(x).$$

Setzen wir $\tilde{O} = \overline{B_{r/2}(0)}$ und $O = f^{-1}(\tilde{O})$, so ist

$$f|_O: O \rightarrow \tilde{O}$$

bijektiv.

Die Abbildung $f|_O^{-1}$ ist stetig, denn für $x_1, x_2 \in O$ gilt

$$\|x_1 - x_2\| = \|g(x_1) - g(x_2) + f(x_1) - f(x_2)\| \leq \frac{1}{2}\|x_1 - x_2\| + \|f(x_1) - f(x_2)\|$$

und somit $\|x_1 - x_2\| \leq 2\|f(x_1) - f(x_2)\|$, so daß für $x_i = f|_O^{-1}(y_i)$, $i = 1, 2$ gilt

$$\|f|_O^{-1}(y_1) - f|_O^{-1}(y_2)\| \leq 2\|y_1 - y_2\|. \quad (*)$$

Um die Differenzierbarkeit von $f|_O^{-1}$ zu zeigen, benutzen wir, daß $Df|_x$ wegen Teil a) von Lemma 4.1 für alle $x \in \overline{B_r(0)}$ invertierbar ist (da $Df|_x = \text{Id} - Dg|_x$ und $\|Dg|_x\| \leq \frac{1}{2}$). Für $x, x_1 \in O$ und $y = f(x)$, $y_1 = f(x_1)$ gilt

$$\begin{aligned} \tilde{R}(y) &:= f|_O^{-1}(y) - f|_O^{-1}(y_1) - (Df|_{x_1})^{-1}(y - y_1) = \\ &= x - x_1 - (Df|_{x_1})^{-1}(f(x) - f(x_1)) = -(Df|_{x_1})^{-1}(R(x)), \end{aligned}$$

wobei $R(x)$ definiert ist durch

$$f(x) - f(x_1) = Df|_{x_1}(x - x_1) + R(x).$$

Differenzierbarkeit von f in x_1 impliziert $\frac{R(x)}{\|x-x_1\|} \rightarrow 0$ für $x \rightarrow x_1$, womit wegen $\|x-x_1\| \leq 2\|y-y_1\|$ (siehe (*)) für $y \rightarrow y_1$ und somit $x \rightarrow x_1$ (da $f|_O^{-1}$ stetig ist) gilt

$$\frac{\|\tilde{R}(y)\|}{\|y-y_1\|} \leq 2\|(Df|_{x_1})^{-1}\| \left(\frac{\|R(x)\|}{\|x-x_1\|} \right) \rightarrow 0.$$

Also ist $f|_O^{-1}$ differenzierbar und (wie auch aus der Kettenregel folgen würde)

$$D(f^{-1})|_y = (Df|_{f^{-1}(y)})^{-1}$$

für alle $y \in \tilde{O}$. Insbesondere ist $y \mapsto D(f^{-1})|_y$ als Verkettung

$$\tilde{O} \xrightarrow{f|_O^{-1}} O \xrightarrow{Df} GL(V) \xrightarrow{(\cdot)^{-1}} GL(V) \quad y \mapsto f^{-1}(y) \mapsto Df|_{f^{-1}(y)} \mapsto (Df|_{f^{-1}(y)})^{-1}$$

von stetigen Abbildungen selbst stetig und $f|_O^{-1}$ ist stetig differenzierbar. Per Induktion erhält man

$$f \in C^k \quad \implies \quad f|_O^{-1} \in C^k$$

(da $A \in GL(V) \mapsto A^{-1} \in GL(V)$ wegen Teil b) von Lemma 4.1 unendlich oft differenzierbar ist). \square

Lemma 4.1. *Sei V ein endlichdimensionaler normierter Vektorraum.*

- a) Für $X \in \text{End}(V)$ mit $\|X\| < 1$ ist $\text{Id} - X$ invertierbar und die Reihe $\sum_{k=0}^{\infty} X^k$ konvergiert gegen $(\text{Id} - X)^{-1}$.
- b) Die Menge $GL(V) \subset \text{End}(V)$ ist offen und

$$i: GL(V) \rightarrow GL(V) \quad A \mapsto A^{-1}$$

ist C^∞ mit Differential $di|_A(H) = -A^{-1}HA^{-1}$.

Beweis. a) Für jede Norm auf einem Vektorraum V erfüllt die Operatornorm (Definition 1.95) wie man leicht sieht

$$\|A \circ B\| \leq \|A\| \cdot \|B\| \quad \text{für alle } A, B \in \text{End}(V).$$

Die Reihe $\sum_{k=0}^{\infty} X^k$ wird damit für alle $X \in \text{End}(V)$ mit $\|X\| < 1$ in der Operatornorm von der geometrischen Reihe $\sum_{k=0}^{\infty} \|X\|^k$ majoriert und konvergiert (denn die Partialsummenfolge $Y_n = \sum_{k=0}^n X^k$ erfüllt für $m > n$)

$$\|Y_m - Y_n\| \leq \|X\|^m + \dots + \|X\|^{n+1} \leq \|X\|^{n+1} \underbrace{(\|X\|^{m-n+1} + \dots + 1)}_{\leq \frac{1}{1-\|X\|}}$$

und ist somit eine Cauchy-Folge). Da für den Grenzwert $Y = \sum_{k=0}^{\infty} X^k$ gilt

$$Y(\text{Id} - X) = (\text{Id} - X)Y = \text{Id},$$

ist $\text{Id} - X$ invertierbar und

$$(\text{Id} - X)^{-1} = \sum_{k=0}^{\infty} X^k.$$

b) $Gl(V) \subset \text{End}(V)$ ist offen, denn ist $A \in Gl(V)$ und $B \in \text{End}(V)$ mit $\|A - B\| < \frac{1}{\|A^{-1}\|}$, so ist

$$B = A - (A - B) = A(\text{Id} - A^{-1}(A - B))$$

wegen Teil a) invertierbar, da $\|A^{-1}(A - B)\| < 1$.

Die Abbildung i mit $A \in Gl(V) \mapsto A^{-1} \in Gl(V)$ ist differenzierbar, denn für H mit $\|H\| < \frac{1}{\|A^{-1}\|}$ ist

$$\begin{aligned} (A + H)^{-1} &= (A(\text{Id} + A^{-1}H))^{-1} = \left(\sum_{k=0}^{\infty} (-1)^k (A^{-1}H)^k \right) A^{-1} = \\ &= A^{-1} - A^{-1}HA^{-1} + \underbrace{\left(\sum_{k=2}^{\infty} (-1)^k (A^{-1}H)^k \right)}_{=:R(H)} A^{-1}, \end{aligned}$$

wobei $\frac{\|R(H)\|}{\|H\|} \rightarrow 0$ für $H \rightarrow 0$. Da die Differenzierbarkeit der Abbildung $A \mapsto A^{-1}$ deren Stetigkeit impliziert, hängt das Differential $Di|_A(H) = -A^{-1}HA^{-1}$ stetig von A ab und i ist stetig differenzierbar. Per Induktion ist i unendlich oft differenzierbar. \square

Wir folgern nun den Satz über implizite Funktionen (Satz 2.72) aus dem Satz über die Umkehrabbildung (Satz 2.75).

Beweis des Satzes über implizite Funktionen (Satz 2.72). Wie in der Formulierung des Satzes über implizite Funktionen sei $f: U \subset V_1 \times V_2 \rightarrow W$ eine C^k -Abbildung. Wir betrachten

$$F: U \subset V_1 \times V_2 \rightarrow V_1 \times W \quad (x, y) \mapsto (x, f(x, y)).$$

Dann ist

$$DF_{|(x,y)}(v_1, v_2) = (v_1, Df_{|(x,y)}(v_1, v_2)) = (v_1, D_1f_{|(x,y)}(v_1) + D_2f_{|(x,y)}(v_2))$$

bzw. in Matrixschreibweise (falls $V_1 = \mathbb{R}^n$ und $V_2 = W = \mathbb{R}^m$)

$$DF_{|(x,y)} = \begin{pmatrix} \text{Id}_{n \times n} & 0 \\ D_1f_{|(x,y)} & D_2f_{|(x,y)} \end{pmatrix}.$$

Da $D_2f_{|(p,q)}: V_2 \rightarrow W$ nach Voraussetzung invertierbar ist, ist auch

$$DF_{|(p,q)}: V_1 \times V_2 \rightarrow V_1 \times W$$

invertierbar und mit dem Satz über die Umkehrabbildung (Satz 2.75) gibt es eine Umgebung $O \subset V_1 \times V_2$ von (p, q) , so daß $F|_O: O \rightarrow \tilde{O}$ ein C^k -Diffeomorphismus auf die offene Menge $\tilde{O} = F(O)$ ist.

Durch Verkleinern von O können wir annehmen, daß $O = U_1 \times U_2$, mit $p \in U_1$ und $q \in U_2$. Da \tilde{O} eine Umgebung von $(p, 0)$ ist, gibt es offene Mengen \tilde{U}_1 und \tilde{U}_2 mit $p \in \tilde{U}_1$ und $0 \in \tilde{U}_2$ sowie $\tilde{U}_1 \times \tilde{U}_2 \subset \tilde{O}$. Wir können oBdA annehmen, daß $U_1 = \tilde{U}_1$ (indem wir die alte Menge U_1 durch die Teilmenge \tilde{U}_1 ersetzen).

Dann gibt es für alle $x \in U_1$ ein eindeutiges $y \in U_2$, so daß

$$F(x, y) = (x, 0) \quad \text{bzw.} \quad f(x, y) = 0.$$

Diese y ist gegeben durch $y = g(x)$, wobei $g: U_1 \rightarrow U_2$ die C^k -Abbildung mit $(x, g(x)) = F|_O^{-1}(x, 0)$ ist. Die Formel für die Ableitung von g folgt aus $f(x, g(x)) = 0$, denn mit der Kettenregel gilt

$$D_1f_{|(x,g(x))} + D_2f_{|(x,g(x))} \circ Dg|_x.$$

□

Bemerkung 4.2. Umgekehrt kann man auch den Satz über die Umkehrabbildung aus dem Satz über implizite Funktionen herleiten (Übungsaufgabe) und den Satz über implizite Funktionen direkt mit dem Banachschen Fixpunktsatz beweisen (siehe z.B. Otto Forster *Analysis 2*).

4.2 Untermannigfaltigkeiten

Eine Untermannigfaltigkeit des \mathbb{R}^N ist eine Teilmenge $M \subset \mathbb{R}^N$, die lokal und modulo Anwendung eines Diffeomorphismus aussieht wie ein linearer Unterraum. Der folgende Satz zeigt die Äquivalenz von verschiedenen Möglichkeiten, Untermannigfaltigkeit zu definieren.

Definition 4.3. • Eine C^k -Abbildung $f: U \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$ heißt *Immersion* (*Submersion*), falls die lineare Abbildung $Df|_p: \mathbb{R}^n \rightarrow \mathbb{R}^m$ für alle $p \in U$ injektiv (surjektiv) ist.

- Eine Teilmenge $M \subset \mathbb{R}^N$ ist eine C^k -Untermannigfaltigkeit der Dimension n , wenn i)–iv) im folgenden Satz 4.4 erfüllt sind.

Satz 4.4. Für eine Teilmenge $M \subset \mathbb{R}^N$ sind äquivalent:

- i) (Man kann M lokal "flachmachen".) Für jeden Punkt $p \in M$ gibt es eine offene Umgebung $U \subset \mathbb{R}^N$ von p und einen C^k -Diffeomorphismus

$$\Phi: U \rightarrow \tilde{U} \subset \mathbb{R}^N$$

(genannt Flachmacher) mit

$$\Phi(M \cap U) = \tilde{U} \cap (\mathbb{R}^n \times \{0\}).$$

- ii) (M ist lokal gleichungsdefiniert.) Für jeden Punkt $p \in M$ gibt es eine offene Umgebung $U \subset \mathbb{R}^N$ von p und eine C^k -Submersion

$$h: U \rightarrow \mathbb{R}^{N-n}$$

mit

$$M \cap U = h^{-1}\{0\}.$$

- iii) (M ist lokal Graph einer Abbildung.) Für jeden Punkt $p \in M$ gibt es eine offene Umgebung $U \subset \mathbb{R}^N$ und einer Permutation $\sigma \in \mathcal{S}_N = \text{Perm}(1, \dots, N)$, so daß

$$U = \{x \in \mathbb{R}^N \mid (x_{\sigma(1)}, \dots, x_{\sigma(n)}) \in U_1, (x_{\sigma(n+1)}, \dots, x_{\sigma(N)}) \in U_2\}$$

für $U_1 \subset \mathbb{R}^n$ und $U_2 \subset \mathbb{R}^{N-n}$ offen und

$$M \cap U = \{x \in U \mid (x_{\sigma(n+1)}, \dots, x_{\sigma(N)}) = g(x_{\sigma(1)}, \dots, x_{\sigma(n)}),$$

wobei $g: U_1 \rightarrow U_2$ eine C^k -Abbildung ist.

iv) (*M* ist lokal parametrisierbar.) Für jeden Punkt $p \in M$ gibt es eine offene Umgebung $U \subset \mathbb{R}^N$ und eine C^k -Immersion $\psi: V \rightarrow U$ auf einer offenen Menge $V \subset \mathbb{R}^n$ mit Werten in $M \cap U$, so daß

$$\psi: V \rightarrow M \cap U \quad \text{bijektiv ist}$$

und eine stetige Inverse hat.

Beweis. $i) \Rightarrow ii)$ Ist $\Phi: U \rightarrow \tilde{U}$ ein C^k -Flachmacher bei $p \in M$, so ist

$$h = pr_{\mathbb{R}^{N-n}} \circ \Phi,$$

wobei $pr_{\mathbb{R}^{N-n}}: \mathbb{R}^N \rightarrow \mathbb{R}^{N-n}$ die Projektion auf die letzten $(N-n)$ -Koordinaten bezeichnet, eine C^k -Submersion wie in $ii)$.

$ii) \Rightarrow iii)$ Ist h eine C^k -Submersion wie in $ii)$ für einen Punkt $p \in M$. Dann kann man durch eine Permutation $\sigma \in \mathcal{S}_n = Perm(1, \dots, n)$ der Koordinaten erreichen, daß die Matrix bestehend aus den partiellen Ableitungen

$$\frac{\partial h}{\partial x_{\sigma(n+1)}}(p), \dots, \frac{\partial h}{\partial x_{\sigma(N)}}(p)$$

invertierbar ist. Nach dem Satz über implizite Funktionen (Satz 2.72) kann man die Gleichung $h(x) = 0$ lokal um p auflösen nach $(x_{\sigma(n+1)}, \dots, x_{\sigma(N)})$, was bedeutet, deren Lösungsmenge lokal zu schreiben als Graph

$$(x_{\sigma(n+1)}, \dots, x_{\sigma(N)}) = g(x_{\sigma(1)}, \dots, x_{\sigma(n)})$$

einer C^k -Abbildung g .

$iii) \Rightarrow iv)$ Ist g eine C^k -Abbildung mit

$$x \in M \cap U \Leftrightarrow \underbrace{(x_{\sigma(n+1)}, \dots, x_{\sigma(N)})}_{\in U_2} = g(\underbrace{(x_{\sigma(1)}, \dots, x_{\sigma(n)})}_{\in U_1}),$$

dann ist $\psi: U_1 \rightarrow U$, $y \mapsto A_{\sigma^{-1}}(y, g(y))$ mit $A_{\sigma^{-1}}(y_1, \dots, y_N) = (y_{\sigma^{-1}(1)}, \dots, y_{\sigma^{-1}(N)})$ eine C^k -Parametrisierung wie in $iv)$.

$iv) \Rightarrow i)$ Ist $\psi: V \rightarrow U \cap M$ eine C^k -Parametrisierung um $p \in M$ und $\tilde{p} \in V$ mit $\psi(\tilde{p}) = p$. Da ψ eine Immersion ist, können wir $v_{n+1}, \dots, v_N \in \mathbb{R}^N$ wählen, so daß

$$\frac{\partial \psi}{\partial y_1}(\tilde{p}), \dots, \frac{\partial \psi}{\partial y_n}(\tilde{p}), v_{n+1}, \dots, v_N$$

eine Basis von \mathbb{R}^N ist. Die C^k -Abbildung

$$\Psi: V \times \mathbb{R}^{N-n} \rightarrow \mathbb{R}^N \quad (y_1, \dots, y_N) \mapsto \psi(y_1, \dots, y_n) + y_{n+1}v_{n+1} + \dots + y_N v_N$$

ist dann wegen des Satzes über die Umkehrabbildung (Satz 2.75) ein Diffeomorphismus von einer offenen Umgebung O von $(\tilde{p}, 0)$ auf eine offene Umgebung \tilde{O} von p , von der wir oBdA $\tilde{O} \subset U$ annehmen können. Da die Inverse von $\psi: V \rightarrow U \cap M$ stetig ist, gibt es eine offene Umgebung $\tilde{O}' \subset \tilde{O}$ von p mit $\psi(O \cap (\mathbb{R}^n \times \{0\})) = \tilde{O}' \cap M$. Dann ist die Inverse von $\Psi|_{O'}: O' \rightarrow \tilde{O}'$ mit $O' = \Psi^{-1}(\tilde{O}')$ ein Flachmacher bei $p \in M$. \square

Beispiel 4.5. 1.) (Teil *iii*) des Satzes eignet sich am besten, wenn man zeigen will, daß eine Menge $M \subset \mathbb{R}^N$ keine Untermannigfaltigkeit ist.) Die Menge

$$M = \{(x, y) \in \mathbb{R}^2 \mid xy = 0\}$$

ist keine Untermannigfaltigkeit, denn für keine offene Umgebung U von $(0, 0)$ kann $M \cap U$ als Graph einer Funktion $y = g(x)$ bzw. $x = g(y)$ dargestellt werden.

2.) (Teil *ii*) des Satzes ist oft am effizientesten, wenn man zeigen will, daß eine Menge $M \subset \mathbb{R}^N$ eine Untermannigfaltigkeit ist.) Die Einheitssphäre

$$S^n = \{(x_0, \dots, x_n) \in \mathbb{R}^{n+1} \mid x_0^2 + \dots + x_n^2 = 1\}$$

ist eine Untermannigfaltigkeit, denn

$$h: \mathbb{R}^{n+1} \setminus \{0\} \rightarrow \mathbb{R} \quad (x_0, \dots, x_n) \mapsto x_0^2 + \dots + x_n^2 - 1$$

ist eine Submersion und $M = h^{-1}(\{0\})$.

(Zu Teil *iii*) des Satzes.) Wegen $\frac{\partial h}{\partial x_i} = 2x_i$ impliziert der Satz über implizite Funktionen (Satz 2.72), daß man S^n weg von den Punkten mit $x_i = 0$ lokal als Graph einer Funktion von $x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n$ schreiben kann (und zwar der Funktion

$$x_i = \pm \sqrt{x_1^2 + \dots + x_{i-1}^2 + x_{i+1}^2 + \dots + x_n^2},$$

siehe auch Beispiel 2.71 für den Fall S^1 .)

(Zu Teil *i*) des Satzes.) Flachmacher für (Teile von) S^1 bzw. S^2 erhält man, indem man die Umkehrabbildungen der Abbildungen Φ in 1.) bzw. 3.) von Beispiel 2.79 betrachtet, welche Polarkoordinaten bzw. sphärische Koordinaten beschreibt.

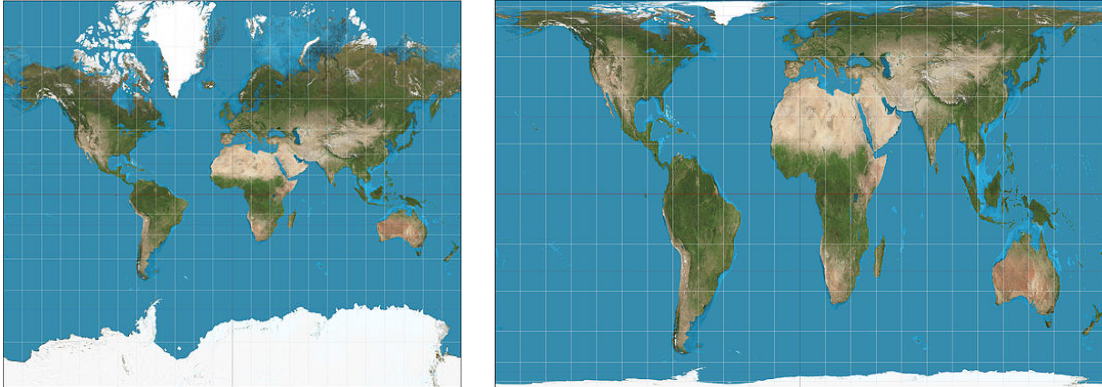
(Zu Teil *iv*) des Satzes.) Die 1–dimensionale Sphäre S^1 besitzt lokale Parametrisierungen

$$\psi:]0, 2\pi[\rightarrow S^1 \quad t \mapsto e^{i(t+\theta)} = (\cos(t+\theta), \sin(t+\theta)),$$

$\theta \in \mathbb{R}$, die “optimal” in dem Sinne sind, daß sie S^1 mit konstanter Geschwindigkeit 1 ablaufen.

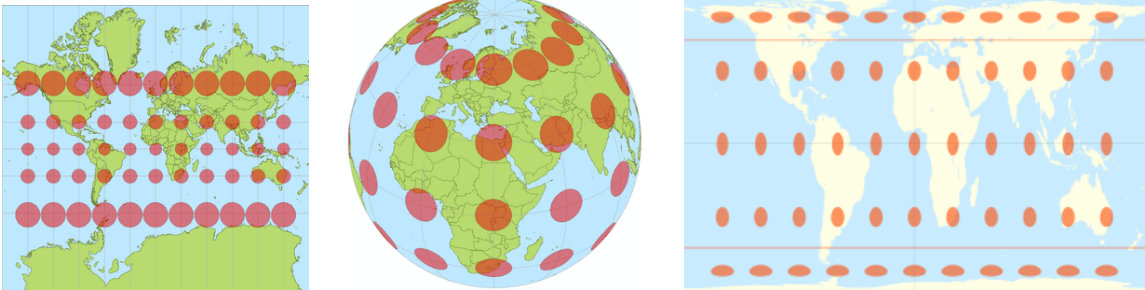
Vergleichbare “optimale” Erdkarten, also Karten von S^2 , zu finden ist nicht so einfach. Ein berühmter Satz von Gauß, von ihm selbst *Theorema Egregium* genannt, besagt, daß es keine längentreue Erdkarte geben kann. Es gibt aber winkeltreue und flächentreue Parametrisierungen (siehe Aufgabe 2 von Blatt 8).

Hier eine winkeltreue (links) und eine flächentreue Erdkarte (rechts), genannt Mercator- bzw. Gall-Peters Projektion:



(Bilder von Daniel R. Strebe, CC BY-SA 3.0)

Die winkeltreue Projektion (links) verzerrt Flächeninhalte im Vergleich zum original (Mitte), die flächentreue Karte (rechts) verzerrt Winkel und damit auch Kreise :



(Bilder von Stefan Kühn, CC BY-SA 3.0 und Eric Gaba, GFDL)

Definition 4.6. Man sagt

- eine *Kurve* ist eine Untermannigfaltigkeit der Dimension 1,
- eine *Fläche* ist eine Untermannigfaltigkeit der Dimension 2 und
- eine *Hyperfläche* ist eine Untermannigfaltigkeit $M^n \subset \mathbb{R}^{n+1}$, deren Dimension n sich um eins von der des umgebenden Raumes \mathbb{R}^{n+1} unterscheidet.

Eine Hyperfläche kann lokal als Nullstellenmenge einer reellwertigen Funktion h beschrieben werden, deren Ableitung keine Nullstellen hat. Beispiele für Kurven bzw. Hyperflächen sind Geraden bzw. lineare Hyperebenen.

Beispiel 4.7. Jeder lineare oder affin lineare Unterraum

$$M = p + V \subset \mathbb{R}^N,$$

wobei $V \subset \mathbb{R}^N$ ein n -dimensionaler Untervektorraum ist, ist eine Untermannigfaltigkeit von Dimension n .

- Ist $A \in GL_N(\mathbb{R})$ mit $V = A(\mathbb{R}^n \times \{0\})$, so ist $\Phi(x) = A^{-1}(x - p)$ ein Flachmacher für M wie in *i*) von Satz 4.4.
- Ist $v_1, \dots, v_n \in V$ eine Basis von V , so ist

$$\psi(t_1, \dots, t_n) = p + t_1 v_1 + \dots + t_n v_n$$

eine Parametrisierung von M wie in *iv*) von Satz 4.4. Die Inverse von ψ ist stetig, weil sie affin linear ist. Ist $n = 1$, so ist M eine Gerade und wir erhalten die “Punkt–Richtungs–Form” aus Beispiel 2.6.

- Ist $\alpha_1, \dots, \alpha_{N-n}$ eine Basis von

$$V^\perp = \{\alpha \in (\mathbb{R}^N)^* \mid \alpha(v) = 0 \text{ für alle } v \in V\},$$

so ist

$$h: \mathbb{R}^N \rightarrow \mathbb{R}^{N-n} \quad x \mapsto (\alpha_1(x - p), \dots, \alpha_{N-n}(x - p))$$

eine Abbildung mit $h^{-1}(\{0\}) = M$ wie in *ii*) von Satz 4.4. Ist $M^n \subset \mathbb{R}^{n+1}$ eine affine Hyperebene und $n \neq 0$ ein Vektor senkrecht auf V , so ist

$$M = \{x \in \mathbb{R}^{n+1} \mid \langle x - p, n \rangle = 0\}.$$

Dies ist die “Punkt–Normalen–Form” der Gleichung einer Hyperebene.

Definition 4.8. Ist $M \subset \mathbb{R}^N$ eine n -dimensionale Untermannigfaltigkeit, so ist der *Tangententialraum* $T_p M$ im Punkt $p \in M$ definiert als der Untervektorraum

$$T_p M = \ker(Dh|_p),$$

wobei h eine in einer Umgebung von p definierte Submersion ist, deren Nullstellenmenge M lokal beschreibt wie in *ii*) von Satz 4.4.

Beispiel 4.9. • Der Tangentialraum an einen affinen Unterraum

$$M = p + V \subset \mathbb{R}^N$$

ist der zugrundeliegende Vektorraum V .

- Der Tangentialraum an die Einheitssphäre S^n im Punkt $p \in S^n$ ist

$$T_p S^n = p^\perp = \{x \in \mathbb{R}^{n+1} \mid \langle x, p \rangle = 0\},$$

denn beschreibt man S^n als Nullstellenmenge von $h(x) = \langle x, x \rangle - 1$, so ist $Dh|_p(v) = 2\langle p, v \rangle$ und $\ker(Dh|_p) = p^\perp$. (BILD!)

Bemerkung 4.10. • Der Tangentialraum $T_p M$ ist ein Untervektorraum, wird jedoch meist visualisiert als der affine Unterraum $p + T_p M$, denn man erhält, indem man $T_p M$ in p verschiebt.

- Da h eine Submersion ist, ist die Dimension vom $T_p M$ gleich der Dimension der Untermannigfaltigkeit.
- Daß die Definition von $T_p M$ nicht von der Wahl von h abhängt, folgt aus der alternativen Charakterisierung von $T_p M$ durch das folgende Lemma 4.11.

Lemma 4.11. *Ist ψ eine lokale Parametrisierung einer Untermannigfaltigkeit $M \subset \mathbb{R}^N$ wie in iv) von Satz 4.4. Dann ist deren Tangentialraum im Punkt $p = \psi(\tilde{p})$ gegeben durch*

$$T_p M = \text{Im}(D\psi|_{\tilde{p}}).$$

Beweis. Sei h eine Submersion, deren Nullstellenmenge M in einer Umgebung von p beschreibt wie in ii) von Satz 4.4. Dann gilt

$$h \circ \psi = 0$$

und somit (mit der Kettenregel)

$$Dh|_p \circ D\psi|_{\tilde{p}} = 0.$$

Da $Dh|_p: \mathbb{R}^N \rightarrow \mathbb{R}^{N-n}$ surjektiv ist und $D\psi|_{\tilde{p}}: \mathbb{R}^n \rightarrow \mathbb{R}^N$ injektiv ist, gilt

$$\ker(Dh|_p) = \text{Im}(D\psi|_{\tilde{p}}),$$

was zu zeigen war. (Insbesondere ist $T_p M$ unabhängig von der Wahl von h und ψ .) □

Bemerkung 4.12. Für eine Untermannigfaltigkeit $M \subset \mathbb{R}^N$ von Dimension n gilt:

- die Spaltenvektoren

$$\frac{\partial \psi}{\partial y_1}(\tilde{p}), \dots, \frac{\partial \psi}{\partial y_n}(\tilde{p})$$

der Ableitung $D\psi|_{\tilde{p}}$ einer lokalen Parametrisierung ψ von M im Punkt $p = \psi(\tilde{p})$ (wie in *iv*) von Satz 4.4) bilden eine Basis von T_pM . (Ändert man die lokale Parametrisierung, so erhält man eine andere Basis. Im Allgemeinen hat der Vektorraum T_pM keine ausgezeichnete Basis.)

- die Zeilenvektoren

$$D(h_1)|_p, \dots, D(h_{N-n})|_p$$

der Ableitung $Dh|_p$ einer Submersion $h = (h_1, \dots, h_{N-n})$, deren Nullstellenmenge M in einer Umgebung von p beschreibt (wie in *ii*) von Satz 4.4), bilden eine Basis des Vektorraumes der auf T_pM verschwindenden Linearformen.

Die Vektoren

$$\text{grad}(h_1)(p), \dots, \text{grad}(h_m)(p)$$

bilden eine Basis des sogenannten *Normalenraums*

$$(T_pM)^\perp = \{v \in \mathbb{R}^N \mid \langle v, w \rangle = 0 \text{ für alle } w \in T_pM\}$$

von $M \subset \mathbb{R}^N$ in p .

Beispiel 4.13. Sei $f: U \subset \mathbb{R}^n \rightarrow \mathbb{R}$ eine reellwertige stetig differenzierbare Funktion auf einer offenen Menge $U \subset \mathbb{R}^n$.

- Die Kritischen Punkte von f , also die Punkte mit $Df|_p = 0$, sind genau die Punkte, in denen der Tangentialraum an den Graph von f

$$M = \{(x, f(x)) \in \mathbb{R}^{n+1} \mid x \in U\}$$

“horizontal”, also gleich $\mathbb{R}^n \times \{0\}$, ist. Denn für die Parametrisierung $\psi(x) = (x, f(x))$ von M ist nach Bemerkung 4.12

$$\frac{\partial \psi}{\partial x_1}(x) = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ \frac{\partial f}{\partial x_1}(x) \end{pmatrix}, \quad \dots, \quad \frac{\partial \psi}{\partial x_n}(x) = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ \frac{\partial f}{\partial x_n}(x) \end{pmatrix}$$

eine Basis von T_pM im Punkt $p = \psi(x)$.

- Weg von den kritischen Punkten sind die Niveaumengen von f , also die Mengen $\{x \mid f(x) = c\}$ für $c \in \mathbb{R}$, Untermannigfaltigkeiten (wegen Satz 4.4 oder auch direkt mit dem Satz über implizite Funktionen, also Satz 2.72). Der Gradient $\text{grad}(f)(p)$ steht in jedem nicht-kritischen Punkt senkrecht auf die entsprechende Niveauhyperebene (und ist eine Basis von deren Normalenraum, vgl. Bemerkung 4.12.)

4.3 Extrema mit Nebenbedingung

Satz 4.14. Seien $f: U \rightarrow \mathbb{R}$ und $h = (h_1, \dots, h_m): U \rightarrow \mathbb{R}^m$ stetig differenzierbare Abbildungen auf einer offenen Menge $U \subset \mathbb{R}^N$. Hat die Einschränkung $f|_M$ auf der Menge

$$M = \{x \in U \mid h(x) = 0\}$$

ein lokales Extremum in $p \in M$ und gilt $\text{Rang}(Dh|_p) = m$, so gibt es $\lambda_1, \dots, \lambda_m \in \mathbb{R}$ mit

$$Df|_p = \lambda_1 D(h_1)|_p + \dots + \lambda_m D(h_m)|_p. \quad (*)$$

Bemerkung 4.15. • Die Gleichung (*) wird oft in der äquivalenten Form

$$\text{grad}(f)(p) = \lambda_1 \text{grad}(h_1)(p) + \dots + \lambda_m \text{grad}(h_m)(p)$$

benutzt.

- Die $\lambda_1, \dots, \lambda_m$ nennt man *Lagrange Multiplikatoren*.
- In der Situation von Satz 4.14 spricht man auch von einem lokalem “Extremum unter der Nebenbedingung $h = 0$ ”.

Beweis von Satz 4.14. Wir geben zwei Versionen des Beweises.

1. Version (ohne Benutzung von Abschnitt 4.2) OBdA können wir annehmen (indem wir gegebenenfalls die Koordinaten von \mathbb{R}^N permutieren), daß die $m \times m$ -Matrix

$$\begin{pmatrix} \frac{\partial h_1}{\partial x_{N-m+1}}(p) & \dots & \frac{\partial h_1}{\partial x_N}(p) \\ \vdots & & \vdots \\ \frac{\partial h_m}{\partial x_{N-m+1}}(p) & \dots & \frac{\partial h_m}{\partial x_N}(p) \end{pmatrix} \quad (\#)$$

invertierbar ist. Nach dem Satz über implizite Funktionen (Satz 2.72) kann man die Lösungsmenge der Gleichung $h(x) = 0$ dann lokal um den Punkt p schreiben als Graph einer Abbildung $(x_{N-m+1}, \dots, x_N) = g(x_1, \dots, x_{N-m})$. Dann hat die Verkettung

$$f \circ \psi$$

mit $\psi(x_1, \dots, x_{N-m}) = (x_1, \dots, x_{N-m}, g(x_1, \dots, x_{N-m}))$ ein lokales Extremum im Punkt \tilde{p} mit $\psi(\tilde{p}) = p$ und es gilt (wegen Satz 2.61 und der Kettenregel)

$$Df|_p \circ D\psi|_{\tilde{p}} = 0. \quad (\#\#)$$

Andererseits gilt wegen $h \circ \psi = 0$ (mit der Kettenregel)

$$Dh|_p \circ D\psi|_{\tilde{p}} = 0.$$

Nun sind die Spalten von $D\psi|_{\tilde{p}}$ linear unabhängig und erzeugen einen Untervektorraum von Dimension $N - m$. Die Zeilen von $Dh|_p$ bilden eine Basis des

Vektorraums der Linearformen, die auf diesem $N - n$ -dimensionalen Unterraum verschwinden. Da wegen ($\#\#$) auch die Linearform $Df|_p$ auf diesem Untervektorraum verschwindet, ist $Df|_p$ eine Linearkombination der Zeilen von $Dh|_p$.

2. Version (mit Benutzung von Abschnitt 4.2) OBaA können wir durch Verkleinern von U annehmen, daß h eine Submersion ist, denn ist $Dh|_p$ surjektiv, so ist nach Permutation der Koordinaten die Matrix ($\#$) invertierbar, so daß, weil die Menge der invertierbaren Matrizen offen ist (z.B. wegen Lemma 4.1 oder wegen der Stetigkeit der Determinante), $Dh|_q$ surjektiv ist für q in einer Umgebung von p .

Damit ist $M = h^{-1}(\{0\})$ eine Untermannigfaltigkeit. Sei ψ eine lokale Parametrisierung von M mit $p = \psi(\tilde{p})$ wie in *iv*) von Satz 4.4. Dann ist wegen Satz 2.61 und der Kettenregel

$$Df|_p \circ D\psi|_{\tilde{p}} = 0 \quad (\#\#)$$

und $Df|_p$ verschwindet auf dem Tangentialraum $T_p M = \text{Im}(D\psi|_{\tilde{p}})$ von M in p . Da die Zeilen von $Dh|_p$ eine Basis des Vektorraums der auf $T_p M$ verschwindenden Linearformen bilden, ist $Df|_p$ eine Linearkombination der Zeilen von $Dh|_p$. (Äquivalent und etwas geometrischer besagt ($\#\#$), daß $\text{grad}(f)(p)$ senkrecht auf $T_p M$ steht. Da $\text{grad}(h_1)(p), \dots, \text{grad}(h_m)(p)$ eine Basis des Normalenraums $(T_p M)^\perp$ bildet, gilt die Behauptung.) \square

Beispiel 4.16. Wir betrachten die Extremwerte von

$$f(x, y, z) = xyz$$

unter der Nebenbedingung

$$h(x, y, z) = 0 \quad \text{mit} \quad h(x, y, z) = x^2 + y^2 + z^2 - 1.$$

Da $h^{-1}(\{0\}) = S^2$ kompakt ist, nimmt die stetige Funktion f unter dieser Nebenbedingung tatsächlich ein Maximum und ein Minimum an. Um deren genaue Lage mit Satz 4.14 zu bestimmen, berechnen wir:

$$Df = (yz, xz, xy) \quad Dh = (2x, 2y, 2z).$$

Gesucht sind nun Punkte $p = (x, y, z) \in S^2$, in denen $Df|_p$ und $Dh|_p$ linear Abhängig sind, also gilt $Df|_p = \lambda Dh|_p$ mit $\lambda \in \mathbb{R}$. Das ist äquivalent

$$yz = 2\lambda x \quad xz = 2\lambda y \quad xy = 2\lambda z.$$

Da wir oBdA annehmen dürfen, daß $x, y, z \neq 0$ (sonst wäre $f(x, y, z) = 0$, was weder Maximum noch Minimum ist, wie wir unten sehen werden), können wir die 1. Gleichung mit x multiplizieren, die 2. mit y und die 3. mit z und dann die drei Gleichungen summieren, was

$$3xyz = 2\lambda$$

ergibt. Setzt man das wieder in die drei Gleichungen ein, so erhält man

$$yz = 3x^2yz \quad xz = 3xy^2z \quad xy = 3xyz^2$$

und somit $x, y, z = \pm \frac{1}{\sqrt{3}}$. Damit erhalten wir 4 Punkte (die mit einer geraden Anzahl von +) mit

$$f\left(\pm \frac{1}{\sqrt{3}}, \pm \frac{1}{\sqrt{3}}, \pm \frac{1}{\sqrt{3}}\right) = \frac{1}{3\sqrt{3}}$$

sowie 4 Punkte (die mit einer ungeraden Anzahl von +) mit

$$f\left(\pm \frac{1}{\sqrt{3}}, \pm \frac{1}{\sqrt{3}}, \pm \frac{1}{\sqrt{3}}\right) = -\frac{1}{3\sqrt{3}}.$$

Diese 8 Punkte sind die Maxima bzw. Minima von f unter der Nebenbedingung $h = 0$.

Geometrisch entsprechen diese 8 Punkte den Eckpunkten eines Würfels. Dieser Würfel ist der achsenparallele Quader größtmöglichem Volumens mit Eckpunkten auf der Sphäre S^2 . (Denn $|f(x, y, z)|$ ist bis auf einen Vorfaktor genau gleich dem Volumen des Achsenparallelen Quaders mit Eckpunkten $(\pm x, \pm y, \pm z)$.)