

# Analysis II

Sebastian Baader

Frühlingssemester 2021

## Über diese Vorlesung

Diese Vorlesung besteht aus drei Teilen, die in das Studium von Funktionen in mehreren reellen Variablen einführen soll. Der ganze Inhalt dieser Vorlesung (und mehr) ist in [1] zu finden, konkret in den Kapiteln XIV, XV für gewöhnliche Differentialgleichungen, XX für die Differentialrechnung, und XXI, XXIV für Gradientenfelder und Differentialformen.

Im Kapitel I geht es um “gewöhnliche Differentialgleichungen”. Wir studieren dort Kurven  $\gamma: \mathbb{R} \rightarrow \mathbb{R}^n$ , deren Ableitung durch die Position  $\gamma(t)$  bestimmt ist. Wir werden Vektorfelder einführen um das präziser zu formulieren, und diese dann untersuchen. Im Kapitel II geht es um differenzierbare Ableitungen  $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$  (was sich von Kurven dadurch unterscheidet, dass auch der Definitionsbereich dieser Funktionen mehrdimensional sein darf). Dieses Kapitel wird den grössten Teil dieser Vorlesung formen. Im Kapitel III lernen wir Gradientenfelder und Differentialformen kennen. Dort studieren wir Abbildungen  $f: \mathbb{R}^n \rightarrow \mathbb{R}$ . Die leitende Frage in diesem Kapitel wird sein, welche Vektorfelder  $X: \mathbb{R}^n \rightarrow \mathbb{R}^n$  Gradientenfelder sind.

## Über dieses Dokument

Das ist eine Mitschrift der Vorlesung “Analysis 2” von Prof. Dr. Sebastian Baader im Frühlingssemester 2021. Du darfst sie so verwenden, wie sie dir am meisten beim Verständnis des Materials hilft. Verantwortlich dafür was hier drin steht ist Levi Ryffel. Denke daran dass der Dozent dieses Dokument nicht schreibt (und vielleicht auch nicht liest). Ihn trifft keine Verantwortung, falls Unsinn steht.

## Dein Beitrag zu den Notizen

Diese Vorlesungsnotizen werden in Echtzeit während der Vorlesung mitgeschrieben und werden deshalb viele Probleme enthalten. Damit sind allerlei Missgeschicke gemeint wie zum Beispiel Symbolverwechslungen, unpräzise Aussagen und Argumente, alternative Rechtschreibung und Grammatik, oder unattraktives Layout. Falls dir so etwas auffällt, auch wenn es dich nicht stark stört, und auch wenn du es als etwas subjektiv empfindest, poste doch auf

<https://github.com/raw-bacon/ana2-notes>,

ein “Issue”, oder sende eine E-Mail an [levi.ryffel@math.unibe.ch](mailto:levi.ryffel@math.unibe.ch). Auf demselben Weg kannst du Wünsche und Verbesserungsvorschläge zu dieser Mitschrift anbringen.

# Inhaltsverzeichnis

<b>I</b>	<b>Gewöhnliche Differentialgleichungen</b>	<b>3</b>
1	Normen auf reellen Vektorräumen . . . . .	3
2	Stetigkeit . . . . .	7
3	Der Fixpunktsatz von Banach . . . . .	12
4	Kurven und Vektorfelder . . . . .	17
5	Der Satz von Cauchy-Lipschitz-Picard-Lindelöf . . . . .	22
6	Lineare Differentialgleichungen . . . . .	26
7	Differentialgleichungen in einer Variable . . . . .	29
8	Rektifizierbare Kurven . . . . .	31
<b>II</b>	<b>Differentialrechnung</b>	<b>36</b>
1	Differenzierbarkeit . . . . .	36
2	Die Kettenregel . . . . .	43
3	Der Umkehrsatz . . . . .	50
4	Der Satz über implizite Funktionen . . . . .	58
5	Die Hessische Form . . . . .	63
<b>III</b>	<b>Gradientenfelder und Differentialformen</b>	<b>69</b>

# Kapitel I

## Gewöhnliche Differentialgleichungen

Bevor wir Differentialgleichungen studieren können, müssen wir den Raum  $\mathbb{R}^n$  besser verstehen. Wir erinnern uns nun an das Studium von Funktionen in einer Variable. Dort machen viele Definitionen vom Absolutbetrag Gebrauch.

### Beispiele.

(1) Konvergenz einer Folge  $a: \mathbb{N} \rightarrow \mathbb{R}$ , in Symbolen

$$\lim_{n \rightarrow \infty} a(n) = \alpha \in \mathbb{R}$$

heisst Folgendes. Für alle  $\varepsilon > 0$  existiert  $N \in \mathbb{N}$ , sodass für alle  $n \geq N$  gilt, dass  $|a(n) - \alpha| \leq \varepsilon$ .

(2) Stetigkeit einer Funktion  $f: \mathbb{R} \rightarrow \mathbb{R}$  im Punkt  $p \in \mathbb{R}$  heisst, dass für alle  $\varepsilon > 0$  ein  $\delta > 0$  existiert, so dass für alle  $q \in \mathbb{R}$  mit  $|q - p| \leq \delta$  gilt, dass  $|f(q) - f(p)| \leq \varepsilon$ .

Beide dieser sehr zentralen Konzepte machen kritischen Gebrauch des Absolutbetrags. In  $\mathbb{R}^n$  gibt es aber keinen kanonischen Ersatz für diesen. Dies motiviert unseren ersten Abschnitt in diesem Kapitel.

## 1 Normen auf reellen Vektorräumen

Die Normaxiome greifen die wichtigsten Eigenschaften des Absolutbetrags in  $\mathbb{R}$  auf und verallgemeinern diese, so dass wir in allgemeinen reellen Vektorräumen Konzepte wie Konvergenz und Stetigkeit formalisieren können.

**Definition.** Sei  $V$  ein Vektorraum über  $\mathbb{R}$ . Eine Abbildung  $\|\cdot\|: V \rightarrow \mathbb{R}$  heisst *Norm* auf  $V$ , falls folgende Eigenschaften erfüllt werden.

- (i) (Strikte Positivität) Für alle  $v \in V$  gilt  $\|v\| \geq 0$  und  $\|v\| = 0$  genau dann, wenn  $v = 0$ .

(ii) (Homogenität) Für alle  $v \in V$  und alle  $\lambda \in \mathbb{R}$  gilt  $\|\lambda v\| = |\lambda| \cdot \|v\|$ .

(iii) (Dreiecksungleichung) Für alle  $v, w \in V$  gilt  $\|v + w\| \leq \|v\| + \|w\|$ .

Eigenschaft (ii) hätten wir auch folgendermassen formulieren können: Die Norm  $\|\cdot\|$  auf einem eindimensionalen Unterraum von  $V$  verhält sich (bis auf Streckung) genau so wie der Absolutbetrag auf  $\mathbb{R}$ .

**Definition.** Sei  $\|\cdot\|$  eine Norm auf einem reellen Vektorraum  $V$ . Die Menge

$$B_1 = \{v \in V \mid \|v\| \leq 1\} \subset V$$

heisst *Norm-Einheitsball*.

**Beispiele.** Sei  $V = \mathbb{R}^n$  mit Standardbasis  $e_1, \dots, e_n$ . Einen Vektor  $v \in V$  können wir dann ausdrücken durch  $v = v_1 e_1 + \dots + v_n e_n$  mit  $v_i \in \mathbb{R}$ .

(1) Die *Summennorm* ist die Norm

$$\|v\|_1 = |v_1| + \dots + |v_n|.$$

Wir prüfen nun die Normaxiome.

(i) Dank den Eigenschaften des Absolutbetrags auf  $\mathbb{R}$  haben wir sofort  $\|v\|_1 \geq 0$  und  $\|v\|_1 = 0$  genau dann, wenn alle  $v_i$  null sind.

(ii) Berechne  $\|\lambda v\|_1 = |\lambda v_1| + \dots + |\lambda v_n| = |\lambda| \cdot \|v\|_1$ .

(iii) Berechne

$$\begin{aligned} \|v + w\|_1 &= |v_1 + w_1| + \dots + |v_n + w_n| \\ &\leq |v_1| + |w_1| + \dots + |v_n| + |w_n| \\ &= \|v\|_1 + \|w\|_1. \end{aligned}$$

(2) Die *Maximumnorm* ist die Norm

$$\|v\|_\infty = \max\{|v_1|, \dots, |v_n|\}.$$

Wir prüfen wieder die Normaxiome.

(i) Wir haben  $\|v\|_\infty \geq 0$  und auch  $\|v\|_\infty = 0$  genau dann, wenn alle  $v_i$  null sind.

(ii) Es gilt  $\|\lambda v\|_\infty = \max\{|\lambda v_1|, \dots, |\lambda v_n|\} = |\lambda| \cdot \|v\|_\infty$ .

(iii) Berechne

$$\begin{aligned} \|v + w\|_\infty &= \max\{|v_1 + w_1|, \dots, |v_n + w_n|\} \\ &\leq \max\{|v_1|, \dots, |v_n|\} + \max\{|w_1|, \dots, |w_n|\} \\ &= \|v\|_\infty + \|w\|_\infty, \end{aligned}$$

da jeweils  $|v_i + w_i| \leq |v_i| + |w_i|$  gilt.

(3) die *euklidische Norm* ist die Norm

$$\|v\|_2 = \sqrt{v_1^2 + \cdots + v_n^2}.$$

Auch für diese Norm prüfen wir die Axiome.

- (i) Es gilt  $\|v\|_2 \geq 0$  und  $\|v\|_2 = 0$  genau dann, wenn  $v_1^2 + \cdots + v_n^2 = 0$  gilt, was äquivalent dazu ist, dass alle  $v_i$  null sind.
- (ii) Berechne  $\|\lambda v\|_2 = |\lambda| \cdot \|v\|_2$ , da  $\sqrt{\lambda^2} = |\lambda|$  gilt.
- (iii) Hier stoßen wir zum ersten Mal auf Schwierigkeiten. Wir werden das im Lemma unten zeigen.

(4) Folgendes Beispiel rechtfertigt die Notation für obige Normen. Sei  $p \geq 1$  für  $p \geq 1$ . Die  $p$ -Norm ist

$$\|v\|_p = \sqrt[p]{|v_1|^p + \cdots + |v_n|^p}.$$

Die Dreiecksungleichung  $\|v + w\|_p \leq \|v\|_p + \|w\|_p$  heisst *Minkowski-Ungleichung*, die aus der Konkavität von  $\log$  folgt. Siehe hier Abschnitt 59.3 in [1]. Die Normen  $\|\cdot\|_1$  und  $\|\cdot\|_2$  sind Spezialfälle dieser Familie von Normen. Für alle  $p \geq 1$  gilt, dass  $\|v\|_\infty \leq \|v\|_p \leq \sqrt[p]{n} \|v\|_\infty$ . Im Grenzwert  $p \rightarrow \infty$  erhalten wir

$$\lim_{p \rightarrow \infty} \|v\|_p = \|v\|_\infty$$

da  $\lim_{p \rightarrow \infty} \sqrt[p]{n} = 1$ .

Die Normbälle der ersten drei Normen im Fall  $n = 2$  sind in Abbildung I.1 zu sehen.



Abbildung I.1: Normbälle der Normen  $\|\cdot\|_1, \|\cdot\|_\infty, \|\cdot\|_2$

## Normen aus Skalarprodukten

**Definition.** Ein *Skalarprodukt* auf einem reellen Vektorraum  $V$  ist eine strikt positive, symmetrische, bilineare Abbildung  $\langle \cdot, \cdot \rangle: V \times V \rightarrow \mathbb{R}$ . Das heisst,

- (i) für alle  $v \in V$  gilt  $\langle v, v \rangle \geq 0$  und  $\langle v, v \rangle = 0$  genau dann, wenn  $v = 0$ ,

- (ii) für alle  $v, w \in V$  gilt  $\langle v, w \rangle = \langle w, v \rangle$ ,
- (iii) für alle  $u, v, w \in V$  und  $t \in \mathbb{R}$  gilt  $\langle u, v + tw \rangle = \langle u, v \rangle + t\langle u, w \rangle$ .

**Lemma.** Sei  $\langle \cdot, \cdot \rangle$  ein Skalarprodukt auf  $V$ . Dann ist die Abbildung

$$\begin{aligned} \|\cdot\|: V &\rightarrow \mathbb{R} \\ v &\mapsto \sqrt{\langle v, v \rangle} \end{aligned}$$

eine Norm auf  $V$ .

*Beweis.* Wir prüfen die Normaxiome folgendermassen.

- (i) Es gilt  $\|v\| = \sqrt{\langle v, v \rangle} \geq 0$  und  $\|v\| = 0$  genau dann, wenn  $\langle v, v \rangle = 0$ , das heisst,  $v = 0$  gilt.
- (ii) Berechne  $\|\lambda v\| = \sqrt{\langle \lambda v, \lambda v \rangle} = \sqrt{\lambda^2 \cdot \langle v, v \rangle} = |\lambda| \cdot \|v\|$ .
- (iii) Seien  $v, w \in V$ . Wir wollen zeigen, dass  $\sqrt{\langle v + w, v + w \rangle} \leq \sqrt{\langle v, v \rangle} + \sqrt{\langle w, w \rangle}$ . Da Quadrieren auf  $\mathbb{R}_{\geq 0}$  monoton ist, reicht es zu zeigen, dass

$$\langle v + w, v + w \rangle \leq \langle v, v \rangle + \langle w, w \rangle + 2 \cdot \sqrt{\langle v, v \rangle} \cdot \sqrt{\langle w, w \rangle}.$$

Ausmultiplizieren der linken Seite und Subtraktion der Terme, die dann auf beiden Seiten erscheinen liefert, dass die Dreiecksungleichung für  $\|\cdot\|$  äquivalent zur *Cauchy-Schwarz Ungleichung*

$$\langle v, w \rangle \leq \sqrt{\langle v, v \rangle} \cdot \sqrt{\langle w, w \rangle}$$

ist. Wir beweisen nun diese. Betrachte dazu die Funktion  $h: \mathbb{R} \rightarrow \mathbb{R}$ , die durch

$$h(t) = \langle v + tw, v + tw \rangle \geq 0$$

definiert ist. Es gilt also  $h(t) = t^2 \langle w, w \rangle + 2t \langle v, w \rangle + \langle v, v \rangle$ . Dies beschreibt eine Parabel mit Diskriminante

$$D = b^2 - 4ac = 4\langle v, w \rangle^2 - 4\langle w, w \rangle \cdot \langle v, v \rangle.$$

Aus  $h(t) \geq 0$  für alle  $t \in \mathbb{R}$  folgt, dass  $D \leq 0$  ist. Wir schliessen, dass  $\langle v, w \rangle^2 \leq \langle w, w \rangle \cdot \langle v, v \rangle$  gelten muss.  $\square$

**Beispiele.**

- (1) Sei  $V = \mathbb{R}^n$  und

$$\langle v, w \rangle = v_1 w_1 + \cdots + v_n w_n$$

das *Standardskalarprodukt*. Dann ist

$$\|v\| = \sqrt{v_1^2 + \cdots + v_n^2} = \|v\|_2$$

die euklidische Norm.

- (2) Sei  $V = C[0, 1]$  der Raum von stetigen Funktionen  $f: [0, 1] \rightarrow \mathbb{R}$ . Das ist ein unendlichdimensionaler Vektorraum, da die Funktionen  $1, x, x^2, \dots$  linear unabhängig sind. Für  $f, g \in V$  definieren wir

$$\langle f, g \rangle = \int_0^1 f(x) \cdot g(x) dx.$$

Die Skalarproduktaxiome sind leicht zu überprüfen.

**Bemerkung.** Nicht jede Norm auf  $V$  stammt von einem Skalarprodukt. In den Übungen wird gezeigt, dass jede von einem Skalarprodukt induzierte Norm  $\|\cdot\|$  die *Parallelogrammidentität*

$$\|v + w\|^2 + \|v - w\|^2 = 2(\|v\|^2 + \|w\|^2)$$

erfüllt, und dass die Normen  $\|\cdot\|_1$  und  $\|\cdot\|_\infty$  diese nicht erfüllen.

## 2 Stetigkeit

**Definition.** Seien  $V$  und  $W$  reelle Vektorräume mit Normen  $\|\cdot\|_V$  und  $\|\cdot\|_W$ . Eine Abbildung  $f: V \rightarrow W$  heisst *stetig im Punkt*  $p \in V$ , falls für alle  $\varepsilon > 0$  ein  $\delta > 0$  existiert, so dass für alle  $q \in V$  mit  $\|q - p\|_V \leq \delta$  folgt, dass  $\|f(q) - f(p)\|_W \leq \varepsilon$ . Wir sagen, dass  $f$  *stetig* ist, falls  $f$  in allen Punkten von  $V$  stetig ist.

**Beispiel.** Im Fall  $V = W = \mathbb{R}$  und  $\|\cdot\| = |\cdot|$  erhalten wir den üblichen Stetigkeitsbegriff.

Wir untersuchen nun, wie der Stetigkeitsbegriff von den Normen auf  $V$  und  $W$  abhängt. Betrachte zunächst folgenden günstigen Fall. Sei  $V = W = \mathbb{R}$ . Sei  $\|\cdot\|$  eine Norm auf  $\mathbb{R}$ . Für alle  $x \in \mathbb{R}$  gilt dann

$$\|x\| = \|x \cdot 1\| = |x| \cdot \|1\|.$$

Sei nun  $f: \mathbb{R} \rightarrow \mathbb{R}$  bezüglich der Norm  $|\cdot|$  im Punkt  $p \in \mathbb{R}$  stetig. Wir zeigen nun, dass  $f$  auch bezüglich der Norm  $\|\cdot\|$  im Punkt  $p$  stetig ist. Wähle  $\delta > 0$  so, dass für alle  $q \in \mathbb{R}$  mit  $|q - p| \leq \delta/\|1\|$  gilt, dass  $|f(q) - f(p)| \leq \varepsilon/\|1\|$ . Das geht, da  $\|1\| > 0$  aus  $1 \neq 0$  folgt. Sei nun  $q \in \mathbb{R}$  mit  $\|q - p\| \leq \delta$ . Dann gilt  $|q - p| \leq \delta/\|1\|$ . Wir schliessen, dass  $|f(q) - f(p)| \leq \varepsilon/\|1\|$ , also  $\|f(q) - f(p)\| = |f(q) - f(p)| \cdot \|1\| \leq \varepsilon$ . Insgesamt hängt also der Stetigkeitsbegriff auf  $\mathbb{R}$  nicht von der Norm auf  $\mathbb{R}$  ab.

Es gibt aber auch einen ungünstigen Fall. Sei  $\mathbb{R}^\infty$  der Raum aller Folgen  $(x_1, x_2, \dots) \in \mathbb{R}^\mathbb{N}$  für welche  $N > 0$  existiert, so dass für alle  $n \geq N$  gilt, dass  $x_n = 0$ . Betrachte auf  $\mathbb{R}^\infty$  die beiden Normen

$$\begin{aligned} \|v\|_2 &= \sqrt{v_1^2 + v_2^2 + \dots} \\ \|v\|_\infty &= \max\{|v_1|, |v_2|, \dots\}. \end{aligned}$$



Es gilt für alle  $v \in \mathbb{R}^\infty$ , dass  $\|v\|_\infty \leq \|v\|_2$ . Sei  $V = (\mathbb{R}^\infty, \|\cdot\|_2)$  und  $W = (\mathbb{R}^\infty, \|\cdot\|_\infty)$ . Die Identitätsabbildung

$$\text{Id}: \mathbb{R}^\infty \rightarrow \mathbb{R}^\infty$$

$$v \mapsto v$$

kann auf vier Arten als Abbildung zwischen normierten Räumen betrachtet werden, nämlich

- $f_1: V \rightarrow V$ ,
- $f_2: W \rightarrow W$ ,
- $f_3: V \rightarrow W$ ,
- $f_4: W \rightarrow V$ .

Wir bemerken, dass  $f_1$ ,  $f_2$  und  $f_3$  stetig sind, indem wir  $\delta = \varepsilon$  wählen. Aber  $f_4$  ist nicht stetig!. Betrachte dazu die Folge von Punkten  $p_n = (1/n, 1/n, \dots, 1/n, 0, \dots) \in \mathbb{R}^\infty$ , wobei das Glied  $1/n$  gerade  $n^2$  mal vorkommt. Es gilt  $\|p_n\|_\infty = 1/n$  und  $\|p_n\|_2 = 1$ . Insbesondere gilt

$$\|f_4(p_n) - f_4(0)\|_2 = \|p_n\|_2 = 1$$

und  $\|p_n\|_\infty = 1/n$ , also ist  $f_4$  im Punkt  $p = 0$  nicht stetig.

Zusammengefasst erkennen wir, dass auf  $\mathbb{R}^\infty$  das Verhältnis der Normen  $\|\cdot\|_2$  und  $\|\cdot\|_\infty$  beliebig hohe Werte annimmt. Im Sinne folgender Definition bedeutet das, dass in  $\mathbb{R}^\infty$  diese beiden normen nicht äquivalent sind.

**Definition.** Zwei Normen  $\|\cdot\|$ ,  $\widetilde{\|\cdot\|}$  heißen *äquivalent*, falls es Konstanten  $0 < c \leq C$  gibt, so dass für alle  $v \in V \setminus \{0\}$  gilt, dass

$$c \leq \frac{\widetilde{\|v\|}}{\|v\|} \leq C.$$

Bildlich können wir uns die Äquivalenz von Normen so vorstellen, dass der  $\widetilde{\|\cdot\|}$ -Normball sowohl eine skalierte Version vom  $\|\cdot\|$ -Normball enthält, wie auch in einer skalierten Version vom  $\|\cdot\|$ -Normball enthalten ist.

### Beispiele.

(1) Sei  $\|\cdot\|$  eine Norm auf  $\mathbb{R}$ . Dann gilt für alle  $x \in \mathbb{R}$ , dass  $\|x\| = \|1\| \cdot |x|$ , also sind  $\|\cdot\|$  und  $|\cdot|$  äquivalent.

(2) Für alle  $v \in \mathbb{R}^n$  gilt, dass

$$\|v\|_\infty \leq \|v\|_2 \leq \sqrt{n} \cdot \|v\|_\infty.$$

also sind die beiden Normen  $\|\cdot\|_\infty$  und  $\|\cdot\|_2$  äquivalent auf  $\mathbb{R}^n$ .

(3) Auf  $\mathbb{R}^\infty$  sind die Normen  $\|\cdot\|_\infty$  und  $\|\cdot\|_2$  nicht äquivalent.

**Bemerkung.** Die Äquivalenz von Normen ist eine Äquivalenzrelation. Die Symmetrie und die Reflexivität sind leicht zu überprüfen. Für die Transitivität bemerke, dass aus  $a \leq x/y \leq A$  und  $b \leq y/z \leq B$  folgt, dass  $ab \leq x/z \leq AB$ .

**Theorem.** Auf  $\mathbb{R}^n$  sind alle Normen äquivalent.

*Beweis.* Es reicht zu zeigen, dass jede Norm  $\|\cdot\|$  auf  $\mathbb{R}^n$  zur euklidischen Norm  $\|\cdot\|_2$  äquivalent ist. In einem ersten Schritt zeigen wir, dass eine Konstante  $C > 0$  existiert, so dass  $\|\cdot\|/\|\cdot\|_2 \leq C$  gilt. Sei  $v = v_1 e_1 + \cdots + v_n e_n$  mit  $v_i \in \mathbb{R}$ . Schätze ab, dass

$$\begin{aligned}\|v\| &\leq \|v_1 e_1\| + \cdots + \|v_n e_n\| \\ &= |v_1| \cdot \|e_1\| + \cdots + |v_n| \cdot \|e_n\| \\ &\leq \|v\|_2 \cdot (\|e_1\| + \cdots + \|e_n\|).\end{aligned}$$

Wir erhalten das gewünschte Resultat, indem wir  $C = \|e_1\| + \cdots + \|e_n\|$  setzen, dann gilt nämlich  $\|v\| \leq C \cdot \|v\|_2$ .

Wir suchen nun im zweiten Schritt eine Konstante  $c > 0$ , für die  $\|\cdot\|/\|\cdot\|_2 \geq c$  gilt. Nimm widerspruchswise an, dass das Verhältnis  $\|\cdot\|/\|\cdot\|_2$  beliebig kleine Werte annimmt. Dann gibt es eine Folge von Vektoren  $v_k \in \mathbb{R}^n \setminus \{0\}$  mit  $\|v_k\|/\|v_k\|_2 \leq 1/k$ . Für die normierten Vektoren  $w_k = v_k/\|v_k\|_2 \in \mathbb{R}^n \setminus \{0\}$  gilt dann, dass  $\|w_k\|_2 = 1$  und  $\|w_k\| \leq 1/k$ . Somit liegen alle Koordinaten von  $w_k$  im Intervall  $[-1, 1]$ . Also existiert eine Teilfolge der  $w_k$ , so dass die erste Koordinatenfolge mit Grenzwert in  $[-1, 1]$  konvergiert. Wir iterieren dieses Verfahren  $n$ -mal, und erhalten eine Teilfolge  $w_{k_i}$  der  $w_k$ , welche in allen Koordinaten konvergiert. Setze  $w = \lim_{i \rightarrow \infty} w_{k_i}$ . Wir behaupten nun, dass  $\|w\|_2 = 1$  und  $\|w\| = 0$  gilt, was natürlich den Widerspruch  $w = 0$  impliziert. Bemerke dazu, dass für alle  $i \in \mathbb{N}$  gilt, dass

$$\begin{aligned}1 - \|w - w_{k_i}\|_2 &= \|w_{k_i}\|_2 - \|w - w_{k_i}\|_2 \\ &\leq \|w\|_2 \\ &\leq \|w - w_{k_i}\|_2 + \|w_{k_i}\|_2 \\ &= 1 + \|w - w_{k_i}\|_2.\end{aligned}$$

Da die  $w_{k_i}$  koordinatenweise gegen  $w$  konvergieren, folgt, dass

$$\lim_{i \rightarrow \infty} \|w - w_{k_i}\|_2 = 0$$

gilt. Schätze weiterhin ab, dass

$$\|w\| \leq \|w_{k_i}\| + \|w - w_{k_i}\|,$$

also existiert nach Schritt 1 eine Konstante  $C > 0$ , sodass  $\|w\| \leq 1/k_i + C \cdot \|w - w_{k_i}\|_2$  gilt, woraus im Grenzwert  $i \rightarrow \infty$  folgt, dass  $\|w\| = 0$  gilt.  $\square$

**Korollar 1.** Der Stetigkeitsbegriff für Abbildungen  $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$  hängt nicht von den Normen auf  $\mathbb{R}^n$  und  $\mathbb{R}^m$  ab.

Der Beweis hier lässt sich durch einfache Adjustierung der  $\delta$  und  $\varepsilon$  führen. Siehe dazu unseren Beweis für den Fakt, dass alle Normen auf  $\mathbb{R}$  zum selben Stetigkeitsbegriff führen.

## Lineare Abbildungen

**Frage.** Seien  $V, W$  normierte Vektorräume und  $A: V \rightarrow W$  linear. Ist  $A$  dann stetig?

Die Antwort ist im allgemeinen nein, was unsere Abbildung  $f_4$  oben zeigt. Wir untersuchen nun noch ein weiteres Beispiel.

**Beispiel.** Sei  $V = \mathbb{R}^\infty$  mit der Maximumsnorm  $\|\cdot\|_\infty$ . Definiere  $A: \mathbb{R}^\infty \rightarrow \mathbb{R}^\infty$  als die lineare Erweiterung der Abbildung  $e_k \mapsto ke_k$ . Setze  $q_k = e_k/k \in \mathbb{R}^\infty$ . Es gilt, dass  $\|q_k - 0\|_\infty = \|q_k\|_\infty = 1/k$  und  $\|A(q_k) - A(0)\|_\infty = \|e_k - 0\|_\infty = 1$ . Somit ist  $A$  im Nullpunkt (und auch in allen anderen Punkten) nicht stetig.

**Definition.** Seien  $\|\cdot\|_V, \|\cdot\|_W$  Normen auf normierten Vektorräumen  $V, W$ . Sei  $A: V \rightarrow W$  linear. Die *Operatornorm* von  $A$  ist

$$\|A\|_{\text{op}} = \sup \left\{ \frac{\|Av\|_W}{\|v\|_V} \mid v \in V \setminus \{0\} \right\} \in \mathbb{R} \cup \{\infty\}.$$

**Beispiele.**

(1) Betrachte die Abbildung

$$\begin{aligned} A: \mathbb{R}^\infty &\rightarrow \mathbb{R}^\infty \\ e_k &\mapsto ke_k \end{aligned}$$

auf  $\mathbb{R}^\infty$  mit der Norm  $\|\cdot\|_\infty$ . Dann gilt  $\|A\|_{\text{op}} = \infty$ .

(2) Sei  $A: \mathbb{R}^2 \rightarrow \mathbb{R}^2$  auf  $\mathbb{R}^2$  mit der euklidischen Norm  $\|\cdot\|_2$  die Abbildung, die durch die Matrix

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

gegeben ist. Bemerke folgende Eigenschaften.

(i)  $A = 0$  genau dann, wenn  $\|A\|_{\text{op}} = 0$ .

(ii) Betrachte

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

Sei  $v = (x, y)$ . Es gilt dann  $Av = (y, 0)$ , also gilt

$$\frac{\|Av\|_2}{\|v\|_2} = \frac{y}{\sqrt{x^2 + y^2}} \leq 1.$$

Für  $x = 0$  erhalten wir 1 für diesen Quotienten, also gilt  $\|A\|_{\text{op}} = 1$ .

(iii) Betrachte die Matrix

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

für  $d \geq 1$ . Um  $\|A\|_{\text{op}}$  zu bestimmen reicht es, Vektoren der Form  $v = (1, 0)$  und  $v = (x, 1)$  zu betrachten. Es gilt  $A(1, 0) = (1, 0)$ , also ist  $\|A\|_{\text{op}} \geq 1$ . Weiter ist  $A(x, 1) = (x, d)$ . Wir schliessen, dass für  $v = (x, 1)$  gilt, dass

$$\frac{\|Av\|_2}{\|v\|_2} = \frac{\sqrt{x^2 + d^2}}{\sqrt{x^2 + 1^2}} \leq d.$$

Aber  $d$  wird im Fall  $x = 0$  angenommen, also folgt  $\|A\|_{\text{op}} = d$ .

**Proposition 1.** Sei  $A: V \rightarrow W$  eine lineare Abbildung zwischen normierten Vektorräumen. Dann ist  $A$  stetig, genau dann, wenn  $\|A\|_{\text{op}} < \infty$  gilt.

*Beweis.* Seien  $p \in V$  und  $\varepsilon > 0$  vorgegeben. Setze

$$\delta = \frac{\varepsilon}{\|A\|_{\text{op}} + 1}.$$

Sei nun  $q \in V$  mit  $\|q - p\|_V \leq \delta$ . Dann gilt

$$\begin{aligned} \|A(q) - A(p)\|_W &= \|A(q - p)\|_W \\ &\leq \|A\|_{\text{op}} \cdot \|q - p\|_V \\ &\leq \|A\|_{\text{op}} \cdot \frac{\varepsilon}{\|A\|_{\text{op}} + 1} \\ &\leq \varepsilon. \end{aligned}$$

Somit ist  $A$  stetig im Punkt  $p \in V$ . Für die Umkehrung, siehe Serie 2. □

**Spezialfall.** Sei  $A: \mathbb{R}^n \rightarrow \mathbb{R}^m$  linear. Schätze  $\|A\|_{\text{op}}$  bezüglich der Maximumsnorm  $\|\cdot\|_{\infty}$  auf  $\mathbb{R}^n$  und  $\mathbb{R}^m$  wie folgt ab. Schreibe

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}$$

als Matrix, das heisst  $A(e_i) = \sum_{j=1}^m a_{ij}e_j$ . Setze  $a = \max\{|a_{ij}| \mid 1 \leq i \leq m, 1 \leq j \leq n\}$ . Für  $v \in \mathbb{R}^n$  gilt dann  $\|Av\|_{\infty} \leq n \cdot a \cdot \|v\|_{\infty}$ . Insbesondere ist  $\|A\|_{\text{op}} \leq n \cdot a$  endlich. Wir schliessen, dass  $A: \mathbb{R}^n \rightarrow \mathbb{R}^m$  stetig ist.

**Korollar 2.** Lineare Abbildungen  $A: \mathbb{R}^n \rightarrow \mathbb{R}^m$  sind stetig.

*Beweis.* Wir haben soeben gezeigt, dass  $A$  bezüglich der Maximumsnorm stetig ist. Aus Korollar 1 folgt, dass  $A$  bezüglich jeder Norm stetig ist. □

### 3 Der Fixpunktsatz von Banach

**Motivation.** Sei  $V$  ein normierter reeller Vektorraum, und sei  $A: V \rightarrow V$  linear und *kontrahierend*, das heisst, es gilt  $\|A\|_{\text{op}} < 1$ . Wir wollen die Abbildung  $\text{Id} - A: V \rightarrow V$  wie folgt invertieren:

$$(\text{Id} - A)^{-1} = \text{Id} + A + A^2 + A^3 + \dots$$

Als Analogie dazu erinnern wir uns an den Fakt, dass für  $q \in \mathbb{R}$  mit  $|q| < 1$  gilt, dass

$$\frac{1}{1 - q} = 1 + q + q^2 + \dots$$

Für alle  $Y \in V$  suchen wir ein eindeutiges  $X \in V$  mit  $(\text{Id} - A)(X) = Y$ , also  $X - A(X) = Y$ . Definiere

$$\begin{aligned} f: V &\rightarrow V \\ X &\mapsto A(X) + Y. \end{aligned}$$

Gesucht ist also eine (offenbar eindeutige) Lösung der Fixpunktgleichung  $f(X) = X$ .

**Bemerkung.** Die Bedingung  $\|A\|_{\text{op}} < 1$  ist stärker als die Bedingung  $\|Av\|_V < \|v\|_V$  für alle  $v \in V \setminus \{0\}$ . Betrachte dazu  $V = \mathbb{R}^\infty$  mit der Maximumnorm  $\|\cdot\|_\infty$ . Definiere  $A: \mathbb{R}^\infty \rightarrow \mathbb{R}^\infty$  als die lineare Erweiterung der Abbildung

$$e_k \mapsto \frac{k}{k+1} e_k.$$

Wir stellen fest, dass  $\|Av\|_\infty < \|v\|_\infty$  für alle  $v \in \mathbb{R}^\infty \setminus \{0\}$  gilt, aber  $\|A\|_{\text{op}} = 1$ , da

$$\sup \left\{ \frac{k}{k+1} \mid k \in \mathbb{N} \right\} = 1$$

gilt. Somit ist  $A$  nicht kontrahierend.

Wir stossen auch auf ein Problem beim Invertieren von

$$\begin{aligned} \text{Id} - A: \mathbb{R}^\infty &\rightarrow \mathbb{R}^\infty \\ e_k &\mapsto \left(1 - \frac{k}{k+1}\right) e_k = \frac{1}{k+1} e_k. \end{aligned}$$

Die inverse Abbildung

$$\begin{aligned} (\text{Id} - A)^{-1}: \mathbb{R}^\infty &\rightarrow \mathbb{R}^\infty \\ e_k &\mapsto (k+1)e_k \end{aligned}$$

ist also nicht stetig, da  $\|(\text{Id} - A)^{-1}\|_{\text{op}} = +\infty$ .

## Metrische Räume

**Definition.** Sei  $X$  eine Menge. Eine *Metrik* auf  $X$  ist eine Abbildung  $d: X \times X \rightarrow \mathbb{R}$  mit folgenden Eigenschaften.

- (i) *Positivität:* Für alle  $p, q \in X$  gilt  $d(p, q) \geq 0$ , und  $d(p, q) = 0$  genau dann, wenn  $p = q$ .
- (ii) *Symmetrie.* Für alle  $p, q \in X$  gilt  $d(q, p) = d(p, q)$ .
- (iii) *Dreiecksungleichung.* Für alle  $a, b, c \in X$  gilt  $d(a, c) \leq d(a, b) + d(b, c)$ .

### Beispiele.

- (1) Sei  $V$  ein reeller Vektorraum mit Norm  $\|\cdot\|$ . Definiere

$$\begin{aligned} d: V \times V &\rightarrow \mathbb{R} \\ (p, q) &\mapsto \|q - p\|. \end{aligned}$$

Das erste Axiom für Metrische Räume folgt direkt aus dem ersten Axiom für normierte Räume. Zur Symmetrie bemerke, dass

$$d(p, q) = \|q - p\| = \|(-1)(p - q)\| = |-1| \cdot \|p - q\| = \|p - q\| = d(q, p).$$

Die Dreiecksungleichung erfordert keinen Trick.

- (2) Sei  $X$  beliebig. Definiere auf  $X$  die *diskrete Metrik* wie folgt:

$$d(p, q) = \begin{cases} 0 & q = p \\ 1 & q \neq p \end{cases}.$$

Das einzige interessante Axiom ist die Dreiecksungleichung. Falls  $d(a, c) = 0$  gilt, dann gilt sowieso  $d(a, c) \leq d(a, b) + d(b, c)$ . Der einzige Fall, den wir ausschliessen müssen, ist also  $d(a, c) = 1$  und  $d(a, b) = d(b, c) = 0$ . Aber aus  $d(a, b) = 0$  folgt  $a = b$ , und aus  $d(b, c) = 0$  folgt  $b = c$  und somit auch  $a = c$ , was  $d(a, c) = 0$  impliziert. Dies ist ein Widerspruch.

- (3) Folgende Metrik ist enorm wichtig in der Zahlentheorie. Wir werden sie aber bloss in den Übungen antreffen. Sei  $p \in \mathbb{N}$  eine Primzahl. Definiere die *p-adische Metrik* auf  $\mathbb{Z}$  wie folgt:

$$d_p(n, m) = \begin{cases} 0 & n = m \\ p^{-x} & n \neq m, \end{cases}$$

wobei  $x \in \mathbb{N}$  maximal mit der Eigenschaft ist, dass  $p^x$  die Zahl  $n - m$  teilt. Insbesondere gilt  $d_p(n, n + 1) = 1 = p^{-0}$ , da in diesem Fall  $x = 0$  gilt. Weiter ist  $d_p(0, p^n) = p^{-n}$ , ist also klein für grosse  $n$ .

**Definitionen.** Sei  $X$  ein metrischer Raum mit Metrik  $d$ .

- Eine Folge  $(a_n)_{n \in \mathbb{N}}$  in  $X$  heisst *konvergent* mit Grenzwert  $a \in X$ , falls für alle  $\varepsilon$  ein  $N \in \mathbb{N}$  existiert, so dass für alle  $n \in \mathbb{N}$  mit  $n \geq N$  gilt, dass  $d(a_n, a) \leq \varepsilon$ .
- Eine Folge  $(a_n)_{n \in \mathbb{N}}$  in  $X$  heisst *Cauchyfolge*, falls für alle  $\varepsilon > 0$  ein  $N \in \mathbb{N}$  existiert, so dass für alle  $n, m \geq N$  gilt, dass  $d(a_n, a_m) \leq \varepsilon$ .
- Der metrische Raum  $X$  heisst *vollständig*, falls alle Cauchyfolgen in  $X$  konvergieren.

Ein vollständiger normierter Vektorraum heisst *Banachraum*.

**Beispiele.**

- (1) Der metrische Raum  $\mathbb{R}$  mit  $d(p, q) = |p - q|$  ist vollständig.
- (2) (a) Der metrische Raum  $X = (0, 1) \subset \mathbb{R}$  mit der von  $\mathbb{R}$  induzierten Metrik  $d(p, q) = |q - p|$  ist unvollständig.  
(b) Ebenso ist  $X = \mathbb{Q} \subset \mathbb{R}$  mit der von  $\mathbb{R}$  induzierten Metrik unvollständig.
- (3) Sei  $X = \mathbb{Z}$  mit der  $p$ -adischen Metrik  $d_p$ . Betrachte die Folge  $a_n = p^n$ . Dies ist eine Cauchyfolge mit Grenzwert  $0 \in \mathbb{Z}$ . Man könnte nun vermuten, dass  $\mathbb{Z}$  mit dieser Metrik vollständig ist. Das ist jedoch nicht der Fall. Betrachte dazu die Folge  $b_n = 1 + p + p^2 + \dots + p^n$ . Dies ist eine Cauchyfolge, leider aber ohne Grenzwert in  $\mathbb{Z}$ . Das heisst, dass  $\mathbb{Z}$  mit der  $p$ -adischen Metrik unvollständig ist. Siehe dazu auch die Übungen.

**Definition.** Sei  $(X, d)$  ein metrischer Raum. Eine Abbildung  $f: X \rightarrow X$  heisst *kontrahierend*, falls es eine positive Konstante  $k < 1$  gibt, so dass für alle  $p, q \in X$  gilt, dass  $d(f(p), f(q)) \leq k \cdot d(p, q)$ .

Unser Ziel in diesem Abschnitt ist es, folgenden wichtigen Satz vom polnischen Mathematiker Stefan Banach (1892–1945) zu beweisen.

**Theorem** (Fixpunktsatz von Banach 1922, Heuser 111.11). *Sei  $(X, d)$  ein vollständiger nicht-leerer metrischer Raum. Dann hat jede kontrahierende Abbildung  $f: X \rightarrow X$  einen eindeutigen Fixpunkt, das heisst es existiert genau ein Punkt  $p \in X$  mit  $f(p) = p$ .*

**Bemerkungen.**

- (1) Es ist wichtig, dass  $k < 1$  gilt. Betrachte zum Beispiel folgende Funktionen.
  - (a) Betrachte die Abbildung

$$\begin{aligned} f: \mathbb{R} &\rightarrow \mathbb{R} \\ x &\mapsto x + 1. \end{aligned}$$

Dann gilt  $|f(y) - f(x)| = 1 \cdot |y - x|$ , aber  $f$  hat keinen Fixpunkt.

(b) Die Identität

$$\begin{aligned} f: \mathbb{R} &\rightarrow \mathbb{R} \\ x &\mapsto x \end{aligned}$$

hat unendlich viele Fixpunkte.

(2) Auch die Vollständigkeit ist wichtig. Betrachte hier zum Beispiel die Funktion

$$\begin{aligned} f: \mathbb{R} \setminus \{0\} &\rightarrow \mathbb{R} \setminus \{0\} \\ x &\mapsto x/2 \end{aligned}$$

Die Abbildung  $f$  ist kontrahierend mit  $k = 1/2$ , hat aber keinen Fixpunkt in  $\mathbb{R} \setminus \{0\}$ . In den Übungen werden wir ein weniger künstliches Beispiel untersuchen.

(3) Kontrahierende Abbildungen  $f: X \rightarrow X$  sind wie alle Lipschitz-stetigen Abbildungen  $g: X \rightarrow X$  gleichmäßig stetig: Sei  $L \geq 0$  so, dass für alle  $p, q \in X$  gilt, dass  $d(g(p), g(q)) \leq L \cdot d(p, q)$ . Sei  $\varepsilon > 0$  vorgegeben. Setze  $\delta = \varepsilon/(L + 1)$ . Dann gilt für alle  $p, q \in X$  mit  $d(p, q) \leq \delta$ , dass  $d(f(p), f(q)) \leq \varepsilon$ .

**Lemma.** Seien  $(X_1, d_1)$  und  $(X_2, d_2)$  metrische Räume. Sei  $f: X_1 \rightarrow X_2$  eine beliebige Abbildung. Dann ist  $f$  stetig im Punkt  $p$ , genau dann, wenn  $f$  folgenstetig in  $p$  ist. Folgenstetig heisst, dass für alle konvergenten Folgen  $(a_n)_{n \in \mathbb{N}}$  in  $X_1$  mit  $\lim_{n \rightarrow \infty} a_n = p$  gilt, dass  $\lim_{n \rightarrow \infty} f(a_n) = f(p)$ .

*Beweis.* Für die Richtung “ $\Rightarrow$ ”, sei  $f$  stetig in  $p$  und  $(a_n)_{n \in \mathbb{N}}$  konvergent mit Grenzwert  $\lim_{n \rightarrow \infty} a_n = p$ . Sei  $\varepsilon > 0$  vorgegeben. Wähle  $\delta > 0$  so, dass wenn  $d_1(q, p) \leq \delta$  gilt, dann auch  $d_2(f(q), f(p)) \leq \varepsilon$ . Dann existiert  $N \in \mathbb{N}$  so, dass für  $n \geq N$  gilt, dass  $d(a_n, p) \leq \delta$ . Somit gilt auch für alle  $n \geq N$ , dass  $d_2(f(a_n), f(p)) \leq \varepsilon$ . Somit ist  $f$  folgenstetig.

Umgekehrt, für “ $\Leftarrow$ ”, sei  $f$  nicht stetig in  $p$ . Dann existiert  $\varepsilon > 0$  so, dass für alle  $n \in \mathbb{N}$  ein  $a_n \in X_1$  existiert mit  $d_1(a_n, p) \leq 1/n$ , aber  $d_2(f(a_n), f(p)) > \varepsilon$ . Nach Konstruktion gilt  $\lim_{n \rightarrow \infty} a_n = p$ , aber  $\lim_{n \rightarrow \infty} f(a_n) \neq f(p)$ . Folglich ist  $f$  nicht folgenstetig in  $p$ .  $\square$

*Beweis vom Fixpunktsatz.* Wir zeigen erst die Eindeutigkeit, was ganz schnell geht. Seien  $p, q \in X$  Fixpunkte von  $f$ . Dann gilt  $d(f(p), f(q)) \leq k \cdot d(p, q)$ . Aus  $k < 1$  folgt dann  $d(p, q) = 0$ , also  $q = p$ .

Für die Existenz, wähle  $a_0 \in X$ . Dieser Schritt ist der Grund für die Annahme, dass  $X$  nicht leer ist. Definiere eine Folge in  $X$  durch die Vorschrift  $a_{n+1} = f(a_n)$ . Es gilt dann  $a_1 = f(a_0)$ ,  $a_2 = f^2(a_0) = f(f(a_0))$ , und allgemein  $a_n = f^n(a_0)$ . Die so konstruierte Folge  $(a_n)_{n \in \mathbb{N}}$  ist eine Cauchyfolge in  $X$ . Wir zeigen das später und nehmen vorerst an, dass das stimmt. Da  $X$  vollständig ist, existiert der Grenzwert  $p = \lim_{n \rightarrow \infty} a_n$  in  $X$ . Es gilt dann  $f(p) = p$ . Tatsächlich, berechne

$$f(p) = \lim_{n \rightarrow \infty} f(a_n) = \lim_{n \rightarrow \infty} a_{n+1} = p.$$



Es bleibt also bloss zu zeigen, dass  $(a_n)_{n \in \mathbb{N}}$  eine Cauchyfolge ist. Sei dazu  $\varepsilon > 0$  vorgegeben. Für  $n, m \in \mathbb{N}$  mit  $n \geq m$  schätzen wir ab, dass

$$\begin{aligned} d(a_n, a_m) &= d(f(a_{n-1}), f(a_{m-1})) \\ &\leq k \cdot d(a_{n-1}, a_{m-1}) \\ &\leq k^2 \cdot d(a_{n-2}, a_{m-2}) \\ &\leq \dots \\ &\leq k^m \cdot d(a_{n-m}, a_0). \end{aligned}$$

Setze  $\ell = n - m$ . Schätze weiterhin ab, dass

$$\begin{aligned} d(a_\ell, a_0) &\leq d(a_\ell, a_{\ell-1}) + d(a_{\ell-1}, a_{\ell-2}) + \dots + d(a_1, a_0) \\ &\leq k^{\ell-1} \cdot d(a_1, a_0) + k^{\ell-2} \cdot d(a_1, a_0) + \dots + d(a_1, a_0) \\ &= (1 + k + k^2 + \dots + k^{\ell-1}) \cdot d(a_1, a_0) \\ &\leq \frac{1}{1-k} \cdot d(a_1, a_0). \end{aligned}$$

Insgesamt erhalten wir

$$d(a_n, a_m) \leq k^m \cdot \frac{1}{1-k} \cdot d(a_1, a_0).$$

Wähle  $N \in \mathbb{N}$  so, dass  $k^N \cdot 1/(1-k) \cdot d(a_1, a_0) \leq \varepsilon$ . Solch ein  $N$  existiert, da nach unseren Annahmen an  $k$  gilt, dass  $\lim_{n \rightarrow \infty} k^n = 0$ . Es gilt dann für alle  $n, m \geq N$ , dass  $d(a_n, a_m) \leq \varepsilon$ .  $\square$

Wir werden den Fixpunktsatz auf Banachräume anwenden um Sätze zur Existenz und Eindeutigkeit gewisser Differentialgleichungen zu zeigen.

**Beispiel.** Sei  $K \subset \mathbb{R}$  kompakt, das heisst  $K$  ist beschränkt und abgeschlossen. Definiere  $C(K)$  als die Menge aller stetigen Abbildungen  $f: K \rightarrow \mathbb{R}$ . Man kann sich leicht überlegen, dass  $C(K)$  ein reeller Vektorraum ist. Wir statten  $C(K)$  mit der Norm

$$\|f\|_\infty = \max \{|f(x)| \mid x \in K\} < +\infty$$

aus. Der Satz von Weierstrass besagt, dass für eine Folge  $f_n: K \rightarrow \mathbb{R}$  stetiger Funktionen, welche auf  $K$  gleichmässig gegen eine Funktion  $f: K \rightarrow \mathbb{R}$  konvergiert,  $f$  stetig ist. Gleichmässige Konvergenz heisst, dass für alle  $\varepsilon > 0$  ein  $N \in \mathbb{N}$  existiert, so dass für alle  $n \geq N$  und alle  $p \in K$  gilt, dass  $|f_n(p) - f(p)| \leq \varepsilon$ . In unseren Worten heisst das  $\|f_n - f\|_\infty \leq \varepsilon$ . Der Satz von Weierstrass besagt also, dass der Raum  $C(K)$  mit der Maximumnorm vollständig ist.

## 4 Kurven und Vektorfelder

**Definition.** Eine *Kurve* in  $\mathbb{R}^n$  ist das Bild einer stetigen Abbildung  $\gamma: \mathbb{R} \rightarrow \mathbb{R}^n$ . Die Abbildung  $\gamma$  heisst *Parametrisierung* der Kurve  $\gamma$ . Wir werden die Unterscheidung zwischen Kurven und Parametrisierungen nicht immer explizit machen.

Wir untersuchen nun kurz den Stetigkeitsbegriff auf Kurven. Stetigkeit heisst, dass in jedem  $t \in \mathbb{R}$  die  $(\varepsilon, \delta)$ -Bedingung für  $\gamma$  bezüglich beliebiger Normen auf  $\mathbb{R}$  und  $\mathbb{R}^n$  erfüllt ist, insbesondere bezüglich der Normen  $|\cdot|$  und  $\|\cdot\|_\infty$ . Schreibe nun

$$\gamma(t) = \sum_{i=1}^n \gamma_i(t) \cdot e_i,$$

wobei  $e_i$  der  $i$ -te Basisvektor in  $\mathbb{R}^n$  ist. In diesem Fall ist  $\gamma_i(t) = \langle \gamma(t), e_i \rangle$  die  $i$ -te Komponentenfunktion von  $\gamma$ . Für alle  $s, t \in \mathbb{R}$  gilt  $\|\gamma(s) - \gamma(t)\|_\infty \leq \varepsilon$  genau dann, wenn  $|\gamma_i(s) - \gamma_i(t)| \leq \varepsilon$  für alle  $i$ . Wir folgern, dass  $\gamma$  stetig in  $t \in \mathbb{R}$  ist, genau dann wenn alle Komponentenfunktionen  $\gamma_i$  stetig in  $t$  sind.

**Beispiel.**

(1) Betrachte die Funktion

$$\begin{aligned} \gamma: \mathbb{R} &\rightarrow \mathbb{R}^2 \\ t &\mapsto (\cos(t), \sin(t)). \end{aligned}$$

Dann hat  $\gamma$  stetige Komponentenfunktionen und ist daher die Parametrisierung einer Kurve. In diesem Fall ist die beschriebene Kurve der Einheitskreis, siehe Abbildung I.2. Die Pfeile erklären wir später.

(2) Die Funktion

$$\begin{aligned} \gamma: \mathbb{R} &\rightarrow \mathbb{R}^2 \\ t &\mapsto (t^2, t^3 - t) \end{aligned}$$

ist stetig und somit die Parametrisierung einer Kurve, dargestellt in Abbildung I.2. Diese Kurve einen “singulären Punkt”, herbeigeführt durch die Eigenschaft, dass  $\gamma(1) = \gamma(-1) = 1$  gilt.

**Definition.** Eine parametrisierte Kurve  $\gamma: \mathbb{R} \rightarrow \mathbb{R}^n$  heisst *differenzierbar*, falls für alle  $t \in \mathbb{R}$  die Ableitung

$$\dot{\gamma}(t) = \lim_{h \rightarrow 0} \frac{\gamma(t+h) - \gamma(t)}{h}$$

existiert. Der Vektor  $\dot{\gamma}(t)$  heisst *Geschwindigkeitsvektor*.



Abbildung I.2: Die Bilder der Kurven  $\gamma$  aus den Beispielen

**Bemerkung.** Für alle  $t, h \in \mathbb{R}$  mit  $h \neq 0$  gilt

$$\frac{\gamma(t+h) - \gamma(t)}{h} = \sum_{k=1}^n \frac{\gamma_k(t+h) - \gamma_k(t)}{h} \cdot e_k$$

Wir folgern, dass  $\dot{\gamma}(t) \in \mathbb{R}^n$  existiert, genau dann, wenn alle Ableitungen  $\dot{\gamma}_k(t) \in \mathbb{R}$  existieren. In anderen Worten: Eine Kurve  $\gamma: \mathbb{R} \rightarrow \mathbb{R}^n$  ist differenzierbar, genau dann, wenn alle  $\gamma_k: \mathbb{R} \rightarrow \mathbb{R}$  differenzierbar sind.

**Beispiele.**

(1) Die Kurve

$$\begin{aligned} \gamma: \mathbb{R} &\rightarrow \mathbb{R}^2 \\ t &\mapsto (\cos(t), \sin(t)) \end{aligned}$$

ist differenzierbar mit  $\dot{\gamma}(t) = (-\sin(t), \cos(t))$ . Die grauen Pfeile in Abbildung I.2 stellen die Ableitung von  $\gamma$  dar.

(2) Die Kurve

$$\begin{aligned} \gamma: \mathbb{R} &\rightarrow \mathbb{R}^2 \\ t &\mapsto (t^2, t^3 - t) \end{aligned}$$

ist differenzierbar mit  $\dot{\gamma}(t) = (2t, 3t^2 - 1)$ . Wir berechnen

- $\gamma(0) = (0, 0)$  und  $\dot{\gamma}(0) = (0, -1)$ ,
- $\gamma(1) = (1, 0)$  und  $\dot{\gamma}(1) = (2, 2)$ ,
- $\gamma(-1) = (1, 0)$  und  $\dot{\gamma}(-1) = (-2, 2)$ ,

- $\gamma(1/\sqrt{3}) = (1/2, -2/3 \cdot 1/\sqrt{3})$  und  $\dot{\gamma}(1/\sqrt{3}) = (2/\sqrt{3}, 0)$ ,
- $\gamma(-1/\sqrt{3}) = (1/3, 2/3 \cdot 1/\sqrt{3})$  und  $\dot{\gamma}(-1/\sqrt{3}) = (-2/\sqrt{3}, 0)$ .

Für grosse  $t$  ist  $y$  ungefähr  $x^{3/2}$ . Diese Information reicht, um das Bild von  $\gamma$  wie in Abbildung I.2 zu erraten. Die Ableitungen sind in grau (mit  $1/4$  gestreckt) eingezeichnet.

**Definition.** Sei  $U \subset \mathbb{R}^n$  offen. Ein *Vektorfeld* auf  $U$  ist eine stetige Abbildung

$$\begin{aligned} X: U &\rightarrow \mathbb{R}^n \\ q &\mapsto X(q). \end{aligned}$$

Jedes Vektorfeld  $X$  auf  $U$  liefert eine *gewöhnliche Differentialgleichung*  $\dot{\gamma}(t) = X(\gamma(t))$ . Dies ist wie folgt zu interpretieren. Sei  $p \in U$  fest. Gesucht ist  $T > 0$  und eine differenzierbare *Lösungskurve*  $\gamma: (-T, T) \rightarrow U$  zur Anfangsbedingung  $\gamma(0) = p$ . Falls  $X = A: \mathbb{R}^n \rightarrow \mathbb{R}^n$  linear ist, dann heisst die Differentialgleichung  $\dot{\gamma}(t) = A(\gamma(t))$  *linear*.

**Geometrische Interpretation.** Ein Vektorfeld  $X: U \rightarrow \mathbb{R}^2$  liefert in jedem Punkt  $q \in U$  einen Vektor  $X(q) \in \mathbb{R}^n$ . Siehe Abbildung I.3. In jedem Punkt der Kurve  $\gamma(t) \in U$  stimmt der Geschwindigkeitsvektor  $\dot{\gamma}(t) \in \mathbb{R}^n$  mit dem Vektor  $X(\gamma(t))$  überein.



Abbildung I.3: Ein Vektorfeld  $X$  auf  $U \subset \mathbb{R}^2$  mit einer Lösungskurve  $\gamma$  in schwarz

### Beispiele.

- (1) Betrachte das lineare Vektorfeld

$$\begin{aligned} X: \mathbb{R} &\rightarrow \mathbb{R}^n \\ q &\mapsto 0. \end{aligned}$$

Sei  $p \in \mathbb{R}^n$ . Dann ist  $\gamma(t) = p$  die eindeutige Lösung der Differentialgleichung  $\gamma(0) = p$  und  $\dot{\gamma}(t) = X(\gamma(t))$ .

- (2) Betrachte das lineare Vektorfeld

$$\begin{aligned} A: \mathbb{R}^2 &\rightarrow \mathbb{R}^2 \\ \begin{pmatrix} x \\ y \end{pmatrix} &\mapsto \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -y \\ x \end{pmatrix}, \end{aligned}$$

siehe Abbildung I.4. Betrachte nochmals die Kurve

$$\begin{aligned} \gamma: \mathbb{R} &\rightarrow \mathbb{R}^2 \\ t &\mapsto \begin{pmatrix} \cos(t) \\ \sin(t) \end{pmatrix}. \end{aligned}$$

Es gilt

$$\dot{\gamma}(t) = \begin{pmatrix} -\sin(t) \\ \cos(t) \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \cos(t) \\ \sin(t) \end{pmatrix} = A(\gamma(t)).$$

Also ist die Kurve  $\gamma: \mathbb{R} \rightarrow \mathbb{R}^2$  eine Lösung der linearen Differentialgleichung  $\dot{\gamma}(t) = A(\gamma(t))$  zur Anfangsbedingung  $\gamma(0) = e_1 \in \mathbb{R}^2$ . Allgemeiner ist die Funktion

$$\begin{aligned} \gamma: \mathbb{R} &\rightarrow \mathbb{R}^2 \\ t &\mapsto \begin{pmatrix} r \cos(t + \varphi) \\ r \sin(t + \varphi) \end{pmatrix} \end{aligned}$$

eine Lösung dieser Differentialgleichung. Durch geschickte Wahl von  $\varphi$  und  $r$  deckt das jede Anfangsbedingung ab.

- (3) Die Kurve  $\gamma(t) = (t^2, t^3 - t)$  ist nicht die Lösung einer gewöhnlichen Differentialgleichung. Der Grund dafür ist, dass  $\gamma$  den Punkt  $(1, 0)$  zweimal mit unterschiedlichen Geschwindigkeitsvektoren durchläuft: Es gilt  $\gamma(1) = \gamma(-1)$ , aber  $\dot{\gamma}(1) \neq \dot{\gamma}(-1)$ . Falls nämlich  $X: \mathbb{R}^2 \rightarrow \mathbb{R}^2$  existiert mit  $\dot{\gamma}(t) = X(\gamma(t))$ , dann gälte

$$\begin{aligned} X(e_1) &= X(\gamma(1)) = \dot{\gamma}(1) = (2, 2), \\ X(e_1) &= X(\gamma(-1)) = \dot{\gamma}(-1) = (-2, 2). \end{aligned}$$

Das ist aber nicht möglich, da  $2 \neq -2$ .



Abbildung I.4: Das Vektorfeld  $A(x, y) = (-y, x)$  und das Vektorfeld  $A(x, y) = (y, 0)$

(4) Betrachte das lineare Vektorfeld

$$A: \mathbb{R}^2 \rightarrow \mathbb{R}^2$$

$$\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} y \\ 0 \end{pmatrix}.$$

Sei  $p = (x_0, y_0)$  vorgegeben. Als Ansatz für  $\gamma$  schreiben wir  $\gamma(t) = (x(t), y(t))$ . Aus  $\dot{\gamma}(t) = (y(t), 0)$  erhalten wir  $\dot{y}(t) = 0$ , also ist  $y(t) = y_0$  konstant. Weiterhin ist  $\dot{x}(t) = y(t) = y_0$ , also ist  $x(t) = y_0 t + x_0$ . Zusammengefasst erhalten wir die Lösungskurve

$$\gamma(t) = \begin{pmatrix} x_0 + y_0 t \\ y_0 \end{pmatrix}.$$

(5) Sei  $A: \mathbb{R}^n \rightarrow \mathbb{R}^n$  beliebig und linear. Sei  $v \in \mathbb{R}^n$  ein Eigenvektor zum Eigenwert  $\lambda \in \mathbb{R}$ , das heißt  $A(v) = \lambda v$ . Definiere

$$\gamma: \mathbb{R} \rightarrow \mathbb{R}^n$$

$$t \mapsto e^{\lambda t} \cdot v.$$

Es gilt  $\gamma(0) = v$  und

$$\dot{\gamma}(t) = e^{\lambda t} \cdot \lambda \cdot v = e^{\lambda t} \cdot A(v) = A(e^{\lambda t} \cdot v).$$

Also ist  $\gamma$  eine Lösungskurve. Wir haben aber verwendet, dass  $v$  ein Eigenvektor von  $A$  ist. Die anderen Lösungskurven von  $A$  könnten komplizierter sein.

(6) Sei  $A: \mathbb{R}^n \rightarrow \mathbb{R}^n$  linear und *diagonalisierbar* über  $\mathbb{R}$ , das heißt es existieren *Eigenwerte*  $\lambda_1, \dots, \lambda_n \in \mathbb{R}$  und eine Basis aus *Eigenvektoren*  $\{v_1, \dots, v_n\}$  von  $\mathbb{R}^n$  mit  $A(v_i) = \lambda_i v_i$ . Sei  $p \in \mathbb{R}^n$  vorgegeben. Schreibe

$$p = \sum_{i=1}^n x_i v_i$$

mit  $x_i \in \mathbb{R}$ . Definiere

$$\begin{aligned}\gamma: \mathbb{R} &\rightarrow \mathbb{R}^n \\ t &\mapsto e^{\lambda_1 t} \cdot x_1 \cdot v_1 + \cdots + e^{\lambda_n t} \cdot x_n \cdot v_n.\end{aligned}$$

Es gilt  $\gamma(0) = p$  und

$$\dot{\gamma}(t) = \sum_{i=1}^n \lambda_i e^{\lambda_i t} x_i v_i = A(\gamma(t)).$$

Es existiert also eine Lösungskurve  $\gamma(t)$  für beliebige Anfangsbedingungen.

(7) Sei

$$\begin{aligned}f: \mathbb{R} &\rightarrow \mathbb{R} \\ x &\mapsto 2\sqrt{|x|}\end{aligned}$$

ein Vektorfeld auf  $\mathbb{R}$ . Betrachte die Anfangsbedingung  $p = 0$ . Die konstante Nullkurve ist eine Lösungskurve. Betrachte nun die Kurve

$$\begin{aligned}\gamma: \mathbb{R} &\rightarrow \mathbb{R} \\ t &\mapsto \begin{cases} 0 & t \leq 0 \\ t^2 & t \geq 0. \end{cases}\end{aligned}$$

Es gilt  $\gamma(0) = 0$  und

$$\dot{\gamma}(t) = \begin{cases} 0 & t \leq 0 \\ 2t & t \geq 0. \end{cases}$$

Es gilt aber auch

$$f(\gamma(t)) = \begin{cases} 2\sqrt{|0|} = 0 & t \leq 0 \\ 2\sqrt{|t^2|} = 2t & t \geq 0. \end{cases}$$

Das heisst  $\gamma(t)$  ist auch eine Lösung zur selben Anfangsbedingung. Das “Problem” hier ist, dass  $f: \mathbb{R} \rightarrow \mathbb{R}$  nicht Lipschitz-stetig ist.

## 5 Der Satz von Cauchy-Lipschitz-Picard-Lindelöf

**Definition.** Sei  $U \subset \mathbb{R}^n$  offen. Eine Abbildung  $X: U \rightarrow \mathbb{R}^n$  heisst *Lipschitz-stetig* mit Konstante  $k \geq 0$ , falls für alle  $p, q \in U$  die Ungleichung

$$\|X(q) - X(p)\|_2 \leq k \cdot \|q - p\|_2$$

gilt.

**Bemerkung.** Die Wahl der Norm  $\|\cdot\|_2$  ist für diese Definition irrelevant, da alle Normen auf  $\mathbb{R}^n$  äquivalent sind. Die Konstante  $k \geq 0$  hängt aber von der gewählten Norm ab.

**Theorem** (Cauchy-Lipschitz-Picard-Lindelöf). Sei  $U \subset \mathbb{R}^n$  offen und  $X: U \rightarrow \mathbb{R}^n$  Lipschitz-stetig mit Konstante  $k \geq 0$ . Sei  $p \in U$  vorgegeben. Dann existiert  $T > 0$  und eine eindeutige differenzierbare Kurve  $\gamma: (-T, T) \rightarrow U$  mit  $\gamma(0) = p$ , so dass für alle  $t \in (-T, T)$  gilt, dass  $\dot{\gamma}(t) = X(\gamma(t))$ .

**Bemerkung.**

1. Falls  $U = \mathbb{R}^n$  ist, kann  $T > 0$  beliebig gewählt werden. Siehe dazu Übungsserie 4.
2. Falls  $U = \mathbb{R}^n$  und  $k = 0$ , dann gilt für alle  $p, q \in \mathbb{R}^n$ , dass  $\|X(p) - X(q)\|_2 \leq 0 \cdot \|p - q\|_2 = 0$ . Das Vektorfeld  $X$  ist also konstant. Dann ist die Lösungskurve

$$\begin{aligned}\gamma: \mathbb{R} &\rightarrow \mathbb{R}^n \\ t &\mapsto p + t \cdot X(p).\end{aligned}$$

Vergleiche [1], Satz 117.1. Ein wichtiger Unterschied in der Version, die Heuser formuliert, ist es, dass das Vektorfeld  $f$  dort auch von der Zeit  $x$  abhängen darf. Wir werden nun einige Vorbereitungen zum Beweis treffen. Dazu werden wir zunächst Kurven genauer betrachten.

**Definition.** Sei  $\alpha: [a, b] \rightarrow \mathbb{R}^n$  stetig. Schreibe

$$\alpha(t) = \sum_{i=1}^n \alpha_i(t) e_i,$$

wobei  $\alpha_i: [a, b] \rightarrow \mathbb{R}$  stetige Funktionen sind. Wir schreiben dann

$$\int_a^b \alpha(s) ds = \sum_{i=1}^n \left( \int_a^b \alpha_i(s) ds \right) \cdot e_i.$$

**Proposition 2.** Seien  $\alpha: [a, b] \rightarrow \mathbb{R}$  stetig und  $\gamma: (a, b) \rightarrow \mathbb{R}$  stetig differenzierbar. Dann gilt:

(i) für alle  $t \in (a, b)$  gilt

$$\frac{d}{dt} \int_a^t \alpha(s) ds = \alpha(t),$$

(ii) unter der Annahme, dass  $[0, t] \subset (a, b)$ , gilt

$$\gamma(t) - \gamma(0) = \int_0^t \dot{\gamma}(s) ds.$$

*Beweis.* Sowohl die Differentiation als auch die Integration von Kurven lässt sich komponentenweise ausführen. Deshalb folgen die beiden Eigenschaften aus den entsprechenden Sätzen der Differential- und Integralrechnung für Funktionen auf  $\mathbb{R}$ .  $\square$



Sei  $C([a, b], \mathbb{R}^n)$  der Raum aller stetiger Funktionen  $\alpha: [a, b] \rightarrow \mathbb{R}^n$ . Für eine Funktion  $\alpha \in C([a, b], \mathbb{R}^n)$  definieren wir

$$\|\alpha\|_\infty = \max \{ \|\alpha(t)\|_2 \mid t \in [a, b] \} \geq 0.$$

**Lemma 1.** *Der Raum  $C([a, b], \mathbb{R}^n)$  ist bezüglich der Norm  $\|\cdot\|_\infty$  vollständig.*

*Beweis.* Sei  $(\alpha_m)_{m \in \mathbb{N}}$  eine Cauchyfolge in  $C([a, b], \mathbb{R})$  bezüglich der Norm  $\|\cdot\|_\infty$ . Dann erhalten wir in jeder Komponente  $i \leq n$  eine Cauchyfolge  $\langle \alpha_m, e_i \rangle$  in  $C([a, b], \mathbb{R})$  (bezüglich der Norm  $\|\cdot\|_\infty$ ). Nach Weierstrass existiert für alle  $i \leq n$  der Grenzwert  $\lim_{m \rightarrow \infty} \langle \alpha_m, e_i \rangle$  in  $C([a, b], \mathbb{R})$ . Es folgt wie im Beweis von Banachs Fixpunktsatz, dass der Grenzwert

$$\lim_{m \rightarrow \infty} \alpha_m = \sum_{i=1}^n \lim_{m \rightarrow \infty} \langle \alpha_m, e_i \rangle \cdot e_i$$

in  $C([a, b], \mathbb{R}^n)$  existiert. Also ist  $C([a, b], \mathbb{R})$  mit der Norm  $\|\cdot\|_\infty$  vollständig.  $\square$

Der Beweis des Theorems reduziert sich auf ein Fixpunktproblem, welches wir mit Banachs Fixpunktsatz lösen können. Dazu verwenden wir folgende Abschätzung.

**Lemma 2.** *Sei  $\alpha: [a, b] \rightarrow \mathbb{R}^n$  stetig und  $t \in \mathbb{R}$  mit  $[0, t] \subset [a, b]$ . Dann gilt*

$$\left\| \int_0^t \alpha(s) ds \right\|_2 \leq \int_0^t \|\alpha(s)\|_2 ds.$$

*Beweis.* Setze  $v = \int_0^t \alpha(s) ds$  Berechne

$$\begin{aligned} \|v\|_2^2 &= \langle v, v \rangle \\ &= \left\langle v, \int_0^t \alpha(s) ds \right\rangle \\ &= \int_0^t \langle v, \alpha(s) \rangle ds \\ &\leq \int_0^t \|v\|_2 \cdot \|\alpha(s)\|_2 ds \end{aligned}$$

nach komponentenweiser Integration und einer Anwendung der Cauchy-Schwarz Ungleichung. Teilung durch  $\|v\|_2$  liefert

$$\|v\|_2 \leq \int_0^t \|\alpha(s)\|_2 ds. \quad \square$$

*Beweis von Cauchy-Lipschitz-Picard-Lindelöf.* Wir betrachten nur den Fall, dass  $U = \mathbb{R}^n$  gilt. Allgemeine  $U$  zu betrachten benötigt nicht viele zusätzliche Überlegungen. Sei  $T > 0$ . Betrachte  $V = C([-T, T], \mathbb{R}^n)$  mit der Norm

$$\|\alpha\|_\infty = \max \{ \|\alpha(t)\|_2 \mid t \in [-T, T] \}.$$

Definiere nun

$$\begin{aligned} P: V &\rightarrow V \\ \alpha &\mapsto P(\alpha) \end{aligned}$$

durch

$$P(\alpha)(t) = p + \int_0^t X(\alpha(s)) ds$$

für alle  $t \in [-T, T]$ . Wir bemerken, dass  $X \circ \alpha$  stetig ist, da  $\alpha$  und  $X$  beide stetig sind. Deshalb ist  $P(\alpha)$  auf  $(-T, T)$  differenzierbar. Weiterhin haben wir bereits gesehen, dass für  $t \in (-T, T)$  gilt, dass

$$\frac{d}{dt}P(\alpha)(t) = X(\alpha(t)).$$

Nimm nun an, dass  $P$  einen Fixpunkt  $\gamma \in V$  hat (wir schreiben  $\gamma$  für diesen Punkt, da die Punkte in  $V$  Kurven sind). Dann gilt für alle  $t \in (-T, T)$ , dass

$$\dot{\gamma}(t) = \frac{d}{dt}\gamma(t) = \frac{d}{dt}P(\gamma)(t) = X(\gamma(t)).$$

Ausserdem ist

$$\gamma(0) = P(\gamma)(0) = p + \int_0^0 X(\gamma(s)) ds = p,$$

also ist  $\gamma$  eine Lösungskurve unserer Differentialgleichung mit  $\gamma(0) = p$ . Umgekehrt, sei  $\gamma \in V$  so, dass  $\gamma|_{(-T, T)}$  eine Lösungskurve der Differentialgleichung  $\dot{\gamma}(t) = X(\gamma(t))$  mit  $\gamma(0) = p$  ist. Dann gilt für alle  $t \in (-T, T)$ , dass

$$P(\gamma(t)) = p + \int_0^t X(\gamma(s)) ds = p + \int_0^t \dot{\gamma}(s) ds = p + \gamma(t) - \gamma(0) = \gamma(t).$$

Aus der Stetigkeit von  $\gamma$  und  $P(\gamma)$  folgt auch  $P(\gamma)(\pm T) = \gamma(\pm T)$ . Also ist  $\gamma$  ein Fixpunkt von  $P$ . Zusammengefasst erhalten wir, dass die Fixpunkte von  $P$  genau die Lösungskurven der Differentialgleichung sind.

Um Banachs Fixpunktsatz anzuwenden zeigen wir nun, dass für alle  $T < 1/k$  die Abbildung  $P$  kontrahierend ist. Dazu sei  $t \in [-T, T]$  beliebig, und  $\alpha, \beta \in V$ . Schätze ab, dass

$$\begin{aligned} \|P(\alpha(t)) - P(\beta)(t)\|_2 &= \left\| \int_0^t X(\alpha(s)) - X(\beta(s)) ds \right\|_2 \\ &\leq \int_0^t \|X(\alpha(s)) - X(\beta(s))\|_2 ds \\ &\leq k \int_0^t \|\alpha(s) - \beta(s)\|_2 ds \\ &\leq k \cdot T \cdot \|\alpha - \beta\|_\infty. \end{aligned}$$

Wir schliessen, dass  $\|P(\alpha) - P(\beta)\|_\infty \leq k \cdot T \cdot \|\alpha - \beta\|_\infty$ , also  $P$  kontrahierend ist. Wir schliessen aus der Vollständigkeit von  $V$  mit Hilfe des Fixpunktsatzes von Banach, dass  $P$  einen eindeutigen Fixpunkt  $\gamma \in V$  hat. Also ist  $\gamma \in V$  die eindeutige Lösung zur Differentialgleichung  $\dot{\gamma}(t) = X(\gamma(t))$  mit  $\gamma(0) = p$ .  $\square$

## 6 Lineare Differentialgleichungen

Sei  $A: \mathbb{R}^n \rightarrow \mathbb{R}^n$  linear. Aus Abschnitt 2 wissen wir, dass die Operatornorm  $\|A\|_{\text{op}}$  endlich ist. Insbesondere ist  $A$  Lipschitz-stetig mit Konstante  $k = \|A\|_{\text{op}}$ . Tatsächlich gilt für alle  $v, w \in \mathbb{R}^n$ , dass

$$\|A(w) - A(v)\|_2 \leq \|A\|_{\text{op}} \cdot \|v - w\|_2.$$

Wir folgern, dass die Differentialgleichung  $\dot{\gamma}(t) = A(\gamma(t))$  zu jeder Anfangsbedingung  $p \in \mathbb{R}^n$  und jeder Zeit  $T < 1/k$  eine eindeutige Lösungskurve  $\gamma: (-T, T) \rightarrow \mathbb{R}^n$  hat. Sei  $V = C([-T, T], \mathbb{R}^n)$ . Wir hatten  $\gamma$  als Fixpunkt eines kontrahierenden Operators

$$P: V \rightarrow V$$

konstruiert, wobei

$$P(\alpha)(t) = p + \int_0^t A(\alpha(s)) \, ds.$$

Sei  $\gamma_0 \in V$  beliebig. Definiere rekursiv  $\gamma_{n+1} = P(\gamma_n) \in V$ . Dann gilt  $\gamma = \lim_{n \rightarrow \infty} \gamma_n$ . Wähle nun  $\gamma_0(t) = p$  konstant. Wir berechnen nun einige Glieder der Folge  $(\gamma_n)_{n \in \mathbb{N}}$ . Das erste Glied ist

$$\begin{aligned} \gamma_1(t) &= P(\gamma_0)(t) \\ &= p + \int_0^t A(\gamma_0(s)) \, ds \\ &= p + t \cdot A(p). \end{aligned}$$

Mit der Linearität von  $A$  berechnen wir

$$\begin{aligned} \gamma_2(t) &= p + \int_0^t A(\gamma_1(s)) \, ds \\ &= p + \int_0^t A(p + s \cdot A(p)) \, ds \\ &= p + \int_0^t A(p) + s \cdot A^2(p) \, ds \\ &= p + t \cdot A(p) + \frac{t^2}{2} A^2(p). \end{aligned}$$

Um unsere Vermutung, wie diese Folge weitergeht, zu bestätigen, berechnen wir noch einen letzten Term explizit:

$$\begin{aligned} \gamma_3(t) &= P(\gamma_2)(t) \\ &= p + \int_0^t A(p + s \cdot A(p) + s^2/2 \cdot A^2(p)) \, ds \\ &= p + t \cdot A(p) + \frac{t^2}{2} A^2(p) + \frac{t^3}{3!} A^3(p). \end{aligned}$$

Induktiv erhalten wir für alle  $n \in \mathbb{N}$ , dass

$$\gamma_n(t) = \sum_{k=0}^n \frac{t^k}{k!} A^k(p)$$

(mit der Konvention, dass  $A^0$  die Identitätsabbildung ist).

**Definition.** Sei  $A: \mathbb{R}^n \rightarrow \mathbb{R}^n$  linear. Die *Endomorphismenexponentialabbildung* ist

$$\exp(tA) = \sum_{k=0}^{\infty} \frac{t^k}{k!} A^k.$$

**Bemerkung.** Aus den Beweisen der Theoreme von Banach und Cauchy-Lipschitz-Picard-Lindelöf wissen wir bloss, dass dieser Grenzwert für alle  $|t| < 1/\|A\|_{\text{op}}$  existiert. Tatsächlich existiert der Grenzwert für alle  $t \in \mathbb{R}$ , unabhängig von  $A$ . Benutze dazu, dass für die Operatornorm die Ungleichung  $\|A^k\|_{\text{op}} \leq \|A\|_{\text{op}}^k$  gilt. Dann folgt nämlich  $\|e^{tA}\|_{\text{op}} \leq e^{|t| \cdot \|A\|_{\text{op}}}$ . Also ist die Abbildung  $\exp(tA)$  ein Endomorphismus von  $\mathbb{R}^n$  für alle  $t \in \mathbb{R}$ .

Wir schliessen, dass die Lösung zur Differentialgleichung  $\dot{\gamma}(t) = A(\gamma(t))$  mit  $\gamma(0) = p$  durch

$$\gamma(t) = e^{tA} \cdot p$$

gegeben ist.

**Beispiele.** Sei  $A: \mathbb{R}^2 \rightarrow \mathbb{R}^2$  linear, repräsentiert bezüglich der Standardbasis durch eine Matrix

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

(1) Betrachte

$$A = \begin{pmatrix} a & 0 \\ 0 & d \end{pmatrix} = \text{diag}(a, d).$$

Berechne für  $k \in \mathbb{N}$ , dass  $A^k = \text{diag}(a^k, d^k)$ , Also ist  $e^{tA} = \text{diag}(e^{ta}, e^{td})$ . Wir betrachten den Spezialfall  $a = 1, d = -1$ . Dann ist  $e^{tA} = \text{diag}(e^t, e^{-t})$ . Einige Lösungskurven sind in Tabelle I.1 aufgelistet und in Abbildung I.5 gezeichnet.

Anfangsbedingung $p \in \mathbb{R}^2$	Lösungskurve $\gamma(t)$
$(\pm 1, 0)$	$(\pm e^t, 0)$
$(0, \pm 1)$	$(0, \pm e^{-t})$
$\pm(1, 1)$	$\pm(e^t, e^{-t})$
$\pm(1, -1)$	$\pm(e^t, -e^{-t})$

Tabelle I.1: Einige Lösungen der Differentialgleichung zur Matrix  $A = \text{diag}(1, -1)$ .



Abbildung I.5: Einige Lösungen der Differentialgleichung zur Matrix  $A = \text{diag}(1, -1)$ .

(2) Betrachte die Matrix

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

und bemerke, dass  $A^2 = 0$  gilt. Folglich ist

$$e^{tA} = \begin{pmatrix} 1 & -t \\ 0 & 1 \end{pmatrix}.$$

Die Lösungskurven von  $A$  sind konstant auf der  $x$ -Achse, und ausserhalb sind sie parallel dazu, wobei ihre Geschwindigkeit mit zunehmender  $y$ -Komponente linear wächst.

(3) Die Matrix

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

liefert ein rotiertes Bild des obigen Beispiels.

(4) Die Matrix

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

ist eine Drehung um den Winkel  $\pi/2$ . Die Potenzen von  $A$  sind  $A^2 = -1$ ,  $A^3 = -A$ , und  $A^4 = 1$ . Dann geht es periodisch weiter. Berechne

$$e^{tA} = \begin{pmatrix} 1 - t^2/2! + t^4/4! - \dots & -t + t^3/3! - t^5/5! + \dots \\ t - t^3/3! + t^5/5! - \dots & 1 - t^2/2! + t^4/4! - \dots \end{pmatrix} = \begin{pmatrix} \cos(t) & -\sin(t) \\ \sin(t) & \cos(t) \end{pmatrix}.$$

Wir haben also nun hergeleitet, dass die Lösungskurven dieses Vektorfelds konzentrische Kreise sind. Wir sehen auch, dass im allgemeinen  $e^{t(A+B)} = e^{tA} \cdot e^{tB}$  nicht gilt (das gilt nur, wenn  $AB = BA$ ).

**Bemerkung.** Die Matrix  $e^{tA}$  lässt sich für diagonalisierbare Matrizen  $A \in \mathbb{R}^{n \times n}$  relativ einfach berechnen. Sei  $A$  eine solche Matrix, das heisst, es existiert  $P \in \mathbb{R}^{n \times n}$  mit  $\det(P) \neq 0$  und  $PAP^{-1} = D = \text{diag}(\lambda_1, \dots, \lambda_n)$  ist diagonal. Es gilt dann

$$A^k = (P^{-1}DP)^k = P^{-1}D^kP,$$

also folgt  $e^{tA} = P^{-1}e^{tD}P$ , wobei  $e^{tD} = \text{diag}(e^{t\lambda_1}, \dots, e^{t\lambda_n})$ . Dies funktioniert genau so, falls  $A$  über  $\mathbb{C}$  diagonalisierbar ist. Für nicht diagonalisierbare  $A$  wird die Jordan-Normalform zur Berechnung von  $e^{tA}$  verwendet. Schreibe dazu  $A = D + N$  für eine diagonalisierbare Matrix  $D$  und eine nilpotente Matrix  $N$  mit  $DN = ND$ .

## 7 Differentialgleichungen in einer Variable

Sei  $f: \mathbb{R} \rightarrow \mathbb{R}$  stetig (aber nicht unbedingt linear). Wir erhalten eine Differentialgleichung für  $y: (-T, T) \rightarrow \mathbb{R}$  mit  $y(0) = y_0 \in \mathbb{R}$ , nämlich  $\dot{y} = f(y(t))$ . Falls  $f(y_0) = 0$ , dann ist  $y(t) = y_0$  eine konstante Lösung. Falls  $f(y_0) \neq 0$ , schreibe (lokal) um, dass

$$\dot{y}(t) \frac{1}{f(y(t))} = 1.$$

**Lemma.** Mit den Voraussetzungen oben, sei  $g: \mathbb{R} \rightarrow \mathbb{R}$  eine Stammfunktion von  $1/f$ , das heisst  $g'(x) = 1/f(x)$ . Dann gilt  $g(y(t)) = t + c$ .

*Beweis.* Berechne

$$\begin{aligned} \frac{d}{dt} g(y(t)) &= g'(y(t)) \cdot \frac{d}{dt} y(t) \\ &= \frac{1}{f(y(t))} \dot{y}(t) \\ &= 1. \end{aligned}$$

□

Mit etwas Glück lässt sich diese Formel nach  $y(t)$  auflösen.

### Beispiele.

- (1) Betrachte  $f(x) = x$ , das heisst, wir untersuchen die Differentialgleichung  $\dot{y}(t) = y(t)$ . Sei  $y_0 > 0$  und  $g(x) = \log(x)$ . Dann sind die Voraussetzungen vom Lemma erfüllt. Es gilt also  $\log(y(t)) = t + c$ , also  $y(t) = e^{t+c} = e^t \cdot e^c$ . Mit  $y(0) = e^c = y_0$  folgt  $y(t) = y_0 e^t$ .
- (2) Sei  $f(x) = x^2$ , das heisst wir untersuchen  $\dot{y}(t) = (y(t))^2$ . Sei  $y_0 \geq 0$  und  $g(x) = -1/x$ . Aus  $-1/y(t) = t + c$  erhalten wir  $y(t) = -1/(t + c)$ . Mit  $y(0) = -1/c = y_0$  erhalten wir

$$y(t) = \frac{1}{1/y_0 - t}.$$

Im Spezialfall  $y_0 = 1$  sehen wir, dass die Lösung  $y(t) = \frac{1}{1-t}$  in endlicher Zeit divergiert: es gilt

$$\lim_{t \rightarrow 1} \frac{1}{1-t} = +\infty.$$

- (3) Sei  $f(x) = 2\sqrt{x}$ . Dann ist  $\dot{y}(t) = 2\sqrt{y(t)}$ . Wir haben  $g(x) = \sqrt{x}$ , also ist  $\sqrt{y(t)} = t + c$ . Wir schliessen, dass  $y(t) = (t+c)^2$  eine Lösung ist. Mit  $y(0) = y_0$  folgt  $y(t) = (t + \sqrt{y_0})^2$ . Im Grenzfall  $y_0 = 0$  erhalten wir  $y(t) = t^2$ . Es gibt aber noch eine alternative Lösung zur selben Anfangsbedingung, nämlich  $y(t) = 0$ . Diese Differentialgleichung modelliert das Abflussverhalten einer Badewanne. Dies kann einen physikalischen Hinweis darauf geben, dass die Lösungen nicht eindeutig sind. Tatsächlich weiss man, wenn man eine leere Badewanne antrifft nicht, wann sie zum letzten Mal voll war.

### Bemerkungen.

- (i) Die Lösungskurven brauchen nicht für alle Zeiten zu existieren. Im Beispiel zwei zum Beispiel galt  $\lim_{t \rightarrow 1} y(t) = \infty$ .
- (ii) Die Lösungskurven brauchen bei gegebener Anfangsbedingung nicht eindeutig zu sein, siehe Beispiel (3). Für die Eindeutigkeit ist die Lipschitz-Stetigkeit des Vektorfelds essentiell.
- (iii) Für die Existenz von Lösungen reicht die Stetigkeit des Vektorfelds. Diese Tatsache ist als "Satz von Peano" bekannt.

Seien  $a_0, a_1, \dots, a_{n-1} \in \mathbb{R}$ . Wir möchten die *lineare homogene Differentialgleichung  $n$ -ter Ordnung*

$$y^{(n)}(t) + a_{n-1}y^{(n-1)}(t) + \dots + a_2y''(t) + a_1y'(t) + a_0y(t) = 0$$

mit Anfangsbedingungen  $y(0) = a_0, y'(0) = y_1, \dots, y^{(n-1)}(0) = y_{n-1}$ . Konkret gesucht ist eine  $n$ -mal differenzierbare Lösungsfunktion  $y: \mathbb{R} \rightarrow \mathbb{R}$  die diese Bedingungen erfüllt. Unsere Methode ist es, die Differentialgleichung auf eine lineare Differentialgleichung erster Ordnung auf  $\mathbb{R}^n$  zurückzuführen. Führe dazu Koordinaten  $y, y', y'', \dots, y^{(n-1)}$  auf  $\mathbb{R}^n$  ein. Schreibe  $\gamma(t) = (y(t), y'(t), \dots, y^{(n-1)}(t))$ ,

$$A = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ & & \ddots & & \vdots \\ & & & 0 & 1 \\ -a_0 & -a_1 & -a_2 & \dots & -a_{n-1} \end{pmatrix}.$$

Dann liest sich unsere Differentialgleichung als  $\dot{\gamma}(t) = A \cdot \gamma(t)$ . Also ist jede lineare homogene Differentialgleichung  $n$ -ter Ordnung auf  $\mathbb{R}$  äquivalent zu einer linearen homogenen Differentialgleichung erster Ordnung auf  $\mathbb{R}^n$ . Die Anfangsbedingung in unserer Formulierung lautet  $\gamma(0) = (y_0, y_1, \dots, y_{n-1})$ . Die bekannte Lösung  $\gamma(t) = e^{tA} \cdot \gamma(0)$  liefert die gewünschte Lösung  $y(t)$  durch herauslesen der ersten Komponente von  $\gamma$ .

### Beispiele.

- (1) Die Differentialgleichung  $\ddot{y}(t) + y(t) = 0$  ist als *freie Schwingung* bekannt. In der Physik ist  $\ddot{y} = -g/\ell \cdot y$ , wobei  $g$  die Gravitationsbeschleunigung und  $\ell$  die Länge des Pendels ist. Die Funktion  $y$  beschreibt den Auslenkungswinkel. Wir betrachten den Fall  $g/\ell = 1$ . Setze  $\gamma(t) = (y(t), \dot{y}(t))$ . Dann ist

$$\dot{\gamma}(t) = (\dot{y}(t)\ddot{y}(t)) = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} (y(t)\dot{y}(t)).$$

Wenn wir die Anfangsbedingungen  $y_0 = 0$  und  $y_1 = 1$  betrachten, dann ist die Lösung

$$\gamma(t) = \begin{pmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} \sin t \\ \cos t \end{pmatrix}$$

- (2) Die Differentialgleichung  $\ddot{y}(t) + \dot{y}(t) + y(t) = 0$  ist als *gedämpfte Schwingung* bekannt. In der Physik trifft man diese Gleichung als  $\ddot{y}(t) = -ky(t) - \theta\dot{y}(t)$  an, wobei  $\theta$  ein Reibungsterm ist. Setze  $\gamma(t) = (y(t), \dot{y}(t))$  wie oben. Wir erhalten die Differentialgleichung

$$\dot{\gamma}(t) = \begin{pmatrix} 0 & 1 \\ -1 & -1 \end{pmatrix} \begin{pmatrix} y(t) \\ \dot{y}(t) \end{pmatrix}.$$

Um die explizite Lösung  $e^{tA}$  für

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

zu bestimmen, berechne das charakteristische Polynom  $\chi_A(t) = t^2 + t + 1$  und dessen Eigenwerte und Eigenvektoren  $v_1$  und  $v_2$  um  $A$  zu diagonalisieren. Vergleiche Serie 4.

## 8 Rektifizierbare Kurven

Das Material in diesem Abschnitt ist im Abschnitt 177 in [1] nachzuschlagen.

**Definition.** Eine Kurve  $\gamma: [0, T] \rightarrow \mathbb{R}^n$  heisst *stetig differenzierbar*, falls ein offenes Intervall  $(a, b)$  existiert, welches  $[0, T]$  enthält, und eine stetig differenzierbare Kurve  $\bar{\gamma}: (a, b) \rightarrow \mathbb{R}^n$  mit  $\bar{\gamma}|_{[0, T]} = \gamma$  existiert. Wir setzen

$$L(\gamma) = \int_0^T \|\dot{\gamma}(s)\|_2 ds.$$

**Beispiel.** Betrachte die Kurve

$$\begin{aligned} \gamma: [0, 2\pi] &\rightarrow \mathbb{R}^2 \\ t &\mapsto (\cos(t), \sin(t)). \end{aligned}$$

Berechne  $\dot{\gamma}(s) = (-\sin(s), \cos(s))$  und  $\|\dot{\gamma}(s)\|_2 = 1$  für alle  $s \in \mathbb{R}$ . Wir folgern, dass

$$L(\gamma) = \int_0^{2\pi} 1 ds = 2\pi.$$



Eine Kurve braucht nicht stetig differenzierbar zu sein, um eine endliche “Länge” zu haben.

**Beispiele.** Folgende Beispiele sollen Zeigen, dass unser  $L(\gamma)$  als Längenbegriff nicht allgemein genug ist.

(1) Betrachte die Kurve

$$\begin{aligned}\gamma: [0, 1] &\rightarrow \mathbb{R}^2 \\ t &\mapsto (t, \sqrt{t}).\end{aligned}$$

Dann ist  $\gamma$  im Nullpunkt nicht stetig differenzierbar: es gilt

$$\lim_{s \rightarrow 0} \|\dot{\gamma}(s)\|_2 = \lim_{s \rightarrow 0} \sqrt{1 + 1/4s} = +\infty.$$

Somit ist  $\dot{\gamma}$  nicht Riemann-integrierbar. Dennoch hat  $\gamma$  eine endliche “Länge” (siehe auch Abbildung I.6). Der Grund dazu ist folgender. Sei  $\varepsilon > 0$ . Dann gilt

$$\int_{\varepsilon}^1 \|\dot{\gamma}(s)\|_2 ds = \int_{\varepsilon}^1 \sqrt{1 + 1/4s} ds \leq \int_{\varepsilon}^1 \sqrt{5/4s} ds \leq \sqrt{5}(1 - \sqrt{\varepsilon}) \leq \sqrt{5}.$$

(2) Betrachte die Kurve  $\gamma: [0, 1] \rightarrow \mathbb{R}^2$  aus Abbildung I.6 welche durch stückweise gerade Abschnitte deren Länge durch Terme einer geometrischen Reihe  $\sum (1/2)^n$  gegeben sind. Wir geben keine expliziten Formeln dafür an. Die Kurve  $\gamma$  ist in abzählbar vielen Punkten nicht differenzierbar, hat aber “Länge” 1.

(3) Es gibt sogar Kurven endlicher “Länge”, welche in überabzählbar vielen Punkten nicht differenzierbar sind. Eine Konstruktion inspiriert durch die “Kochsche Schneeflocke” liefert solch eine Kurve.

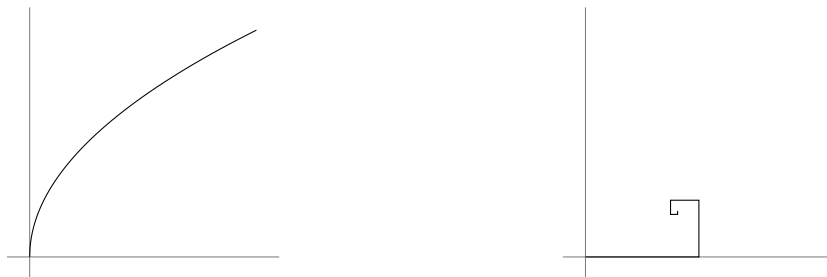


Abbildung I.6: Die Kurven aus den Beispielen (1) und (2)

**Definition.** Sei  $\gamma: [0, 1] \rightarrow \mathbb{R}^n$  stetig. Eine *Partition* des Intervalls  $[0, 1]$  ist eine endliche Folge von Zeitparametern  $0 = t_0 < t_1 < \dots < t_N = 1$ . Wir notieren diese Partition als  $P = \{t_0, t_1, \dots, t_N\}$ . Setze

$$L(\gamma; P) = \sum_{k=1}^N \|\gamma(t_k) - \gamma(t_{k-1})\|_2.$$

Dies ist in Abbildung I.7 veranschaulicht.

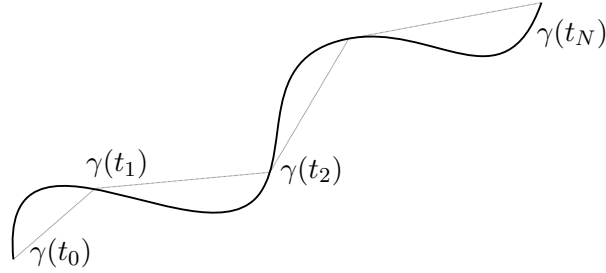


Abbildung I.7: Eine Partition der Länge  $N = 3$

**Bemerkung.** Falls eine Partition  $P_2$  feiner ist als  $P_1$ , das heisst  $P_1 \subset P_2$ , dann gilt  $L(\gamma; P_1) \leq L(\gamma; P_2)$ . Der Grund dafür ist die Dreiecksungleichung.

**Definition.** Sei  $\gamma: [0, 1] \rightarrow \mathbb{R}^n$  stetig. Dann ist die *Länge* von  $\gamma$  das Supremum von  $L(\gamma; P)$ , wobei  $P$  über alle Partitionen von  $[0, 1]$  läuft. Falls die Länge von  $\gamma$  endlich ist, dann heisst  $\gamma$  *rektifizierbar*.

### Beispiele.

- (1) Sei  $\gamma: [0, 1] \rightarrow \mathbb{R}^n$  Lipschitz-stetig mit Konstante  $k \geq 0$ . Dann gilt für alle Zeiten  $t_{k-1}, t_k \in [0, 1]$ , dass  $\|\gamma(t_k) - \gamma(t_{k-1})\|_2 \leq k \cdot |t_k - t_{k-1}|$ . Für jede Partition  $P$  von  $[0, 1]$  gilt also

$$L(\gamma; P) = \sum_{k=1}^N \|\gamma(t_k) - \gamma(t_{k-1})\|_2 \leq k.$$

Also ist  $\gamma$  rektifizierbar mit Länge höchstens  $k$ .

- (2) Betrachte die Kurve

$$\gamma: [0, 1] \rightarrow \mathbb{R}^2$$

$$t \mapsto \begin{cases} (0, 0) & t = 0, \\ (t, t \cos(\pi/t)) & t > 0. \end{cases}$$

Betrachte die Partition  $P = \{0, 1/N, 1/(N-1), \dots, 1/2, 1\}$  für festes  $N \in \mathbb{N}$ . Dann ist  $L(\gamma; P) \geq 1 + 1/2 + \dots + 1/N$ . Vergleiche dazu Serie 5. Es folgt, dass die Länge von  $\gamma$  nicht endlich ist, da die harmonische Reihe divergiert.

**Theorem.** Sei  $\gamma: [0, 1] \rightarrow \mathbb{R}^n$  stetig differenzierbar. Dann ist  $\gamma$  rektifizierbar, und die Länge von  $\gamma$  ist

$$L(\gamma) = \int_0^1 \|\dot{\gamma}(s)\|_2 ds.$$

*Beweis.* Wir zeigen zunächst, dass die Länge von  $\gamma$  höchstens  $L(\gamma)$  ist. Sei dazu  $P = \{t_0, t_1, \dots, t_N\}$  eine Partition von  $[0, 1]$ . Für alle  $a, b \in [0, 1]$  mit  $a \leq b$  gilt

$$\gamma(b) - \gamma(a) = \int_a^b \dot{\gamma}(s) ds.$$

Es folgt, dass

$$\|\gamma(b) - \gamma(a)\|_2 = \left\| \int_a^b \dot{\gamma}(s) ds \right\|_2 \leq \int_a^b \|\dot{\gamma}(s)\|_2 ds.$$

Siehe dazu die Integralabschätzung in der Vorbereitung zum Beweis von Theorem von Cauchy-Lipschitz-Picard-Lindelöf. Wir folgern, dass

$$L(\gamma; P) = \sum_{k=1}^N \|\gamma(t_k) - \gamma(t_{k-1})\|_2 \leq \sum_{k=1}^N \int_{t_{k-1}}^{t_k} \|\dot{\gamma}(s)\|_2 ds = \int_0^1 \|\dot{\gamma}(s)\|_2 ds = L(\gamma).$$

Die Länge von  $\gamma$  ist also höchstens  $L(\gamma) < \infty$ , und  $\gamma$  ist rektifizierbar.

Als nächstes ist die Gleichheit zu zeigen. Sei dazu  $\varepsilon > 0$ . Wir konstruieren eine Partition  $P$  mit  $L(\gamma; P) \geq L(\gamma) - 2\varepsilon$ . Daraus wird folgen, dass das Supremum über alle Partitionen genau  $L(\gamma)$  ist. Da  $\dot{\gamma}: [0, 1] \rightarrow \mathbb{R}^n$  stetig ist, und  $[0, 1]$  kompakt ist folgt, dass  $\dot{\gamma}$  gleichmäßig stetig ist. Es existiert also  $\delta > 0$  so, dass für alle  $a, b \in [0, 1]$  mit  $|b - a| \leq \delta$  gilt, dass  $\|\dot{\gamma}(b) - \dot{\gamma}(a)\|_2 \leq \varepsilon$ . Wähle nun  $N \in \mathbb{N}$  mit  $1/N \leq \delta$ . Betrachte die Partition  $P = \{0, 1/N, 2/N, \dots, (N-1)/N, 1\}$ . Sei  $k \in \{1, \dots, N\}$  beliebig. Dann gilt für  $s \in [t_{k-1}, t_k]$ , dass

$$\|\dot{\gamma}(s)\|_2 \leq \|\dot{\gamma}(t_{k-1})\|_2 + \|\dot{\gamma}(s) - \dot{\gamma}(t_{k-1})\|_2 \leq \|\dot{\gamma}(t_{k-1})\|_2 + \varepsilon,$$

da  $|s - t_{k-1}| \leq 1/N$ . Es gilt also, dass

$$\begin{aligned} \int_{t_{k-1}}^{t_k} \|\dot{\gamma}(s)\|_2 ds &\leq \int_{t_{k-1}}^{t_k} \|\dot{\gamma}(t_{k-1})\|_2 ds + \int_{t_{k-1}}^{t_k} \|\dot{\gamma}(s) - \dot{\gamma}(t_{k-1})\|_2 ds \\ &\leq 1/N \cdot \|\dot{\gamma}(t_{k-1})\|_2 + \varepsilon/N. \end{aligned}$$

Berechne weiterhin, dass

$$\begin{aligned} \left\| \int_{t_{k-1}}^{t_k} \dot{\gamma}(s) ds \right\|_2 &= \left\| \int_{t_{k-1}}^{t_k} \dot{\gamma}(t_{k-1}) ds + \int_{t_{k-1}}^{t_k} \dot{\gamma}(s) - \dot{\gamma}(t_{k-1}) ds \right\|_2 \\ &\geq \left\| \int_{t_{k-1}}^{t_k} \dot{\gamma}(t_{k-1}) ds \right\|_2 - \left\| \int_{t_{k-1}}^{t_k} \dot{\gamma}(s) - \dot{\gamma}(t_{k-1}) ds \right\|_2 \\ &\geq \frac{1}{N} \cdot \|\dot{\gamma}(t_{k-1})\|_2 - \frac{\varepsilon}{N}. \end{aligned}$$

Es folgt, dass

$$\begin{aligned} \|\gamma(t_k) - \gamma(t_{k-1})\|_2 &\geq \frac{\|\dot{\gamma}(t_{k-1})\|_2}{N} - \frac{\varepsilon}{N} \\ &\geq \int_{t_{k-1}}^{t_k} \|\dot{\gamma}(s)\|_2 ds - \frac{2\varepsilon}{N}. \end{aligned}$$

Aufsummieren liefert, dass

$$L(\gamma; P) \geq \int_0^1 \|\dot{\gamma}(s)\|_2 ds - 2\varepsilon.$$

□

**Beispiel** (Schwarzscher Stiefel). Wir versuchen, den Inhalt einer Fläche durch eine “feine” Triangulierung zu approximieren. Leider geht der naive Ansatz schief! Dazu untersuchen wir einen Zylinder der Höhe  $h$  mit Radius 1. Wähle  $N \in \mathbb{N}$  gross, und Punkte  $P_1, \dots, P_N$  auf der Höhe 0, welche ein regelmässiges  $N$ -Eck bilden. Wähle versetzte Punkte  $Q_1, \dots, Q_N$  auf der Höhe  $h$ , welche auch ein regelmässiges  $N$ -Eck bilden. Siehe Abbildung I.8. Wir erhalten eine Triangulierung mit  $2N$  Dreiecken.



Abbildung I.8: Die Punkte  $P_i$  und  $Q_i$  von oben betrachtet und die daraus resultierende Triangulierung

Sei  $\Delta$  eines der Dreiecke. Die Fläche von  $\Delta$  ist ungefähr  $\pi/N \cdot 1$ . Die Gesamtfläche ist also ungefähr  $2\pi$ , was genau die Fläche des Zylinders ist. Sei aber nun  $h$  sehr klein, zum Beispiel  $1/N^3$ . Dann ist die Höhe von  $\Delta$  ungefähr proportional  $1/N^2$  (und nicht wie erwartet  $1/N^3$ ). Das liegt daran, dass  $\cos(1/N)$  etwa  $1 - 1/(2N^2)$  ist. Die Gesamtfläche ist dann etwa proportional zu  $1/N^2$ . Teilen wir nun die ursprüngliche Fläche in  $N^3$  Unterzylinder, so ist der Gesamtflächeninhalt der Dreiecke proportional zu  $N^3/N^2 = N$ . Unsere Definition der Rektifizierbarkeit ist somit für Flächen ungebracht. Ein besserer Flächenbegriff wird in der Differentialgeometrie studiert.

# Kapitel II

## Differentialrechnung

Dieses Kapitel stellt den Hauptteil dieser Vorlesung dar. Auch in [1] ist das korrespondierende Kapitel XX bei weitem das grösste. Der Inhalt ist das Studium differenzierbarer Abbildungen  $f: \mathbb{R}^m \rightarrow \mathbb{R}^n$ . Die erste Schwierigkeit daran ist, dass der Ausdruck

$$\lim_{h \rightarrow 0} \frac{f(p+h) - f(p)}{h}$$

für  $m \geq 2$  keinen Sinn macht: Vektoren kann man nicht definieren.

### 1 Differenzierbarkeit

**Definition.** Sei  $U \subset \mathbb{R}^m$  offen. Eine Abbildung  $f: U \rightarrow \mathbb{R}^n$  heisst *differenzierbar* im Punkt  $p \in U$ , falls folgendes existiert:

- (i) eine lineare Abbildung  $(Df)_p: \mathbb{R}^m \rightarrow \mathbb{R}^n$ , genannt *Differential* von  $f$  an der Stelle  $p$ ,
- (ii) für alle  $h \in \mathbb{R}^m$  mit  $p+h \in U$  ein *Restterm*  $(Rf)_p(h) \in \mathbb{R}^n$ , der relativ klein in  $\|h\|_2$  ist, so dass für alle  $h \in \mathbb{R}^m$  mit  $p+h \in U$  gilt, dass

$$f(p+h) = f(p) + (Df)_p(h) + (Rf)_p(h).$$

Der Unterteilung von  $f$  in diese drei Summanden sagt man *Dreigliedentwicklung*. Die Forderung, dass  $(Rf)_p(h)$  *relativ klein* in  $\|h\|_2$  ist, bedeutet, dass

$$\lim_{h \rightarrow 0} \frac{\|(Rf)_p(h)\|_2}{\|h\|_2} = 0.$$

In anderen Worten existiert für alle  $\varepsilon > 0$  ein  $\delta > 0$ , so dass für alle  $h \in \mathbb{R}^m$  mit  $\|h\|_2 \leq \delta$  gilt, dass  $\|(Rf)_p(h)\|_2 \leq \varepsilon \cdot \|h\|_2$ .

**Bemerkung.** Das Differential ist a priori nur für diejenigen  $h$  definiert, für die  $p + h$  noch in  $U$  liegt. Da  $U$  aber offen ist, existiert aber eine Basis von  $\mathbb{R}^m$  aus Vektoren  $h_i$  für die  $p + h_i$  noch in  $U$  liegt. Da lineare Abbildungen durch die Bilder der Basisvektoren eindeutig bestimmt sind, lässt sich das Differential auf ganz  $\mathbb{R}^m$  erweitern.

**Beispiel.** Sei  $L: \mathbb{R}^m \rightarrow \mathbb{R}^n$  linear. Dann gilt für alle  $p, h \in \mathbb{R}^m$ , dass

$$L(p + h) = L(p) + L(h).$$

Dies ist eine Dreigliedentwicklung für  $L$  an der Stelle  $p$  mit Restterm  $(RL)_p(h) = 0$ . Tatsächlich ist die Abbildung  $(DL)_p = L$  linear. Die informale Erklärung dafür ist, dass  $L$  die “beste lineare Approximation” von  $L$  ist, und das an jeder Stelle.

**Beispiele.**

(1) Betrachte die Funktion

$$\begin{aligned} f: \mathbb{R}^m &\rightarrow \mathbb{R} \\ p &\mapsto \langle p, p \rangle. \end{aligned}$$

Für alle  $p, h \in \mathbb{R}^m$  gilt, dass

$$f(p + h) = \langle p + h, p + h \rangle = \langle p, p \rangle + 2\langle p, h \rangle + \langle h, h \rangle.$$

Dies ist eine Dreigliedentwicklung für  $f$  mit Differential  $(Df)_p(h) = 2\langle p, h \rangle$  und Restterm  $(Rf)_p(h) = \langle h, h \rangle$ . Tatsächlich ist  $(Df)_p$  linear in  $h$ , da Skalarprodukte bilinear sind, und

$$(Rf)_p(h) = \langle h, h \rangle = \|h\|_2^2$$

ist relativ klein in  $\|h\|_2$ . Zum Beweis dafür sei  $\varepsilon > 0$  und setze  $\delta = \varepsilon$ . Für alle  $h \in \mathbb{R}^m$  mit  $\|h\|_2 \leq \delta$  gilt dann, dass  $|(Rf)_p(h)| = \|h\|_2^2 \leq \varepsilon \cdot \|h\|_2$ .

(2) Betrachte die Abbildung

$$\begin{aligned} m: \mathbb{R}^2 &\rightarrow \mathbb{R} \\ (x, y) &\mapsto xy. \end{aligned}$$

Seien  $p = (p_1, p_2)$  und  $h = (h_1, h_2)$  in  $\mathbb{R}^2$ . Dann gilt

$$m(p + h) = (p_1 + h_1) \cdot (p_2 + h_2) = p_1 p_2 + (p_1 h_2 + p_2 h_1) + h_1 h_2.$$

Dies ist eine Dreigliedentwicklung für  $m$  bei  $p$  mit  $(Dm)_p(h) = p_1 h_2 + p_2 h_1$  und  $(Rm)_p(h) = h_1 h_2$ . Da

$$(Dm)_p(h) = \begin{pmatrix} p_2 & p_1 \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \end{pmatrix}$$

gilt, ist  $(Dm)_p$  linear. Weiter ist der Restterm relativ klein in  $\|h\|_2$ , da

$$|(Rm)_p(h)| = |h_1| \cdot |h_2| \leq \|h\|_2 \cdot \|h\|_2 = \|h\|_2^2,$$

und  $\|h\|_2^2$  ist nach Beispiel (1) relativ klein in  $\|h\|_2$ .

Wir treffen nun die Kettenregel als Quelle für viele Beispiele an. Wir werden sie später in diesem Kapitel beweisen.

**Theorem** (Kettenregel). *Seien  $U \subset \mathbb{R}^m$  und  $V \subset \mathbb{R}^k$  offen, und sei  $f: U \rightarrow \mathbb{R}^k$  differenzierbar bei  $p \in U$  mit  $f(U) \subset V$ , sowie  $g: V \rightarrow \mathbb{R}^n$  bei  $f(p) \in V$  differenzierbar. Dann ist die Komposition  $g \circ f: U \rightarrow \mathbb{R}^n$  differenzierbar bei  $p$ , und es gilt*

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

**Beispiel.** Betrachte die Abbildung

$$\begin{aligned} \ell: \mathbb{R}^m &\rightarrow \mathbb{R} \\ p &\mapsto \sqrt{\langle p, p \rangle}. \end{aligned}$$

In Koordinaten  $x_1, \dots, x_m$  ist

$$\ell(x_1, \dots, x_m) = \sqrt{x_1^2 + \dots + x_m^2}.$$

Sei  $U = \mathbb{R}^m \setminus \{0\} \subset \mathbb{R}^m$  und  $V = \mathbb{R}_{>0}$ . Beide dieser Mengen sind offen. Setze weiterhin

$$\begin{aligned} f: U &\rightarrow \mathbb{R} \\ p &\mapsto \langle p, p \rangle \end{aligned}$$

und

$$\begin{aligned} g: \mathbb{R}_{>0} &\rightarrow \mathbb{R} \\ t &\mapsto \sqrt{t}. \end{aligned}$$

Für alle  $p \in U$  gilt  $\ell(p) = g(f(p))$ . Die Funktion  $f$  ist differenzierbar in allen Punkten  $p \in U$  (eigentlich sogar in allen  $p \in \mathbb{R}^n$ ). Ausserdem ist  $g$  differenzierbar an allen Stellen  $t > 0$ . Aus der Analysis I wissen wir, dass

$$(Dg)_t(h) = g'(t) \cdot h = \frac{1}{2\sqrt{t}} \cdot h.$$

Die Kettenregel liefert, dass  $\ell = g \circ f$  an jeder Stelle  $p \in U$  differenzierbar ist, und es gilt

$$(D\ell)_p(h) = (Dg)_{f(p)}((Df)_p(h)) = \frac{1}{2\sqrt{\langle p, p \rangle}} \cdot 2\langle p, h \rangle = \langle p/\|p\|_2, h \rangle.$$

Der Vektor  $p/\|p\|_2$  ist der Vektor mit Länge 1, der in die selbe Richtung zeigt wie  $p$ . Er beschreibt den “Zuwachs” der Funktion  $\ell$ .

**Definition.** Sei  $f: \mathbb{R}^m \rightarrow \mathbb{R}^n$  eine Abbildung. Dann existieren eindeutige *Komponentenfunktionen*  $f_k: \mathbb{R}^m \rightarrow \mathbb{R}$  für  $1 \leq k \leq n$  mit folgender Eigenschaft: für alle  $p \in \mathbb{R}^m$  gilt, dass

$$f(p) = \sum_{k=1}^n f_k(p) \cdot e_k.$$

**Lemma.** Eine Abbildung  $f: \mathbb{R}^m \rightarrow \mathbb{R}^n$  ist differenzierbar, genau dann wenn alle Komponentenfunktionen  $f_k: \mathbb{R}^m \rightarrow \mathbb{R}$  bei  $p$  differenzierbar sind.

*Beweis.* Um zu zeigen, dass die Komponentenfunktionen einer differenzierbaren Abbildung  $f: \mathbb{R}^m \rightarrow \mathbb{R}^n$  selbst differenzierbar sind, sei

$$f(p+h) = f(p) + (Df)_p(h) + (Rf)_p(h)$$

eine Dreigliedentwicklung von  $f$ . Für  $1 \leq k \leq n$  setze  $(Df_k)_p(h) = \langle (Df)_p(h), e_k \rangle$  und  $(Rf_k)_p(h) = \langle (Rf)_p(h), e_k \rangle$ . Dann gilt für alle  $k \leq n$ , dass

$$f_k(p+h) = f_k(p) + (Df_k)_p(h) + (Rf_k)_p(h).$$

Weiterhin ist

(i) die Funktion

$$\begin{aligned} (Df_k)_p: \mathbb{R}^m &\rightarrow \mathbb{R} \\ h &\mapsto \langle (Df)_p(h), e_k \rangle \end{aligned}$$

als Verknüpfung einer linearen Funktion mit einer Projektion linear,

(ii) der Restterm  $(Rf_k)_p(h)$  relativ klein in  $\|h\|_2$ , da  $|(Rf_k)_p(h)| \leq \|(Rf)_p(h)\|_2$  gilt.

Umgekehrt nehmen wir an, dass die Komponentenfunktionen  $f_k: \mathbb{R}^m \rightarrow \mathbb{R}$  einer Funktion  $f: \mathbb{R}^m \rightarrow \mathbb{R}^n$  differenzierbar sind. Für alle  $k$  gibt es dann eine Dreigliedentwicklung  $f_k(p+h) = f_k(p) + (Df_k)_p(h) + (Rf_k)_p(h)$ . Setze

$$(Df)_p(h) = \sum_{k=1}^n (Df_k)_p(h) \cdot e_k$$

und

$$(Rf)_p(h) = \sum_{k=1}^n (Rf_k)_p(h) \cdot e_k.$$

Dann gilt  $f(p+h) = f(p) + (Df)_p(h) + (Rf)_p(h)$ . Weiterhin ist

(i) die Abbildung

$$\begin{aligned} (Df)_p: \mathbb{R}^m &\rightarrow \mathbb{R}^n \\ h &\mapsto \sum_{k=1}^n (Df_k)_p(h) \cdot e_k \end{aligned}$$

linear da Summen linearer Abbildungen linear sind, und



(ii)  $(Rf)_p(h)$  ist relativ klein in  $\|h\|_2$ , da

$$\|(Rf)_p(h)\|_2 \leq \sum_{k=1}^n |(Rf_k)_p(h)|.$$

Dies folgt aus der allgemeinen Ungleichung, dass die euklidische Norm immer kleiner ist als die Summennorm. Weiterhin verwenden wir, dass endliche Summen relativ kleiner Funktionen selbst relativ klein ist. Dazu teile  $\varepsilon$  durch  $n$  um ein geeignetes  $\delta$  als Minimum  $n$  verschiedener  $\delta$  zu bekommen.  $\square$

**Frage.** Können wir die Differenzierbarkeit von Abbildungen  $f: \mathbb{R}^m \rightarrow \mathbb{R}$  weiter auf die Differenzierbarkeit von Funktionen  $g: \mathbb{R} \rightarrow \mathbb{R}$  zurückführen? Konkreter: Ist  $f: \mathbb{R}^m \rightarrow \mathbb{R}$  bei  $p = 0$  differenzierbar, falls die Einschränkung von  $f$  auf allen Koordinatenachsen differenzierbar ist?

Die Antwort hier ist leider “nein”, wie in folgendem Beispiel zu erkennen ist.

**Beispiel.** Betrachte die Funktion

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

Die Einschränkung von  $f$  auf beide Achsen ist identisch null, also differenzierbar. Aber  $f$  ist nicht einmal stetig im Nullpunkt:

$$\lim_{n \rightarrow \infty} f(1/n, 1/n) = \frac{1}{2} \neq f(0, 0).$$

Vergleiche auch Serie 6 für ein Beispiel einer Funktion, die trotz Stetigkeit im Nullpunkt dort nicht differenzierbar ist.

**Bemerkung.** Falls  $f$  an der Stelle  $p \in U$  eine Dreigliedentwicklung besitzt, dann ist diese eindeutig. Insbesondere ist das Differential  $(Df)_p: \mathbb{R}^m \rightarrow \mathbb{R}^n$  wohldefiniert. Der Grund für die Eindeutigkeit des Differentials liegt darin, dass nur die Nullfunktion gleichzeitig linear und relativ klein ist. Man kann zeigen, dass die Differenz zweier potentiellen Differentialen derselben Funktion diese Eigenschaften erfüllt. Die Details dazu werden in Serie 7 behandelt.

**Definition.** Das Differential  $(Df)_p: \mathbb{R}^m \rightarrow \mathbb{R}^n$  einer differenzierbaren Funktion  $f$  ist linear, hat also bezüglich der Standardbasen auf  $\mathbb{R}^m$  und  $\mathbb{R}^n$  eine Abbildungsmatrix  $(Jf)_p \in \mathbb{R}^{n \times m}$ , genannt *Jakobimatrix* von  $f$  an der Stelle  $p \in U$ .

Wir bestimmen nun die Einträge der Jakobimatrix  $(Jf)_p$ . Schreibe wie üblich

$$f(p) = \sum_{k=1}^n f_k(p) e_k.$$

Die Koeffizienten von  $(Jf)_p$  sind dann

$$\begin{aligned} ((Jf)_p)_{ij} &= \langle (Df)_p(e_j), e_i \rangle \\ &= \left\langle \sum_{k=1}^n (Df_k)_p(e_j) e_k, e_i \right\rangle \\ &= (Df_i)_p(e_j). \end{aligned}$$

**Definition.** Sei  $U \subset \mathbb{R}^m$  offen und  $f: U \rightarrow \mathbb{R}^n$  differenzierbar. Die  $j$ -te *partielle Ableitung* von  $f_i: \mathbb{R}^m \rightarrow \mathbb{R}$  ist gegeben durch

$$\frac{\partial f_i}{\partial x_j}(p) = (Df_i)_p(e_j) \in \mathbb{R}.$$

Gemäss dieser Definition ist dann

$$(Jf)_p = \begin{pmatrix} \partial f_1/\partial x_1(p) & \partial f_1/\partial x_2(p) & \cdots & \partial f_1/\partial x_m(p) \\ \partial f_2/\partial x_1(p) & \partial f_2/\partial x_2(p) & \cdots & \partial f_2/\partial x_m(p) \\ \vdots & \vdots & \ddots & \vdots \\ \partial f_n/\partial x_1(p) & \partial f_n/\partial x_2(p) & \cdots & \partial f_n/\partial x_m(p) \end{pmatrix}.$$

Im Spezialfall  $n = 1$ , das heisst, dass  $f: \mathbb{R}^m \rightarrow \mathbb{R}$  skalare Werte annimmt, erhalten wir einen Zeilenvektor  $(Jf)_p$ . Der transponierte Vektor

$$(\nabla f)_p = \begin{pmatrix} \partial f/\partial x_1(p) \\ \vdots \\ \partial f/\partial x_m(p) \end{pmatrix} \in \mathbb{R}^m$$

heisst *Gradient* von  $f$  bei  $p$ . Es gilt

$$(Df)_p(v) = \langle (\nabla f)_p, v \rangle.$$

**Beispiele.**

(1) Betrachte die Funktion

$$\begin{aligned} f: \mathbb{R}^m &\rightarrow \mathbb{R} \\ p &\mapsto \langle p, p \rangle. \end{aligned}$$

In Koordinaten ist

$$f(x_1, x_2, \dots, x_m) = x_1^2 + x_2^2 + \cdots + x_m^2$$

und es gilt

$$(\nabla f)_p = \begin{pmatrix} 2x_1 \\ 2x_2 \\ \vdots \\ 2x_m \end{pmatrix} = 2p.$$

(2) Betrachte die Funktion

$$\begin{aligned} f: \mathbb{R}^m &\rightarrow \mathbb{R} \\ p &\mapsto \sqrt{\langle p, p \rangle} = \|p\|_2. \end{aligned}$$

Dann gilt

$$(Df)_p(v) = \left\langle \frac{p}{\|p\|_2}, v \right\rangle,$$

also ist

$$(\nabla f)_p = \frac{p}{\|p\|_2}.$$

In Koordinaten liest sich das als

$$(\nabla f)_{(x_1, \dots, x_m)} = \frac{1}{\sqrt{x_1^2 + \dots + x_m^2}} \begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix},$$

wobei  $f(x_1, \dots, x_m) = \sqrt{x_1^2 + \dots + x_m^2}$ .

(3) Betrachte die Funktion

$$\begin{aligned} f: \mathbb{R}^2 &\rightarrow \mathbb{R} \\ (x, y) &\mapsto xy. \end{aligned}$$

Es gilt dann

$$(Df)_p(v) = yv_1 + xv_2.$$

Also ist

$$(\nabla f)_{(x,y)} = \begin{pmatrix} y \\ x \end{pmatrix}.$$

Mithilfe folgender Proposition können wir  $(Df)_p(v)$  für beliebige  $v \in \mathbb{R}^n$  ausrechnen. Insbesondere liefert sie Möglichkeiten für die Berechnung der Partiellen Ableitungen  $\partial f / \partial x_i(p) = (Df)_p(e_i)$ .

**Proposition.** Sei  $f: \mathbb{R}^m \rightarrow \mathbb{R}$  differenzierbar an der Stelle  $p \in \mathbb{R}^m$ . Dann gilt für alle  $v \in \mathbb{R}^m$ , dass

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

*Beweis.* Falls  $v = 0$  gilt, dann gilt

$$(Df)_p(0) = 0 = \lim_{t \rightarrow 0} \frac{f(p) - f(p)}{t}.$$

Für  $v \neq 0$  folgt aus der Dreigliedertentwicklung bei  $p$ , dass

$$f(p + tv) - f(p) = (Df)_p(tv) + (Rf)_p(tv).$$

Mithilfe der Linearität folgern wir, dass

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

Sei  $\varepsilon > 0$  vorgegeben. Wähle  $\delta > 0$  so, dass für alle  $h$  mit  $\|h\|_2 \leq \delta$  gilt, dass  $\|(Rf)_p(h)\|_2 \leq \varepsilon \cdot \|h\|_2 / \|v\|_2$ . Für  $t \leq \delta / \|v\|_2$  gilt dann, dass  $\|tv\|_2 \leq \delta$ , also gilt auch, dass  $\|(Rf)_p(tv)\|_2 \leq \varepsilon \cdot \|tv\|_2 / \|v\|_2$ . Wir schliessen, dass

$$\left\| \frac{(Rf)_p(tv)}{t} \right\| \leq \frac{\varepsilon \cdot \|v\|_2}{\|v\|_2} = \varepsilon.$$

In anderen Worten ist

$$\lim_{t \rightarrow 0} \frac{(Rf)_p(tv)}{t} = 0.$$

□

**Bemerkung.** Die Existenz des Grenzwerts

$$\lim_{t \rightarrow 0} \frac{f(p+tv) - f(p)}{t}$$

für alle  $v \in \mathbb{R}^m$  impliziert nicht die Differenzierbarkeit von  $f$  an der Stelle  $p$ .

**Beispiel.** Die Funktion

$$\begin{aligned} f: \mathbb{R}^2 &\rightarrow \mathbb{R} \\ (x, y) &\mapsto \frac{xy(x-y)}{x^2 + y^2} \end{aligned}$$

ist nicht differenzierbar bei  $p = 0$ , obwohl die Funktion auf allen eindimensionalen Untervektorräumen von  $\mathbb{R}^2$  linear ist. Siehe Serie 6 für die Details.

## 2 Die Kettenregel

**Theorem** (Kettenregel). Seien  $U \subset \mathbb{R}^m$  und  $V \subset \mathbb{R}^k$  offen, und sei  $f: U \rightarrow \mathbb{R}^k$  differenzierbar bei  $p \in U$  mit  $f(U) \subset V$ , sowie  $g: V \rightarrow \mathbb{R}^n$  bei  $f(p) \in V$  differenzierbar. Dann ist die Komposition  $g \circ f: U \rightarrow \mathbb{R}^n$  differenzierbar bei  $p$ , und es gilt

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

*Beweis.* Wir leiten eine Dreigliedentwicklung für  $g \circ f$  bei  $p \in U$  her. Sei  $h \in \mathbb{R}^m$  mit  $p + h \in U$ . Berechne mit der Dreigliedentwicklung von  $f$ , dass

$$\begin{aligned} g \circ f(p+h) &= g(f(p+h)) \\ &= g(f(p)) + (Dg)_{f(p)}(h) + (Rf)_p(h). \end{aligned}$$

Für  $\bar{h} = (Df)_p(h) + (Rf)_p(h)$  gilt dann  $f(p) + \bar{h} = f(p+h) \in V$ , da  $p+h \in U$ . Es gilt also mit Hilfe der Dreigliedentwicklung von  $g$ , dass

$$\begin{aligned} g \circ f(p+h) &= g(f(p) + \bar{h}) \\ &= g(f(p)) + (Dg)_{f(p)}(\bar{h}) + (Rg)_{f(p)}(\bar{h}) \\ &= g(f(p)) + (Dg)_{f(p)}((Df)_p(h)) + S_1 + S_2, \end{aligned}$$

wobei  $S_1 = (Dg)_{f(p)}((Rf)_p(h))$  und  $S_2 = (Rg)_{f(p)}((Df)_p(h) + (Rf)_p(h))$ . Wir zeigen nun, dass  $S_1$  und  $S_2$  relativ klein in  $\|h\|_2$  sind. Es folgt dann nämlich, dass

$$g(f(p+h)) = g(f(p)) + (Dg)_{f(p)}((Df)_p(h)) + S_1 + S_2$$

eine Dreigliedentwicklung für die Komposition  $g \circ f$  an der Stelle  $p$  ist, was das Theorem impliziert. Die Behauptung folgt aus untenstehendem Lemma.  $\square$

**Lemma.** Seien  $A_1: \mathbb{R}^m \rightarrow \mathbb{R}^k$  und  $A_2: \mathbb{R}^k \rightarrow \mathbb{R}^n$  linear, so wie  $R_1: \mathbb{R}^m \rightarrow \mathbb{R}^k$  und  $R_2: \mathbb{R}^k \rightarrow \mathbb{R}^n$  relativ klein in  $\|h\|_2$ . Dann sind die Ausdrücke

- (i)  $A_2(R_1(h))$
- (ii)  $R_2(A_1(h)) + R_1(h)$

relativ klein in  $\|h\|_2$ .

*Beweis.* Wir erinnern uns zunächst an die Operatornorm

$$\|A\|_{\text{op}} = \sup \left\{ \frac{\|A(v)\|_2}{\|v\|_2} \mid v \in \mathbb{R}^m \setminus \{0\} \right\}$$

einer linearen Abbildung  $A: \mathbb{R}^m \rightarrow \mathbb{R}^n$ . Setze  $a_1 = \|A_1\|_{\text{op}}$  und  $a_2 = \|A_2\|_{\text{op}}$ . Falls  $A_2 = 0$ , dann ist  $A_2(R_1(h)) = 0$ , was relativ klein in  $\|h\|_2$  ist.

- (i) Falls  $A_2 \neq 0$  gilt, dann gilt  $a_2 > 0$ . Sei  $\varepsilon > 0$  vorgegeben. Wähle ein  $\delta > 0$  so, dass für  $h \in \mathbb{R}^m$  mit  $\|h\|_2 \leq \delta$  gilt, dass  $\varepsilon/a_2 \|h\|_2$  folgt. Dann gilt

$$\|A_2(R_1(h))\|_2 \leq a_2 \cdot \|R_1(h)\|_2 \leq \varepsilon \cdot \|h\|_2.$$

- (ii) Sei  $\varepsilon > 0$  vorgegeben. Wähle  $\delta_1 > 0$  so, dass für alle  $h \in \mathbb{R}^m$  mit  $\|h\|_2 \leq \delta_1$  gilt, dass  $\|R_1(h)\|_2 \leq \|h\|_2$ . Dann folgt, dass

$$\begin{aligned} \|A_1(h) + R_1(h)\|_2 &\leq \|A_1(h)\|_2 + \|R_1(h)\|_2 \\ &\leq a_1 \|h\|_2 + \|h\|_2 \\ &= (a_1 + 1) \|h\|_2. \end{aligned}$$

Wähle  $\delta_2 > 0$  so dass für  $v \in \mathbb{R}^k$  und  $\|v\|_2 \leq \delta_2$  gilt, dass  $\|R_2(v)\|_2 \leq \varepsilon/(a_1+1) \cdot \|v\|_2$ . Wähle nun  $\delta = \min \{\delta_1, \delta_2/(a_1+1)\}$ . Für alle  $h \in \mathbb{R}^m$  mit  $\|h\|_2 \leq \delta$  gilt dann, dass

$$\|A_1(h) + R_1(h)\|_2 \leq (a_1 + 1) \|h\|_2 \leq \delta_2$$

Setze nun  $v = A_1(h) + R_1(h)$ . Wir folgern wegen  $\|v\|_2 \leq \delta_2$ , dass

$$\begin{aligned}\|R_2(A_1(h) + R_1(h))\| &= \|R_2(v)\|_2 \\ &\leq \varepsilon/(a_1 + 1) \cdot \|v\|_2 \\ &\leq \varepsilon \cdot \|h\|_2.\end{aligned}$$

Also ist  $R_2(A_1(h) + R_1(h))$  relativ klein in  $\|h\|_2$ . □

## Anwendungen der Kettenregel

### Dimension eins

Sei  $f: \mathbb{R} \rightarrow \mathbb{R}$  an der Stelle  $x \in \mathbb{R}$  differenzierbar. Dann gilt für alle  $h \in \mathbb{R}$ , dass

$$(Df)_x(h) = h \cdot (Df)_x(1) = h \cdot \lim_{t \rightarrow 0} \frac{f(x + t \cdot 1) - f(x)}{t}.$$

Daraus folgt, dass  $(Df)_x(h) = h \cdot f'(x)$ . Insbesondere ist  $f'(x) = (Df)_x(1)$ . Seien nun  $f: \mathbb{R} \rightarrow \mathbb{R}$  und  $g: \mathbb{R} \rightarrow \mathbb{R}$  an den Stellen  $x \in \mathbb{R}$  beziehungsweise  $f(x) \in \mathbb{R}$  differenzierbar. Dann ist  $g \circ f: \mathbb{R} \rightarrow \mathbb{R}$  an der Stelle  $x$  differenzierbar und es gilt

$$\begin{aligned}(g \circ f)'(x) &= (D(g \circ f))_x(1) \\ &= (Dg)_{f(x)}((Df)_x(1)) \\ &= (Dg)_{f(x)}(f'(x)) \\ &= f'(x) \cdot g'(f(x)).\end{aligned}$$

Wir haben somit die eindimensionale Kettenregel

$$(g \circ f)'(x) = g'(f(x)) \cdot f'(x)$$

erfolgreich aus der höherdimensionalen Version hergeleitet.

### Die Produktregel

Seien  $f, g: \mathbb{R}^n \rightarrow \mathbb{R}$  an der Stelle  $p \in \mathbb{R}^n$  differenzierbar. Das Produkt  $fg: \mathbb{R}^n \rightarrow \mathbb{R}$  lässt sich als Komposition  $m \circ h$  darstellen, wobei

$$\begin{aligned}h: \mathbb{R}^n &\rightarrow \mathbb{R}^2 \\ q &\mapsto (f(q), g(q))\end{aligned}$$

und

$$\begin{aligned}m: \mathbb{R}^2 &\rightarrow \mathbb{R} \\ (x, y) &\mapsto xy.\end{aligned}$$

Wir erinnern uns, dass

$$(Dh)_p(v) = ((Df)_p(v), (Dg)_p(v))$$

und

$$(Dm)_{(x,y)}(h_1, h_2) = y \cdot h_1 + x \cdot h_2.$$

Daraus folgt, dass

$$\begin{aligned} (D(fg))_p(v) &= (D(m+h))_p(v) \\ &= (Dm)_{h(p)}((Dh)_p(v)) \\ &= (Dm)_{(f(p), g(p))}((Df)_p(v), (Dg)_p(v)) \\ &= g(p) \cdot (Df)_p(v) + f(p) \cdot (Dg)_p(v). \end{aligned}$$

Kurz erhalten wir die *Produktregel*

$$(D(fg))_p = g(p) \cdot (Df)_p + f(p) \cdot (Dg)_p.$$

**Bemerkung.** Die Summe  $f + g: \mathbb{R}^n \rightarrow \mathbb{R}$  ist ebenso differenzierbar bei  $p \in \mathbb{R}^n$  und es gilt

$$(D(f+g))_p = (Df)_p + (Dg)_p.$$

Es folgt, dass alle polynomialen Abbildungen  $\mathbb{R}^n \rightarrow \mathbb{R}$  auf ganz  $\mathbb{R}^n$  differenzierbar sind. Dies kann durch Induktion über den Grad unter Benutzung der Produkt- und Additionsregeln gezeigt werden. Die Verankerung ist lediglich, dass konstante und lineare Abbildungen  $\mathbb{R}^n \rightarrow \mathbb{R}$  auf  $\mathbb{R}^n$  differenzierbar sind. Insbesondere sind dies alle Koordinatenfunktionen  $x_i: \mathbb{R}^n \rightarrow \mathbb{R}$ .

## Extrema von Abbildungen

**Definition.** Sei  $U \subset \mathbb{R}^n$  offen und  $f: U \rightarrow \mathbb{R}$  eine Abbildung. Dann hat  $f$  an der Stelle  $p \in U$  ein *lokales Minimum* (beziehungsweise ein *lokales Maximum*) falls ein  $r > 0$  existiert mit

- (i)  $B_p(r) = \{q \in \mathbb{R}^n \mid \|q - p\|_2 < r\} \subset U$ ,
- (ii) für alle  $q \in B_p(r)$  gilt  $f(q) \geq f(p)$  (beziehungsweise  $f(q) \leq f(p)$ ).

**Proposition.** Sei  $f: U \rightarrow \mathbb{R}$  differenzierbar an der Stelle  $p \in U$ , welche ein lokales Minimum (beziehungsweise Maximum) von  $f$  ist. Dann gilt  $(Df)_p = 0$ .

*Beweis.* Sei  $v \in \mathbb{R}^n$  beliebig. Wähle  $\delta > 0$  so, dass  $B_p(\delta \cdot \|v\|_2) \subset U$  gilt. Definiere eine Kurve

$$\begin{aligned} \gamma: (-\delta, \delta) &\rightarrow \mathbb{R}^n \\ t &\mapsto p + tv. \end{aligned}$$

Es gilt  $\gamma(0) = p$ ,  $\gamma'(0) = v$ , und das Bild von  $\gamma$  liegt im Ball  $B_p(\delta \cdot \|v\|_2)$ . Betrachte die Komposition

$$\begin{aligned} h: (-\delta, \delta) &\rightarrow \mathbb{R} \\ t &\mapsto f(\gamma(t)). \end{aligned}$$

Nach Annahme hat die Funktion  $h$  an der Stelle  $t = 0$  ein lokales Minimum (beziehungsweise Maximum). Wir bemerken mit der Kettenregel, dass  $h$  an der Stelle  $t = 0$  differenzierbar ist. Es folgt aus Diskussion in der Analysis I, dass  $h'(0) = 0$ . Andererseits gilt

$$h'(0) = (Dh)_0(1) = (Df)_{\gamma(0)}((D\gamma)_0(1)) = (Df)_{\gamma(0)}(\gamma'(0)) = (Df)_p(v). \quad \square$$

**Beispiel.** Die Funktion

$$\begin{aligned} \ell: \mathbb{R}^n &\rightarrow \mathbb{R} \\ p &\mapsto \langle p, p \rangle \end{aligned}$$

hat ein (globales) Minimum bei  $p = 0$ . Die Beobachtung, dass  $(\nabla \ell)_0 = 0$  gilt, ist kompatibel mit obiger Proposition.

## Gradient und Niveaulinien

Seien  $\gamma: \mathbb{R} \rightarrow \mathbb{R}^n$  und  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  differenzierbar. Dann ist die Komposition  $h = f \circ \gamma: \mathbb{R} \rightarrow \mathbb{R}$  differenzierbar. Es gilt für alle  $t \in \mathbb{R}$ , dass

$$\begin{aligned} h'(t) &= (Dh)_t(1) = (Df)_{\gamma(t)}((D\gamma)_t(1)) \\ &= (Df)_{\gamma(t)}(\dot{\gamma}(t)) \\ &= \langle (\nabla f)_{\gamma(t)}, \dot{\gamma}(t) \rangle. \end{aligned}$$

**Proposition.** Sei  $\gamma: \mathbb{R} \rightarrow \mathbb{R}^n$  eine differenzierbare Kurve und  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  eine differenzierbare Funktion. Dann ist  $h = f \circ \gamma$  konstant, genau dann, wenn für alle  $t \in \mathbb{R}$  der Gradient  $(\nabla f)_{\gamma(t)}$  senkrecht auf  $\dot{\gamma}(t)$  steht.

*Beweis.* Die Funktion  $h$  ist konstant, genau dann, wenn  $h'(t) = 0$  für alle  $t \in \mathbb{R}$  gilt. Dies wiederum ist äquivalent dazu zu fordern, dass

$$\langle (\nabla f)_{\gamma(t)}, \dot{\gamma}(t) \rangle = 0. \quad \square$$

**Definition.** Das Bild einer Kurve  $\gamma: \mathbb{R} \rightarrow \mathbb{R}^n$  heisst *Niveaulinie* von  $f$ , falls  $f \circ \gamma(t)$  konstant ist.

Der Gradient von  $f$  steht somit senkrecht auf den Niveaulinien, siehe Abbildung II.1

**Beispiel.** Betrachte die Funktion

$$\begin{aligned} f: \mathbb{R}^2 &\rightarrow \mathbb{R} \\ p &\mapsto \langle p, p \rangle \end{aligned}$$





Abbildung II.1: Eine Niveaulinie

und die Kurve

$$\begin{aligned}\gamma: \mathbb{R} &\rightarrow \mathbb{R}^2 \\ t &\mapsto (\cos(t), \sin(t)).\end{aligned}$$

Es gilt für alle  $t \in \mathbb{R}$ , dass

$$\langle (\nabla f)_{\gamma(t)}, \dot{\gamma}(t) \rangle = \langle 2\gamma(t), \dot{\gamma}(t) \rangle = 0.$$

Tatsächlich sind die Kurven

$$\begin{aligned}\gamma_r: \mathbb{R} &\rightarrow \mathbb{R}^2 \\ t &\mapsto (r \cos(t), r \sin(t))\end{aligned}$$

für alle  $r > 0$  Niveaulinien von  $f$ .

### Der Mittelwertsatz

Sei  $a, b \in \mathbb{R}^n$  mit  $a \neq b$ . Definiere  $I_{a,b} = \{a + t(b - a) \mid t \in [0, 1]\}$ . Die Menge  $I_{a,b}$  besteht aus allen Punkten die auf dem Geradenabschnitt zwischen  $a$  und  $b$  liegen.

**Theorem 1** (Mittelwertsatz). *Sei  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  differenzierbar und  $a, b \in \mathbb{R}^n$  mit  $a \neq b$ . Dann existiert  $p \in I_{a,b} \setminus \{a, b\}$  mit*

$$(Df)_p(b - a) = f(b) - f(a).$$

### Spezialfälle.

- (i) Sei  $f: \mathbb{R} \rightarrow \mathbb{R}$  (das heisst  $n = 1$ ). Seien  $a, b \in \mathbb{R}$  mit  $a < b$ . Dann existiert  $x \in I_{a,b} \setminus \{a, b\} = (a, b)$  mit  $(Df)_{x(b-a)=f(b)-f(a)}$ . Es gilt also

$$f'(x) = \frac{f(b) - f(a)}{b - a}.$$

Wir haben also die eindimensionale Version des Mittelwertsatzes erfolgreich aus der momentanen Version extrahiert.

(ii) Nehme an, dass  $f(b) = f(a)$  gilt. Dann existiert  $p \in I_{a,b} \setminus \{a, b\}$  mit

$$\langle (\nabla f)_p, b - a \rangle = (Df)_p(b - a) = 0,$$

das heisst  $(\nabla f)_p$  steht senkrecht auf  $b - a$ . Betrachte zum Beispiel

$$\begin{aligned} f: \mathbb{R}^2 &\rightarrow \mathbb{R} \\ p &\mapsto \langle p, p \rangle \end{aligned}$$

und seien  $a, b \in \mathbb{R}^2$  mit  $\|a\|_2 = \|b\|_2$ . Für  $p = (a + b)/2$  gilt dann  $(\nabla f)_p = 2p$ , was senkrecht auf  $b - a$  steht.

*Beweis des Mittelwertsatzes.* Wir leiten das Resultat aus dem Satz von Rolle aus der Analysis I her. Betrachte die Hilfsfunktion

$$\begin{aligned} h: [0, 1] &\rightarrow \mathbb{R} \\ t &\mapsto f(a + t(b - a)) - f(a) - t(f(b) - f(a)). \end{aligned}$$

Berechne

$$h(0) = h(1) = 0.$$

Für alle  $t \in (0, 1)$  gilt dank der Kettenregel, dass

$$h'(t) = (Df)_{a+t(b-a)}(b - a) - (f(b) - f(a)),$$

da  $(Df)_{\gamma(t)}(\dot{\gamma}(t)) = (Df)_{a+t(b-a)}(b - a)$ . Nach dem Satz von Rolle existiert  $t_0 \in (0, 1)$  mit  $h'(t_0) = 0$ . Setze  $p = a + t_0(b - a) \in I_{a,b} \setminus \{a, b\}$ . Dann gilt

$$0 = h'(t_0) = (Df)_{a+t_0(b-a)}(b - a) - (f(b) - f(a)).$$

Es folgt, dass  $(Df)_p(b - a) = f(b) - f(a)$ . □

## Diffeomorphismen

**Definition.** Eine differenzierbare bijektive Abbildung  $f: \mathbb{R}^n \rightarrow \mathbb{R}^n$  heisst *Diffeomorphismus*, falls  $f$  eine differenzierbare Umkehrabbildung  $g: \mathbb{R}^n \rightarrow \mathbb{R}^n$  hat. Das heisst, dass  $g \circ f = f \circ g = \text{Id}_{\mathbb{R}^n}$ .

Für einen Diffeomorphismus  $f: \mathbb{R}^n \rightarrow \mathbb{R}^n$  mit Umkehrabbildung  $g$  erhalten wir dann

$$(Dg)_{f(p)} \circ (Df)_p = (D\text{Id}_{\mathbb{R}^n})_p = \text{Id}_{\mathbb{R}^n}.$$

Ähnlich gilt auch  $(Df)_p \circ (Dg)_{f(p)} = \text{Id}_{\mathbb{R}^n}$ . Folglich sind die linearen Abbildungen  $(Df)_p$  und  $(Dg)_{f(p)}$  zueinander inverse Vektorraumisomorphismen von  $\mathbb{R}^n$ . Insbesondere gilt  $\det(Df)_p \neq 0$ .

### 3 Der Umkehrsatz

In [1] ist der Umkehrsatz im Abschnitt 171 diskutiert. Wir haben oben gesehen, dass Differentiale differenzierbare Abbildungen mit differenzierbarer Umkehrfunktion Vektorraumisomorphismen sind. In diesem Abschnitt werden wir die Umkehrung dieser Aussage untersuchen.

**Beispiel.** Sei  $A: \mathbb{R}^n \rightarrow \mathbb{R}^n$  linear mit  $\det(A) = \det((DA)_p) \neq 0$ . Dann ist  $A$  eine differenzierbare Funktion mit differenzierbarer Umkehrabbildung  $A^{-1}$ .

Wir erinnern uns nun an ein eindimensionales Phänomen, welches wir bereits in der Analysis I untersucht haben. Sei  $f: \mathbb{R} \rightarrow \mathbb{R}$  differenzierbar mit  $f'(x) > 0$  für alle  $x \in \mathbb{R}$ . Dann ist  $f$  injektiv. Tatsächlich gilt für alle  $a, b \in \mathbb{R}$  mit  $a < b$ , dass

$$f(b) - f(a) = \int_a^b f'(x) dx > 0.$$

Insbesondere gilt  $f(b) \neq f(a)$ .

**Frage.** Sei  $f: \mathbb{R}^n \rightarrow \mathbb{R}^n$  stetig differenzierbar, so dass für alle  $p \in \mathbb{R}^n$  gilt, dass das Differential  $(Df)_p$  in allen Punkten  $p \in \mathbb{R}^n$  ein Isomorphismus ist, das heisst  $\det((Df)_p) \neq 0$ . Ist dann  $f$  injektiv?

Die Antwort dazu ist nein, wie folgendes Beispiel zeigt.

**Beispiel.** Betrachte die *komplexe Exponentialfunktion*

$$\begin{aligned} f: \mathbb{R}^2 &\rightarrow \mathbb{R}^2 \\ (x, y) &\mapsto (e^x \cos y, e^x \sin y). \end{aligned}$$

Dann ist  $f$  stetig differenzierbar, da die Komponentenfunktionen  $f_1(x, y) = e^x \cos y$  und  $f_2(x, y) = e^x \sin y$  differenzierbar sind. Das folgt aus Betrachtungen in der Analysis I und der Produktregel. Berechne für  $p = (x, y)$  die Jakobimatrix

$$(Jf)_p = \begin{pmatrix} \frac{\partial f_1}{\partial x} & \frac{\partial f_1}{\partial y} \\ \frac{\partial f_2}{\partial x} & \frac{\partial f_2}{\partial y} \end{pmatrix} = \begin{pmatrix} e^y \cos y & -e^x \sin y \\ e^x \sin y & e^x \cos y \end{pmatrix}.$$

Aus  $\cos^2 + \sin^2 = 1$  folgt, dass  $\det((Jf)_p) = e^x > 0$  für alle  $p = (x, y)$ . Aber  $f$  ist nicht injektiv, da  $f(x, y) = f(x, y + 2\pi)$  gilt.

#### Stetige Differenzierbarkeit

Sei  $U \subset \mathbb{R}^n$  offen und  $f: U \rightarrow \mathbb{R}^m$  differenzierbar. Dann erhalten wir für alle  $p \in U$  eine lineare Abbildung  $(Df)_p: \mathbb{R}^n \rightarrow \mathbb{R}^m$ , das heisst, dass  $(Df)_p \in \text{Hom}(\mathbb{R}^n, \mathbb{R}^m)$ , wobei  $\text{Hom}(\mathbb{R}^n, \mathbb{R}^m)$  aus allen linearen Abbildungen  $L: \mathbb{R}^n \rightarrow \mathbb{R}^m$  besteht. Auf  $\text{Hom}(\mathbb{R}^n, \mathbb{R}^m)$  haben wir eine Norm, nämlich die Operatornorm  $\|\cdot\|_{\text{op}}$ .

**Definition.** Sei  $U \subset \mathbb{R}^n$  offen. Eine Abbildung  $f: U \rightarrow \mathbb{R}^m$  heisst *stetig differenzierbar* oder  $C^1$ , falls  $f$  differenzierbar ist, und die Abbildung

$$\begin{aligned} df: U &\rightarrow \text{Hom}(\mathbb{R}^n, \mathbb{R}^m) \\ p &\mapsto (Df)_p \end{aligned}$$

bezüglich der Operatornorm auf  $\text{Hom}(\mathbb{R}^n, \mathbb{R}^m)$  stetig ist. Ausgepackt heisst das, dass für alle  $p \in U$  und alle  $\varepsilon > 0$  ein  $\delta > 0$  existiert, so dass

- (i)  $B_p(\delta) \subset U$ ,
- (ii) für alle  $q \in B_p(\delta)$  gilt, dass  $\|(Df)_q - (Df)_p\|_{\text{op}} \leq \varepsilon$ .

In Koordinaten heisst das, dass die Einträge der Jakobimatrix  $(Jf)_p \in \mathbb{R}^{m \times n}$  stetig von  $p$  abhängen. In anderen Worten ist für alle  $i \leq m$  und  $j \leq n$  die partielle Ableitungsfunktion

$$\begin{aligned} \frac{\partial f_i}{\partial x_j}: U &\rightarrow \mathbb{R} \\ p &\mapsto \frac{\partial f_i}{\partial x_j}(p) = \lim_{t \rightarrow 0} \frac{f_i(p + te_j) - f_i(p)}{t} \end{aligned}$$

stetig.

### Bemerkungen.

- (1) Die Existenz der partiellen Ableitungsfunktionen reicht nicht, um die Differenzierbarkeit von  $f$  zu garantieren. Das Beispiel dazu war die Funktion

$$\begin{aligned} f: \mathbb{R}^2 &\rightarrow \mathbb{R} \\ (x, y) &\mapsto \begin{cases} 0 & (x, y) = (0, 0), \\ \frac{xy}{x^2+y^2} & (x, y) \neq (0, 0). \end{cases} \end{aligned}$$

Diese Funktion  $f$  hat in allen Punkten  $p = (x, y)$  partielle Ableitungen, ist aber nicht einmal stetig bei  $p = (0, 0)$ .

- (2) Es existiert dennoch ein Differenzierbarkeitskriterium, welches auf der Stetigkeit der partiellen Ableitungen beruht.

**Proposition 1.** Sei  $U \subset \mathbb{R}^n$  offen und  $f: U \rightarrow \mathbb{R}$  eine Abbildung, so dass für alle  $j \leq n$  und alle  $p \in U$  die partielle Ableitung  $\partial f / \partial x_j(p) \in \mathbb{R}$  existiert, und die Funktion

$$\begin{aligned} \frac{\partial f}{\partial x_j}: U &\rightarrow \mathbb{R} \\ p &\mapsto \frac{\partial f}{\partial x_j}(p) \end{aligned}$$

stetig ist. Dann ist  $f$  stetig differenzierbar und insbesondere differenzierbar.

Wir werden Proposition 1 nicht beweisen. Der Beweis kann in [1] bei 163.1 nachgeschlagen werden.

### Beispiele.

(1) Alle polynomialen Abbildungen  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  haben stetige partielle Ableitungen, sind also stetig differenzierbar.

(2) Die Funktion

$$\begin{aligned} f: \mathbb{R}^2 &\rightarrow \mathbb{R}^2 \\ (x, y) &\mapsto (e^x \cos y, e^x \sin y) \end{aligned}$$

hat stetige partielle Ableitungen und ist somit stetig differenzierbar.

**Proposition 2** (Lokale Injektivität). *Sei  $U \subset \mathbb{R}^n$  offen,  $f: U \rightarrow \mathbb{R}^n$  stetig differenzierbar, und  $p \in U$  so dass  $(Df)_p: \mathbb{R}^n \rightarrow \mathbb{R}^n$  ein Isomorphismus ist. Dann existiert  $\varepsilon > 0$  mit*

- (i)  $B_p(\varepsilon) \subset U$ ,
- (ii) *die Einschränkung von  $f$  auf  $B_p(\varepsilon)$  ist injektiv.*

*Beweis.* Nach Annahme ist  $A = (Df)_p$  ein Isomorphismus. Betrachte  $\bar{f}$  die Komposition  $\bar{f} = A^{-1} \circ f: U \rightarrow \mathbb{R}^n$ . Für alle  $q \in U$  gilt nach der Kettenregel, dass

$$(D\bar{f})_q = (DA^{-1})_{f(q)} \circ (Df)_q = A^{-1} \circ (Df)_q.$$

In anderen Worten ist  $\bar{f}$  auch stetig differenzierbar und es gilt  $(D\bar{f})_p = \text{Id}_{\mathbb{R}^n}$ .

Wähle nun  $\varepsilon > 0$  so, dass

- (i)  $B_p(\varepsilon) \subset U$ ,
- (ii) für alle  $q \in B_p(\varepsilon)$  gilt, dass  $\|(D\bar{f})_q - (D\bar{f})_p\|_{\text{op}} \leq 1/2$ .

Wir behaupten nun, dass die Einschränkung  $\bar{f}|_{B_p(\varepsilon)}: B_p(\varepsilon) \rightarrow \mathbb{R}^n$  injektiv ist. Daraus wird folgen, dass  $f|_{B_p(\varepsilon)} = A \circ \bar{f}|_{B_p(\varepsilon)}$  auch injektiv ist, da  $A$  selbst bijektiv ist. Es bleibt also lediglich die Behauptung zu zeigen.

Dazu seien  $q, \tilde{q} \in B_p(\varepsilon)$  mit  $q \neq \tilde{q}$ . Wir wollen zeigen, dass  $\bar{f}(\tilde{q}) \neq \bar{f}(q)$ . Dazu betrachte die Kurve

$$\begin{aligned} \gamma: [0, 1] &\rightarrow B_p(\varepsilon) \\ t &\mapsto q + t(\tilde{q} - q). \end{aligned}$$

Definiere eine Hilfsabbildung

$$\begin{aligned} h: [0, 1] &\rightarrow \mathbb{R}^n \\ t &\mapsto \bar{f}(\gamma(t)) = \bar{f}(q + t(\tilde{q} - q)). \end{aligned}$$

Dann ist  $h$  eine differenzierbare Kurve mit

$$h'(t) = (D\bar{f})_{\gamma(t)}(\dot{\gamma}(t)) = (D\bar{f})_{q+t(\tilde{q}-q)}(\tilde{q}-q) \in \mathbb{R}^n.$$

Berechne nun mithilfe dem Integralsatz für Kurven, dass

$$\bar{f}(\tilde{q}) - \bar{f}(q) = h(1) - h(0) = \int_0^1 h'(t) dt = \int_0^1 (D\bar{f})_{q+t(\tilde{q}-q)}(\tilde{q}-q) dt.$$

Berechne nun

$$\begin{aligned} \bar{f}(\tilde{q}) - \bar{f}(q) &= \int_0^1 (D\bar{f})_p(\tilde{q}-q) dt + \int_0^1 (D\bar{f})_{q+t(\tilde{q}-q)}(\tilde{q}-q) - (D\bar{f})_p(\tilde{q}-q) dt \\ &= \tilde{q} - q + R, \end{aligned}$$

wobei

$$R = \int_0^1 (D\bar{f})_{q+t(\tilde{q}-q)}(\tilde{q}-q) - (D\bar{f})_p(\tilde{q}-q) dt.$$

Schätze ab, dass

$$\|R\|_2 \leq \int_0^1 \|(D\bar{f})_{q+t(\tilde{q}-q)}(\tilde{q}-q) - (D\bar{f})_p(\tilde{q}-q)\|_2 dt.$$

Für alle  $t \in [0, 1]$  gilt nach Wahl von  $\varepsilon$ , dass  $\|(D\bar{f})_{\gamma(t)} - (D\bar{f})_p\|_{\text{op}} \leq 1/2$  gilt, da  $\gamma(t) \in B_p(\varepsilon)$ . Es folgt, dass  $\|(D\bar{f})_{\gamma(t)}(\tilde{q}-q) - (D\bar{f})_p(\tilde{q}-q)\|_2 \leq 1/2 \cdot \|\tilde{q}-q\|_2$ , also auch

$$\|R\|_2 \leq \int_0^1 \frac{1}{2} \cdot \|\tilde{q}-q\|_2 dt \leq \frac{1}{2} \cdot \|\tilde{q}-q\|_2.$$

Aus  $\bar{f}(\tilde{q}) - \bar{f}(q) = \tilde{q} - q + R$  mit  $\|R\|_2 \leq 1/2 \cdot \|\tilde{q}-q\|_2$  folgt, dass

$$\|\bar{f}(\tilde{q}) - \bar{f}(q)\|_2 \geq 1/2 \cdot \|\tilde{q}-q\|_2,$$

und insbesondere ist  $\bar{f}(\tilde{q}) \neq \bar{f}(q)$ . □

**Theorem 2** (Umkehrsatz). Sei  $U \subset \mathbb{R}^n$  offen,  $f: U \rightarrow \mathbb{R}^n$  stetig differenzierbar, und  $p \in U$  ein Punkt mit  $\det((Df)_p) \neq 0$ . Dann existieren offene Mengen  $V, W \subset \mathbb{R}^n$  mit  $p \in V \subset U$  und  $f(p) \in W$ , sowie eine differenzierbare Abbildung  $g: W \rightarrow V$  mit  $f \circ g = \text{Id}_W$  und  $g \circ f|_V = \text{Id}_V$ . Für alle  $y \in W$  gilt  $(Dg)_y = (Df)_{g(y)}^{-1}$ .

Die einzelnen Schritte des Beweises sind folgende.

- (1) Lokale Injektivität (siehe Proposition 2 oben),
- (2) Lokale Surjektivität,
- (3) Differenzierbarkeit der lokalen Umkehrfunktion  $g: W \rightarrow V$ .

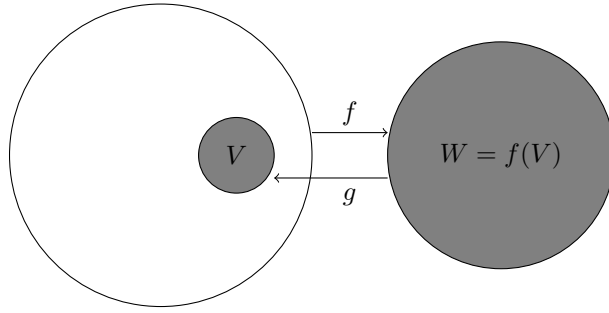


Abbildung II.2: Der Umkehrsatz

Um die beiden letzten Schritte des Beweises auszuführen, treffen wir folgende Vorbereitungen.

(i) Definiere

$$\begin{aligned}\bar{f}: U &\rightarrow \mathbb{R}^n \\ q &\mapsto ((Df)_p^{-1} \circ f)(q).\end{aligned}$$

Dann gilt  $(D\bar{f})_p = (Df)_p^{-1} \circ (Df)_p = \text{Id}_{\mathbb{R}^n}$ . Wähle  $\varepsilon > 0$  so, dass  $B_p(\varepsilon) \subset U$ , und so dass für alle  $q \in B_p(\varepsilon)$  gilt, dass  $\|(D\bar{f})_q - (D\bar{f})_p\|_{\text{op}} \leq 1/2$ . Für alle  $q, \tilde{q} \in B_p(\varepsilon)$  gilt dann, dass

$$\bar{f}(\tilde{q}) - f(q) = \tilde{q} - q + R$$

mit  $\|R\|_2 \leq 1/2 \cdot \|\tilde{q} - q\|_2$ , wie im Beweis von Proposition 2. Insbesondere gilt

$$1/2 \cdot \|\tilde{q} - q\|_2 \leq \|\bar{f}(\tilde{q}) - \bar{f}(q)\|_2 \leq 3/2 \cdot \|\tilde{q} - q\|_2.$$

(ii) Definiere  $\tilde{U} = U - p = \{q \in \mathbb{R}^n \mid q + p \in U\}$  und

$$\begin{aligned}\tilde{f}: \tilde{U} &\rightarrow \mathbb{R}^n \\ q &\mapsto f(p + q) - f(p).\end{aligned}$$

Dann gilt  $\tilde{f}(0) = \bar{f}(p) - \bar{f}(p) = 0$ . Weiterhin gilt  $(D\tilde{f})_0 = \text{Id}_{\mathbb{R}^n}$ . Dies fällt aus der Dreigliedrentwicklung

$$\tilde{f}(0 + h) = \tilde{f}(h) = \bar{f}(p + h) - \bar{f}(p) = (D\bar{f})_p(h) + (R\bar{f})_p(h)$$

Wir werden nun mithilfe dieser Vereinfachungen annehmen, dass

- $p = 0$ ,
- $f(0) = 0$ ,
- $(Df)_0 = \text{Id}_{\mathbb{R}^n}$ .

Mit dem  $\varepsilon > 0$  aus dem Beweis von Proposition 2 gilt, dass  $f|_{B_0(\varepsilon)}$  injektiv ist. In einem nächsten Schritt werden wir zeigen, dass für das selbe  $\varepsilon > 0$  auch gilt, dass  $B_0(\varepsilon/2) \subset f(B_0(\varepsilon))$ , das heisst, für alle  $y \in \mathbb{R}^n$  mit  $\|y\|_2 < \varepsilon/2$  existiert ein  $q \in \mathbb{R}^n$  mit  $\|q\|_2 < \varepsilon$  und  $f(q) = y$ . Nach dem ersten Schritt ist dieses  $q \in B_0(\varepsilon)$  dann eindeutig. Wir werden es mit  $g(y)$  bezeichnen.

Es stellt sich heraus, dass wir  $g$  rekursiv definieren können. Als nullten Ansatz wählen wir

$$g_0(y) = y.$$

Da  $f$  in erster Ordnung die Identität ist, sollte das nicht allzu schlecht sein. Der erste Ansatz ist dann  $g_1(y) = g_0(y) + y - f(g_0(y))$ . Rekursiv setzen wir

$$g_{n+1}(y) = g_n(y) + y - f(g_n(y)).$$

**Bemerkung.** Falls per Zufall  $f(g_n(y)) = y$  für ein  $n \in \mathbb{N}$  gilt, dann folgt auch, dass  $g_{n+1}(y) = g_n(y)$ . Insbesondere ist auch  $f(g_{n+1}(y)) = y$ .

**Behauptung** (Lokale Surjektivität). *Der Grenzwert  $g(y) = \lim_{n \rightarrow \infty} g_n(y) \in B_0(\varepsilon)$  existiert, und  $f(g(y)) = y$ .*

*Beweis.* Durch vollständige Induktion zeigen wir

- (i)  $\|g_n(y)\|_2 < (1 - 1/2^{n+1}) \cdot \varepsilon$
- (ii)  $\|g_{n+1}(y) - g_n(y)\|_2 < \varepsilon/2^{n+2}$ .

Für  $n = 0$  stimmt diese Behauptung, da

$$\|g_0(y)\|_2 = \|y\|_2 < \varepsilon/2 = (1 - 1/2) \cdot \varepsilon$$

und

$$\|g_1(y) - g_0(y)\|_2 = \|y - f(g_0(y))\|_2 = \|y - f(y)\|_2 \leq \|R\|_2 \leq \|y\|_2/2 < \varepsilon/4,$$

wobei  $R$  in der Vorbereitung (i) untersucht wurde. Für den Induktionsschritt, berechne

$$\begin{aligned} \|g_{n+1}(y)\|_2 &\leq \|g_n(y)\|_2 + \|g_{n+1}(y) - g_n(y)\|_2 \\ &\leq (1 - 1/2^{n+1}) \cdot \varepsilon + \varepsilon/2^{n+2} = (1 - 1/2^{n+2})\varepsilon. \end{aligned}$$

Ähnlich berechne

$$\begin{aligned} \|g_{n+2}(y) - g_{n+1}(y)\|_2 &= \|g_{n+1}(y) - f(g_{n+1}(y)) - g_n(y) + f(g_n(y))\|_2 \\ &= \|g_{n+1}(y) - g_n(y) - (f(g_{n+1}(y)) - f(g_n(y)))\|_2 \\ &\leq 1/2 \|g_{n+1}(y) - g_n(y)\|_2 \\ &< 1/2 \cdot \varepsilon \cdot 1/2^{n+2} = 1/2^{n+3}. \end{aligned}$$

Hier haben wir die Abschätzung aus der Vorbereitung (i) sowie die Induktionsannahme verwendet.



Wir folgern, dass  $(g_n(y))_{n \in \mathbb{N}}$  eine Cauchyfolge in  $B_0(\varepsilon)$  ist. Wir schätzen ab, dass

$$\begin{aligned}\|g(y)\|_2 &\leq \|g_0(y)\|_2 + \|g_1(y) - g_0(y)\|_2 + \|g_2(y) - g_1(y)\|_2 + \cdots \\ &< \varepsilon/2 + \varepsilon/4 + \varepsilon/8 + \cdots \\ &= \varepsilon,\end{aligned}$$

also liegt  $g(y)$  in  $B_0(\varepsilon)$ . Berechne nun mithilfe der Stetigkeit von  $f$ , dass

$$\begin{aligned}f(g(y)) &= f\left(\lim_{n \rightarrow \infty} g_n(y)\right) \\ &= \lim_{n \rightarrow \infty} f(g_n(y)).\end{aligned}$$

Aus  $g_{n+1}(y) = g_n(y) + y - f(g_n(y))$  folgt, dass

$$\begin{aligned}f(g(y)) &= \lim_{n \rightarrow \infty} y + g_n(y) - g_{n+1}(y) \\ &= y + g(y) - g(y) = y.\end{aligned}$$

□

Im letzten Schritt des Beweis vom Umkehrsatz zeigen wir, dass die oben konstruierte Umkehrfunktion  $g$  differenzierbar ist. Betrachte dazu die beiden Mengen  $W = B_0(\varepsilon/2)$  und  $V = f^{-1}(W) \cap B_0(\varepsilon) \subset U$ . Aus der lokalen Bijektivität folgt, dass die Einschränkung  $f|_V: V \rightarrow W$  bijektiv ist, mit Umkehrabbildung  $g: W \rightarrow V$ .

**Behauptung.** Die Abbildung  $g: W \rightarrow V$  ist Lipschitz-stetig mit Konstante 2.

*Beweis.* Für alle  $q, \tilde{q} \in B_0(\varepsilon)$  gilt  $1/2 \cdot \|\tilde{q} - q\|_2 \leq \|f(\tilde{q}) - f(q)\|_2$ . Seien nun  $\tilde{y}, y \in W$ . Setze  $\tilde{q} = g(\tilde{y})$  und  $q = g(y) \in V$ . Es folgt, dass

$$1/2 \cdot \|g(\tilde{y}) - g(y)\|_2 \leq \|f(g(\tilde{y})) - f(g(y))\|_2 = \|\tilde{y} - y\|_2,$$

was  $\|g(\tilde{y}) - g(y)\|_2 \leq 2 \cdot \|\tilde{y} - y\|_2$  impliziert.

□

Sei  $y \in W$  und  $k \in \mathbb{R}^n$  mit  $y + k \in W$ . Betrachte den Ansatz

$$g(y + k) = g(y) + (Df)_{g(y)}^{-1}(k) + (Rg)_y(k).$$

für eine Dreigliedertwicklung für  $g$  an der Stelle  $y$ . Zu zeigen ist, dass  $(Rg)_y(k)$  relativ klein in  $\|k\|_2$  ist, dass heisst, dass

$$\lim_{\|k\|_2 \rightarrow 0} \frac{\|(Rg)_y(k)\|_2}{\|k\|_2} = 0.$$

Nach Konstruktion von  $\varepsilon > 0$  gilt für  $g(y) \in V$ , dass

$$\|(Df)_g(y) - (Df)_0\|_{\text{op}} \leq 1/2,$$

wobei  $(Df)_0 = \text{Id}_{\mathbb{R}^n}$ . Es gilt also für alle  $v \in \mathbb{R}^n$ , dass  $\|(Df)_{g(y)}(v) - v\|_2 \leq 1/2 \cdot \|v\|_2$ . Wir schliessen, dass  $1/2 \cdot \|v\|_2 \leq \|(Df)_{g(y)}(v)\|_2$ . Insbesondere ist  $(Df)_{g(y)}$  invertierbar.

Sei nun  $w \in \mathbb{R}^n$ . Für  $v = (Df)_{g(y)}^{-1}(w)$  erhalten wir, dass

$$1/2 \cdot \|(Df)_{g(y)}^{-1}(w)\|_2 \leq \|(Df)_{g(y)}((Df)_{g(y)}^{-1}(w))\|_2 = \|w\|_2,$$

beziehungsweise  $\|(Df)_{g(y)}^{-1}(w)\|_2 \leq 2 \cdot \|w\|_2$ . Folglich gilt  $\|(Df)_{g(y)}^{-1}\|_{\text{op}} \leq 2$ . Wende nun auf unseren Ansatz die Abbildung  $f$  auf beide Seiten an. Die linke Seite wird dann zu  $f(g(y+k)) = y+k$ , und die rechte Seite zu  $f(g(y)+h)$ , wobei  $h = (Df)_{g(y)}^{-1}(k) + (Rg)_y(k)$ . Es gilt

$$\|h\| \leq \|g(y+k) - g(y)\|_2 \leq 2 \cdot \|k\|_2.$$

Die Dreigliedentwicklung von  $f$  bei  $g(y)$  liefert, dass

$$\begin{aligned} f(g(y) + h) &= f(g(y)) + (Df)_{g(y)}(h) + (Rf)_{g(y)}(h) \\ &= y + (Df)_{g(y)}((Df)_{g(y)}^{-1}(k) + (Rg)_y(k)) + (Rf)_{g(y)}(h) \\ &= y + k + (Df)_{g(y)}((Rg)_y(k)) + (Rf)_{g(y)}(h). \end{aligned}$$

Hier haben wir verwendet, dass  $(Df)_{g(y)}^{-1}$  linear ist. Gleichsetzen der beiden Seiten liefert, dass  $y + k = y + k + (Df)_{g(y)}((Rg)_y(k)) + (Rf)_{g(y)}(h)$ , also gilt

$$(Rg)_y(k) = -(Df)_{g(y)}^{-1}((Rf)_{g(y)}(h)).$$

Aus  $\|(Df)_{g(y)}^{-1}\|_{\text{op}} \leq 2$  folgt, dass  $\|(Rg)_y(k)\|_2 \leq 2 \cdot \|(Rf)_{g(y)}(h)\|_2$  gilt. Wir schliessen, dass

$$\begin{aligned} \lim_{\|k\|_2 \rightarrow 0} \frac{\|(Rg)_y(k)\|_2}{\|k\|_2} &\leq \lim_{\|k\|_2 \rightarrow 0} 2 \cdot \frac{\|(Rf)_{g(y)}(h)\|_2}{\|k\|_2} \\ &\leq \lim_{\|h\|_2 \rightarrow 0} 4 \cdot \frac{\|(Rf)_{g(y)}(h)\|_2}{\|h\|_2} = 0, \end{aligned}$$

da  $(Rf)_{g(y)}(h)$  relativ klein in  $\|h\|_2$  ist. Wir haben hier verwendet, dass  $\|h\|_2 \leq 2 \cdot \|k\|_2$  gilt, um  $\|k\|_2 \rightarrow 0$  durch  $\|h\|_2 \rightarrow 0$  zu ersetzen. Wir folgern, dass obiger Ansatz tatsächlich eine Dreigliedentwicklung von  $g$  an der Stelle  $y \in W$  ist. Somit ist  $g$  differenzierbar, und für alle  $y \in W$  gilt, dass  $(Dg)_y = (Df)_{g(y)}^{-1}$ .

Wir werden darauf verzichten, die Transformationen, welche wir zu Beginn auf  $f$  angewendet haben, wieder rückgängig zu machen (obwohl das nicht allzu schwierig wäre). Somit ist der Umkehrsatz vollständig bewiesen.

**Bemerkung.** Die stetige Differenzierbarkeit von  $f$  ist eine essentielle Hypothese im Umkehrsatz, bereits in einer Variable.

**Beispiel.** Betrachte die Funktion

$$f: \mathbb{R} \rightarrow \mathbb{R} \quad x \mapsto \begin{cases} 0 & x = 0, \\ x + 2x^2 \cdot \sin(1/x) & x \neq 0. \end{cases}$$

Die Funktion  $f$  ist differenzierbar, und es gilt

$$f'(x) = \begin{cases} 1 & x = 0, \\ 1 + 4x \cdot \sin(1/x) - 2 \cos(1/x) & x \neq 0. \end{cases}$$

Für alle  $x_n = 1/2\pi n$  gilt, dass  $f'(x_n) = 1 + 0 - 2 = -1$  gilt. Also gilt

$$\lim_{n \rightarrow \infty} f'(x_n) = -1 \neq 1 = f'(0) = f' \left( \lim_{n \rightarrow \infty} x_n \right).$$

Also ist  $f'$  an der Stelle  $x = 0$  nicht stetig, und ausserdem in keiner noch so kleinen offenen Umgebung von  $x = 0$  monoton wachsend. Insbesondere hat  $f$  keine lokale Umkehrfunktion.

## 4 Der Satz über implizite Funktionen

In [1] wird der Satz über implizite Funktionen in den Abschnitten 169 und 170 diskutiert. Der Beweis des Satzes empfiehlt sich nicht, dort nachzulesen. Der Grund dazu ist, dass dort (und auch in den meisten anderen Lehrbüchern) der Umkehrsatz aus dem Satz über implizite Funktionen hergeleitet wird. Wir machen das in dieser Vorlesung umgekehrt.

Sei  $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$  stetig differenzierbar mit  $m < n$ . Für alle  $p \in \mathbb{R}^n$  erhalten wir eine lineare Abbildung  $(Df)_p: \mathbb{R}^n \rightarrow \mathbb{R}^m$ . Die Dimensionsformel aus der linearen Algebra liefert

$$n = \dim \ker (Df)_p + \dim \operatorname{im} (Df)_p,$$

wobei  $\ker (Df)_p$  den Kern und  $\operatorname{im} (Df)_p$  das Bild von  $(Df)_p$  bezeichnet. Aus der Beobachtung, dass  $\dim \operatorname{im} (Df)_p \leq m < n$  folgt, dass  $\dim \ker (Df)_p \geq 1$ . Insbesondere kann  $(Df)_p$  kein Vektorraumisomorphismus sein. Wohl aber kann  $(Df)_p: \mathbb{R}^n \rightarrow \mathbb{R}^m$  surjektiv sein.

Im Spezialfall  $m = 1$ , das heisst  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  ist die Bedingung, dass  $(Df)_p: \mathbb{R}^n \rightarrow \mathbb{R}$  nicht die Nullabbildung ist, oder äquivalent, dass  $(\nabla f)_p$  nicht der Nullvektor ist. Noch konkreter ist  $(Df)_p: \mathbb{R}^n \rightarrow \mathbb{R}$  surjektiv, genau dann wenn eine partielle Ableitung  $\partial f / \partial x_k$  in  $p$  nicht null ist.

**Frage.** Wie verhält sich  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  in der Nähe eines Punkts  $p \in \mathbb{R}^n$  mit  $(\nabla f)_p \neq 0$ ?  
Genauer: Wie sehen die *Niveaumengen*

$$f^{-1}(w) = \{q \in \mathbb{R}^n \mid f(q) = w\} \subset \mathbb{R}^n$$

in der Nähe eines Punkts  $p \in \mathbb{R}^n$  mit  $(\nabla f)_p \neq 0$  aus?

**Beispiele.**

(1) Betrachte die Funktion

$$\begin{aligned} f: \mathbb{R}^2 &\rightarrow \mathbb{R} \\ (x, y) &\mapsto y. \end{aligned}$$

Für alle  $p = (x, y) \in \mathbb{R}^2$  gilt, dass  $\nabla f(p) = (0, 1) \neq 0$ . Für  $w \in \mathbb{R}$  ist

$$f^{-1}(w) = \{(x, y) \in \mathbb{R}^2 \mid y = w\}.$$

In anderen Worten sind alle horizontalen Geraden Niveaumengen.

(2) Betrachte die Funktion

$$\begin{aligned} f: \mathbb{R}^2 &\rightarrow \mathbb{R} \\ (x, y) &\mapsto x^2 + y^2. \end{aligned}$$

Für  $p = (x, y) \in \mathbb{R}^2$  gilt  $(\nabla f)_p = 2p$ . Also ist  $(\nabla f)_p \neq 0$  für alle  $p \neq 0$ . Die Niveaumengen  $f^{-1}(w)$  sind leer für negative  $w$ , der Nullpunkt für  $w = 0$ , und Kreise mit Radius  $\sqrt{w}$  für  $w > 0$ . Es fällt auf, dass die Niveaulinie Nullpunkt  $w = 0$  einer anderen Qualität ist, als die restlichen Niveaulinien. Es ist kein Zufall, dass genau in diesem Punkt der Gradient verschwindet.

(3) Betrachte die Funktion

$$\begin{aligned} f: \mathbb{R}^2 &\rightarrow \mathbb{R} \\ (x, y) &\mapsto y^2 - x^2 = (y - x)(y + x). \end{aligned}$$

Für alle  $p = (x, y) \in \mathbb{R}^2$  ist  $(\nabla f)_p = (-2x, 2y)$  null, genau dann wenn  $p$  auch null ist. Für alle Werte  $w \neq 0$  sind die Niveaulinien  $f^{-1}(w)$  Hyperbeln. Insbesondere sehen sie wie differenzierbare Kurven aus. Aber bei  $w = 0$  geschieht wieder etwas besonderes, nämlich ist die Niveaumenge  $f^{-1}(0)$  ein Kreuz, und nicht das Bild einer differenzierbaren Kurve.

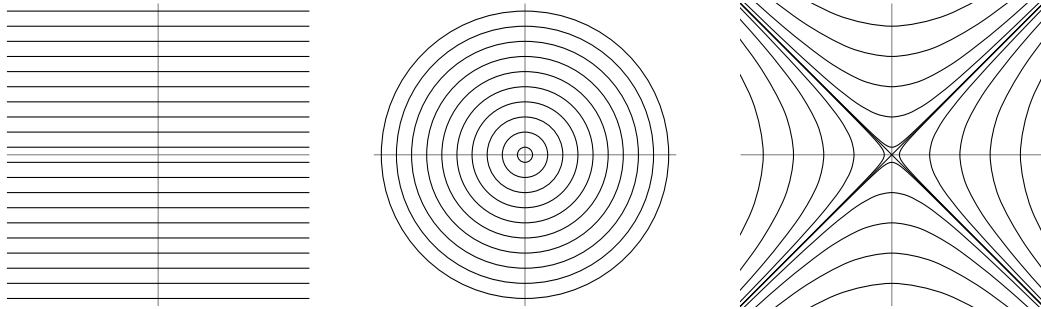


Abbildung II.3: Die Niveaulinien aus den drei Beispielen

**Theorem 3** (Satz über implizite Funktionen, Spezialfall). *Sei  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  stetig differenzierbar und  $p \in \mathbb{R}^n$  so, dass  $(Df)_p: \mathbb{R}^n \rightarrow \mathbb{R}$  surjektiv ist, das heisst, dass  $(\nabla f)_p \neq 0$  ist. Nehme zusätzlich an, dass  $p = 0$ ,  $f(0) = 0$ , und  $\partial f / \partial x_n(0) \neq 0$ . Dann existieren  $\varepsilon > 0$ ,  $\delta > 0$ , und eine differenzierbare Abbildung*

$$g: B_0^{\mathbb{R}^{n-1}}(\varepsilon) \times (-\delta, \delta) \rightarrow \mathbb{R},$$

so dass für alle  $q = (x_1, \dots, x_{n-1}) \in B_0^{\mathbb{R}^{n-1}}(\varepsilon)$  und alle  $w \in (-\delta, \delta)$  gilt, dass

$$f(x_1, \dots, x_{n-1}, g(x_1, \dots, x_{n-1}, w)) = w.$$

Hier ist  $B_0^{\mathbb{R}^{n-1}}(\varepsilon) = \{q \in \mathbb{R}^{n-1} \mid \|q\|_2 < \varepsilon\}$ .

**Geometrische Interpretation.** Bei festem  $q = (x_1, \dots, x_{n-1}) \in B_0^{\mathbb{R}^{n-1}}$  und festem  $w \in (-\delta, \delta)$  hat die Gleichung  $f(x_1, \dots, x_{n-1}, z) = w$  eine Lösung

$$z = g(x_1, \dots, x_{n-1}, w) \in \mathbb{R}.$$

Die Niveaumengen  $f^{-1}(w)$  lassen sich in einer Umgebung des Punkts  $p \in \mathbb{R}^n$  als Graphen von Funktionen betrachten. Die Einschränkung von  $g$  auf  $B_0^{\mathbb{R}^{n-1}} \times \{w\}$  liefert eine Abbildung

$$\begin{aligned} \bar{g}: B_0^{\mathbb{R}^{n-1}}(\varepsilon) &\rightarrow \mathbb{R} \\ q &\mapsto g(q, w). \end{aligned}$$

Die Niveaulinie  $f^{-1}(w)$  ist über  $B_0^{\mathbb{R}^{n-1}}(\varepsilon)$  realisiert als Graph der Funktion  $\bar{g}$ .

**Bemerkung.** Sei  $g: \mathbb{R}^{n-1} \rightarrow \mathbb{R}$  eine Funktion. Die Teilmenge

$$\Gamma(g) = \{(x_1, \dots, x_{n-1}, g(x_1, \dots, x_{n-1})) \mid x_i \in \mathbb{R}\} \subset \mathbb{R}^n$$

heißt *Funktionsgraph* von  $g$ . Obige Situation lässt sich also wie folgt formulieren. Für alle festen  $w \in (-\delta, \delta)$  lässt sich die Menge  $f^{-1}(w)$  lokal als Graph der Funktion

$$\begin{aligned} \bar{g}: B_0^{\mathbb{R}^{n-1}}(\varepsilon) &\rightarrow \mathbb{R} \\ (x_1, \dots, x_{n-1}) &\mapsto g(x_1, \dots, x_{n-1}, w) \end{aligned}$$

betrachten.

**Beispiele.** Wir greifen die Beispiele oben wieder auf.

- (1) Betrachte die Funktion  $f(x, y) = y$ . Gesucht ist eine Lösung der impliziten Gleichung  $f(x, z) = w$ . Wir bemerken, dass  $z = g(x, w) = w$  für alle  $x \in \mathbb{R}$  und alle  $w \in \mathbb{R}$  eine Lösung ist.
- (2) Betrachte die Funktion  $f(x, y) = x^2 + y^2$ . Gesucht ist eine Lösung der impliziten Gleichung  $f(x, z) = w$ , das heißt  $x^2 + z^2 = w$ . Wir bemerken, dass

$$z = g(x, w) = \sqrt{w - x^2}$$

eine Lösung ist, sofern  $w > 0$  und  $x^2 < w$  gilt. Insbesondere ist  $g(x, w)$  für alle  $w \in (1, 2)$  und alle  $x \in (-1, 1)$  definiert. Über dem Intervall  $(-1, 1)$  ist also  $f^{-1}(1)$  als Graph der Funktion  $g(x, 1) = \sqrt{1 - x^2}$  darstellbar.

Sei  $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$  stetig differenzierbar, so dass das Differential  $(Df)_p: \mathbb{R}^n \rightarrow \mathbb{R}^m$  im Punkt  $p \in \mathbb{R}^n$  surjektiv ist. Sei  $K = \ker (Df)_p$ . Dann lautet die Dimensionsformel

$$n = \dim \ker (Df)_p + \dim \operatorname{im} (Df)_p = \dim K + m.$$

Also gilt  $\dim K = n - m$ .

**Theorem 4** (Satz über implizite Funktionen, allgemeine Version). *Sei  $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$  stetig differenzierbar und  $p \in \mathbb{R}^n$  so, dass  $(Df)_p: \mathbb{R}^n \rightarrow \mathbb{R}^m$  surjektiv ist. Zusätzlich nehmen wir an, dass  $p = 0$  und  $f(p) = 0$  ist. Setze  $K = \ker (Df)_p \subset \mathbb{R}^n$ . Wähle in  $\mathbb{R}^n$  einen zu  $K$  komplementären Untervektorraum  $S \subset \mathbb{R}^n$ . Da heisst, dass  $\mathbb{R}^n = K + S$  und  $K \cap S = \{0\}$  gilt. In anderen Worten kann jeder Vektor  $v \in \mathbb{R}^n$  eindeutig als Summe  $v = k + s$  mit  $k \in K$  und  $s \in S$  geschrieben werden. Betrachte die Projektion*

$$\begin{aligned} \pi_K^S: \mathbb{R}^n &\rightarrow K \\ v = k + s &\mapsto k. \end{aligned}$$

Dann existieren  $\varepsilon > 0$ ,  $\delta > 0$ , und eine differenzierbare Abbildung

$$g: B_0^K(\varepsilon) \times B_0^{\mathbb{R}^m} \rightarrow \mathbb{R}^n,$$

so dass für alle  $k \in B_0^K(\varepsilon)$  und alle  $w \in B_0^{\mathbb{R}^m}(\delta)$  gilt, dass  $f(g(k, w)) = w$  und  $\pi_K^S(g(k, w)) = k$ .

**Geometrische Interpretation.** Bei fest vorgegebenem Wert  $w \in \mathbb{R}^m$  mit  $\|w\|_2 < \delta$  existiert über jedem Punkt  $k \in K$  mit  $\|k\|_2 < \varepsilon$  in Richtung  $S$  gesehen ein Punkt  $g(k, w) \in \mathbb{R}^n$ , welcher zur Menge  $f^{-1}(w)$  gehört.

*Beweis von Theorem 4.* Wir beweisen das Theorem via Umkehrsatz. Dazu widmen wir uns erst einer Promotion von  $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$  zu einer Abbildung  $F: \mathbb{R}^n \rightarrow \mathbb{R}^m \times K$ . Definiere dazu  $F(q) = (f(q), \pi_K^S(q))$ . Wir zeigen nun, dass  $(DF)_p: \mathbb{R}^n \rightarrow \mathbb{R}^m \times K$  ein Isomorphismus ist.

- Die Abbildung  $(DF)_p$  ist injektiv. Leite dazu eine Dreigliedentwicklung für  $F$  an der Stelle  $p \in \mathbb{R}^n$  her. Sei  $h \in \mathbb{R}^n$ . Berechne

$$\begin{aligned} F(p + h) &= (f(p + h), \pi_K^S(p + h)) \\ &= (f(p) + (Df)_p(h) + (Rf)_p(h), \pi_K^S(p) + \pi_K^S(h)). \end{aligned}$$

Wir schliessen, dass  $F$  an der Stelle  $p$  differenzierbar ist. Es gilt

$$(DF)_p(h) = ((Df)_p(h), \pi_K^S(h)).$$

Sei nun  $h \in \ker (DF)_p$ , das heisst  $(Df)_p(h) = 0$  und  $\pi_K^S(h) = 0$ . Es gilt  $(Df)_p(h) = 0$  genau dann, wenn  $h \in K$  gilt, und  $\pi_K^S(h) = 0$  genau dann, wenn  $h \in S$  gilt. Da  $K \cap S = \{0\}$  ist, folgt  $h = 0$ .

- $(DF)_p$  ist surjektiv. Aus der Dimensionsformel folgt, dass  $\dim K = n - m$  gilt. Also ist  $\dim(\mathbb{R}^n \times K) = n$ . Die Dimensionsformel für  $(DF)_p: \mathbb{R}^n \rightarrow \mathbb{R}^m \times K$  liefert, dass  $n = \dim \operatorname{im} (DF)_p$  gilt, also muss  $\operatorname{im} (DF)_p = \mathbb{R}^m \times K$  gelten.

Da  $F$  zudem stetig differenzierbar ist, können wir im Punkt  $p = 0 \in \mathbb{R}^n$  den Umkehrsatz anwenden. Dieser liefert offene Teilmengen  $V \subset \mathbb{R}^n$  und  $W \subset \mathbb{R}^m \times K$  mit  $0 \in V, 0 \in W$ , sowie eine differenzierbare Abbildung  $g: W \rightarrow V$ , so dass gilt, dass  $F \circ g = \operatorname{Id}_W$ . Dies heisst für alle  $(w, k) \in W \subset \mathbb{R}^n \times K$ , dass  $F(g(w, k)) = (w, k)$ , also  $f(g(w, k)) = w$  und  $\pi_K^S(g(w, k)) = k$ .

Der einzige Unterschied zur Formulierung von Theorem 4 ist, dass  $g$  auf  $W$  statt auf  $B_0^K(\varepsilon) \times B_0^{\mathbb{R}^m}(\delta)$  definiert ist. Um diesen Unterschied zu beheben, benötigen wir eine Norm auf  $\mathbb{R}^m \times K$  (auch schon nur um über Offenheit von  $W$  zu sprechen, was wir aufgrund der Normäquivalenz auf  $\mathbb{R}^n$  häufig unter den Teppich kehren). Wir arbeiten mit der Norm  $\|(w, k)\| = \sqrt{\|w\|_2^2 + \|k\|_2^2}$ . Diese Wahl liefert auf dem Produkt  $\mathbb{R}^m \times \mathbb{R}^n$  die euklidische Norm. Da  $0 \in W$  und  $W \subset \mathbb{R}^m \times K$  offen ist, existiert  $\tilde{\varepsilon} > 0$  mit  $B_0^{\mathbb{R}^m \times K}(\tilde{\varepsilon}) \subset W$ . Setze nun  $\varepsilon = \delta = \tilde{\varepsilon}/\sqrt{2} > 0$ . Aus  $\varepsilon^2 + \delta^2 = \tilde{\varepsilon}^2$  folgt, dass

$$B_0^{\mathbb{R}^m}(\delta) \times B_0^K(\varepsilon) \subset W$$

gilt. Die Einschränkung der obigen Funktion  $g: W \rightarrow V$  auf dieses Produkt von Bällen liefert die gewünschte implizite Funktion aus Theorem 4.  $\square$

**Definition** (Ausblick Differentialgeometrie). Sei  $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$  stetig differenzierbar und  $w \in \mathbb{R}^m$  so, dass für alle  $q \in f^{-1}(w)$  gilt, dass  $(Df)_q: \mathbb{R}^n \rightarrow \mathbb{R}^m$  surjektiv. Falls  $f^{-1}(w)$  nicht leer ist, dann heisst  $f^{-1}(w) \subset \mathbb{R}^n$  *Untermannigfaltigkeit* der Dimension  $n - m$ . Im Spezialfall  $f: \mathbb{R}^3 \rightarrow \mathbb{R}$  heisst eine solche Untermannigfaltigkeit *Fläche*.

**Beispiel.** Betrachte die Funktion

$$\begin{aligned} f: \mathbb{R}^3 &\rightarrow \mathbb{R} \\ (x, y, z) &\mapsto x^2 + y^2 + z^2. \end{aligned}$$

Dann ist  $(\nabla f)_{(x,y,z)} = 0$  genau dann, wenn  $(x, y, z) = 0$ . Für alle  $w > 0$  ist  $f^{-1}(w)$  eine Kugel mit Radius  $\sqrt{w}$ . Kugeln sind dementsprechend nach obiger Definition Flächen.

*Beweis von Theorem 3.* Wir schreiben  $\tilde{g}$  für die Funktion aus Theorem 3 und zeigen deren Existenz indem wir sie mit Hilfe der Funktion  $g$  aus Theorem 4 konstruieren. Aus  $(Df)_0(e_n) = \partial f / \partial x_n(0) \neq 0$  folgt, dass  $e_n$  nicht in  $K = \ker (Df)_p$  liegt. Wähle  $S = \operatorname{span}\{e_n\} \subset \mathbb{R}^n$ . Es gilt  $\mathbb{R}^n = K \oplus S$ , da  $\dim K = n - 1$  und  $K \cap S = \{0\}$ .

Sei  $g: B_0^K(\varepsilon) \times (-\delta, \delta) \rightarrow \mathbb{R}^n$  die Abbildung aus Theorem 4. Sei weiterhin  $\tilde{\varepsilon} > 0$  so gewählt, dass  $B_0^{\mathbb{R}^{n-1}}(\tilde{\varepsilon}) \subset \pi_{\mathbb{R}^{n-1}}^S(B_0^K(\varepsilon))$ . Sei  $A = (x_1, \dots, x_{n-1}, 0) \in B_0^{\mathbb{R}^{n-1}}(\tilde{\varepsilon}) \times \mathbb{R}$ . Dann existiert ein eindeutiges  $s \in \mathbb{R}$ , so dass  $B = (x_1, \dots, x_{n-1}, s) \in K$ , da  $\mathbb{R}^n = K \oplus S$ . Setze  $C = g(B, w) \in \mathbb{R}^n$ . Dann gilt  $g(B, w) \in f^{-1}(w)$ . Definiere schliesslich

$$\tilde{g}(A, w) = \langle g(B, w), e_n \rangle,$$

die letzte Koordinate von  $C = g(B, w)$ . Der Punkt  $C$  liegt in  $e_n$ -Richtung über dem Punkt  $A = (x_1, \dots, x_{n-1}, 0) \in \mathbb{R}^n$ . Per Definition gilt also

$$C = (x_1, \dots, x_{n-1}, \tilde{g}(x_1, \dots, x_{n-1}, w)) \in f^{-1}(w),$$

das heisst  $f(x_1, \dots, x_{n-1}, \tilde{g}(x_1, \dots, x_{n-1}, w)) = w$ . □

## 5 Die Hessische Form

In [1] ist der Inhalt dieses Abschnitts im Abschnitt 173 zu finden. Wir studieren das Verhalten differenzierbarer Funktionen  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  in der Nähe von *kritischen Punkten*, das heisst Punkten  $p \in \mathbb{R}^n$  mit  $(Df)_p = 0$ , das heisst  $(\nabla f)_p = 0$ .

Im Spezialfall  $n = 1$  gilt folgendes. Sei  $f: \mathbb{R} \rightarrow \mathbb{R}$  zweimal differenzierbar und  $x \in \mathbb{R}$  mit  $f'(x) = 0$  und  $f''(x) > 0$  (respektive  $f''(x) < 0$ ). Dann hat  $f$  an der Stelle  $x$  ein lokales Minimum (respektive ein lokales Maximum).

Im Fall  $n \geq 2$  brauchen wir einen Ersatz für die Bedingung  $f''(x) > 0$ , beziehungsweise benötigen wir ein Analogon zu  $f''$ .

**Definition.** Eine *quadratische Form* auf  $\mathbb{R}^n$  ist eine Abbildung  $q: \mathbb{R}^n \rightarrow \mathbb{R}$  mit folgenden Eigenschaften:

- (i) für alle  $v \in \mathbb{R}^n$  und alle  $\lambda \in \mathbb{R}$  gilt  $q(tv) = t^2 \cdot q(v)$ ,
- (ii) die Abbildung  $B: \mathbb{R}^n \rightarrow \mathbb{R}^n \rightarrow \mathbb{R}$  definiert durch  $B(v, w) = q(v + w) - q(v) - q(w)$  ist bilinear.

**Beispiel.** Sei  $M \in \mathbb{R}^{n \times n}$  eine beliebige symmetrische Matrix. Dann ist  $q(v) = v^T A v$  eine quadratische Form.

Sei  $f: \mathbb{R}^n \rightarrow \mathbb{R}$ , das heisst  $f$  ist differenzierbar, und die Abbildung

$$\begin{aligned} \nabla f: \mathbb{R}^n &\rightarrow \mathbb{R}^n \\ q &\mapsto (\nabla f)_q \end{aligned}$$

ist differenzierbar. Die Komponentenfunktionen von  $\nabla f$  sind die partiellen Ableitungen  $\partial f / \partial x_1, \dots, \partial f / \partial x_n$ . Wir erinnern uns, dass differenzierbare Abbildungen  $g: \mathbb{R}^n \rightarrow \mathbb{R}^n$  mit Komponentenfunktionen  $g_1, \dots, g_n: \mathbb{R}^n \rightarrow \mathbb{R}$  gilt, dass das Differential  $(Dg)_p: \mathbb{R}^n \rightarrow \mathbb{R}^n$  in einem Punkte  $p \in \mathbb{R}^n$  durch die Jacobimatrix

$$(Jg)_p = \begin{pmatrix} \partial g_1 / \partial x_1(p) & \cdots & \partial g_1 / \partial x_n(p) \\ \vdots & \ddots & \vdots \\ \partial g_n / \partial x_1(p) & \cdots & \partial g_n / \partial x_n(p) \end{pmatrix}$$

gegeben ist. Für  $g = \nabla f$  erhalten wir

$$(J\nabla f)_p = \begin{pmatrix} \partial^2 f / \partial x_1^2(p) & \cdots & \partial^2 f / \partial x_n \partial x_1(p) \\ \vdots & \ddots & \vdots \\ \partial^2 f / \partial x_1 \partial x_n(p) & \cdots & \partial^2 f / \partial x_n^2(p) \end{pmatrix}.$$

Die Matrix  $H = (J\nabla f)_p$  heisst *Hessische Matrix* von  $f$  an der Stelle  $p$ .



**Definition.** Die *Hessische Form* von  $f$  an der Stelle  $p$  ist die quadratische Form

$$(Hf)_p: \mathbb{R}^n \rightarrow \mathbb{R}$$

$$v \mapsto v^T \cdot (J\nabla f)_p \cdot v.$$

**Bemerkung.** Sei  $A \in \mathbb{R}^{n \times n}$  eine quadratische Matrix. Dann ist die Abbildung

$$Q: \mathbb{R}^n \rightarrow \mathbb{R}$$

$$v \mapsto v^T A v$$

eine quadratische Form. Tatsächlich gilt  $Q(tv) = t^2 Q(v)$  und

$$(v+w)^T A (v+w) - v^T A v - w^T A w = v^T A w + w^T A v.$$

**Beispiele.**

- (1) Sei  $L: \mathbb{R}^n \rightarrow \mathbb{R}$  linear. Dann gilt für alle  $p \in \mathbb{R}^n$ , dass  $(DL)_p = L$  ist. Insbesondere ist die Abbildung  $\nabla L: \mathbb{R}^n \rightarrow \mathbb{R}^n$  konstant. Somit ist  $(D\nabla L)_p = 0$  und somit auch  $(J\nabla L)_p = 0$  und  $(HL)_p = 0$ .
- (2) Sei  $A = (a_{ij}) \in \mathbb{R}^{n \times n}$  symmetrisch, das heisst  $A^T = A$ , oder  $a_{ij} = a_{ji}$ . Definiere

$$Q: \mathbb{R}^n \rightarrow \mathbb{R}$$

$$v \mapsto v^T A v.$$

In Koordinaten gilt dann für  $v = x_1 e_1 + \dots + x_n e_n$ , dass

$$Q(x_1, \dots, x_n) = \sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i x_j.$$

Berechne für feste Indizes  $i, j \in \{1, \dots, n\}$ , dass  $\partial^2 Q / \partial x_j \partial x_i = 2a_{ij}$ . Wir folgern, dass  $(J\nabla Q)_p = 2A$  für alle  $p \in \mathbb{R}^n$ . Somit gilt  $Q(v) = 1/2 \cdot (HQ)_p(v)$ . Alle quadratischen Polynome (ohne konstante und lineare Terme) auf  $\mathbb{R}^n$  lassen sich als  $v^T A v$  mit  $A^T = A$  beschreiben.

Im Fall  $n = 2$  schreibe

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

und erhalte  $Q(x, y) = ax^2 + 2bxy + cy^2$ .

**Proposition 3** (Viergliedentwicklung). Sei  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  zweimal differenzierbar und  $p \in \mathbb{R}^n$  ein beliebiger Punkt. Dann hat  $f$  an der Stelle  $p$  eine Viergliedentwicklung der Form

$$f(p+h) = f(p) + (Df)_p(h) + 1/2 \cdot (Hf)_p(v) + (Rf)_p(h),$$

wobei  $(Rf)_p(h)$  relativ klein in  $\|h\|_2^2$  ist, das heisst

$$\lim_{h \rightarrow 0} \frac{\|(Rf)_p(h)\|_2}{\|h\|_2^2} = 0.$$

*Beweis.* Seien  $p, h \in \mathbb{R}^n$ . Zur Auswertung der Differenz  $f(p+h) - f(p)$  betrachten wir den Weg

$$\begin{aligned}\gamma: [0, 1] &\rightarrow \mathbb{R}^n \\ t &\mapsto p + th.\end{aligned}$$

Es gilt  $\gamma(0) = p$  und  $\gamma(1) = p + h$ . Für alle  $t \in [0, 1]$  gilt weiterhin  $\dot{\gamma}(t) = h$ . Berechne

$$\begin{aligned}f(p+h) - f(p) &= f(\gamma(1)) - f(\gamma(0)) \\ &= \int_0^1 \frac{d}{dt} f(\gamma(t)) dt \\ &= \int_0^1 (Df)_{\gamma(t)}(\dot{\gamma}(t)) dt \\ &= \int_0^1 (Df)_{p+th}(h) dt \\ &= \int_0^1 \langle \nabla f(p+th), h \rangle dt.\end{aligned}$$

Nach der Dreigliedrentwicklung von  $\nabla f$  an der Stelle  $p$  gilt

$$\nabla f(p+k) = \nabla f(p) + (J\nabla f)_p \cdot k + (R\nabla f)_p(k).$$

Sei nun  $\varepsilon > 0$  vorgegeben. Dann existiert  $\delta > 0$ , so dass für alle  $k \in \mathbb{R}^n$  mit  $\|k\|_2 \leq \delta$  gilt, dass  $\|(R\nabla f)_p(k)\|_2 \leq \varepsilon \cdot \|k\|_2$ . Für alle  $h \in \mathbb{R}^n$  mit  $\|h\|_2 \leq \delta$  und alle  $t \in [0, 1]$  gilt also, dass  $\|(R\nabla f)_p(th)\| \leq \varepsilon \cdot \|h\|_2$ . Wir folgern für  $k = th$ , dass

$$\begin{aligned}f(p+h) - f(p) &= \int_0^1 \langle (\nabla f)_{p+th}, h \rangle dt \\ &= \int_0^1 \langle (\nabla f)_p, h \rangle dt + \int_0^1 \langle (D\nabla f)_p(th), h \rangle dt + \int_0^1 \langle (R\nabla f)_p(th), h \rangle dt.\end{aligned}$$

Berechne, dass

$$\int_0^1 \langle (\nabla f)_p, h \rangle dt = \int_0^1 (Df)_p(h) dt = (Df)_p(h),$$

und auch

$$\int_0^1 \langle (D\nabla f)_p(th), h \rangle dt = \int_0^1 t \cdot h^T (J\nabla f)_p h dt = 1/2 \cdot (Hf)_p(h).$$

Zuletzt schätzen wir ab, dass

$$\begin{aligned}\left\| \int_0^1 \langle (R\nabla f)_p(th), h \rangle dt \right\|_2 &\leq \int_0^1 |\langle (R\nabla f)_p(th), h \rangle| dt \\ &\leq \int_0^1 \|(R\nabla f)_p(th)\|_2 \cdot \|h\|_2 dt \\ &\leq \int_0^1 \varepsilon \cdot \|h\|_2^2 dt \\ &\leq \varepsilon \cdot \|h\|_2.\end{aligned}$$

Somit erfüllt

$$(Rf)_p(h) = \int_0^1 \langle (R\nabla)_p(th), h \rangle dt$$

die gewünschte Bedingung.  $\square$

Sei  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  zweimal differenzierbar. Dann hat  $f$  an der Stelle  $p \in \mathbb{R}^n$  ein *lokales Minimum* (beziehungsweise ein lokales Maximum), falls  $\varepsilon > 0$  existiert, so dass für alle  $q \in B_p(\varepsilon)$  gilt, dass  $f(q) \geq f(p)$  (beziehungsweise  $f(q) \leq f(p)$ ).

**Definition.** Ein Punkt  $p \in \mathbb{R}^n$  heisst *kritischer Punkt* einer differenzierbaren Funktion  $f: \mathbb{R}^n \rightarrow \mathbb{R}$ , falls  $(Df)_p$  nicht surjektiv ist, das heisst  $\nabla f(p) = 0$  gilt.

**Bemerkung.** Wir erinnern uns, dass lokale Minima (beziehungsweise Maxima) kritische Punkte von  $f$  sind. Die Umkehrung gilt im allgemeinen nicht, wie das Beispiel  $f(x) = x^3$  in  $\mathbb{R}$  zeigt.

In der Nähe von kritischen Punkten wird das Verhalten von  $f$  durch die Hessische Form beschrieben, denn dort gilt

$$f(p+h) = f(p) + 1/2 \cdot (Hf)_p(h) + (Rf)_p(h).$$

**Proposition 4** (Satz von Schwarz). *Sei  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  zweimal stetig differenzierbar. Dann gilt für alle  $p \in \mathbb{R}^n$  und alle  $i, j \in \{1, \dots, n\}$ , dass*

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}.$$

**Bemerkung.**

(1) Für Polynome  $f(x, y)$  ist diese Gleichheit offensichtlich, da

$$\partial^2 / \partial x \partial y (x^a y^b) = ab x^{a-1} y^{b-1} = \partial^2 / \partial y \partial x$$

gilt.

(2) Die Bedingung, dass  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  zweimal (nicht unbedingt stetig) differenzierbar ist, reicht, um diese Gleichheit zu zeigen, jedoch nicht die Existenz der zweiten partiellen Ableitungen. Betrachte zum Beispiel die Funktion

$$f: \mathbb{R}^2 \rightarrow \mathbb{R} \\ (x, y) \mapsto xy \frac{x^2 - y^2}{x^2 + y^2}.$$

Dann gilt  $\partial^2 f / \partial x \partial y (0, 0) = 1 \neq -1 = \partial^2 f / \partial y \partial x (0, 0)$ .

(3) In der Sprache der Differentialformen lautet die obige Proposition  $ddf = 0$ . Siehe dazu das Buch *Mathematical Methods in Classical Mechanics* von Arnold, sowie das Buch *Formes Différentielles* von Cartan.

Wir werden Proposition 4 später zeigen. Aus der Linearen Algebra folgt, dass die symmetrische Matrix  $(J\nabla f)_p$  über  $\mathbb{R}$  diagonalisierbar ist. Konkreter existieren sogar  $\lambda_1, \dots, \lambda_n \in \mathbb{R}$  und eine orthogonale Matrix  $P \in \mathbb{R}^{n \times n}$  (das heisst  $P = P^{-1}$ ), so dass  $P^{-1}(J\nabla f)_p P = \text{diag}(\lambda_1, \dots, \lambda_n)$  gilt. Siehe dazu den Satz von Sylvester. Bezüglich einer geeigneten orthogonalen Basis  $\{f_1, \dots, f_n\}$  von  $\mathbb{R}^n$  gilt also folgendes. Für alle Vektoren  $v = y_1 f_1 + \dots + y_n f_n$  gilt  $(Hf)_p(v) = \lambda_1 y_1^2 + \dots + \lambda_n y_n^2$ .

Wir machen nun die Zusatzannahme, dass  $(Hf)_p$  nicht-singulär ist, das heisst alle  $\lambda_k$  sind ungleich null. Dann hat  $f$  an der Stelle  $p$  ein lokales Minimum (beziehungsweise ein lokales Maximum) genau dann, wenn alle  $\lambda_k$  positiv (beziehungsweise negativ) sind.

Dies kann man im Falle  $n = 2$  leicht testen. Schreibe

$$(J\nabla f)_p = \begin{pmatrix} a & b \\ b & c \end{pmatrix}.$$

Dann hat  $f$  bei  $p$  ein

- lokales Minimum, falls  $a > 0$  und  $ac - b^2 > 0$ ,
- lokales Maximum falls  $a < 0$  und  $ac - b^2 > 0$ ,
- Sattelpunkt, falls  $ac - b^2 < 0$ .

Dies erhält man durch die Forderung, dass bei Minima die beiden Eigenwerte positiv sind, bei Maxima die Eigenwerte negativ sind, und bei Sattelpunkten die Eigenwerte unterschiedliches Vorzeichen haben.

*Beweis von Proposition 4.* Wir zeigen den Satz von Schwarz für zweimal stetig differenzierbare Funktionen  $f: \mathbb{R}^2 \rightarrow \mathbb{R}$ . Dieselbe Idee zeigt, dass die Aussage auch für zweimal differenzierbare  $f$  gilt. Diese Version des Beweises ist jedoch schwieriger auszuführen. Konkreter werden wir zeigen, dass

$$\frac{\partial^2 f}{\partial y \partial x}(0,0) = \lim_{x \rightarrow 0} \lim_{y \rightarrow 0} \frac{1}{xy} f(x,y) - f(x,0) - f(0,y) + f(0,0)$$

gilt. Aus Symmetriegründen (vertausche die Rollen von  $x$  und  $y$ ) folgt dann, dass  $\partial^2 f / \partial y \partial x(0,0) = \partial^2 f / \partial x \partial y(0,0)$  gilt.

Wir zeigen zuerst, dass wir uns auf den Fall  $n = 2$  und  $p = 0$  beschränken können. Sei  $\bar{f}: \mathbb{R}^n \rightarrow \mathbb{R}$  zweimal stetig differenzierbar und sei  $p \in \mathbb{R}^n$  fest. Betrachte statt  $f$  die Abbildung  $\bar{f}: \mathbb{R}^n \rightarrow \mathbb{R}$ , definiert durch  $\bar{f}(q) = f(p+q)$ . Dann gilt  $\bar{f}(0) = f(p)$ . Weiterhin ist  $\partial^2 \bar{f} / \partial x_j \partial x_i(0) = \partial^2 f / \partial x_j \partial x_i(p)$ . Deshalb reicht es, den Punkte  $p = 0$  zu betrachten. Um die Notation weiter zu vereinfachen, nenn wir die Koordinaten  $x_i, x_j$  neu  $x, y$  und beschränken uns auf Abbildungen  $f: \mathbb{R}^2 \rightarrow \mathbb{R}$ .

Seien  $x, y \in \mathbb{R}$  fix. Definiere  $h_y: \mathbb{R} \rightarrow \mathbb{R}$  durch  $h_y(x) = f(x, y) - f(x, 0)$ . Bemerke, dass  $f(x, y) - f(x, 0) - f(0, y) + f(0, 0) = h_y(x) - h_y(0)$ . Nach dem Mittelwertsatz existiert ein  $a \in (0, x)$  mit  $x \cdot \partial h_y / \partial x(a) = h_y(x) - h_y(0)$ . Es folgt, dass

$$f(x, y) - f(x, 0) - f(0, y) + f(0, 0) = x \cdot (\partial f / \partial x(a, y) - \partial f / \partial x(a, 0)) = x \cdot \partial h_y / \partial x(a).$$

Nach dem Mittelwertsatz angewendet auf die Funktion  $y \mapsto \partial f / \partial x(a, y)$  existiert ein  $b \in (0, y)$  mit  $\partial f / \partial x(a, y) - \partial f / \partial x(a, 0) = y \cdot \partial^2 f / \partial y \partial x(a, b)$ . Folglich gilt

$$f(x, y) - f(x, 0) - f(0, y) + f(0, 0) = xy \cdot \partial^2 f / \partial y \partial x(a, b).$$

Im Grenzübergang  $x \rightarrow 0$  und  $y \rightarrow 0$  gilt  $a \rightarrow 0$  und  $b \rightarrow 0$ , also erhalten wir das gewünschte Resultat unter der Verwendung der Stetigkeit von  $\partial^2 f / \partial y \partial x$ .  $\square$

## Kapitel III

# Gradientenfelder und Differentialformen

# Literaturverzeichnis

- [1] H. Heuser. *Lehrbuch der Analysis*. Mathematische Leitfäden. Vieweg+Teubner Verlag, 2009.
- [2] E. Hairer and G. Wanner. *Analysis by Its History*. Undergraduate Texts in Mathematics / Readings in Mathematics. Springer New York, 2000.