

Vorlesungsnotizen

Mathematik 4

(für den Lehramtsstudiengang Sekundarstufe)

Sommersemester 2022

Thomas Schmidt

Stand: 31. Januar 2023

Mathematik 4: Inhaltsverzeichnis

Inhaltsverzeichnis	1
1 Metrische und topologische Konzepte, allgemeine Stetigkeitsbegriffe	3
1.1 Metrische und normierte Räume	4
1.2 Topologische Grundbegriffe	14
1.3 Kompaktheit und Gleichmäßigkeit	22
2 Differentialrechnung mehrerer Veränderlicher	29
2.1 Ableitungsbegriffe	29
2.2 Extremstellenbestimmung	42
2.3 Der Umkehrsatz und der Satz über implizite Funktionen	48
2.4 Ableitungen zweiter und höherer Ordnung	51
3 Grundlagen der Volumenmessung und der mehrdimensionalen Integration	
3.1 Der Jordan-Inhalt (nicht digital verfügbar)	
3.2 Ausblick zu Riemann-Integral und Jordan-Inhalt (nicht digital verfügbar)	
Literaturverzeichnis	59

Warnung: Diese Notizen umfassen nicht den vollständigen Vorlesungsstoff!
Falls Sie Fehler (jeglicher Art) finden oder sonstige Hinweise haben, bitte ich Sie, mir dies entweder persönlich oder unter thomas.schmidt.math@uni-hamburg.de mitzuteilen.

Kapitel 1

Metrische und topologische Konzepte, allgemeine Stetigkeitsbegriffe

In der Mathematik 4 werden wir **Analysis nicht nur in/auf \mathbb{R} oder \mathbb{C} sondern allgemein in/auf \mathbb{R}^n mit $n \in \mathbb{N}$ betreiben**. Teils bietet es sich dabei an, statt in \mathbb{R}^n **direkt in allgemeineren Mengen \mathcal{X} von Punkten** mit gewisser Zusatzstruktur zu arbeiten. Genauere Definitionen werden wir demnächst geben. Wir reißen die relevanten **Konzepte von Mengen \mathcal{X} mit zugehörigen Strukturen** vorab aber schon einmal kurz und informell an, beginnen dabei mit dem allgemeinsten Konzept und gehen dann schrittweise zu spezielleren über:

- In **topologischen Räumen \mathcal{X}** mit der Struktur einer Topologie lassen sich nur *qualitative Aussagen* zur Lage von Punkten und Mengen zueinander tätigen. Zum Beispiel macht die Frage nach der **Konvergenz** einer Folge $(x_n)_{n \in \mathbb{N}}$ gegen x in \mathcal{X} als Ja-Nein-Aussage Sinn. Wir werden in dieser Vorlesung einige topologische Konzepte wie etwa den **Rand von Mengen** in \mathcal{X} benötigen. Es wird aber ausreichen, diese Begriffe im etwas spezielleren metrischen Fall einzuführen, während wir auf die Definition und die allgemeine Theorie topologischer Räume verzichten können und werden.
- In **metrischen Räumen \mathcal{X}** mit der Struktur einer Metrik ist zusätzlich zu topologischen Begriffen der **Abstand von zwei Punkten** (oder Mengen) *quantitativ* als nicht-negative reelle Zahl erklärt. Im Allgemeinen besitzt \mathcal{X} hier aber keine Vektorraumstruktur, und man kann nicht durch Anlegen eines Vektors von einem Punkt zu einem anderen gelangen. Schlimmer noch darf man in allgemeinen metrischen Räumen überhaupt nicht an eine (kürzeste) Verbindung von zwei Punkten denken oder sich gar den Abstand als die Länge einer solchen vorstellen.
- **Normierte Räume \mathcal{X}** mit der Struktur einer Norm wurden in der Mathematik 3 bereits kurz erwähnt und werden in Folge genauer besprochen. Sie weisen zusätzlich zur topologischen und metrischen Struktur tatsächlich eine **Vektorraumstruktur** auf. Vektoraddition und Skalarmultiplikation sind daher allgemein möglich. Vorstellungen wie die vom Ortsvektor eines Punktes und vom Anlegen eines Vektors an einen Punkt sind sinnvoll, und der **Abstand zweier Punkte** kann allgemein als **Länge des Verbindungsvektors** interpretiert werden. Elementargeometrische Sätze gelten in normierten Räumen aber nicht

unbedingt, zum Beispiel kann es in einem normierten Raum ein nicht-entartetes Dreieck mit Seiten der Längen 1, 2 und 3 geben.

- **Skalarprodukträume** \mathcal{X} mit der Struktur eines Skalarprodukts wurden in der Mathematik 3 bereits weitgehend behandelt. In ihnen kann man über das bereits Erwähnte hinaus **Orthogonalität** zweier Vektoren und den **Winkel** zwischen solchen erklären, und es gelten sehr viele aus der Elementargeometrie bekannte Sachverhalte wie etwa des Satz des Pythagoras und der Kosinussatz.

Tatsächlich erweisen sich **alle vier gerade aufgeführten Konzepte** als **universell für weite Teile der Mathematik**. Präzise Definitionen und Einzelheiten zu den Eigenschaften folgen nun nach und nach.

1.1 Metrische und normierte Räume

Wir geben nun die Definition eines metrischen Raums und wiederholen im direkten Vergleich auch die eines normierten Raums, wobei wir **im Folgenden immer** $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$ verstehen.

Definition (metrische Räume, Metrik). Sei \mathcal{X} eine Menge. Eine Abbildung $d: \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ heißt eine **Metrik** auf \mathcal{X} und das Paar (\mathcal{X}, d) ein **metrischer Raum**, wenn folgende drei Axiome erfüllt sind:

- (M1) **(Strikte) Positivität:** $d(x, y) \geq 0$ für alle $x, y \in \mathcal{X}$ mit Gleichheit genau im Fall $x = y$,
- (M2) **Symmetrie:** $d(x, y) = d(y, x)$ für alle $x, y \in \mathcal{X}$,
- (M3) **Dreiecksungleichung:** $d(x, z) \leq d(x, y) + d(y, z)$ für alle $x, y, z \in \mathcal{X}$.

Definition (normierte Räume, Norm). Sei \mathcal{X} ein \mathbb{K} -Vektorraum. Eine Abbildung $\|\cdot\|: \mathcal{X} \rightarrow \mathbb{R}$ heißt eine **Norm** auf \mathcal{X} und das Paar $(\mathcal{X}, \|\cdot\|)$ ein **normierter Raum** über \mathbb{K} , wenn folgende drei Axiome erfüllt sind:

- (N1) **(Strikte) Positivität:** $\|x\| \geq 0$ für alle $x \in \mathcal{X}$ mit Gleichheit genau im Fall $x = 0_{\mathcal{X}}$,
- (N2) **(Absolut-)Homogenität:** $\|sx\| = |s| \|x\|$ für alle $x \in \mathcal{X}$ und $s \in \mathbb{K}$,
- (N3) **Dreiecksungleichung:** $\|x+y\| \leq \|x\| + \|y\|$ für alle $x, y \in \mathcal{X}$.

Erläuterungen und Beispiele zu diesen Definitionen und ihrer Bedeutung folgen.

Bemerkungen (zu metrischen und normierten Räumen).

- (0) Gibt es eine Standard-Wahl der Metrik oder Norm oder ist diese aus dem Kontext klar, so bezeichnet man statt (\mathcal{X}, d) oder $(\mathcal{X}, \|\cdot\|)$ auch \mathcal{X} allein als metrischen oder normierten Raum. Bei Bedarf notiert man die zugehörige Metrik oder Norm als $d_{\mathcal{X}}$ oder $\|\cdot\|_{\mathcal{X}}$.
- (1) **Interpretation:** Bei einer **Metrik** d stellt man sich $d(x, y)$ als **Abstand zweier Punkte** x und y vor, bei einer **Norm** $\|\cdot\|$ interpretiert man $\|x\|$ als **Länge des Vektors** x .

- (2) Jede *Teilmenge* A eines metrischen Raums \mathcal{X} ist mit der Einschränkung $d_{\mathcal{X}}|_{A \times A}$ der Metrik selbst ein metrischer Raum. Analog ist jeder \mathbb{K} -*Untervektorraum* U eines normierten Raums \mathcal{X} über \mathbb{K} mit der Einschränkung $\|\cdot\|_{\mathcal{X}}|_U$ der Norm selbst ein normierter Raum über \mathbb{K} . Durch die **Möglichkeit des Übergangs zu beliebigen Teilmengen** ergibt sich bei **metrischen Räumen** hieraus **deutlich mehr Flexibilität** als bei normierten Räumen.
- (3) **Jede Norm $\|\cdot\|$ auf \mathcal{X} induziert eine Metrik d auf \mathcal{X} und auf jeder beliebigen Teilmenge von \mathcal{X} durch Festlegung des **Abstands zweier Punkte als Länge des Verbindungsvektors**, das heißt durch**

$$d(x, y) := \|y - x\| \quad \text{für } x, y \in \mathcal{X}.$$

Der kurze Nachweis der Axiome (M1), (M2), (M3) bei dieser Festlegung wird in der Vorlesung ausgeführt. In diesem Sinn **ist jeder normierte Raum und jede Teilmenge eines normierten Raums ein metrischer Raum**.

- (4) Umgekehrt wird *nicht* jede Metrik auf einem \mathbb{K} -Vektorraum \mathcal{X} von einer Norm induziert. Tatsächlich sind die Metriken d auf einem \mathbb{K} -Vektorraum \mathcal{X} , die von einer Norm herrühren, genau die, die mit der linearen Struktur von \mathcal{X} im Sinn von Homogenität

$$d(sx, sy) = |s|d(x, y) \quad \text{für alle } x, y \in \mathcal{X} \text{ und } s \in \mathbb{K}$$

und Translationsinvarianz

$$d(x+z, y+z) = d(x, y) \quad \text{für alle } x, y, z \in \mathcal{X}$$

verträglich sind. Man erhält in diesem Fall die induzierende Norm $\|\cdot\|$ aus $\|x\| = d(0_{\mathcal{X}}, x)$ für $x \in \mathcal{X}$. Der Beweis dieser Behauptungen benutzt nur die Definitionen und wird in den Lernwerkstätten behandelt.

Beispiele (von normierten Räumen).

- (0) Die wohl grundlegendste Norm ist die Betragsfunktion, mit der \mathbb{R} **und \mathbb{C} normierte Räume** werden.
- (1) Übliche **Normen auf \mathbb{R}^n und \mathbb{C}^n** mit $n \in \mathbb{N}$ sind die **p -Normen $|\cdot|_p$** , die durch

$$|x|_p := \left(\sum_{i=1}^n |x_i|^p \right)^{\frac{1}{p}} \quad \text{für } x = (x_1, x_2, \dots, x_n) \in \mathbb{K}^n \text{ und } p \in [1, \infty)$$

sowie

$$|x|_{\infty} := \max\{|x_1|, |x_2|, \dots, |x_n|\} \quad \text{für } x = (x_1, x_2, \dots, x_n) \in \mathbb{K}^n$$

definiert werden. **Falls nichts anderes gesagt wird, verwendet man auf \mathbb{R}^n oder \mathbb{C}^n stets die 2-Norm**. Man schreibt auch nur $|\cdot|$ für $|\cdot|_2$ und nennt $|\cdot|$ die **Euklidische Norm** auf \mathbb{K}^n sowie $|x| = |x|_2$ den **Betrag des Vektors** $x \in \mathbb{K}^n$, denn $|x|$ entspricht nach dem Satz des Pythagoras der **elementargeometrischen Länge** von x .

Der Nachweis der Normaxiome (N1) und (N2) für die p -Normen ist dabei sehr einfach. In den Fällen $p = 1$ und $p = \infty$ ist auch der Nachweis von (N3) problemlos, und im Fall $1 < p < \infty$ gelingt er mit Hilfe der **Hölder-Ungleichung für (endliche) Summen**

$$\left| \sum_{i=1}^n x_i y_i \right| \leq |x|_p |y|_q \quad \text{für konjugierte Exponenten } p, q \in [1, \infty] \text{ mit } \frac{1}{p} + \frac{1}{q} = 1,$$

wobei $x = (x_1, \dots, x_n) \in \mathbb{K}^n$, $y = (y_1, \dots, y_n) \in \mathbb{K}^n$ und $\frac{1}{\infty} = 0$ zu verstehen ist. Der Beweis der Hölder-Ungleichung wird in den Übungen behandelt, der darauf aufbauende Nachweis von (N3) im Fall $1 < p < \infty$ wird in der Vorlesung ausgeführt.

(2) Der **Raum der p -summierbaren Folgen** über \mathbb{K} wird als

$$\ell^p = \ell^p(\mathbb{K}) := \left\{ (x_i)_{i \in \mathbb{N}} \mid (x_i)_{i \in \mathbb{N}} \text{ ist Folge in } \mathbb{K} \text{ mit } \sum_{i=1}^{\infty} |x_i|^p < \infty \right\} \quad \text{für } p \in [1, \infty)$$

definiert. Dieser **unendlich-dimensionale (!) Folgenraum** wird mit der **ℓ^p -Norm**

$$\|x\|_{\ell^p} := \left(\sum_{i=1}^{\infty} |x_i|^p \right)^{\frac{1}{p}} \quad \text{für } x = (x_i)_{i \in \mathbb{N}} \in \ell^p$$

zu einem normierten Raum. Für $p = \infty$ betrachtet man als Analoga den Raum $\ell^\infty = \ell^\infty(\mathbb{K})$ aller beschränkten Folgen in \mathbb{K} und den Raum $c_0 = c_0(\mathbb{K})$ aller Nullfolgen in \mathbb{K} , die beide durch die ℓ^∞ -Norm $\|x\|_{\ell^\infty} := \sup\{|x_i| \mid i \in \mathbb{N}\}$ normiert werden.

Dass ℓ^p überhaupt ein \mathbb{K} -Vektorraum ist und $\|\cdot\|_{\ell^p}$ die Dreiecksungleichung erfüllt, ergibt sich durch Grenzübergang $n \rightarrow \infty$ in der Dreiecksungleichung für die p -Normen auf \mathbb{K}^n . Nebenbei erhält man die **Hölder-Ungleichung für Reihen** (mit Notation wie zuvor)

$$\left| \sum_{i=1}^{\infty} x_i y_i \right| \leq \|x\|_{\ell^p} \|y\|_{\ell^q} \quad \text{für } x \in \ell^p, y \in \ell^q \text{ und } p, q \in [1, \infty] \text{ mit } \frac{1}{p} + \frac{1}{q} = 1.$$

(3) Für $a < b$ in \mathbb{R} wird der **unendlich-dimensionale (!) Funktionenraum** $C^0([a, b], \mathbb{K})$ der stetigen¹ Funktionen $f: [a, b] \rightarrow \mathbb{K}$ mit jeder der **L^p -(Integral-)Normen**

$$\|f\|_{L^p(a,b)} := \left(\int_a^b |f(x)|^p dx \right)^{\frac{1}{p}} \quad \text{zu } p \in [1, \infty)$$

ein normierter Raum. Das Analogon für $p = \infty$ ist die durch $\|f\|_{L^\infty(a,b)} := \sup_{(a,b)} |f|$ erklärte **L^∞ -(Supremums-)Norm**, die neben $C^0([a, b], \mathbb{K})$ auch den Raum *aller* beschränkten Funktionen $[a, b] \rightarrow \mathbb{K}$ normiert. Der Nachweis der Normeigenschaften ist Thema der Übungen und basiert im Fall $1 < p < \infty$ auf der **Hölder-Ungleichung für Integrale**

$$\left| \int_a^b f(x)g(x) dx \right| \leq \|f\|_{L^p(a,b)} \|g\|_{L^q(a,b)} \quad \text{für } p, q \in [1, \infty] \text{ mit } \frac{1}{p} + \frac{1}{q} = 1$$

mit Riemann-integrierbaren Funktionen $f, g: (a, b) \rightarrow \mathbb{K}$ (für die nach Proposition 7.16 der Mathematik 3 auch fg und für $p, q < \infty$ auch $|f|^p$ und $|g|^q$ Riemann-integrierbar sind).

¹Auf dem Raum aller Riemann-integrierbaren Funktionen ist $\|\cdot\|_{L^p(a,b)}$ nur eine sogenannte Halbnorm, bei der $\|f\|_{L^p(a,b)} = 0$ noch für andere Funktionen f neben der Nullfunktion gilt (beispielsweise für Funktionen, die nur an endlich vielen Stellen von Null verschieden sind), aber ansonsten alle Norm-Axiome erfüllt sind. Dies ist der Grund, warum wir uns hier auf stetige Funktionen beschränken, auf denen $\|\cdot\|_{L^p(a,b)}$ eine echte Norm wird.

Beispiele (von metrischen Räumen).

(0) Es sei daran erinnert, dass normierte Räume und deren beliebige Teilmengen mit der induzierten Metrik stets metrische Räume sind. Insofern ergibt sich eine **weite Klasse von Beispielen** für metrische Räume als **beliebige Teilmenge der zuvor betrachteten normierten Räume**. Im Folgenden geht es um andere Typen von Beispielen.

(1) Auf jeder Menge \mathcal{X} wird durch

$$d_0(x, y) := \begin{cases} 1 & \text{falls } x \neq y \\ 0 & \text{falls } x = y \end{cases} \quad \text{für } x, y \in \mathcal{X}$$

eine Metrik, die **diskrete Metrik**, definiert. Bezüglich dieser Metrik, die vor allem als Extrembeispiel dient, haben zwei voneinander verschiedene Punkte den immer gleichen Abstand 1.

(2) Zwei Metriken d_P und d_E auf \mathbb{K}^n werden durch

$$d_P(x, y) := \begin{cases} |x|+|y| & \text{falls } x \neq y \\ 0 & \text{falls } x = y \end{cases} \quad \text{und} \quad d_E(x, y) := \begin{cases} |x|+|y| & \text{falls } x, y \text{ linear unabhängig} \\ |y-x| & \text{falls } x, y \text{ linear abhängig} \end{cases}$$

für $x, y \in \mathbb{K}^n$ definiert. Diese Metriken werden manchmal als (Italienische-) **Post-Metrik** und (Französische-/Britische-) **Eisenbahn-Metrik** bezeichnet, da man bezüglich ihnen immer über den Ursprung (=Rom/Paris/London) gehen muss, außer man bewegt sich im Fall von d_E nur auf demselben Ursprungsstrahl (=radialem Gleisstrang). Analoge Bildungen sind ausgehend von einem beliebigen normierten Raum möglich.

(3) Gelegentlich treten Metriken auf, die über ein „**Prinzip des kürzesten Pfads**“ gebildet werden, was hier anhand von zwei Situationen nur angedeutet sei:

Für eine Teilmenge A eines normierten Raums \mathcal{X} definiert man eine „**innere Metrik**“ d_A der Menge A durch

$$d_A(x, y) := \inf \left\{ \sum_{i=1}^n \|x_i - x_{i-1}\| \mid \begin{array}{l} n \in \mathbb{N}_0, x_0 = x, x_1, x_2, \dots, x_{n-1} \in A, x_n = y \\ [x_{i-1}, x_i] \subset A \text{ für } i = 1, 2, \dots, n \end{array} \right\}$$

für $x, y \in A$, wobei $[a, b] := \{(1-t)a + tb \mid t \in [0, 1]\}$ die Strecke von a nach b in \mathcal{X} bezeichnet. Eine echte Metrik d_A auf A ergibt sich, wenn die Menge im Infimum stets $\neq \emptyset$ ist, eine Eigenschaft von A , die man als Streckenzug-zusammenhängend bezeichnen kann. Ansonsten erhält man eine sogenannte erweiterte Metrik auf A , die auch den Wert ∞ annehmen kann.

Ein sogenannter Graph lässt sich durch eine beliebige Menge V von Knoten („vertices“) und eine Menge $E \subset \{\{v, w\} \mid v, w \in V, v \neq w\}$ von je zwei Knoten verbindenden Kanten („edges“) modellieren. Durch

$$d_{(V,E)}(v, w) := \inf \left\{ n \in \mathbb{N}_0 \mid \begin{array}{l} \exists v_0, v_1, \dots, v_n \in V: \\ v_0 = v, \{v_0, v_1\}, \{v_1, v_2\}, \dots, \{v_{n-1}, v_n\} \in E, v_n = w \end{array} \right\}$$

für $v, w \in V$ erhält man dann eine **Metrik $d_{(V,E)}$ des kürzesten Kantenpfads im Graph (V, E)** , sofern auch hier das Infimum stets $\neq \emptyset$ und der Graph (V, E) in diesem Sinn zusammenhängend ist.

Für anschauliche Beschreibungen von d_A und $d_{(V,E)}$ in guten Fällen wird auf die Vorlesung verwiesen.

Definitionen & Bemerkungen (Konzepte in metrischen Räumen). Viele Konzepte der Analysis lassen sich allgemein für metrische (und damit insbesondere für normierte) Räume erklären, wobei oft starke Analogie zum von \mathbb{R} oder \mathbb{C} bekannten Vorgehen besteht. Wir fassen uns deshalb im Folgenden eher kurz, wobei $(\mathcal{X}, d_{\mathcal{X}})$ und $(\mathcal{Y}, d_{\mathcal{Y}})$ stets metrische Räume seien:

- (1) Die **offene Kugel** und die **abgeschlossene Kugel** in \mathcal{X} mit „Mittelpunkt“ $x_0 \in \mathcal{X}$ und „Radius“ $r \in (0, \infty)$ sind

$$\begin{aligned} B_r(x_0) &:= B_r^{\mathcal{X}}(x_0) := \{x \in \mathcal{X} \mid d_{\mathcal{X}}(x, x_0) < r\}, \\ \bar{B}_r(x_0) &:= \bar{B}_r^{\mathcal{X}}(x_0) := \{x \in \mathcal{X} \mid d_{\mathcal{X}}(x, x_0) \leq r\}. \end{aligned}$$

Die **Sphäre** in \mathcal{X} mit „Mittelpunkt“ $x_0 \in \mathcal{X}$ und „Radius“ $r \in (0, \infty)$ ist

$$S_r(x_0) := S_r^{\mathcal{X}}(x_0) := \bar{B}_r(x_0) \setminus B_r(x_0) = \{x \in \mathcal{X} \mid d_{\mathcal{X}}(x, x_0) = r\}.$$

Für $\mathcal{X} = \mathbb{R}^2$ beziehungsweise $\mathcal{X} = \mathbb{R}^3$ mit der Euklidischen Norm handelt es sich bei $B_r(x_0)$ und $\bar{B}_r(x_0)$ tatsächlich um Kreisscheiben beziehungsweise Kugeln elementargeometrischen Sinn, bei $S_r(x_0)$ um die diese begrenzende Kreislinie beziehungsweise Kugeloberflächen. Für die p -Normen mit $p \neq 2$ auf \mathbb{R}^2 handelt es sich bei den „Kugeln“ $B_r(x_0)$ um „ausgedellte“ (Fall $2 < p < \infty$) oder „eingedellte“ (Fall $1 < p < 2$) Kreisscheiben, die schließlich in Quadrate mit achsenparallelen (Fall $p = \infty$) oder diagonalen (Fall $p = 1$) Seiten übergehen. Für die graphische Illustration wird auf die Vorlesung verwiesen.

In einem allgemeinen metrischen Raum können die „Kugeln“ $B_r(x_0)$ auch noch ganz andere Formen annehmen. Im Extremfall, dass \mathcal{X} einen sogenannten isolierten Punkt $x_0 \in \mathcal{X}$ besitzt, kommt sogar $B_r(x_0) = \{x_0\}$, $S_r(x_0) = \emptyset$ (zumindest) für $0 < r \ll 1$ vor.

- (2) Neben den Abständen $d_{\mathcal{X}}(x, y)$ zweier Punkte $x, y \in \mathcal{X}$ erklärt man auch ...

- den **Abstand**

$$\text{dist}(x, A) := \inf\{d_{\mathcal{X}}(x, y) \mid y \in A\} \in [0, \infty)$$

eines Punktes $x \in \mathcal{X}$ zu einer nicht-leeren **Teilmenge** A von \mathcal{X} ;

- den **Abstand**

$$\text{dist}(A, B) := \inf\{d_{\mathcal{X}}(x, y) \mid x \in A, y \in B\} \in [0, \infty)$$

zweier nicht-leerer **Teilmengen** A und B von \mathcal{X} ;

- den **Durchmesser**

$$\text{diam } A := \sup\{d_{\mathcal{X}}(x, y) \mid x, y \in A\} \in [0, \infty]$$

einer **Teilmenge** von \mathcal{X} (mit der Konvention $\text{diam } \emptyset = 0$).

Weiterhin nennt man $A \subset \mathcal{X}$ **beschränkt**, wenn $\text{diam } A < \infty$ gilt. Äquivalent zu Beschränktheit von A ist $\sup_{x \in A} d_{\mathcal{X}}(x, x_0) < \infty$ für einen beliebigen Referenzpunkt $x_0 \in \mathcal{X}$, was sich auch als $A \subset B_r(x_0)$ für ein $x_0 \in \mathcal{X}$ und ein $r \in (0, \infty)$ ausdrücken lässt.

- (3) Die **Konvergenz von Folgen „in der Metrik“** definiert man genau wie früher die Konvergenz in \mathbb{K} , nur mit der Metrik anstelle des Betrags. Ausgeschrieben führt das für eine Folge $(x_k)_{k \in \mathbb{N}}$ in \mathcal{X} und $x \in \mathcal{X}$ zur Definition

$$\lim_{k \rightarrow \infty} x_k = x \text{ in } \mathcal{X} \iff \forall \varepsilon > 0: \exists k_0 \in \mathbb{N}: \forall k \in \mathbb{N}_{\geq k_0}: d_{\mathcal{X}}(x_k, x) < \varepsilon,$$

wobei man statt $d_{\mathcal{X}}(x_k, x) < \varepsilon$ natürlich auch $x_k \in B_{\varepsilon}(x)$ schreiben kann. Liest man die Definition als Eigenschaft der Folge $(d_{\mathcal{X}}(x_k, x))_{k \in \mathbb{N}}$ in \mathbb{R} (!), so entnimmt man unmittelbar die oft nützliche Charakterisierung

$$\lim_{k \rightarrow \infty} x_k = x \text{ in } \mathcal{X} \iff \lim_{k \rightarrow \infty} d_{\mathcal{X}}(x_k, x) = 0,$$

die Konvergenz in der Metrik auf „normale“ Konvergenz in \mathbb{R} zurückführt und sich ebenfalls als Definition eignen würde. Aus der Charakterisierung wiederum wird klar, dass **konvergente Folgen** $(x_k)_{k \in \mathbb{N}}$ auch im metrischen Kontext **stets beschränkt** sind (tatsächlich so zu verstehen, dass $\{x_k \mid k \in \mathbb{N}\}$ beschränkt im Sinn des vorigen Punkts ist). Alles hier Gesagte gilt natürlich insbesondere für Konvergenz von Folgen „in der Norm“ (dann einfach mit $\|x_k - x\|_{\mathcal{X}}$ anstelle $d_{\mathcal{X}}(x_k, x)$), und ganz analog lassen sich auch **Häufungswerte von Folgen** verallgemeinern.

- (4) Auch **Grenzwerte von Funktionen** $f: D \rightarrow \mathcal{Y}$ mit $D \subset \mathcal{X}$ lassen sich verallgemeinern, indem man einfach die Metrik anstelle des Betrags verwendet. Ausgeschrieben besagt die Grenzwert-Definition für den Grenzübergang aus D gegen einen Berührungspunkt² $x_0 \in \mathcal{X}$ von D und für den (potentiellen) Grenzwert $y_0 \in \mathcal{Y}$ in der ε - δ -Formulierung dann

$$\lim_{D \ni x \rightarrow x_0} f(x) = y_0 \iff \forall \varepsilon > 0: \exists \delta > 0: \forall x \in D: (d_{\mathcal{X}}(x, x_0) < \delta \implies d_{\mathcal{Y}}(f(x), y_0) < \varepsilon),$$

wobei die Implikation alternativ mit Kugeln als $x \in B_{\delta}(x_0) \implies f(x) \in B_{\varepsilon}(y_0)$ geschrieben oder der hintere Teil ab $\forall x$ kurz als die Inklusion $f(D \cap B_{\delta}(x_0)) \subset B_{\varepsilon}(y_0)$ gefasst werden kann. Genau wie in Kapitel 1 der Mathematik 3 sieht man, dass der Grenzwertbegriff äquivalent mit Folgen durch

$$\begin{aligned} & \lim_{D \ni x \rightarrow x_0} f(x) = y_0 \\ & \iff \text{Für alle Folgen } (x_k)_{k \in \mathbb{N}} \text{ in } D: \left(\lim_{k \rightarrow \infty} x_k = x_0 \text{ in } \mathcal{X} \implies \lim_{k \rightarrow \infty} f(x_k) = y_0 \text{ in } \mathcal{Y} \right) \end{aligned}$$

charakterisiert ist. Außerdem ist der Grenzwert y_0 , wenn er denn existiert, natürlich weiterhin eindeutig. Auch diese Definitionen greifen natürlich insbesondere in normierten Räumen, und ganz analog kann man auch **Häufungswerte von Funktionen** $\mathcal{X} \rightarrow \mathcal{Y}$ erklären.

- (5) **Stetigkeit von Funktionen** $f: D \rightarrow \mathcal{Y}$ mit $D \subset \mathcal{X}$ in einem Punkt $x_0 \in D$ erklärt man wie früher durch³ $\lim_{D \ni x \rightarrow x_0} f(x) = f(x_0)$ und ihre Stetigkeit auf D als $\lim_{D \ni x \rightarrow x_0} f(x) = f(x_0)$ für alle $x_0 \in D$, wobei die Grenzwerte im gerade diskutierten Sinn zu verstehen sind. Möchte

²Berührungspunkte können genau wie in der Mathematik 3 über Folgenkonvergenz definiert werden. Wir betrachten das Konzept demnächst aber sowieso noch genauer und gehen deshalb an dieser Stelle nicht ins Detail.

³Genau genommen reicht es nach den getroffenen Definitionen sogar, hier nur die Existenz des Limes zu fordern, denn dann ist dieser automatisch gleich $f(x_0)$. Das hat aber nichts mit metrischen Räumen zu tun, sondern hängt nur an der genauen Form der Grenzwertdefinition und war auch in der Mathematik 3 schon so.

man dies mit ε und δ ausschreiben, so ersetzt man einfach auf der rechten Seite der Grenzwertdefinition y_0 durch $f(x_0)$.

Wie in der Mathematik 3 überträgt sich auch Stetigkeit von komponierbaren Funktionen zwischen metrischen Räumen auf ihre Komposition.

- (6) Schließlich können auch quantitative Stetigkeitsbegriffe, **Hölder-Stetigkeit und Lipschitz-Stetigkeit**, für $f: D \rightarrow \mathcal{Y}$ mit $D \subset \mathcal{X}$ erklärt werden: Hölder-Stetigkeit von f zum Exponent $\alpha \in (0, 1]$ auf D bedeutet

$$d_{\mathcal{Y}}(f(x), f(\tilde{x})) \leq C d_{\mathcal{X}}(x, \tilde{x})^{\alpha} \quad \text{für alle } x, \tilde{x} \in D$$

mit einer festen Hölder-Konstante $C \in [0, \infty)$. Im Fall des Exponenten $\alpha = 1$ spricht man von Lipschitz-Stetigkeit oder auch Dehnungsbeschränktheit und von der Lipschitz-Konstante.

Bemerkungen (zu normierten Räumen). Über die für metrische Räume eingeführten Konzepte hinaus halten wir für einen normierten Raum \mathcal{X} über \mathbb{K} folgende Aussagen fest:

- (1) Für jeden Vektor $x \in \mathcal{X} \setminus \{0_{\mathcal{X}}\}$ hat der zugehörige **normierte Vektor** $\frac{x}{\|x\|_{\mathcal{X}}} := \frac{1}{\|x\|_{\mathcal{X}}}x \in \mathcal{X}$ die Norm 1 und ist damit ein sogenannter Einheitsvektor.
- (2) Es gilt die (verallgemeinerte) **Dreiecksungleichung**

$$\left\| \sum_{i=1}^n s_i x_i \right\|_{\mathcal{X}} \leq \sum_{i=1}^n |s_i| \|x_i\|_{\mathcal{X}} \quad \text{für } n \in \mathbb{N}, s_1, s_2, \dots, s_n \in \mathbb{K}, x_1, x_2, \dots, x_n \in \mathcal{X},$$

aus der man als Spezialfall Konvexität⁴ der Norm $\|\cdot\|_{\mathcal{X}}: \mathcal{X} \rightarrow \mathbb{R}$ abliest. Zudem gilt die **umgekehrte Dreiecksungleichung**

$$\left| \|y\|_{\mathcal{X}} - \|x\|_{\mathcal{X}} \right| \leq \|y - x\|_{\mathcal{X}} \quad \text{für } x, y \in \mathcal{X},$$

gemäß der die Norm $\|\cdot\|_{\mathcal{X}}: \mathcal{X} \rightarrow \mathbb{R}$ Lipschitz-stetig mit Lipschitz-Konstante 1 ist.

- (3) Für konvergente Folgen $(x_k)_{k \in \mathbb{N}}$ und $(y_k)_{k \in \mathbb{N}}$ in \mathcal{X} und eine konvergente Folge $(s_k)_{k \in \mathbb{N}}$ in \mathbb{K} gelten die **Summen- und Faktorregeln**

$$\lim_{k \rightarrow \infty} (x_k + y_k) = \left(\lim_{k \rightarrow \infty} x_k \right) + \left(\lim_{k \rightarrow \infty} y_k \right), \quad \lim_{k \rightarrow \infty} (s_k x_k) = \left(\lim_{k \rightarrow \infty} s_k \right) \left(\lim_{k \rightarrow \infty} x_k \right) \quad \text{in } \mathcal{X}.$$

Analoge Summen- und Faktorregeln gelten für Grenzwerte und die Erhaltung von Stetigkeit bei \mathcal{X} -wertigen Funktionen (auf einem beliebigen metrischen Definitionsbereich).

Einen gewissen Überblick die Normen auf einem Vektorraum kann man sich in guten Fällen mit Hilfe des folgenden Konzepts und des darauf aufbauenden Satzes verschaffen.

Definition (Vergleich von Normen). Seien $\|\cdot\|_1$ und $\|\cdot\|_2$ Normen auf **einem** \mathbb{K} -Vektorraum \mathcal{X} . Dann heißt $\|\cdot\|_1$ *schwächer als* $\|\cdot\|_2$ und dementsprechend $\|\cdot\|_2$ *stärker als* $\|\cdot\|_1$, wenn es eine Konstante $C \in [0, \infty)$ mit $\|x\|_1 \leq C\|x\|_2$ für alle $x \in \mathcal{X}$ gibt. Ist $\|\cdot\|_1$ zugleich schwächer und stärker als $\|\cdot\|_2$, so heißen $\|\cdot\|_1$ und $\|\cdot\|_2$ **äquivalent**.

⁴In der Mathematik 3 wurde Konvexität \mathbb{R} -wertiger Funktionen auf einem Intervall I über (Konvexitäts-) Ungleichungen definiert. Diese Definition kann analog auf \mathbb{R} -wertige Funktionen mit (einer konvexen Menge in) einem beliebigen \mathbb{K} -Vektorraum als Definitionsbereich übertragen werden und macht dann insbesondere für jede Norm $\|\cdot\|_{\mathcal{X}}$ als \mathbb{R} -wertige Funktion auf einem \mathbb{K} -Vektorraum \mathcal{X} Sinn.

Bemerkungen (zur **Äquivalenz von Normen**). Sei \mathcal{X} ein \mathbb{K} -Vektorraum.

(1) Äquivalenz zweier Normen $\|\cdot\|_1$ und $\|\cdot\|_2$ auf \mathcal{X} bedeutet mit anderen Worten, dass

$$C^{-1}\|x\|_2 \leq \|x\|_1 \leq C\|x\|_2 \quad \text{für alle } x \in \mathcal{X}$$

mit einer Konstante $C \in [1, \infty)$ gilt.

(2) Äquivalenz von Normen ist eine Äquivalenzrelation auf der Menge aller Normen auf \mathcal{X} .

(3) **Äquivalente Normen führen zu den gleichen Begriffen von Konvergenz, Grenzwerten und Stetigkeit.** (Kugeln und Hölder-/Lipschitz-Konstanten bezüglich äquivalenter Normen können sich allerdings unterscheiden.)

Im **endlich-dimensionalen Fall** zeigt der nächste Satz, dass man sich um die genaue Wahl der Norm künftig nicht mehr zu viele Gedanken machen muss:

Satz (Äquivalenz aller Normen auf einem endlich-dimensionalen Vektorraum). *Auf einem endlich-dimensionalen \mathbb{K} -Vektorraum \mathcal{X} (beispielsweise auf \mathbb{K}^n) sind alle Normen äquivalent, und die Konvergenz einer Folge in \mathcal{X} bezüglich irgendeiner Norm auf \mathcal{X} ist äquivalent zu ihrer komponentenweisen Konvergenz bezüglich einer beliebigen Basis von \mathcal{X} .*

Dabei bedeutet komponentenweise Konvergenz einer Folge $(x_k)_{k \in \mathbb{N}}$ in \mathcal{X} gegen einen Grenzwert $x_* \in \mathcal{X}$ bezüglich einer \mathbb{K} -Basis e_1, e_2, \dots, e_n von \mathcal{X} , dass $\lim_{k \rightarrow \infty} s_{i,k} = s_{i,*}$ für alle $i \in \{1, 2, \dots, n\}$ gilt, wobei $s_{i,k}, s_{i,*} \in \mathbb{K}$ die (bekanntlich eindeutigen) Koeffizienten in den Basisdarstellungen $x_k = \sum_{i=1}^n s_{i,k} e_i$ und $x_* = \sum_{i=1}^n s_{i,*} e_i$ sind.

Für einen Beweis des Äquivalenz-Satzes unter entscheidender Verwendung des Satzes von Bolzano-Weierstraß wird auf die Vorlesung verwiesen.

Bemerkungen.

- (1) Für endlich-dimensionales \mathcal{X} folgt aus dem Äquivalenz-Satz auch, dass eine \mathcal{X} -wertige Funktion (auf einem beliebigem matriscen Definitionsbereich) genau dann stetig ist, wenn all ihre Komponentenfunktionen bezüglich einer beliebigen Basis dies sind.
- (2) In scharfem Kontrast zum endlich-dimensionalen Fall des Äquivalenz-Satzes gibt es auf einem **unendlich-dimensionalen** \mathbb{K} -Vektorraum stets inäquivalente Normen. Beispielsweise wird für die Folgeräume zu Exponenten $p < q$ in $[1, \infty]$ in den Übungen gezeigt, dass

$$\ell^p \subset \ell^q \quad \text{mit} \quad \|\cdot\|_{\ell^q} \leq \|\cdot\|_{\ell^p} \text{ auf } \ell^p$$

gilt. Eine umgekehrte Abschätzung des Typs $\|\cdot\|_{\ell^p} \leq C\|\cdot\|_{\ell^q}$ gilt aber eben **nicht**.

Als letzten Typ von Räumen sprechen wir in diesem Kapitel Skalarprodukträume an und ergänzen die schon in der Mathematik 3 im Rahmen der (bi)linearen Algebra behandelte Theorie dieser Räume um Beispiele und geometrische Aspekte.

Erinnerung (an **Skalarprodukträume**). Ein **Skalarprodukt** $\langle \cdot, \cdot \rangle: \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{K}$ auf einem \mathbb{K} -Vektorraum \mathcal{X} wurde in der Mathematik 3 als eine positiv definite, symmetrische Bilinearform (Fall $\mathbb{K} = \mathbb{R}$) beziehungsweise eine positiv definite, hermitesche Sesquilinearform (Fall $\mathbb{K} = \mathbb{C}$) eingeführt und macht das Paar $(\mathcal{X}, \langle \cdot, \cdot \rangle)$ zu einem **Skalarproduktraum** über \mathbb{K} . Ein

Skalarprodukt $\langle \cdot, \cdot \rangle$ auf \mathcal{X} induziert durch $\|x\| := \sqrt{\langle x, x \rangle}$ für $x \in \mathcal{X}$ eine Norm $\|\cdot\|$ auf \mathcal{X} , und mit dieser gilt die **Cauchy-Schwarz-Ungleichung**

$$|\langle x, y \rangle| \leq \|x\| \|y\| \quad \text{für alle } x, y \in \mathcal{X}$$

mit Gleichheit genau bei \mathbb{K} -linear Abhängigkeit der Vektoren x und y .

Bemerkungen und Beispiele (zu Skalarprodukten).

- (0) Wir notieren für Skalarprodukte gleichbedeutend mit $\langle x, y \rangle$ auch $\langle x, y \rangle_{\mathcal{X}}$, $x \cdot y$ oder $x \cdot_{\mathcal{X}} y$.
- (1) Das wohl grundlegendste Beispiel eines Skalarprodukts ist das **Euklidische Skalarprodukt auf \mathbb{K}^n** , das durch

$$\langle x, y \rangle_{\mathbb{K}^n} := \sum_{i=1}^n \overline{x_i} y_i \quad \text{für } x, y \in \mathbb{K}^n$$

gegeben ist. Dieses Skalarprodukt induziert die Euklidische Norm, mit anderen Worten die 2-Norm, auf \mathbb{K}^n . Weitere natürliche Beispiele sind das **ℓ^2 -Skalarprodukt** auf dem Folgenraum ℓ^2 , gegeben durch

$$\langle x, y \rangle_{\ell^2} := \sum_{i=1}^{\infty} \overline{x_i} y_i \quad \text{für } x = (x_i)_{i \in \mathbb{N}}, y = (y_i)_{i \in \mathbb{N}} \in \ell^2,$$

und das **L^2 -Skalarprodukt** auf $C^0([a, b], \mathbb{K})$ mit $a < b$ in \mathbb{R} , gegeben durch

$$\langle f, g \rangle_{L^2(a,b)} := \int_a^b \overline{f(x)} g(x) dx \quad \text{für } f, g \in C^0([a, b], \mathbb{K}).$$

Diese Skalarprodukte induzieren die ℓ^2 - beziehungsweise die L^2 -Norm. Im Fall $\mathbb{K} = \mathbb{R}$ kann die Konjugation $\overline{(\cdot)}$ hier jeweils entfallen, und die Cauchy-Schwarz-Ungleichung entspricht bei diesen Beispielen dem Spezialfall $p = q = 2$ der jeweiligen Hölder-Ungleichung.

- (2) Für das Euklidische Skalarprodukt auf \mathbb{R}^n gilt

$$\langle u, v \rangle = |u| |v| \cos \sphericalangle(u, v) \quad \text{für } u, v \in \mathbb{R}^n \setminus \{0\}$$

mit dem vorerst elementargeometrisch aufgefassten Winkel $\sphericalangle(u, v) \in [0, \pi]$ zwischen den Vektoren u und v . Um diese Formel einzusehen, betrachtet man zunächst den Fall $u = u_1 e_1$ mit $u_1 > 0$, argumentiert anhand des rechtwinkligen Dreiecks mit den Ecken 0 , $v_1 e_1$ und v , dass $v_1 = |v| \cos \sphericalangle(e_1, v) = |v| \cos \sphericalangle(u, v)$ gilt, und schließt auf $\langle u, v \rangle = u_1 v_1 = |u| v_1 = |u| |v| \cos \sphericalangle(u, v)$. Der allgemeine Fall lässt sich durch eine Drehung, die e_1 in $\frac{u}{|u|}$ überführt und als orthogonale Abbildung Skalarprodukt und induzierte Norm erhält, darauf zurückführen.

Motiviert durch das Vorausgehende erhebt man in einem beliebigen Skalarproduktraum $(\mathcal{X}, \langle \cdot, \cdot \rangle)$ über dem Grundkörper \mathbb{R} die nach dem Winkel aufgelöste Formel

$$\sphericalangle(u, v) := \arccos \frac{\langle u, v \rangle}{\|u\| \|v\|} \in [0, \pi] \quad \text{für } u, v \in \mathcal{X} \setminus \{0\}$$

zur **Definition des Winkels zwischen zwei Vektoren** u und v (wobei $\|\cdot\|$ die von $\langle \cdot, \cdot \rangle$ induzierte Norm bezeichnet, nach Cauchy-Schwarz-Ungleichung $\frac{\langle u, v \rangle}{\|u\| \|v\|} \in [-1, 1]$ gilt

und daher $\arccos \frac{\langle u, v \rangle}{\|u\| \|v\|}$ überhaupt definiert ist). Speziell im Fall $\langle u, v \rangle = 0$ erhält man $\sphericalangle(u, v) = \frac{\pi}{2}$ und nennt u und v dann (**zueinander**) **orthogonal**. (Über dem Grundkörper \mathbb{C} übrigens bleibt von diesen Konzepten nur der schon in der Mathematik 3 verwendete Begriff der Orthogonalität im Sinn von $\langle u, v \rangle = 0$ sinnvoll.)

- (3) Mit Winkeln im Sinn des vorigen Punkts lassen sich viele elementargeometrische Sachverhalte allgemein in Skalarprodukträumen formulieren und teils schnell und elegant beweisen. Wir führen dies nun kurz am Beispiel des Kosinussatzes (der bekanntlich den Satz des Pythagoras für nicht notwendig rechtwinklige Dreiecke erweitert) vor.

Kosinussatz: Ist \mathcal{X} ein Skalarproduktraum über \mathbb{R} , so gilt

$$\|u-v\|^2 = \|u\|^2 + \|v\|^2 - 2\|u\| \|v\| \cos \sphericalangle(u, v) \quad \text{für alle } u, v \in \mathcal{X} \setminus \{0\}$$

(was perfekt dem elementargeometrischen Kosinussatz im Dreieck mit Ecken 0 , u , v und Seitenlängen $\|u\|$, $\|v\|$, $\|u-v\|$ entspricht).

Beweis. Eigenschaften des Skalarprodukts und die Formel des vorausgehenden Punkts ergeben $\|u-v\|^2 = \langle u-v, u-v \rangle = \langle u, u \rangle + \langle v, v \rangle - 2\langle u, v \rangle = \|u\|^2 + \|v\|^2 - 2\|u\| \|v\| \cos \sphericalangle(u, v)$. \square

Schließlich weiten wir noch den im Zusammenhang mit \mathbb{R} und \mathbb{C} schon lange bekannten Begriff der Vollständigkeit auf allgemeine Räume aus.

Definition (allgemeine Cauchy-Folgen, vollständige Räume). Wir nennen eine Folge $(x_k)_{k \in \mathbb{N}}$ in einem metrischen Raum \mathcal{X} eine **Cauchy-Folge** in \mathcal{X} , wenn

$$\forall \varepsilon > 0: \exists k_0 \in \mathbb{N}: \forall k, \ell \in \mathbb{N}_{\geq k_0}: d_{\mathcal{X}}(x_k, x_\ell) < \varepsilon.$$

gilt. Darauf aufbauend bezeichnen wir einen metrischen Raum \mathcal{X} als **vollständig**, wenn jede Cauchy-Folge in \mathcal{X} gegen einen Grenzwert in \mathcal{X} konvergiert. Einen vollständigen normierten Raum nennen wir auch **Banachraum**, einen vollständigen Skalarproduktraum **Hilbertraum**.

Bemerkungen und Beispiele (zu Vollständigkeit).

- (1) **Vollständigkeit ist für die Analysis essentiell.** Auf unvollständigen Räumen kann man meist nur wenig sinnvolle Analysis betreiben.
- (2) Ähnlich zur in der Mathematik 2 erwähnten Konstruktion von \mathbb{R} aus \mathbb{Q} kann man auch für jeden normierten oder metrischen Raum \mathcal{X} durch eine Konstruktion des Typs „Cauchy-Folgen modulo Nullfolgen“ die sogenannte **Vervollständigung** $\widehat{\mathcal{X}}$ bilden. Im Fall von normierten Räumen handelt es sich dabei tatsächlich um eine Faktorbildung mit \mathbb{K} -Vektorräumen. In allgemeinen metrischen Räume \mathcal{X} hat man Konzepte wie „Null“, „Nullfolge“, „Unterraum“ nicht zur Verfügung, kann aber dennoch eine analoge Faktorisierung nach einer Äquivalenzrelation durchführen. Durch Vervollständigung lässt sich insgesamt jeder normierte Raum \mathcal{X} mit einer Teilmenge eines Banachraums $\widehat{\mathcal{X}}$, jeder metrische Raum \mathcal{X} mit einer Teilmenge eines vollständigen metrischen Raums $\widehat{\mathcal{X}}$ (isometrisch) identifizieren.
- (3) Die schon diskutierten **Standard-Beispiele normierter Räume sind** (abgesehen von eher technischen Problemen im Fall des Funktionenraums $C^0([a, b], \mathbb{K})$) alle **vollständig**:
 - Der endlich-dimensionale Raum \mathbb{K}^n ist mit jeder p -Norm, $p \in [1, \infty]$, (und auch mit jeder anderen Norm) ein Banachraum, mit der 2-Norm sogar ein Hilbertraum.

- Für jedes $p \in [1, \infty]$ ist der Folgenraum $\ell^p(\mathbb{K})$ mit der ℓ^p -Norm ein Banachraum, für $p = 2$ sogar ein Hilbertraum. Auch der Raum $c_0(\mathbb{K})$ der Nullfolgen ist mit der ℓ^∞ -Norm ein Banachraum.
- Für $a < b$ in \mathbb{R} ist der Funktionenraum $C^0([a, b], \mathbb{K})$ ein Banachraum mit der L^∞ -Norm. Mit den L^p -Normen zu $p \in [1, \infty)$ dagegen ist $C^0([a, b], \mathbb{K})$ nicht vollständig (und auch der Raum aller beschränkten, Riemann-integrierbaren Funktionen ist mit der entsprechenden Halbnorm nicht vollständig). Bei diesem Mangel an Vollständigkeit handelt es sich tatsächlich um einen Defekt der *Riemannschen* Integrationstheorie. Vervollständigung führt auf Räume der überlegenen *Lebesgueschen* Integrationstheorie.

Tatsächlich folgt die Vollständigkeit in all diesen Beispielen aus der Vollständigkeit des Grundkörpers \mathbb{K} . Details der hierfür noch benötigten Argumentationen werden zumindest für einzelne Fälle in den Übungen behandelt.

Abschließend halten wir fest, dass sich der aus der Mathematik 2 bekannte Kontraktionssatz praktisch wörtlich auf vollständige metrische Räume übertragen lässt:

Satz (Banachscher Fixpunktsatz, Kontraktionssatz). *Sei \mathcal{X} ein vollständiger metrischer Raum. Dann besitzt jede strikte Kontraktion $f: \mathcal{X} \rightarrow \mathcal{X}$ (d.h. jede Lipschitz-stetige Abbildung f von \mathcal{X} in sich mit Lipschitz-Konstante echt kleiner als 1) genau einen Fixpunkt $x_* \in \mathcal{X}$, und für beliebiges $x_0 \in \mathcal{X}$ konvergiert die durch $x_{k+1} := f(x_k)$ für $k \in \mathbb{N}_0$ definierte Iterationsfolge $(x_k)_{k \in \mathbb{N}_0}$ gegen x_* .*

Zum Beweis des Satzes kann man ganz analog zur Mathematik 2 vorgehen und muss in der Argumentation tatsächlich nur die Beträge von Differenzen durch Abstände im Sinn der Metrik $d_{\mathcal{X}}$ ersetzen. Auch die in der Mathematik 2 angegebenen Fehlerabschätzungen erhält man völlig analog.

1.2 Topologische Grundbegriffe

Wie zu Kapitelanfang schon angedeutet beschreiben topologische Konzepte die Lage von Punkten und Mengen zueinander auf qualitative Weise und hängen von quantitativen geometrischen Größen wie Abständen, Längen und Winkeln *nicht* ab. Wir werden topologische Begriffe im Folgenden unter Verwendung der offenen Kugeln $B_r(x) = \{y \in \mathcal{X} \mid d_{\mathcal{X}}(x, y) < r\}$ in metrischen Räumen \mathcal{X} einführen. Auch wenn dabei mit der Metrik $d_{\mathcal{X}}$ zunächst ein quantitatives Konzept zugrunde gelegt wird, erhalten wir als Resultat dennoch rein qualitative Begriffe, denn es wird tatsächlich nur auf das System aller Kugeln (um gegebene Mittelpunkte) ankommen und eben nicht auf quantitative Eigenschaften wie etwa den Radius der einzelnen Kugeln.

Wir beginnen mit folgenden grundlegenden Definitionen.

Definition (topologische Punktlagen, Inneres, Abschluss, Rand). *Sei \mathcal{X} ein metrischer Raum. Man erklärt folgende Lagen eines Punktes $x \in \mathcal{X}$ zu einer Menge $A \subset \mathcal{X}$:*

- x **innerer Punkt** von A : $\iff \exists r > 0: B_r(x) \subset A$,
- x **äußerer Punkt** von A : $\iff \exists r > 0: B_r(x) \cap A = \emptyset$,
- x **Berührungspunkt** von A : $\iff \forall r > 0: B_r(x) \cap A \neq \emptyset$,
- x **Randpunkt** von A : $\iff \forall r > 0: B_r(x) \cap A \neq \emptyset \neq B_r(x) \setminus A$,

- x **Häufungspunkt** von A : $\iff \forall r > 0$: $(B_r(x) \cap A$ enthält ∞ viele⁵ Elemente),
- x **isolierter Punkt** von A : $\iff \exists r > 0$: $B_r(x) \cap A = \{x\}$.

Darauf aufbauend definiert man als aus einer Menge $A \subset \mathcal{X}$ abgeleitete Punktmengen ...

- das **Innere** \mathring{A} von A (in \mathcal{X}) als Menge aller inneren Punkte von A ,
- den **Abschluss** \bar{A} von A (in \mathcal{X}) als Menge aller Berührungspunkte von A ,
- den **Rand** ∂A von A (in \mathcal{X}) als Menge aller Randpunkte von A .

Bemerkungen (zu **Innerem**, **Abschluss** und **Rand**). Sei A eine feste Teilmenge eines metrischen Raums \mathcal{X} .

- (1) Die **inneren Punkte**, **Randpunkte** und **äußeren Punkte** von A teilen den Raum \mathcal{X} in **drei disjunkte Mengen** auf. Die **inneren Punkte** und **Randpunkte** von A ergeben dabei zusammen genau die **Berührungspunkte** von A , es gilt also

$$\boxed{\bar{A} = \mathring{A} \cup \partial A}$$

Auch die Häufungspunkte und isolierten Punkte von A ergeben (im hiesigen Rahmen metrischer Räume) zusammen genau die Berührungspunkte von A und liefern neben der in innere Punkte und Randpunkte noch eine etwas andere Unterteilung der Berührungspunkte in zwei disjunkte Mengen. Diese Eigenschaften ergeben sich durch Vergleich der Definitionen (bezüglich Häufungspunkten und isolierten Punkten unter Berücksichtigung der Fußnote 5).

- (2) Aus den Definitionen folgt auch (vergleiche mit den Übungen)

$$\boxed{\mathring{A} \subset A \subset \bar{A}},$$

was zusammen mit dem vorigen Punkt die Identitäten $\mathring{A} = A \setminus \partial A$ und $\bar{A} = A \cup \partial A$ impliziert.

Beispiele (zu **Innerem**, **Abschluss** und **Rand**).

- (1) Das Rechteck $R := (-1, 2] \times [1, 3]$ hat in \mathbb{R}^2 das Innere $\mathring{R} = (-1, 2) \times (1, 3)$, den Abschluss $\bar{R} = [-1, 2] \times [1, 3]$ und den Rand $\partial R = (\{-1, 2\} \times [1, 3]) \cup ([-1, 2] \times \{1, 3\})$. Die Häufungspunkte von R sind genau die Berührungspunkte, also die Punkte in \bar{R} . Isolierte Punkte von R gibt es nicht.
- (2) Die Menge $S := \{\frac{1}{k} \mid k \in \mathbb{N}\}$ der Stammbrüche hat in \mathbb{R} das Innere $\mathring{S} = \emptyset$, den Abschluss $\bar{S} = S \cup \{0\}$ und den Rand $\partial S = S \cup \{0\}$. Dabei ist 0 einziger Häufungspunkt von S , alle Stammbrüche sind isolierte Punkte von S .

(Natürlich haben wir auf \mathbb{R} und \mathbb{R}^2 hier die durch Betrag bzw. 2-Norm induzierte Standardmetrik zugrunde gelegt. Wegen der Äquivalenz aller Normen auf den endlich-dimensionalen Räumen \mathbb{R} und \mathbb{R}^2 , würde jede andere durch eine Norm induzierte Metrik aber nichts ändern.)

⁵Man kann in der Definition eines Häufungspunkts auch nur fordern, dass $B_r(x) \cap A$ für jedes $r > 0$ mindestens zwei Elemente enthält. Wie in den Übungen gezeigt wird, sind es dann automatisch ∞ viele Elemente, so dass diese Variante der Definition äquivalent ist. Ebenso kann man in der Definition eines isolierten Punkts auch nur fordern, dass $B_r(x) \cap A$ für ein $r > 0$ nur endlich viele Elemente enthält. Auch dies ist äquivalent.

Es folgt eine weitere sehr grundlegende Definition.

Definition (offene Mengen und abgeschlossene Mengen). Sei \mathcal{X} ein metrischer Raum. Eine Teilmenge A von \mathcal{X} heißt ...

- **offen (in \mathcal{X})**, wenn $\overset{\circ}{A} = A$ ist,
- **abgeschlossen (in \mathcal{X})**, wenn $\overline{A} = A$ ist.

Bemerkungen (zu offenen Mengen und abgeschlossenen Mengen). Sei A eine Teilmenge eines metrischen Raums \mathcal{X} mit Komplement $A^c = \mathcal{X} \setminus A$.

- (1) Wie in den Übungen bewiesen wird, ist das Innere $\overset{\circ}{A}$ **stets offen** und der Abschluss \overline{A} **stets abgeschlossen**.
- (2) Aus den Bemerkungen zu Innerem, Abschluss und Rand ergeben sich verschiedene, allesamt wichtige **Charakterisierungen der offenen und der abgeschlossenen Mengen**:

$$A \text{ offen} \iff A \subset \overset{\circ}{A} \iff A \cap \partial A = \emptyset \iff \forall x \in A: \exists r > 0: B_r(x) \subset A,$$

$$A \text{ abgeschlossen} \iff \overline{A} = A \iff \partial A \subset A \iff \forall x \in A^c: \exists r > 0: B_r(x) \cap A = \emptyset.$$

Außerdem **entsprechen offene und abgeschlossene Mengen einander durch Komplementbildung**, es gelten also

$$A \text{ offen} \iff A^c \text{ abgeschlossen} \quad \text{und} \quad A \text{ abgeschlossen} \iff A^c \text{ offen}.$$

- (3) Anders als man aufgrund der Umgangssprache vermuten könnte, sind die topologischen Konzepte „**offen**“ und „**abgeschlossen**“ **keine logischen Gegenteile (Achtung!)**. Vielmehr gibt es sehr viele Mengen, die weder offen noch abgeschlossen sind, etwa die Mengen R und S der vorigen Beispiele.
- (4) Ist $x \in \overset{\circ}{A}$ innerer Punkt von A , so bezeichnet man A auch als **Umgebung** von A (in \mathcal{X}). Bei dieser Bezeichnung denkt man typischerweise daran, dass man zwischen x und A im Sinn von $x \in O \subset A$ noch eine in \mathcal{X} offene Menge O , konkret etwa $O = \overset{\circ}{A}$ oder $O = B_r(x)$ mit ausreichend kleinem $r > 0$, schieben kann. Oft arbeitet man auch nur mit offenen Umgebungen oder stellt sich Umgebungen selbst als kleine Kugeln vor.

Beispiele (für offene Mengen und abgeschlossene Mengen).

- (0) Für jeden metrischen Raum \mathcal{X} sind \emptyset und \mathcal{X} **zugleich offen und abgeschlossen** in \mathcal{X} .
- (1) Für Randpunkte $-\infty < a < b < \infty$ sind die **offenen Intervalle** (a, b) , (a, ∞) , $(-\infty, b)$ offen in \mathbb{R} , die **abgeschlossen Intervalle**⁶ $[a, b]$, $[a, \infty)$, $(-\infty, b]$ abgeschlossen in \mathbb{R} und die halboffenen Intervalle $[a, b)$, $(a, b]$ weder offen noch abgeschlossen in \mathbb{R} .

⁶Hierbei können unbeschränkte Intervalle, die einen uneigentlichen Randpunkt $\pm\infty$ nicht einschließen, dennoch abgeschlossen sein, denn der uneigentliche Randpunkt gehört gar nicht zum Grundraum \mathbb{R} und damit auch nicht zum Rand oder Abschluss des Intervalls in \mathbb{R} . In den erweiterten reellen Zahlen $\overline{\mathbb{R}} = \mathbb{R} \cup \{\infty, -\infty\}$ mit einer geeigneten Metrik wie $d_{\overline{\mathbb{R}}}(x, y) := |\arctan y - \arctan x|$ allerdings sind solche Intervalle nicht mehr abgeschlossen.

- (2) Für jeden metrischen Raum \mathcal{X} , jedes $x_0 \in \mathcal{X}$ und jedes $r \in (0, \infty)$ ist die **offene Kugel** $B_r(x_0) = \{x \in \mathcal{X} \mid d_{\mathcal{X}}(x, x_0) < r\}$ offen in \mathcal{X} , während die **abgeschlossene Kugel** $\bar{B}_r(x_0) = \{x \in \mathcal{X} \mid d_{\mathcal{X}}(x, x_0) \leq r\}$ abgeschlossen in \mathcal{X} ist. (In normierten Räumen ist übrigens auch stets $\overline{B_r(x_0)} = \bar{B}_r(x_0)$, im allgemeinen metrischen Fall muss nur $\overline{B_r(x_0)} \subset \bar{B}_r(x_0)$ gelten.)

Offenheit und Abgeschlossenheit bleiben bei gewissen Mengenoperationen erhalten:

Proposition (über Vereinigungen und Schnitte offener und abgeschlossener Mengen).
Seien \mathcal{X} ein metrischer Raum, O_i offene und A_i abgeschlossene Teilmengen von \mathcal{X} , $k \in \mathbb{N}$ und I eine beliebige Indexmenge. Dann gelten:

- (a) **Beliebige Vereinigungen** $\bigcup_{i \in I} O_i$ offener O_i sind **wieder offen** in \mathcal{X} .
 (b) **Endliche Durchschnitte** $\bigcap_{i=1}^k O_i$ offener O_i sind **wieder offen** in \mathcal{X} .
 (c) **Beliebige Durchschnitte** $\bigcap_{i \in I} A_i$ abgeschlossener A_i sind **wieder abgeschlossen** in \mathcal{X} .
 (d) **Endliche Vereinigungen** $\bigcup_{i=1}^k A_i$ abgeschlossener A_i sind **wieder abgeschlossen** in \mathcal{X} .

Man spricht auch davon, dass das System der offenen/abgeschlossenen Teilmengen von \mathcal{X} bezüglich beliebigen/endlichen Vereinigungen/Durchschnitten abgeschlossen ist.

Der kurze Beweis der Proposition ist Thema der Vorlesung.

Wir kommen nun zum richtigen Verständnis topologischer Konzepte in Teilmengen.

Bemerkung (topologische Konzepte in Teilmengen). Sind \mathcal{X} ein metrischer Raum und T eine Teilmenge von \mathcal{X} , so machen alle topologischen Konzepte genau wie „in \mathcal{X} “ auch „in T “ Sinn, denn T ist (mit der Einschränkung der Metrik) ja selbst ein metrischer Raum. Man erhält für Teilmengen und Punkte in T bei Betrachtung „in T “ aber teils andere Konzepte als bei Betrachtung „in \mathcal{X} “, zum Beispiel weil „in T “ die Kugeln nur aus Punkten von T bestehen und als Randpunkte nur Punkte von T zählen. Die Präzisierung, in welcher Grundmenge ein Konzept angewandt wird, kann daher wichtig sein. Um die Grundmenge T bei solchem Bedarf hervorzuheben, spricht man etwa vom **relativen Inneren** in T , vom **relativen Abschluss** in T , von **relativ offen** in T und von **relativ abgeschlossen** in T . Das Hinzusetzen von „relativ“ ändert dabei aber tatsächlich nichts an den Definitionen, sondern betont nur deren Anwendung in der Teilmenge T .

Beispiel (für eine **relativ offene Menge** und eine **relativ abgeschlossene Menge**). Das Liniensegment $L := (-7, 3) \times \{0\}$ ist ...

- weder offen noch abgeschlossen **in ganz** \mathbb{R}^2 mit Innerem $\overset{\circ}{L} = \emptyset$ sowie Abschluss und Rand $\bar{L} = \partial L = [-7, 3] \times \{0\}$,
- relativ offen **in** $\mathbb{R} \times \{0\}$ mit relativem Innerem $(-7, 3) \times \{0\}$, relativem Abschluss $[-7, 3] \times \{0\}$ und Rand $\{-7, 3\} \times \{0\}$,
- relativ abgeschlossen **in** $(-7, 3) \times \mathbb{R}$ mit relativem Innerem \emptyset , relativem Abschluss und Rand $(-7, 3) \times \{0\}$.

(Natürlich haben wir \mathbb{R}^2 und seine Teilmengen hier wieder mit der Standard-Metrik betrachtet.)

Proposition (Charakterisierung relativ offener und relativ abgeschlossener Mengen). Für einen metrischen Raum \mathcal{X} und Teilmengen $A \subset T \subset \mathcal{X}$ gelten:

- (i) Genau dann ist A relativ offen in T , wenn es eine in \mathcal{X} offene Menge E mit $A = E \cap T$ gibt.
- (ii) Genau dann ist A relativ abgeschlossen in T , wenn es eine in \mathcal{X} abgeschlossene Menge E mit $A = E \cap T$ gibt.

Der Beweis der Proposition ist Thema der Vorlesung. Die dort gegebene Argumentation basiert auf der Beobachtung $\overline{A}^T = \overline{A}^{\mathcal{X}} \cap T$ für die Abschlüsse \overline{A}^T und $\overline{A}^{\mathcal{X}}$ von $A \subset T$ in T und \mathcal{X} und führt zunächst zur Äquivalenz (ii). Die Äquivalenz (i) folgt dann durch Komplementbildung.

Schließlich führen wir auf Folgenkonvergenz basierende Varianten des Abschlusses und der Abgeschlossenheit ein:

Definitionen (Folgen-Abschluss, Folgen-Abgeschlossenheit). Seien \mathcal{X} ein metrischer Raum und $A \subset \mathcal{X}$. Der Folgen-Abschluss von A in \mathcal{X} ist definiert als

$$\left\{ \lim_{k \rightarrow \infty} x_k \mid (x_k)_{k \in \mathbb{N}} \text{ Folge in } A, \lim_{k \rightarrow \infty} x_k \text{ existiert in } \mathcal{X} \right\}.$$

Stimmt der Folgen-Abschluss von A in \mathcal{X} mit A selbst überein, so heißt die Menge A **Folgen-abgeschlossen (in \mathcal{X})**.

Tatsächlich wird sich aus der nächsten Proposition ergeben, dass diese Varianten der Konzepte (im hiesigen metrischen Kontext) mit den schon betrachteten übereinstimmen.

Proposition (Folgen-Charakterisierung von Berührungspunkten und Häufungspunkten). Für einen metrischen Raum \mathcal{X} , eine Teilmenge $A \subset \mathcal{X}$ und einen Punkt $x \in \mathcal{X}$ gelten:

- (i) Genau dann ist x ein Berührungspunkt von A , wenn eine Folge $(x_k)_{k \in \mathbb{N}}$ in A mit $\lim_{k \rightarrow \infty} x_k = x$ existiert.
- (ii) Genau dann ist x ein Häufungspunkt von A , wenn eine Folge $(x_k)_{k \in \mathbb{N}}$ in $A \setminus \{x\}$ mit $\lim_{k \rightarrow \infty} x_k = x$ existiert.

Der Beweis der Proposition ist Thema der Übungen.

Aus dem ersten Teil der Proposition ergibt sich wie angekündigt direkt:

Korollar. Für eine Menge A in einem metrischen Raum \mathcal{X} **stimmt der Folgen-Abschluss von A in \mathcal{X} mit dem Abschluss \overline{A} von A in \mathcal{X} überein**, und insbesondere ist A genau dann Folgen-abgeschlossen in \mathcal{X} , wenn A abgeschlossen in \mathcal{X} ist. \square

Gerade das im Korollar festgestellte Zusammenfallen der Konzepte rechtfertigt aber den häufig verwendeten Schluss

$$x_k \xrightarrow[k \rightarrow \infty]{} x \text{ in } \mathcal{X} \text{ und } x_k \in A \text{ für } k \gg 1 \implies x \in \overline{A}$$

(und sogar den Schluss auf $x \in A$ bei abgeschlossenem A). Zudem wird das Korollar in der Vorlesung zum Beweis der folgenden Korrespondenz verwendet:

Satz. Für einen *vollständigen* metrischen Raum \mathcal{X} und eine Teilmenge A von \mathcal{X} gilt

$$A \text{ abgeschlossen in } \mathcal{X} \iff A \text{ vollständig}$$

(wobei Vollständigkeit von A als eigener metrischer Raum gemeint ist).

Tatsächlich ist auch Stetigkeit von Funktionen ein rein topologischer Begriff, was sich an folgendem Sachverhalt besonders deutlich ablesen lässt.

Satz (über **Abbildungseigenschaften stetiger Funktionen**). Seien \mathcal{X} und \mathcal{Y} metrische Räume und $f: D \rightarrow \mathcal{Y}$ eine Funktion auf $D \subset \mathcal{X}$. Dann gilt:

$$\begin{aligned} f \text{ stetig auf } D &\iff f^{-1}(A) \text{ relativ offen in } D \text{ für alle offenen } A \subset \mathcal{Y} \\ &\iff f^{-1}(A) \text{ relativ abgeschlossen in } D \text{ für alle abgeschlossenen } A \subset \mathcal{Y} \end{aligned}$$

mit den **Urbildern** $f^{-1}(A) = \{x \in D \mid f(x) \in A\}$.

Bemerkungen (zu **Abbildungseigenschaften stetiger Funktionen**). Seien \mathcal{X} und \mathcal{Y} metrische Räume und $D \subset \mathcal{X}$.

(1) Die **vielleicht wichtigsten Konsequenzen des vorigen Satzes** sind folgende Prinzipien:

- **Durch** endliche viele **strikte Ungleichungen** mit stetigen $f_i, g_i: D \rightarrow \mathbb{R}$ **definierte Mengen** $\{x \in D \mid f_i(x) < g_i(x) \text{ für alle } i \in \{1, 2, \dots, k\}\}$ **sind stets offen** in D . Dies gilt auch, wenn „<“ für einige oder alle i durch „>“ oder „ \neq “ ersetzt wird.
- **Durch schwache Ungleichungen** mit stetigen $f_i, g_i: D \rightarrow \mathbb{R}$ **definierte Mengen** $\{x \in D \mid f_i(x) \leq g_i(x) \text{ für alle } i \in I\}$ mit beliebiger Indexmenge I **sind stets abgeschlossen** in D . Dies gilt auch, wenn „ \leq “ für einige oder alle i durch „ \geq “ oder „ $=$ “ ersetzt wird, insbesondere **auch bei Gleichungen**.

Konkret ist zum Beispiel das Dreieck $\{x \in \mathbb{R}^2 \mid x_1 < 0, x_2 > 0, x_2 - x_1 < 1\}$ offen in \mathbb{R}^2 , die Sphäre $S_r(x_0) := \{x \in \mathcal{X} \mid d_{\mathcal{X}}(x, x_0) = r\}$ mit $x_0 \in \mathcal{X}$, $r \in (0, \infty)$ stets abgeschlossen in \mathcal{X} .

Zur Begründung des ersten Prinzips schreibt man die dort betrachtete Menge in der Form $\bigcap_{i=1}^k (g_i - f_i)^{-1}((0, \infty))$. Gemäß dem Satz sind dann die Urbilder $(g_i - f_i)^{-1}((0, \infty))$ des offenen Intervalls $(0, \infty)$ unter den stetigen Funktionen $g_i - f_i$ wieder offen und nach der etwas früheren Proposition auch ihr endlicher Durchschnitt. Ähnlich erhält man das zweite Prinzip aus der Darstellung $\bigcap_{i \in I} (g_i - f_i)^{-1}([0, \infty))$ mit dem abgeschlossenen Intervall $[0, \infty)$. Bei anderen (Un-)Gleichheitszeichen kann man natürlich analog vorgehen.

- (2) Ist bei $f: D \rightarrow \mathcal{Y}$ für alle relativ offenen $A \subset D$ das Bild $f(A)$ offen in \mathcal{Y} , so nennt man f eine **offene Abbildung**. Ist dagegen für alle relativ abgeschlossenen $A \subset D$ das Bild $f(A)$ abgeschlossen in \mathcal{Y} , so heißt f eine **abgeschlossene Abbildung**. Für bijektive f folgt aus dem vorigen Satz, dass Offenheit und Abgeschlossenheit von f jeweils äquivalent zur Stetigkeit der Umkehrabbildung f^{-1} sind. Für allgemeine f besteht kein ähnlich einfacher Zusammenhang mit Stetigkeit.
- (3) Eine stetige Bijektion mit stetiger Umkehrabbildung (oder äquivalent eine stetige offene Bijektion) nennt man auch einen **Homöomorphismus**. Ein solcher erhält alle topologischen Eigenschaften, und seine Existenz bedeutet, dass der Definitions- und Zielbereich einander in topologischer Hinsicht gleichen.

Es folgen einige weitere topologische Begriffe:

Definition (dichte, nirgends dichte und diskrete Mengen). Sei \mathcal{X} ein metrischer Raum. Eine Teilmenge A von \mathcal{X} heißt ...

- **dicht** in \mathcal{X} , wenn $\overline{A} = \mathcal{X}$ gilt,
- **nirgends dicht** in \mathcal{X} , wenn $\overset{\circ}{\overline{A}} = \emptyset$ gilt (also ihr Abschluss keine inneren Punkte besitzt),
- **diskret** in \mathcal{X} , wenn A keine Häufungspunkte in \mathcal{X} hat (oder äquivalent, wenn alle Punkte in \overline{A} isolierte Punkte von \overline{A} sind).

Beispiele (für dichte, nirgends dichte und diskrete Mengen). Jeweils mit der Standardmetrik auf \mathbb{R}^n , \mathbb{C}^n , \mathbb{R} können wir sagen:

- (1) Als **absolute Standard-Beispiele dichter Mengen** sind \mathbb{Q}^n und $\mathbb{R}^n \setminus \mathbb{Q}^n$ beide **dicht** in \mathbb{R}^n .
- (2) \mathbb{Z}^n ist nirgends dicht und diskret in \mathbb{R}^n .
- (3) Drei Beispiele für nirgends dichte, aber nicht diskrete Mengen sind erstens die schon betrachtete Menge $S = \{\frac{1}{k} \mid k \in \mathbb{N}\}$ der Stammbrüche in \mathbb{R} , zweitens \mathbb{R}^n in \mathbb{C}^n und drittens die aus den Übungen zur Mathematik 3 bekannte Cantor-Menge der mittleren Drittel in \mathbb{R} . (Die letzteren beiden Mengen sind zudem auch abgeschlossen.)

Bemerkung. Im Normalfall (nämlich immer, wenn der umgebende Raum \mathcal{X} keine isolierten Punkte besitzt) ist jede in \mathcal{X} diskrete Menge auch nirgends dicht in \mathcal{X} .

Schließlich definieren wir in diesem Abschnitt noch zwei Zusammenhangsbegriffe, deren erster auf der sehr grundlegenden Idee des Verbindens durch stetige Wege/Kurven/Pfade basiert.

Definition (Wege/Kurven/Pfade). Sei \mathcal{X} ein metrischer Raum. Als **Weg**, **Kurve** oder **Pfad** in A bezeichnet man eine **stetige Abbildung** $c: I \rightarrow \mathcal{X}$ eines Intervalls I nach \mathcal{X} . Im Fall $I = [a, b]$ eines kompaktes Intervalls mit Randpunkten $a \leq b$ in \mathbb{R} heißen $c(a) \in \mathcal{X}$ und $c(b) \in \mathcal{X}$ die **Endpunkte** von c .

Definitionen (Zusammenhang, Wegzusammenhang, Zusammenhangskomponenten). Seien \mathcal{X} ein metrischer Raum und A eine Teilmenge von \mathcal{X} .

- (I) Man nennt A **wegzusammenhängend**, wenn es zu je zwei Punkten $x, y \in A$ stets einen Weg in A mit Endpunkten x und y gibt.
- (II) Man nennt A **zusammenhängend**, wenn für jede disjunkte Zerlegung $A = A_1 \dot{\cup} A_2$ von A in relativ offene Teilmengen A_1, A_2 von A stets $A_1 = \emptyset$ oder $A_2 = \emptyset$ gilt.
- (III) Die **(Weg-)Zusammenhangskomponenten** von A sind die (bezüglich Mengeneinklusion) maximalen (weg)zusammenhängenden Teilmengen von A .

Bemerkungen und Beispiele (zu den Zusammenhangsbegriffen).

- (1) Zur **Motivation des Zusammenhangsbegriffs** in Teil (II) der Definition kann man an den Fall denken, dass $A = A_1 \dot{\cup} A_2$ die disjunkte Vereinigung zweier Mengen $A_1 \neq \emptyset \neq A_2$ mit positivem Abstand $\text{dist}(A_1, A_2) > 0$ ist. Eine solche Menge A , die in zwei separate relativ offene Teilmengen A_1 und A_2 zerfällt, möchte man sicherlich *nicht* als zusammenhängend betrachten. Der Sinn der Definition ist nun in der Tat, derartige Zerlegungen (und auch etwas allgemeinere ohne positiven Abstand) bis auf die generell möglichen, trivialen Zerlegungen $A = A \dot{\cup} \emptyset$ und $A = \emptyset \dot{\cup} A$ auszuschließen.

- (2) Die Zusammenhangsbegriffe sind **intrinsische Begriffe**, das heißt, (Weg-)Zusammenhang und (Weg-)Zusammenhangskomponenten von A hängen nur von A selbst (als eigener metrischer Raum mit der auf A eingeschränkten Metrik) ab und nicht weiter vom umgebenden Raum \mathcal{X} . (Dies ist aus der Definitionen, die nur Wege in A und relativ offene Teilmengen von A benutzen, direkt klar.)
- (3) **Bei zusammenhängendem A sind \emptyset und A die einzigen Teilmengen von A , die gleichzeitig offen und abgeschlossen in A sind.** Diese Tatsache ist eine direkte Folge der Definition und ermöglicht es des Öfteren, eine Eigenschaft für alle Punkte einer zusammenhängenden Menge A mit einer eleganten Methode nachzuweisen: Man zeigt, dass die Punkte mit der betreffenden Eigenschaft eine nicht-leere, sowohl offene als auch abgeschlossene Teilmenge bilden, um dann schließen zu können, dass diese Teilmenge zwingend ganz A ist.
- (4) Aus den Ordnungseigenschaften von \mathbb{R} und dem Zwischenwertsatz folgt, dass die Eigenschaften Zusammenhang und Wegzusammenhang **für Teilmengen von \mathbb{R}** (mit Standard-Metrik) äquivalent und beide genau für Intervalle erfüllt sind. **Mit anderen Worten ist ein Intervall nichts anderes als eine zusammenhängende Teilmenge von \mathbb{R} .**
- (5) Allgemein ist jede wegzusammenhängende Menge zusammenhängend. Dass diese beiden Begriffe anders als in \mathbb{R} schon in \mathbb{R}^2 nicht mehr äquivalent ist, zeigt das Beispiel der zusammenhängenden, nicht wegzusammenhängenden Menge $\{(x, \sin \frac{1}{x}) \mid x \in (0, \infty)\} \cup \{(0, 0)\}$ in \mathbb{R}^2 .
- (6) Für offene Teilmengen von \mathbb{K}^N sind Zusammenhang und Wegzusammenhang aber doch wieder äquivalent, und solche Mengen besitzen höchstens abzählbar viele (Weg-)Zusammenhangskomponenten, die ebenfalls offen in \mathbb{K}^N sind.
- (7) Die Cantor-Menge der mittleren Drittel C dagegen ist ein Beispiel einer (abgeschlossenen, nicht offenen) Teilmenge von \mathbb{R} mit überabzählbar vielen (Weg-)Zusammenhangskomponenten, nämlich allen ihren ein-elementigen Teilmengen $\{x\}$ mit $x \in C$.

Auch die Beweise der Behauptungen in (4)–(7) sind elementar, auf die Ausführung von Details hierzu verzichten wir aber.

Als wichtige Eigenschaft des Zusammenhangs heben wir abschließend hervor:

Proposition („stetiges Bild erhält Zusammenhang“). Für metrische Räume \mathcal{X} , \mathcal{Y} und eine Abbildung $f: A \rightarrow \mathcal{Y}$ von $A \subset \mathcal{X}$ nach \mathcal{Y} gelten:

- (i) Ist A wegzusammenhängend und f stetig auf A , so ist $f(A)$ ebenfalls wegzusammenhängend.
- (ii) Ist A zusammenhängend und f stetig auf A , so ist $f(A)$ ebenfalls zusammenhängend.

Bemerkung. Betrachtet man Intervalle im Licht der vorigen Bemerkung (4) als (weg)zusammenhängende Teilmengen von \mathbb{R} , so verallgemeinert die Proposition die durch den Zwischenwertsatz garantierte Erhaltung der Intervall-Eigenschaft bei stetigem Bild auf Fälle mit viel allgemeinerem Definitions- und Zielbereich.

Der Beweis der Proposition anhand der Definitionen wird in der Vorlesung geführt.

1.3 Kompaktheit und Gleichmäßigkeit

Einordnung (Kompaktheit). Ein weiteres, **sehr grundlegendes topologisches Konzept** ist die **Kompaktheit von Mengen**. Kompaktheit ist zum einen **entscheidend** für Existenzresultate (an ganz verschiedenen Stellen in der Mathematik), in dieser Vorlesung vor allem **für Existenz von Extrema** und einen **allgemeinen Extremalsatz**. Zum anderen erlaubt Kompaktheit verschiedene **Lokal-Global-Schlüsse** und ermöglicht **Gleichmäßigkeitsaussagen**, wozu wir aber erst gegen Ende dieses Abschnitts kommen.

Im anfänglichen Verlauf dieses Abschnitts werden wir nun ...

- zwei verschiedene Definitionen von Kompaktheit einführen,
- die allgemeine Äquivalenz der Definitionen sowie **in \mathbb{K}^n** die wichtige Charakterisierung

$$A \text{ kompakt} \iff A \text{ abgeschlossen und beschränkt} \quad \text{für } A \subset \mathbb{K}^n$$

beweisen

- und uns dann dem Extremalsatz zuwenden.

Wie angekündigt beginnen wir mit den Definitionen.

Definitionen (Kompaktheit). Seien \mathcal{X} ein metrischer Raum und $A \subset \mathcal{X}$.

- (I) Eine Familie $(U_i)_{i \in I}$ von Teilmengen $U_i \subset \mathcal{X}$ mit Indexmenge $I \neq \emptyset$ nennt man eine **Überdeckung** von A , wenn $A \subset \bigcup_{i \in I} U_i$ gilt. Sind bei einer Überdeckung $(U_i)_{i \in I}$ von A alle U_i mit $i \in I$ offen in \mathcal{X} , so spricht man von einer **offenen Überdeckung** von A .

Die Menge A heißt **(Überdeckungs-)kompakt**, wenn es für jede (!) offene Überdeckung $(U_i)_{i \in I}$ von A ein $k \in \mathbb{N}$ und $i_1, i_2, \dots, i_k \in I$ mit $A \subset U_{i_1} \cup U_{i_2} \cup \dots \cup U_{i_k}$ gibt. Man beschreibt diesen Sachverhalt auch so, dass jede offene Überdeckung $(U_i)_{i \in I}$ von A eine **endliche Teilüberdeckung** $(U_{i_j})_{j \in \{1, 2, \dots, k\}}$ enthält.

- (II) Die Menge A heißt **Folgen-kompakt**, wenn jede Folge in A eine gegen einen Grenzwert in A konvergente Teilfolge besitzt.
- (III) Die Menge A heißt **relativ kompakt** beziehungsweise **relativ Folgen-kompakt** in \mathcal{X} , wenn ihr in \mathcal{X} gebildeter Abschluss \bar{A} kompakt beziehungsweise Folgen-kompakt ist.

Bemerkungen (zu Kompaktheit). Seien \mathcal{X} ein metrischer Raum und $A \subset \mathcal{X}$.

- (0) Für den Moment interpretieren wir Kompaktheit ohne weiteren Zusatz als Überdeckungs-Kompaktheit. Demnächst wird sich die Unterscheidung zwischen Überdeckungs- und Folgen-Kompaktheit wegen der (dann bewiesenen) Äquivalenz dieser Begriffe sowieso erledigen.
- (1) Überdeckungs- und Folgen-Kompaktheit von A sind **intrinsische Eigenschaften**, hängen also nur von A selbst (als eigener metrischer Raum) ab und nicht vom umgebenden Raum \mathcal{X} . Im Fall der Folgen-Kompaktheit ist dies aus der Definition, bei der ja nur Folgen *in* A auftreten, offensichtlich. Im Fall der Überdeckungs-Kompaktheit ergibt es sich aus der Beobachtung, dass die Definition äquivalent mit Überdeckungen $(O_i)_{i \in I}$ von A durch in A relativ offene Teilmengen $O_i \subset A$ (anstelle der in \mathcal{X} offenen $U_i \subset \mathcal{X}$) formuliert werden kann. Dass man einerseits von den U_i zu $O_i := U_i \cap A$ übergehen, andererseits die O_i stets als $O_i = U_i \cap A$ schreiben kann, garantiert dabei die in Abschnitt 1.2 bewiesene Charakterisierung relativ offener Mengen.

- (2) **Wichtig** ist sowohl bei der Definition der Überdeckungs- als auch der Folgen-Kompaktheit der **All-Quantor**, der eben **für jede offene Überdeckung** von A bzw. **für jede Folge** in A die Existenz einer endlichen Teilüberdeckung bzw. einer konvergenten Teilfolge fordert.
- (3) Ein **erstes Bild von Kompaktheit** kann man aus der **Frage** gewinnen, **wie Kompaktheit überhaupt scheitern kann**. Eine Antwort auf diese Frage ist, dass schon die Existenz eines Punktes $x_0 \in \bar{A} \setminus A$ ausreicht:

Zu einem solchen Punkt $x_0 \in \bar{A} \setminus A$ gibt es nämlich eine Folge $(x_k)_{k \in \mathbb{N}}$ in A mit $\lim_{k \rightarrow \infty} x_k = x_0$ in \mathcal{X} , und eine solche Folge besitzt keine in A konvergente Teilfolge (denn der einzig mögliche Grenzwert x_0 liegt nicht in A). Also ist A in dieser Situation nicht Folgen-kompakt.

Auch bedeutet $x_0 \in \bar{A} \setminus A$, dass mit $A \subset \mathcal{X} \setminus \{x_0\} = \bigcup_{i=1}^{\infty} (\mathcal{X} \setminus \bar{B}_{1/i}(x_0))$ eine offene Überdeckung $(\mathcal{X} \setminus \bar{B}_{1/i}(x_0))_{i \in \mathbb{N}}$ von A vorliegt, diese wegen $\bar{B}_{1/i}(x_0) \cap A \neq \emptyset$ für alle $i \in \mathbb{N}$ aber keine endliche Teilüberdeckung enthält. Also ist A auch nicht Überdeckungs-kompakt.

Insgesamt ist mit diesen Argumenten gezeigt, dass A im Fall $\bar{A} \not\subset A$ weder Folgen- noch Überdeckungs-kompakt ist. Durch Kontraposition ergibt sich daraus **grundlegende Beobachtung**, gültig für beide Kompaktheitsbegriffe,

$$\boxed{A \text{ kompakt} \implies A \text{ abgeschlossen in } \mathcal{X}}.$$

- (4) Weitere **Grundeigenschaften kompakter Mengen**, ebenfalls gültig für beide Kompaktheitsbegriffe, sind:
- (a) **Jede endliche Menge $E \subset \mathcal{X}$ ist kompakt.**
Dies wird in den Lernwerkstätten gezeigt (sowohl mit Überdeckungen als auch mit Folgen).
- (b) **Für kompakte Mengen $H, K \subset \mathcal{X}$ ist auch die Vereinigung $H \cup K$ kompakt.**
Dies wird in den Übungen bewiesen (sowohl mit Überdeckungen als auch mit Folgen).
- (c) **Für eine kompakte Menge $K \subset \mathcal{X}$ ist auch jede in K abgeschlossene Teilmenge $A \subset K$ kompakt.**
Dies wird in der Vorlesung bewiesen (sowohl mit Überdeckungen als auch mit Folgen).
- (d) Für metrische Räume \mathcal{X}_1 und \mathcal{X}_2 und die Produktmetrik $d_{\mathcal{X}_1 \times \mathcal{X}_2}$ auf $\mathcal{X}_1 \times \mathcal{X}_2$, erklärt durch $d_{\mathcal{X}_1 \times \mathcal{X}_2}((x_1, x_2), (y_1, y_2)) := d_{\mathcal{X}_1}(x_1, y_1) + d_{\mathcal{X}_2}(x_2, y_2)$ für $(x_1, x_2), (y_1, y_2) \in \mathcal{X}_1 \times \mathcal{X}_2$:
Für kompakte Mengen $K_1 \subset \mathcal{X}_1$ und $K_2 \subset \mathcal{X}_2$ ist auch das kartesische Produkt $K_1 \times K_2$ kompakt.
Für Folgen-Kompaktheit wird der Beweis dieser Eigenschaft in den Übungen behandelt. Ein direkter Beweis mit Überdeckungen ist möglich, aber etwas komplizierter, so dass man sich hierfür besser auf die demnächst angegangene Äquivalenz der Begriffe beruft.

Mit dem nun folgenden Hauptsatz erhalten wir wie angekündigt die allgemeine Äquivalenz der beiden Kompaktheitsbegriffe und die wichtige Charakterisierung kompakter Mengen in \mathbb{K}^n als abgeschlossene und beschränkte Mengen.

Hauptsatz (zur Charakterisierung kompakter Mengen).

- (I) *Für einen metrischen Raum \mathcal{X} und $A \subset \mathcal{X}$ gilt:*

$$\boxed{A \text{ kompakt} \iff A \text{ Folgen-kompakt} \implies A \text{ abgeschlossen in } \mathcal{X} \text{ und beschränkt}}.$$

(II) In jedem **endlich-dimensionalen** normierten Raum \mathcal{X} , insbesondere in $\mathcal{X} = \mathbb{K}^n$, gelten auch die folgenden, noch wichtigeren Umkehrungen:

- **Satz von Heine-Borel:** Jede abgeschlossene und beschränkte Teilmenge von \mathcal{X} ist kompakt.
- **Satz von Bolzano-Weierstraß:** Jede abgeschlossene und beschränkte Teilmenge von \mathcal{X} ist Folgen-kompakt (d.h. jede Folge in einer abgeschlossenen und beschränkten Menge $A \subset \mathcal{X}$ besitzt eine in A konvergente Teilfolge).

Der nicht ganz einfache Beweis der Hauptsatzes ist Thema der Vorlesung.

Insgesamt brauchen wir nach dem Hauptsatz **nicht mehr zwischen Kompaktheit und Folgen-Kompaktheit unterscheiden** und haben **kompakte Mengen in endlich-dimensionalen Räumen wie \mathbb{K}^n vollständig charakterisiert als abgeschlossene und beschränkte Mengen**. Speziell sind die schon in der Mathematik 1 als kompakt benannten Intervalle $[a, b]$ mit $-\infty < a \leq b < \infty$ in der Tat kompakte Mengen.

Bemerkungen (zu den Sätzen von Heine-Borel und Bolzano-Weierstraß).

- (1) Die in Teil (II) des Hauptsatzes formulierten **Sätze von Heine-Borel und Bolzano-Weierstraß** gelten tatsächlich in **keinem ∞ -dimensionalen normierten Raum \mathcal{X}** . Dies allgemein zu beweisen, übersteigt hier den Rahmen und gehört eher in eine Vorlesung über Funktionalanalysis. Zumindes für $\mathcal{X} = \ell^p$ mit $p \in [1, \infty]$ erkennt man es aber direkt daran, dass die Folge $(e_k)_{k \in \mathbb{N}}$ in ℓ^p mit $e_k := (0, 0, \dots, 0, 1, 0, 0, \dots) \in \ell^p$ (Eintrag 1 an k -ter Stelle) keine in ℓ^p konvergente Teilfolge besitzt.
- (2) Eine Abschwächung der Sätze von Heine-Borel und Bolzano-Weierstraß, bei der Beschränktheit durch eine etwas andere Bedingung, die sogenannte Total-Beschränktheit, ersetzt wird, bleibt aber tatsächlich in jedem metrischen Raum gültig.

Bevor wir nun den für diese Vorlesung entscheidenden Extremalsatz auf kompakten Mengen behandeln, halten wir vorbereitend eine weitere Abbildungseigenschaft stetiger Funktionen fest.

Satz (zum Prinzip „stetiges Bild erhält Kompaktheit“). Seien \mathcal{X}, \mathcal{Y} metrische Räume. Ist K kompakte Teilmenge von \mathcal{X} und $f: K \rightarrow \mathcal{Y}$ stetig auf K , so ist auch $f(K)$ kompakt.

Zwei Beweise, einer mit Überdeckungen, einer mit Folgen, werden in der Vorlesung geführt.

Die für den Moment entscheidende Konsequenz der Kompaktheit folgt.

Hauptsatz (allgemeiner Extremalsatz). Sei \mathcal{X} ein metrischer Raum. Ist K eine nicht-leere, **kompakte** Teilmenge von \mathcal{X} und ist $f: K \rightarrow \mathbb{R}$ **stetig** auf K , so **existieren eine Minimalstelle und eine Maximalstelle** von f auf K , d.h. es existieren $x_* \in K$ und $x^* \in K$ mit $f(x_*) \leq f(x) \leq f(x^*)$ für alle $x \in K$.

Die Existenz einer Minimal-/Maximalstelle ist dabei wie üblich als Existenz *mindestens* einer Minimal-/Maximalstelle zu verstehen, und es geht es um *globale* Minimal-/Maximalstelle auf ganz K in dem im Satz angegebenen Sinn — nicht zu verwechseln mit *lokalen* Minimal-/Maximalstellen, zu denen wir im nächsten Kapitel noch einmal kommen.

Die **zentrale Bedeutung des Extremalsatzes** für uns liegt darin, mit diesem Satz (oder seiner unten folgenden Erweiterung) **im ersten Schritt jeder Extremwertbestimmung überhaupt die Existenz von Extremstellen und Extremwerten zu sichern**. Genaueres zu diesem Vorgehen und Beispiele aus der Rechenpraxis folgen im nächsten Kapitel.

Beweis des allgemeinen Extremalsatzes. Nach dem vorigen Satz ist $f(K)$ kompakte Teilmenge von \mathbb{R} und nach dem Hauptsatz zur Charakterisierung kompakter Mengen somit abgeschlossen in \mathbb{R} und beschränkt. Es folgt, dass $\inf_K f = \inf(f(K))$ Element von $f(K)$ ist, dass also ein $x_* \in K$ mit $f(x_*) = \inf_K f$ existiert. Dieses x_* ist eine Minimalstelle. Analog begründet man die Existenz einer Maximalstelle x^* . \square

Alternativ zum gegebenen Beweis mit der Abbildungseigenschaft lässt sich der allgemeine Extremalsatz auch wie in der Mathematik 3 mit Minimal- und Maximalfolgen beweisen.

Anwendungen (des Extremalsatzes).

(1) Stetige Funktionen auf kompakten Mengen sind stets beschränkt.

Dies ist für \mathbb{R} -wertige Funktionen f aus obiger Formulierung des Extremalsatzes klar, denn offensichtlich sind $f(x_*)$ eine untere, $f(x^*)$ eine obere Schranke für f auf K . Allgemeiner gilt die Aussage aber auch für stetiges $f: K \rightarrow \mathcal{Y}$ auf kompaktem $K \subset \mathcal{X}$ mit beliebigen metrischen Räumen \mathcal{X} und \mathcal{Y} in Sinn von $\text{diam}(f(K)) < \infty$. Das folgt aus der Kompaktheit von $f(K)$ oder alternativ aus der Beschränktheit der \mathbb{R} -wertigen, stetigen Funktion $K \rightarrow \mathbb{R}$, $x \mapsto d_{\mathcal{Y}}(f(x), y_0)$ gemäß dem Vorigen.

(2) Ist K eine nicht-leere, kompakte Teilmenge eines metrischen Raums \mathcal{X} , so existieren stets Punkte $x_0, y_0 \in K$ mit $d_{\mathcal{X}}(x_0, y_0) = \text{diam } K$, und insbesondere ist $\text{diam } K < \infty$.

Ein Beweis hierfür wird in der Vorlesung angerissen und in den Übungen im Detail behandelt.

Bei fehlender Kompaktheit des Definitionsbereichs greift der bisher formulierte Extremalsatz nicht. Das nächste Resultat zeigt aber, dass man zumindest für Mengen A in \mathbb{K}^n **fehlende Kompaktheit durch geeignetes Funktionsverhalten „bei ∞ “ und/oder bei nicht zu A gehörigen Randpunkten kompensieren** kann.

Erweiterung (des Extremalsatzes). Ist A nicht-leere Teilmenge von \mathbb{K}^n und ist $f: A \rightarrow \mathbb{R}$ **stetig** auf A mit $\liminf_{A \ni x \rightarrow x_0} f(x) > \inf_A f$ für alle $x_0 \in (\partial A) \setminus A$ und bei unbeschränktem A zudem mit $\liminf_{\substack{|x| \rightarrow \infty \\ x \in A}} f(x) > \inf_A f$, so **existiert eine Minimalstelle** von f auf A .

(Dabei ist $\liminf_{A \ni x \rightarrow x_0} f(x) = \min\{\liminf_{k \rightarrow \infty} f(x_k) \mid (x_k)_{k \in \mathbb{N}} \text{ Folge in } A, \lim_{k \rightarrow \infty} x_k = x_0\} \in \overline{\mathbb{R}}$. Der analog erklärte Grenzübergang $|x| \rightarrow \infty$ lässt sich im Fall $A \subset \mathbb{R}$ auch in $x \rightarrow \infty$ und $x \rightarrow -\infty$ aufteilen.)

Analog existiert unter der Voraussetzung $\limsup_{A \ni x \rightarrow x_0} f(x) < \sup_A f$ für alle $x_0 \in (\partial A) \setminus A$ (bei unbeschränktem A zuzüglich $\limsup_{x \in A, |x| \rightarrow \infty} f(x) < \sup_A f$) stets eine Maximalstelle von f auf A , und **vor allem** sind diese Aussagen natürlich dann **anwendbar, wenn die Limes inferior bzw. superior von $f(x)$ gleich ∞ bzw. $-\infty$** (und damit übrigens Limes) sind, womit die benötigten Ungleichungen trivial gelten.

Zum Beweis der Erweiterung modifiziert man den Minimal-/Maximalfolgen-Beweis des Extremalsatzes. Näheres ist Thema der Übungen.

Beispiele (zur Erweiterung des Extremalsatzes).

- (1) Für die stetige Funktion $f: [1, \infty) \rightarrow \mathbb{R}$ mit $f(x) := \frac{1}{x}$ gilt $\lim_{x \rightarrow \infty} f(x) = 0 = \inf_{[1, \infty)} f < \sup_{[1, \infty)} f = 1$. Die Bedingung der Erweiterung für Minimalstellen ist also *nicht* erfüllt, die analoge für Maximalstellen dagegen schon, und tatsächlich besitzt diese Funktion f *keine* Minimalstelle auf $[1, \infty)$, aber die Maximalstelle 1 auf $[1, \infty)$. (In diesem Beispiel lässt sich aber die (Nicht-)Existenz natürlich auch elementar ohne das allgemeine Resultat klären. Wir illustrieren hiermit nur, dass und wieso die geforderte Bedingung nötig ist.)

- (2) Für die stetige Funktion $f: \mathbb{R} \rightarrow \mathbb{R}$ mit $f(x) := e^x - x^3$ zeigt separate Betrachtung von $x \rightarrow \infty$ und $x \rightarrow -\infty$, dass mit $\lim_{|x| \rightarrow \infty} f(x) = \infty > \inf_{\mathbb{R}} f$ die Bedingung der Erweiterung für Minimalstellen erfüllt ist (ohne, dass man $\inf_{\mathbb{R}} f$ hier kennt oder kennen müsste). Es folgt somit die Existenz einer Minimalstelle von f auf \mathbb{R} , und hier handelt es sich tatsächlich um eine typische Anwendung des allgemeinen Sachverhalts, denn explizit oder ganz elementar bekommt man eine Minimalstelle in diesem Fall nicht.

Anwendung (der Erweiterung des Extremsatzes). Sind K eine nicht-leere, kompakte und A eine nicht-leere, abgeschlossene Teilmenge von \mathbb{K}^n , so existieren stets zueinander nächste Punkte $x_0 \in K$ und $y_0 \in A$ mit $|y_0 - x_0| = \text{dist}(A, K)$. Insbesondere gilt im Fall $A \cap K = \emptyset$ somit $\text{dist}(A, K) > 0$ (und diese letzte Aussage bleibt mit anderer Begründung tatsächlich auch in einem beliebigen metrischen Raum statt \mathbb{K}^n richtig).

Die Herleitung dieser Aussagen ist Thema der Übungen.

Im letzten Teil dieses Abschnitts (und des Kapitels) **sammeln** wir nun **Aussagen mit/zu Kompaktheit, gleichmäßiger Stetigkeit, gleichmäßiger Konvergenz und Lokal-Global-Schlüssen**. Wir beginnen mit dem allgemeinen Konzept gleichmäßiger Stetigkeit.

Definition (gleichmäßige Stetigkeit). Seien \mathcal{X} und \mathcal{Y} metrische Räume sowie $D \subset \mathcal{X}$. Dann heißt eine Funktion $f: D \rightarrow \mathcal{Y}$ **gleichmäßig stetig** auf D , wenn für sie die Bedingung $\forall \varepsilon > 0: \exists \delta > 0: \forall x, \tilde{x} \in D: (d_{\mathcal{X}}(x, \tilde{x}) < \delta \implies d_{\mathcal{Y}}(f(x), f(\tilde{x})) < \varepsilon)$ erfüllt ist.

Bemerkungen (zu gleichmäßiger Stetigkeit). Seien \mathcal{X}, \mathcal{Y} metrische Räume und $f: D \rightarrow \mathcal{Y}$ eine Funktion auf $D \subset \mathcal{X}$.

- (1) **Entscheidend** ist, dass bei gleichmäßiger Stetigkeit anders als bei „normaler“ Stetigkeit das δ **nur von ε , aber nicht von x abhängen darf**. Dies ist in der Definition dadurch kodiert, dass der Quantor $\forall x$ erst nach dem Quantor $\exists \delta$ folgt.
- (2) **Lipschitz-stetige Funktionen** und allgemeiner auch **Hölder-stetige Funktionen sind stets gleichmäßig stetig**, denn aus der Abschätzung $d_{\mathcal{Y}}(f(x), f(\tilde{x})) \leq C d_{\mathcal{X}}(x, \tilde{x})^\alpha$ ergibt sich, dass man das für gleichmäßige Stetigkeit benötigte $\delta > 0$ zu einem $\varepsilon > 0$ im Fall $C > 0$ als $\delta = (\varepsilon/C)^{1/\alpha}$ (und im Fall $C = 0$ beliebig) wählen kann.
- (3) Tatsächlich ist gleichmäßige Stetigkeit von f auf D äquivalent zur Existenz eines Stetigkeitsmoduls für f , d.h. einer Funktion $\omega: [0, \infty) \rightarrow [0, \infty]$ mit $\lim_{t \searrow 0} \omega(t) = \omega(0) = 0$, so dass die Abschätzung $d_{\mathcal{Y}}(f(x), f(\tilde{x})) \leq \omega(d_{\mathcal{X}}(x, \tilde{x}))$ für alle $x, \tilde{x} \in D$ gilt.

Zum Beweis dieser Äquivalenz verifizieren wir die Hin- und die Rück-Implikation separat:

Besitzt f einen Stetigkeitsmodul ω , so gibt es wegen der Bedingung $\lim_{t \searrow 0} \omega(t) = \omega(0) = 0$ zu jedem $\varepsilon > 0$ ein $\delta > 0$, so dass für $t \in [0, \infty)$ aus $t < \delta$ schon $\omega(t) < \varepsilon$ folgt. Insgesamt ergibt sich für $x, \tilde{x} \in D$ aus $d_{\mathcal{X}}(x, \tilde{x}) < \delta$ daher $d_{\mathcal{Y}}(f(x), f(\tilde{x})) \leq \omega(d_{\mathcal{X}}(x, \tilde{x})) < \varepsilon$. Also ist f gleichmäßig stetig.

Ist f gleichmäßig stetig, so wählen wir beliebige positive $\varepsilon_1 \geq \varepsilon_2 \geq \varepsilon_3 \geq \dots$ mit $\lim_{k \rightarrow \infty} \varepsilon_k = 0$ und dazu sukzessive positive $\delta_1 \geq \delta_2 \geq \delta_3 \geq \dots$, so dass für jedes $k \in \mathbb{N}$ die Implikation aus der Definition gleichmäßiger Stetigkeit mit ε_k, δ_k anstelle ε, δ gilt. Wir setzen $\varepsilon_0 := \infty$ und definieren $\omega: [0, \infty) \rightarrow [0, \infty]$ durch $\omega(0) := 0$ und $\omega(t) := \sup\{\varepsilon_k \mid k \in \mathbb{N}_0, \delta_{k+1} \leq t\}$ für $t > 0$. Damit erhalten wir $\lim_{t \searrow 0} \omega(t) = 0$, denn für $\varepsilon > 0$ gibt es ein $k_* \in \mathbb{N}$ mit $\varepsilon_{k_*} < \varepsilon$, und für alle $t \in (0, \delta_{k_*})$ gilt wegen $\delta_1 \geq \delta_2 \geq \dots \geq \delta_{k_*} > t$ dann $\omega(t) \leq \varepsilon_{k_*} < \varepsilon$. Für $x, \tilde{x} \in D$ unterscheiden wir nun den trivialen Fall $d_{\mathcal{Y}}(f(x), f(\tilde{x})) = 0 \leq \omega(d_{\mathcal{X}}(x, \tilde{x}))$ und den Fall $d_{\mathcal{Y}}(f(x), f(\tilde{x})) > 0$. In letzterem gibt es ein $k_0 \in \mathbb{N}_0$ mit $\varepsilon_{k_0+1} \leq d_{\mathcal{Y}}(f(x), f(\tilde{x})) \leq \varepsilon_{k_0}$, woraus durch Kontraposition der definierenden Implikation $\delta_{k_0+1} \leq d_{\mathcal{X}}(x, \tilde{x})$ folgt. Nach Wahl von ω erhalten wir $\omega(d_{\mathcal{X}}(x, \tilde{x})) \geq \varepsilon_{k_0}$ und insgesamt $d_{\mathcal{Y}}(f(x), f(\tilde{x})) \leq \varepsilon_{k_0} \leq \omega(d_{\mathcal{X}}(x, \tilde{x}))$. Damit ist ω im benötigten Sinn ein Stetigkeitsmodul für f .

Die erste, gelegentlich nützliche Konsequenz aus gleichmäßiger Stetigkeit ist stetige Fortsetzbarkeit:

Satz (über stetige Fortsetzbarkeit). *Seien \mathcal{X} ein metrischer Raum, \mathcal{Y} ein vollständiger metrischer Raum sowie $D \subset \mathcal{X}$. Ist $f: D \rightarrow \mathcal{Y}$ **gleichmäßig** stetig auf D , so gibt es genau eine stetige Funktion $\bar{f}: \bar{D} \rightarrow \mathcal{Y}$ mit $\bar{f}|_D = f$, mit anderen Worten besitzt f dann genau eine stetige Fortsetzung auf den Abschluss \bar{D} von D in \mathcal{X} .*

Der Beweis des Satzes wird in der Vorlesung ausgeführt und nutzt entscheidend die Vollständigkeit von \mathcal{Y} , um die Funktionswerte von \bar{f} auf \bar{D} als Grenzwerte der Funktionswerte von f auf D zu erhalten.

In teilweiser Umkehrung der vorausgehenden Satzes ist gleichmäßige Stetigkeit für Funktionen auf einem kompakten (insbesondere abgeschlossenen) Definitionsbereich generell gesichert:

Satz (gleichmäßige Stetigkeit auf Kompakta). *Seien \mathcal{X} und \mathcal{Y} metrische Räume. Ist K eine **kompakte** Teilmenge von \mathcal{X} und ist $f: K \rightarrow \mathcal{Y}$ **stetig** auf K , so ist f schon **gleichmäßig stetig** auf K .*

Ein analoger Sachverhalt für den Fall $\mathcal{X} = \mathcal{Y} = \mathbb{R}$ wurde bereits in der Mathematik 3 durch ein typisches indirektes Gleichmäßigkeits-Argument mit einer „Folge von Gegenbeispielen“ bewiesen. Diese Argumentation kann mit naheliegenden Anpassungen übertragen werden und wurde auch in der aktuellen Vorlesung noch einmal ausgeführt.

Die nächsten beiden Proposition liefern weitere Beispiele für **typische Gleichmäßigkeitsaussagen**, die sich **auf kompakten Mengen** ergeben.

Proposition. *Seien \mathcal{X} und \mathcal{Y} metrische Räume. Ist K eine **kompakte** Teilmenge von \mathcal{X} und ist $f: K \rightarrow \mathcal{Y}$ **lokal beschränkt** auf K (d.h. für jedes $x \in K$ gibt es ein $r > 0$, so dass f auf $B_r(x) \cap K$ beschränkt ist), so ist f schon **beschränkt** auf ganz K .*

Der Beweis erfolgt durch einen Lokal-Global-Schluss, entweder direkt mit Überdeckungen oder indirekt mit einer „Folge von Gegenbeispielen“. Beide Varianten werden in den Übungen behandelt.

Definition (gleichmäßige Konvergenz). *Seien D eine Menge, \mathcal{Y} ein metrischer Raum sowie sowie $f_k, f: D \rightarrow \mathcal{Y}$ Funktionen auf D . Dann sagen wir, dass f_k bei $k \rightarrow \infty$ **gleichmäßig** auf D gegen f **konvergieren**, wenn $\lim_{k \rightarrow \infty} \sup_{x \in D} d_{\mathcal{Y}}(f_k(x), f(x)) = 0$ gilt.*

Proposition. *Seien \mathcal{X} und \mathcal{Y} metrische Räume sowie $f_k, f: K \rightarrow \mathcal{Y}$ Funktionen auf $K \subset \mathcal{X}$. Ist K **kompakt** und konvergieren f_k bei $k \rightarrow \infty$ **lokal gleichmäßig** auf K gegen f (d.h. für jedes $x \in K$ gibt es ein $r > 0$, so dass f_k bei $k \rightarrow \infty$ gleichmäßig auf $B_r(x) \cap K$ gegen f konvergieren), so konvergieren f_k bei $k \rightarrow \infty$ schon **gleichmäßig** auf ganz K gegen f .*

Auch hier kann der Beweis durch einen Lokal-Global-Schluss direkt mit Überdeckungen oder indirekt mit einer „Folge von Gegenbeispielen“ geführt werden. Die Überdeckungs-Variante der Argumentation wurde in der Vorlesung ausgeführt.

Satz (Stetigkeit der Grenzfunktion bei gleichmäßiger Konvergenz). *Seien \mathcal{X} und \mathcal{Y} metrische Räume sowie $f_k, f: D \rightarrow \mathcal{Y}$ Funktionen auf $D \subset \mathcal{X}$. Ist f_k für jedes $k \in \mathbb{N}$ **stetig** auf D und **konvergieren** f_k bei $k \rightarrow \infty$ **gleichmäßig** auf D gegen f , so ist auch die Grenzfunktion f **stetig** auf D .*

Der Beweis des Satzes erfolgt durch naheliegende Anpassung des in der Mathematik 3 für den Fall $\mathcal{X} = \mathcal{Y} = \mathbb{R}$ behandelten $\frac{\epsilon}{3}$ -Arguments.

Kapitel 2

Differentialrechnung mehrerer Veränderlicher

2.1 Ableitungsbegriffe

Wir führen nun verschiedene Ableitungsbegriffe für Funktionen mehrerer reeller Variablen ein. Teils sind diese sogar für Funktionen zwischen normierten Räumen sinnvoll. Im Vordergrund stehen aber stets Definitionsbereiche in \mathbb{R}^n und das Ziel \mathbb{R}^m und somit Funktionen von $n \in \mathbb{N}$ einzelnen reellen Variablen und mit $m \in \mathbb{N}$ reell-wertigem Komponentenfunktionen.

Wir beginnen mit dem Ableiten einer Funktion mehrerer Variablen in der naheliegenden Weise, dass wir alle anderen Variablen festhalten und nach einer ausgewählten Variablen „ganz normal“ ableiten. Im Kontext mehrerer Variablen führt man hierfür verschiedene Schreibweisen mit dem Symbol ∂ ein (welches teils die Rolle des bei einer Variablen üblicheren d übernimmt):

Definition (partielle Ableitungen, Funktionalmatrizen). Seien $f: A \rightarrow \mathcal{Y}$ eine Funktion von $A \subset \mathbb{R}^n$ in einen normierten Raum \mathcal{Y} und $a = (a_1, a_2, \dots, a_n) \in A$.

- (I) Für ein $j \in \{1, 2, \dots, n\}$ heißt f an der Stelle a /in a /bei a **nach der j -ten Variablen differenzierbar**, falls folgende Ableitung in der Einzelvariable $s \in \mathbb{R}$ bzw. $t \in \mathbb{R}$ existiert:

$$\begin{aligned}\partial_j f(a) &:= (f(a_1, \dots, a_{j-1}, \cdot, a_{j+1}, \dots, a_n))'(a_j) = \left. \frac{d}{ds} \right|_{s=a_j} f(a_1, \dots, a_{j-1}, s, a_{j+1}, \dots, a_n) \\ &= \left. \frac{d}{dt} \right|_{t=0} f(a+te_j) = \lim_{\substack{\mathbb{R} \ni h \rightarrow 0 \\ h \neq 0}} \frac{f(a+he_j) - f(a)}{h} \in \mathcal{Y}.\end{aligned}$$

Hier ist e_j der j -te kanonische Basisvektor in \mathbb{R}^n , $a+te_j = (a_1, \dots, a_{j-1}, a_j+t, a_{j+1}, \dots, a_n)$ genau wie $a+he_j = (a_1, \dots, a_{j-1}, a_j+h, a_{j+1}, \dots, a_n)$, und der Limes der Differenzenquotienten wird im normierten Raum \mathcal{Y} gebildet. Wenn $\partial_j f(a)$ existiert, nennt man $\partial_j f(a)$ die **j -te partielle Ableitung** von f an der Stelle a /in a /bei a .

Gleichbedeutend mit $\partial_j f(a)$ notiert man, vor allem, wenn f als $f(x_1, x_2, \dots, x_n)$ geschrieben wird, $\frac{\partial}{\partial x_j} f(a)$, $\frac{\partial f}{\partial x_j}(a)$ und $\partial_{x_j} f(a)$. Bei anderer Benennung der Variablen, zum Beispiel als (x, y, z) im Fall $n = 3$, verwendet man analoge Notationen wie etwa $\frac{\partial}{\partial x}$, $\frac{\partial}{\partial y}$, $\frac{\partial}{\partial z}$.

(II) Ist f im \mathbb{R}^m -wertigen Fall $\mathcal{Y} = \mathbb{R}^m$ bei a nach allen n Variablen differenzierbar, so heißt die $(m \times n)$ -Matrix

$$Df(a) := \left(\partial_1 f(a) \mid \partial_2 f(a) \mid \cdots \mid \partial_{n-1} f(a) \mid \partial_n f(a) \right)$$

$$= \begin{pmatrix} \partial_1 f_1(a) & \partial_2 f_1(a) & \cdots & \partial_{n-1} f_1(a) & \partial_n f_1(a) \\ \partial_1 f_2(a) & \partial_2 f_2(a) & \cdots & \partial_{n-1} f_2(a) & \partial_n f_2(a) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \partial_1 f_{m-1}(a) & \partial_2 f_{m-1}(a) & \cdots & \partial_{n-1} f_{m-1}(a) & \partial_n f_{m-1}(a) \\ \partial_1 f_m(a) & \partial_2 f_m(a) & \cdots & \partial_{n-1} f_m(a) & \partial_n f_m(a) \end{pmatrix} \in \mathbb{R}^{m \times n}$$

die **Funktionalmatrix** oder **Jacobi-Matrix** von f an der Stelle a /in a /bei a .

Weitere Erläuterungen zu partiellen Ableitungen und Funktionalmatrizen stellen wir etwas zurück, werden darauf aber schon bald in etwas allgemeineren Kontext zurückkommen.

Bemerkung (zur Ableitung von Kurven). Für $n = 1$ und ein Intervall $A = I \subset \mathbb{R}$ reduziert sich die Definition auf die Ableitung einer Kurve $c: I \rightarrow \mathcal{Y}$. In diesem Fall ist die Ableitung $c'(a) := \partial_1 c(a) \in \mathcal{Y}$ in $a \in \overset{\circ}{I}$ nach der einzig vorhandenen (ersten) Variablen der bekannte Limes $\lim_{\substack{h \rightarrow 0 \\ h \neq 0}} \frac{c(a+h) - c(a)}{h}$ der Differenzenquotienten. Gegenüber der Mathematik 3 ist jetzt nur die Möglichkeit hinzugekommen, die Konvergenz der Differenzenquotienten und das Ableitungskonzept auch für Werte in $\mathcal{Y} = \mathbb{R}^m$ oder gar einem normierten Raum \mathcal{Y} zu erklären. **Geometrisch** ist $c'(a) \in \mathcal{Y}$ als **Tangentialvektor an die (Spur der) Kurve c** im Punkt $c(a) \in \mathcal{Y}$ zu interpretieren (dessen Länge $|c'(a)|$ einer Durchlaufgeschwindigkeit entspricht).

Fasst man die j -te partielle Ableitung als Ableitung in Richtung des Vektors e_j auf, so liegt nahe, dass man wie in die Richtung e_j auch in eine ganz beliebige Richtung ableiten kann:

Definition (Richtungsableitungen). Seien $f: A \rightarrow \mathcal{Y}$ eine Funktion von einer Teilmenge A eines normierten Raums \mathcal{X} in einen weiteren normierten Raum \mathcal{Y} und $a \in \overset{\circ}{A}$. Für $v \in \mathcal{X}$ heißt f an der Stelle a /in a /bei a **in Richtung v differenzierbar**, falls folgende Ableitung in der Einzelvariable $t \in \mathbb{R}$ existiert:

$$\partial_v f(a) := \frac{d}{dt} \Big|_{t=0} f(a+tv) = \lim_{\substack{\mathbb{R} \ni h \rightarrow 0 \\ h \neq 0}} \frac{f(a+hv) - f(a)}{h} \in \mathcal{Y}.$$

Wenn $\partial_v f(a)$ existiert, nennt man $\partial_v f(a)$ die **Richtungsableitung** von f an der Stelle a /in a /bei a **in Richtung v** .

Erweiterung (Richtungsableiten entlang Vektorfeldern). Für $f, A, \mathcal{X}, \mathcal{Y}, a$ wie in der vorigen Definition versteht man unter der Richtungsableitung von f an der Stelle a entlang eines Vektorfelds¹ $V: A \rightarrow \mathcal{X}$ einfach die Richtungsableitung

$$\partial_V f(a) := \partial_{V(a)} f(a) \in \mathcal{Y},$$

¹Als Vektorfelder bezeichnet man Abbildungen V von einer Teilmenge von \mathbb{R}^n in denselben \mathbb{R}^n oder allgemeiner von einer Teilmenge eines normierten Raums in denselben normierten Raum. Man stellt sich vor, dass an jeden Punkt x des Definitionsbereichs der Vektor $V(x)$ angeklebt wird.

sofern diese existiert. Häufiger betrachtete Spezialfälle sind die Ableitung $\partial_X f$ nach dem Ortsvektorfeld X mit $X(x) := x$ und die **radiale Ableitung** $\partial_{\text{rad}} f := \partial_R f$ nach dem radialen Einheits-Vektorfeld R auf $A \subset \mathcal{X} \setminus \{0\}$ mit $R(x) := \frac{x}{\|x\|_{\mathcal{X}}}$.

Bemerkungen (zu Richtungsableitungen und partiellen Ableitungen).

- (0) **Partielle Ableitungen sind spezielle Richtungsableitungen.** Genauer ist für Definitionsbereiche in $\mathcal{X} = \mathbb{R}^n$ und $j \in \{1, 2, \dots, n\}$ die partielle Ableitung $\partial_j f(a)$ nach der j -ten Variablen nicht anderes als die Richtungsableitung $\partial_{e_j} f(a)$ in Richtung $e_j \in \mathbb{R}^n$. Deshalb gilt alles im Folgenden für Richtungsableitungen Gesagte genauso für partielle Ableitungen.
- (1) Eine **Richtungsableitung** $\partial_v f$ von $f: A \rightarrow \mathcal{Y}$ wird oft nicht nur an einer Stelle a betrachtet, sondern **selbst als Funktion** $\partial_v f: A \rightarrow \mathcal{Y}$ **aufgefasst**, falls sie (z.B.) an allen Stellen eines offenen Definitionsbereichs A existiert. Dabei hat die Richtungsableitung $\partial_v f$ genau wie f selbst Werte in \mathcal{Y} , **Richtungsableiten ändert also den Zielraum nicht.** Für $\mathcal{Y} = \mathbb{R}^m$ hat die Richtungsableitung eines \mathbb{R}^m -wertigen $f = \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_m \end{pmatrix}$ die Form $\partial_v f = \begin{pmatrix} \partial_v f_1 \\ \partial_v f_2 \\ \vdots \\ \partial_v f_m \end{pmatrix}$, sie darf also komponentenweise berechnet werden.
- (2) Für \mathbb{R} -wertiges $f: A \rightarrow \mathbb{R}$ auf $A \subset \mathbb{R}^n$ entspricht die **Richtungsableitung** $\partial_v f(a)$ mit einem Einheitsvektor $v \in \mathbb{R}^n$, $|v| = 1$ **geometrisch der Steigung des Graphen von f in Richtung v .** In der Vorlesung wurde dies für den Fall $n = 2$, in dem man sich den Graph von f als Fläche in \mathbb{R}^3 vorstellen kann, bildlich veranschaulicht.
- (3) **Ableitungsregeln für Richtungsableitungen** erhält man bei Werten in \mathbb{R} aus den entsprechenden Regeln der Mathematik 3, bei Werten in \mathbb{R}^m dann aus Bemerkung (1) und allgemein für Werte in einem normierten Raum durch naheliegendes Umschreiben der von früher bekannten Beweise. Konkret gelten (für Funktionen f, g auf dem gleichem Definitionsbereich und mit für die jeweilige Regel geeigneten Werten):

- **Summen- und Faktorregel** $\partial_v(sf + tg) = s\partial_v f + t\partial_v g$ für $s, t \in \mathbb{K}$ (womit Richtungsableiten eine \mathbb{K} -lineare Operation ist),
- **Produktregel** $\partial_v(f \cdot g) = (\partial_v f) \cdot g + f \cdot (\partial_v g)$ (wobei \cdot die Multiplikation reeller oder komplexer Zahlen, die Skalarmultiplikation einer Zahl mit einem Vektor, das Skalarprodukt zweier Vektoren oder allgemein eine stetige bilineare Abbildung $\cdot: \mathcal{Y} \times \tilde{\mathcal{Y}} \rightarrow \mathcal{Z}$ zwischen normierten Räumen $\mathcal{Y}, \tilde{\mathcal{Y}}, \mathcal{Z}$ sein kann),
- **Quotientenregel** $\partial_v\left(\frac{f}{g}\right) = \frac{(\partial_v f)g - f(\partial_v g)}{g^2}$ (wenn g Werte in $\mathbb{K} \setminus \{0\}$ und f Werte in \mathbb{K}^m oder einem normierten Raum über \mathbb{K} hat).

Etwas subtiler gestalten sich bei mehreren Variablen allerdings die Kettenregel und die Regel für die Ableitung der Umkehrfunktion. Zu diesen kommen wir daher erst etwas später.

Beispiele (zu partiellen Ableitungen und Richtungsableitungen).

- (1) Bei Benennung der Einzel-Variablen in \mathbb{R}^2 als (x, y) beziehungsweise in \mathbb{R}^3 als (x, y, z) gelten

$$\begin{aligned} \frac{\partial}{\partial x} e^{yx^3+y} &= 3yx^2 e^{yx^3+y}, & \frac{\partial}{\partial y} e^{yx^3+y} &= (x^3+1)e^{yx^3+y}, \\ \frac{\partial}{\partial x} x^2 y^4 z &= 2xy^4 z, & \frac{\partial}{\partial y} x^2 y^4 z &= 4x^2 y^3 z, & \frac{\partial}{\partial z} x^2 y^4 z &= x^2 y^4. \end{aligned}$$

- (2) Für $f: \mathbb{R}^2 \rightarrow \mathbb{R}$ mit $f(x, y) := \sin(2x) \cos^2(y)$ berechnet man die Richtungsableitung an der Stelle $a := (\frac{\pi}{2}, \pi) \in \mathbb{R}^2$ in Richtung $v := (-2, 3) \in \mathbb{R}^2$ direkt anhand der Definition als

$$\begin{aligned} \partial_v f(a) &= \left. \frac{d}{dt} \right|_{t=0} f(a+tv) = \left. \frac{d}{dt} \right|_{t=0} f\left(\frac{\pi}{2}-2t, \pi+3t\right) = \left. \frac{d}{dt} \right|_{t=0} \sin(\pi-4t) \cos^2(\pi+3t) \\ &= \left[-4 \cos(\pi-4t) \cos^2(\pi+3t) - 6 \sin(\pi-4t) \cos(\pi+3t) \sin(\pi+3t) \right]_{t=0} \\ &= -4 \cos^3(\pi) - 6 \cos(\pi) \sin^2(\pi) = 4. \end{aligned}$$

Demnächst lernen wir aber auch noch etwas andere Verfahrensweisen zur Berechnung von Richtungsableitungen kennen.

- (3) Eine **affin lineare Funktion** $f: \mathbb{R}^n \rightarrow \mathbb{R}$, $x \mapsto a \cdot x + b$ mit $a = (a_1, a_2, \dots, a_n) \in \mathbb{R}^n$ und $b \in \mathbb{R}$ hat für alle $j \in \{1, 2, \dots, n\}$ und $v \in \mathbb{R}^n$ konstante Ableitungen

$$\partial_j f \equiv a_j \quad \text{und} \quad \partial_v f \equiv a \cdot v \quad \text{auf } \mathbb{R}^n.$$

Etwas allgemeiner noch hat eine \mathbb{R}^m -wertige affin lineare Funktion $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$, $x \mapsto Ax + b$ mit $A \in \mathbb{R}^{m \times n}$ und $b \in \mathbb{R}^m$ für alle $j \in \{1, 2, \dots, n\}$ und $v \in \mathbb{R}^n$ konstante Ableitungen

$$\partial_j f \equiv A_j, \quad Df \equiv A \quad \text{und} \quad \partial_v f \equiv Av \quad \text{auf } \mathbb{R}^n,$$

wobei $A_j \in \mathbb{R}^m$ die j -te Spalte von A bezeichnet.

- (4) Für **allgemeine Polynome** p ist die Darstellung

$$p(x) = \sum_{|\alpha| \leq m} c_\alpha x^\alpha \quad \text{für } x \in \mathbb{R}^n$$

in sogenannter **Multiindex-Notation** praktisch. Dabei heißt $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_n) \in (\mathbb{N}_0)^n$ Multiindex der Ordnung $|\alpha| := |\alpha|_1 = \sum_{j=1}^n \alpha_j = \alpha_1 + \alpha_2 + \dots + \alpha_n$ (wobei einfache Betragsstriche ausnahmsweise einmal für die 1-Norm, nicht wie sonst für die 2-Norm stehen), und man vereinbart die Schreibweise $x^\alpha := \prod_{j=1}^n x_j^{\alpha_j} = x_1^{\alpha_1} x_2^{\alpha_2} \cdot \dots \cdot x_n^{\alpha_n}$ für Monome. Als j -te partielle Ableitung mit $j \in \{1, 2, \dots, n\}$ des obigen p erhält man dann

$$\partial_i p(x) = \sum_{\substack{|\alpha| \leq m \\ \alpha_i \geq 1}} c_\alpha \alpha_i x^{\alpha - e_i} \quad \text{für } x \in \mathbb{R}^n.$$

- (5) Bei den rotationssymmetrischen Funktionen

$$p: \mathbb{R}^n \setminus \{0\} \rightarrow \mathbb{R}, \quad x \mapsto |x|^s$$

ergeben sich für die Richtungsableitungen in Richtung $v \in \mathbb{R}^n$, die Ableitung nach dem Ortsvektorfeld X und die radiale Ableitung die Formeln

$$\partial_v p(x) = s|x|^{s-2} x \cdot v, \quad \partial_X p(x) = s|x|^s, \quad \partial_{\text{rad}} p(x) = s|x|^{s-1} \quad \text{für } x \in \mathbb{R}^n \setminus \{0\}.$$

Der folgende, im Kern 1-dimensionale Sachverhalt, erlaubt es, Differenzen von Funktionswerten über (Richtungs-)Ableitungen abzuschätzen.

Satz (Schrankensatz mit Richtungsableitungen). Seien \mathcal{X} und \mathcal{Y} normierte Räume sowie $f: A \rightarrow \mathcal{Y}$ eine Funktion auf einer Teilmenge A von \mathcal{X} . Bleibt für $a \neq b$ in A die Strecke $[a, b] := \{a + s(b-a) \mid s \in [0, 1]\}$ in $\overset{\circ}{A}$ und ist f in jedem Punkt dieser Strecke in Richtung $v := \frac{b-a}{\|b-a\|_{\mathcal{X}}}$ des Verbindungsvektors von a zu b differenzierbar, so gilt

$$\|f(b) - f(a)\|_{\mathcal{Y}} \leq \left(\sup_{[a, b]} \|\partial_v f\|_{\mathcal{Y}} \right) \|b-a\|_{\mathcal{X}}.$$

(Insbesondere gilt dies natürlich für $\mathcal{X} = \mathbb{R}^n$ und $\mathcal{Y} = \mathbb{R}^m$ mit dem Betrag als Norm.)

Für Werte in $\mathcal{Y} = \mathbb{R}$ (und bei Stetigkeit von $\partial_v f$ auf $[a, b]$) wurden kurze Beweise auf Grundlage des Mittelwertsatzes beziehungsweise des Hauptsatzes der Differential- und Integralrechnung in der Vorlesung behandelt.

Ein alternativer Beweis in der vollen Allgemeinheit sei hier der Vollständigkeit halber nachgetragen:

Beweis. Wir betrachten $\alpha := 0$, $\beta := \|b-a\|_{\mathcal{X}}$ und die differenzierbare Funktion $g: [\alpha, \beta] \rightarrow \mathcal{Y}$ mit $g(t) := f(a+tv)$. Wegen $\beta-\alpha = \|b-a\|_{\mathcal{X}}$, $g(\alpha) = f(a)$, $g(\beta) = f(b)$ und $\sup_{[\alpha, \beta]} \|g'\|_{\mathcal{Y}} = \sup_{[a, b]} \|\partial_v f\|_{\mathcal{Y}}$ ist dann nur $\|g(\beta) - g(\alpha)\|_{\mathcal{Y}} \leq C(\beta-\alpha)$ mit Schranke $C := \sup_{[\alpha, \beta]} \|g'\|_{\mathcal{Y}}$ zu zeigen. Um diese Abschätzung einzusehen, führen wir für beliebiges $\varepsilon > 0$ die Menge $L_\varepsilon := \{t \in [\alpha, \beta] \mid |g(t) - g(\alpha)| \leq (C+\varepsilon)(t-\alpha)\}$ ein. Da L_ε den Punkt α enthält und durch schwache Ungleichungen definiert ist, ist L_ε nicht-leer und abgeschlossen. Insbesondere ist $s_\varepsilon := \sup L_\varepsilon \in L_\varepsilon$, es gilt also $|g(s_\varepsilon) - g(\alpha)| \leq (C+\varepsilon)(s_\varepsilon - \alpha)$. Wäre $s_\varepsilon < \beta$, so erhielten wir aus $|g'(s_\varepsilon)| \leq C$ zudem $|g(s_\varepsilon+h) - g(s_\varepsilon)| \leq (C+\varepsilon)h$ für $0 < h \ll 1$. Per Dreiecksungleichung ergäbe sich für $0 < h \ll 1$ dann $|g(s_\varepsilon+h) - g(\alpha)| \leq |g(s_\varepsilon) - g(\alpha)| + |g(s_\varepsilon+h) - g(s_\varepsilon)| \leq (C+\varepsilon)(s_\varepsilon+h-\alpha)$, also $s_\varepsilon+h \in L_\varepsilon$ im Widerspruch zur Wahl $s_\varepsilon = \sup L_\varepsilon$. Also verbleibt $s_\varepsilon = \beta$ als einzige Möglichkeit. Wir erhalten somit $\beta \in L_\varepsilon$, mit anderen Worten $|g(\beta) - g(\alpha)| \leq (C+\varepsilon)(\beta-\alpha)$. Aufgrund der Beliebigkeit von $\varepsilon > 0$ folgt daraus die benötigte Abschätzung $|g(\beta) - g(\alpha)| \leq C(\beta-\alpha)$. \square

Bemerkung und Beispiele. Partielle Ableitungen und Richtungsableitungen einer Funktion f an einer Stelle a reichen zur Beschreibung der Verhaltens von f nahe a nicht unbedingt aus. Dies sieht man an den Beispielfunktionen $f_1, f_2: \mathbb{R}^2 \rightarrow \mathbb{R}$ mit

$$f_1(x_1, x_2) := \begin{cases} 1 & \text{falls } x_1 = x_2 \neq 0 \\ 0 & \text{sonst} \end{cases} \quad \text{und} \quad f_2(x_1, x_2) := \begin{cases} 1 & \text{falls } x_1^2 = x_2 \neq 0 \\ 0 & \text{sonst} \end{cases}$$

für $(x_1, x_2) \in \mathbb{R}^2$, denn bei f_1 existieren beide partiellen Ableitungen an der Stelle $(0, 0)$, bei f_2 sogar alle Richtungsableitungen an der Stelle $(0, 0)$ mit

$$\partial_1 f_1(0, 0) = \partial_2 f_1(0, 0) = 0 \quad \text{und} \quad \partial_v f_2(0, 0) = 0 \text{ für alle } v \in \mathbb{R}^2,$$

aber dennoch sind f_1 und f_2 im Punkt $(0, 0)$ nicht einmal stetig. Auch wenn die Beispiele der Funktionen f_1 und f_2 natürlich eher „künstlich konstruiert“, zeigen sie dennoch, dass man mit partiellen Ableitungen und Richtungsableitungen noch keine ganz zufriedenstellende Theorie mehrdimensionaler Ableitungen erhält.

Eine Verbesserung bringt die nächste Definition:

Notation & Bemerkung (Raum der stetigen linearen Abbildungen). Sind \mathcal{X}, \mathcal{Y} normierte Räume (hier und im Folgenden ohne Einschränkung über \mathbb{R} als Grundkörper), so notieren wir

$$\mathcal{L}(\mathcal{X}, \mathcal{Y}) := \{L \in \text{Hom}_{\mathbb{R}}(\mathcal{X}, \mathcal{Y}) \mid L \text{ stetig auf } \mathcal{X}\}$$

für den Raum der stetigen \mathbb{R} -linearen Abbildungen $\mathcal{X} \rightarrow \mathcal{Y}$.

Im Fall $\dim_{\mathbb{R}} \mathcal{X} = \infty$ ist dabei Stetigkeit eine echte Zusatzforderung an $L \in \text{Hom}_{\mathbb{R}}(\mathcal{X}, \mathcal{Y})$, während im Fall $\dim_{\mathbb{R}} \mathcal{X} < \infty$, etwa für $\mathcal{X} = \mathbb{R}^n$, alle linearen Abbildungen $L \in \text{Hom}_{\mathbb{R}}(\mathcal{X}, \mathcal{Y})$ automatisch stetig sind. (Dies liest man an $L(\sum_{i=1}^n x_i e_i) = \sum_{i=1}^n x_i L(e_i)$ mit einer beliebigen Basis e_1, e_2, \dots, e_n von \mathcal{X} ab.) Insbesondere gilt für $\dim_{\mathbb{R}} \mathcal{X} < \infty$ somit $\mathcal{L}(\mathcal{X}, \mathcal{Y}) = \text{Hom}_{\mathbb{R}}(\mathcal{X}, \mathcal{Y})$.

Definition (totale Differenzierbarkeit, totale Ableitungen). Seien \mathcal{X} und \mathcal{Y} normierte Räume sowie $f: A \rightarrow \mathcal{Y}$ eine Funktion auf $A \subset \mathcal{X}$. Man nennt f **(total) differenzierbar** an der Stelle $a \in \overset{\circ}{A}$, wenn ein $L \in \mathcal{L}(\mathcal{X}, \mathcal{Y})$ mit

$$\lim_{\substack{\mathcal{X} \ni v \rightarrow 0 \\ v \neq 0}} \frac{\|f(a+v) - f(a) - L(v)\|_{\mathcal{Y}}}{\|v\|_{\mathcal{X}}} = 0 \quad \text{bzw. äquivalent} \quad \lim_{\substack{\mathcal{X} \ni x \rightarrow a \\ x \neq a}} \frac{\|f(x) - f(a) - L(x-a)\|_{\mathcal{Y}}}{\|x-a\|_{\mathcal{X}}} = 0$$

existiert. Dabei ist L , falls existent, eindeutig durch f und a bestimmt, man schreibt $f'(a)$ für L und nennt $f'(a) \in \mathcal{L}(\mathcal{X}, \mathcal{Y})$ die **(totale) Ableitung** von f an der Stelle a .

Beweis der eindeutigen Bestimmtheit von L . Erfüllen $L, \tilde{L} \in \mathcal{L}(\mathcal{X}, \mathcal{Y})$ beide die Eigenschaft der Definition, so folgt per Dreiecksungleichung

$$\lim_{\substack{\mathcal{X} \ni v \rightarrow 0 \\ v \neq 0}} \frac{\|L(v) - \tilde{L}(v)\|_{\mathcal{Y}}}{\|v\|_{\mathcal{X}}} \leq \lim_{\substack{\mathcal{X} \ni v \rightarrow 0 \\ v \neq 0}} \frac{\|L(v) - f(a+v) + f(a)\|_{\mathcal{Y}}}{\|v\|_{\mathcal{X}}} + \lim_{\substack{\mathcal{X} \ni v \rightarrow 0 \\ v \neq 0}} \frac{\|f(a+v) - f(a) - \tilde{L}(v)\|_{\mathcal{Y}}}{\|v\|_{\mathcal{X}}} = 0.$$

Wäre $L \neq \tilde{L}$, so existierte ein $v_0 \in \mathcal{X} \setminus \{0\}$ mit $L(v_0) \neq \tilde{L}(v_0)$ und folglich

$$\limsup_{\substack{\mathcal{X} \ni v \rightarrow 0 \\ v \neq 0}} \frac{\|L(v) - \tilde{L}(v)\|_{\mathcal{Y}}}{\|v\|_{\mathcal{X}}} \geq \lim_{\substack{\mathbb{R} \ni t \rightarrow 0 \\ t \neq 0}} \frac{\|L(tv_0) - \tilde{L}(tv_0)\|_{\mathcal{Y}}}{\|tv_0\|_{\mathcal{X}}} = \frac{\|L(v_0) - \tilde{L}(v_0)\|_{\mathcal{Y}}}{\|v_0\|_{\mathcal{X}}} > 0$$

(wobei Linearität von L und \tilde{L} im vorletzten Schritt einging). Da dies der vorigen abgesetzten Formel widerspricht, bleibt als einzige Möglichkeit, dass $L = \tilde{L}$ gilt. \square

Bemerkungen (zu totalen Ableitungen).

(0) Tatsächlich wird die **totale Ableitung $f'(a)$ an einer festen Stelle a als lineare Abbildung** definiert und aufgefasst. Dies bedeutet aber keineswegs, dass die zugehörige Ableitungsfunktion f' , die die Zuordnung $a \mapsto f'(a)$ vornimmt, linear wäre. Vielmehr handelt es sich bei dieser (sofern sie auf offenem A überall existiert) um eine im Allgemeinen nicht-lineare Funktion $f': A \rightarrow \mathcal{L}(\mathcal{X}, \mathcal{Y})$ mit Werten im Raum $\mathcal{L}(\mathcal{X}, \mathcal{Y})$ der stetigen linearen Abbildungen. Mit anderen Worten hängt $f'(a)(v) \in \mathcal{Y}$ im Allgemeinen nicht-linear von $x \in A$, aber stets \mathbb{R} -linear von $v \in \mathcal{X}$ ab.

(1) Im Fall $\mathcal{X} = \mathbb{R}$ einer einzelnen Variablen kann und sollte man lineare Abbildungen $f'(a) \in \mathcal{L}(\mathbb{R}, \mathcal{Y})$ mit Vektoren $f'(a) \in \mathcal{Y}$ so identifizieren, dass² $f'(a)(h) = hf'(a)$ für alle $h \in \mathbb{R}$ gilt. Mit diesem Verständnis verallgemeinert die totale Ableitung die früher definierte „normale“ Ableitung in einer einzelnen Variablen.

(Dies wurde in der Vorlesung auch formal durch kurze Rechnung mit den Definitionen verifiziert.)

²Mit anderen Worten identifiziert man also eine lineare Abbildung $f'(a) \in \mathcal{L}(\mathbb{R}, \mathcal{Y})$ mit dem Bild $f'(a)(1) \in \mathcal{Y}$ von $1 \in \mathbb{R}$ und umgekehrt einen Vektor $f'(a) \in \mathcal{Y}$ mit der Linksmultiplikation $\mathbb{R} \rightarrow \mathcal{Y}$, $h \mapsto f'(a)(h)$.

- (2) Bei mehreren Variablen manifestiert sich die Überlegenheit der totalen Ableitung über partielle Ableitungen und Richtungsableitungen unter anderem darin, dass **aus totaler Differenzierbarkeit** von $f: A \rightarrow \mathcal{Y}$ an einer Stelle a die **Stetigkeit** von f in a folgt.

(Auch hierfür wurde in der Vorlesung ein kurzer Beweis gegeben.)

- (3) Zudem folgen **aus totaler Differenzierbarkeit** von $f: A \rightarrow \mathcal{Y}$ an einer Stelle a auch die **Existenz aller partiellen Ableitungen und Richtungsableitungen** von f in a mit

$$\boxed{\partial_v f(a) = f'(a)(v)} \quad \text{für } v \in \mathcal{X}$$

und somit die **lineare Abhängigkeit** der Richtungsableitungen $\partial_v(a)$ vom **Richtungsvektor** $v \in \mathcal{X}$.

(Diese Behauptungen werden im Rahmen der Übungen verifiziert.)

- (4) Wenn immer **im wichtigsten Fall** $\mathcal{X} = \mathbb{R}^n$, $\mathcal{Y} = \mathbb{R}^m$ die **totale Ableitung** $f'(a) \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ existiert, so wird sie als lineare Abbildung $\mathbb{R}^n \rightarrow \mathbb{R}^m$ **durch die Funktionalmatrix** $Df(a) \in \mathbb{R}^{m \times n}$ **dargestellt** (bezüglich der kanonischen Basen von \mathbb{R}^n und \mathbb{R}^m), es gilt also

$$\boxed{f'(a)(v) = Df(a)v} \quad \text{für alle } v \in \mathcal{X}.$$

(Dies folgt, wie in der Vorlesung genauer erläutert, aus der vorigen Bemerkung.)

Nebenbei wird hiermit auch klar, dass im Fall $\mathcal{X} = \mathbb{R}^n$, $\mathcal{Y} = \mathbb{R}^m$ die totale Ableitung statt als lineare Abbildung $L \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ auch als Matrix $M \in \mathbb{R}^{m \times n}$, die im Existenzfall mit der Funktionalmatrix übereinstimmt, hätte definiert werden können. In der gegebenen Definition wäre dazu nur die Anwendung $L(v)$ der linearen Abbildung L durch die Matrix-Vektor-Multiplikation Mv zu ersetzen.

- (5) Die Definition der totalen Ableitung setzt die für die gesamte Differentialrechnung **grundlegende Idee der (affin) linearen Approximation** um, denn $f'(a)$ ist alternativ als die eindeutige lineare Abbildung in $\mathcal{L}(\mathcal{X}, \mathcal{Y})$ mit

$$\boxed{f(x) = f(a) + f'(a)(x-a) + o(\|x-a\|_{\mathcal{X}})} \quad \text{bei } A \ni x \rightarrow a$$

charakterisiert (wobei das Landau-Symbol $o(\|x-a\|_{\mathcal{X}})$ stellvertretend für einen Term $R_a(x)$ mit $\lim_{\substack{\mathcal{X} \ni x \rightarrow a \\ x \neq a}} \frac{\|R_a(x)\|_{\mathcal{Y}}}{\|x-a\|_{\mathcal{X}}} = 0$ steht). In diesem Sinn handelt es sich bei der affin linearen Abbildung $x \mapsto f(a) + f'(a)(x-a)$ um die beste affin lineare Approximation an f bei a .

(Wie in der Vorlesung erklärt wurde, ist dies lediglich eine Umformulierung der gegebenen Definition.)

- (6) **Geometrisch** kann man sich den Graph $\{(x, f(a) + f'(a)(x-a)) \mid x \in \mathcal{X}\} \subset \mathcal{X} \times \mathcal{Y}$ der affin linearen Approximation $x \mapsto f(a) + f'(a)(x-a)$ als (verallgemeinerten) **Tangententialraum** an den Graph $\{(x, f(x)) \mid x \in A\} \subset \mathcal{X} \times \mathcal{Y}$ der Funktion f im Punkt $(a, f(a)) \in \mathcal{X} \times \mathcal{Y}$ vorstellen. Speziell im Fall $\mathcal{X} = \mathbb{R}^n$, $\mathcal{Y} = \mathbb{R}^m$ handelt es sich um einen n -dimensionalen Tangentialraum an einen n -dimensionalen Graph in \mathbb{R}^{n+m} . Bildlich darstellen kann man dies (zumindest) im Fall $n = 2$, $m = 1$ einer (Tangential-)Ebene im Anschauungsraum \mathbb{R}^3 .

Zur Formulierung des nächsten (und einiger folgender) Resultate benötigen wir zwei Begriffe:

Einschub (zweier Definitionen aus etwas anderen Bereichen).

- (1) Für normierte Räume \mathcal{X} , \mathcal{Y} definiert man die **Operatornorm** einer linearen Abbildung $L \in \text{Hom}_{\mathbb{R}}(\mathcal{X}, \mathcal{Y})$ als

$$\|L\|_{\mathcal{L}(\mathcal{X}, \mathcal{Y})} := \sup_{x \in \mathcal{X} \setminus \{0\}} \frac{\|L(x)\|_{\mathcal{Y}}}{\|x\|_{\mathcal{X}}}$$

Es ist dann nicht schwer zu zeigen³, dass $\|L\|_{\mathcal{L}(\mathcal{X}, \mathcal{Y})} < \infty$ genau für stetiges $L \in \text{Hom}_{\mathbb{R}}(\mathcal{X}, \mathcal{Y})$ gilt und die Operatornorm tatsächlich eine **Norm auf $\mathcal{L}(\mathcal{X}, \mathcal{Y})$** ist.

- (2) Eine Teilmenge A eines \mathbb{K} -Vektorraums \mathcal{X} nennt man **konvex**, wenn für alle Punkt $a, b \in A$ auch deren Verbindungsstrecke $[a, b] := \{a + s(b - a) \mid s \in [0, 1]\}$ ganz in A enthalten ist, wenn also $[a, b] \subset A$ für alle $a, b \in A$ gilt.

(Anschaulich kann man sich, wie in der Vorlesung bildlich erläutert wurde, vorstellen, dass der Rand von A überall nach außen gewölbt ist.)

Insbesondere lässt sich für an einer Stelle $a \in \overset{\circ}{A}$ total differenzierbares $f: A \rightarrow \mathcal{Y}$ auf $A \subset \mathcal{X}$ die Operatornorm $\|f'(a)\|_{\mathcal{L}(\mathcal{X}, \mathcal{Y})} \in [0, \infty)$ der totalen Ableitung $f'(a) \in \mathcal{L}(\mathcal{X}, \mathcal{Y})$ bilden und damit der Schrankensatz auch auf folgende, alternative Weise formulieren.

Satz (Schrankensatz mit der totalen Ableitung). *Seien \mathcal{X} und \mathcal{Y} normierte Räume sowie $f: A \rightarrow \mathcal{Y}$ eine auf einer **konvexen** und offenen Teilmenge A von \mathcal{X} überall total differenzierbare Funktion. Dann gilt*

$$\|f(b) - f(a)\|_{\mathcal{Y}} \leq \left(\sup_A \|f'\|_{\mathcal{L}(\mathcal{X}, \mathcal{Y})} \right) \|b - a\|_{\mathcal{X}} \quad \text{für alle } a, b \in A.$$

Der Satz ergibt sich aus dem Schrankensatz mit Richtungsableitungen, indem man für $x \in A$ und $v \in \mathcal{X}$ mit $\|v\|_{\mathcal{X}} = 1$ mittels Bemerkung (3) zu totalen Ableitungen und der Definition der Operatornorm $\|\partial_v f(x)\|_{\mathcal{Y}} = \|f'(x)v\|_{\mathcal{Y}} \leq \|f'(x)\|_{\mathcal{L}(\mathcal{X}, \mathcal{Y})}$ abschätzt.

Korollar. *Für \mathcal{X} , \mathcal{Y} , A , f wie im vorigen Satz gilt:*

$$f' \text{ beschränkt auf } A \text{ (im Sinn } \sup_A \|f'\|_{\mathcal{L}(\mathcal{X}, \mathcal{Y})} < \infty) \iff f \text{ Lipschitz-stetig auf } A.$$

Dabei ist $\sup_A \|f'\|_{\mathcal{L}(\mathcal{X}, \mathcal{Y})}$ die optimale (d.h. kleinstmögliche) Lipschitz-Konstante für f auf A .

Genauer folgt die Implikation „ \implies “ des Korollars direkt aus dem vorausgehenden Schrankensatz, während „ \impliedby “ in der Vorlesung aus der Definition der totalen Ableitung hergeleitet wurde. Die Übereinstimmung der optimalen Schranke für f' und der optimalen Lipschitz-Konstante für f kommt bei diesen Argumenten mit heraus.

³Tatsächlich argumentiert man wie folgt. Ist $\|L\|_{\mathcal{L}(\mathcal{X}, \mathcal{Y})} < \infty$, so erhält man aus der Linearität von L und der Definition der Operatornorm die Abschätzung $\|L(x) - L(\tilde{x})\|_{\mathcal{Y}} = \|L(x - \tilde{x})\|_{\mathcal{Y}} \leq \|L\|_{\mathcal{L}(\mathcal{X}, \mathcal{Y})} \|x - \tilde{x}\|_{\mathcal{X}}$ für alle $x, \tilde{x} \in \mathcal{X}$ und damit sogar Lipschitz-Stetigkeit von L auf \mathcal{X} . Ist umgekehrt L stetig auf \mathcal{X} , so gibt es wegen der Stetigkeit von L in 0 ein $\delta > 0$ mit $\|L(x)\|_{\mathcal{Y}} \leq 1$ für alle $x \in \overline{B}_{\delta}(0)$. Mit Linearität und Homogenität folgt $\|L(v)\|_{\mathcal{Y}} = \delta^{-1} \|v\|_{\mathcal{X}} \|L(\delta \|v\|_{\mathcal{X}}^{-1} v)\|_{\mathcal{Y}} \leq \delta^{-1} \|v\|_{\mathcal{X}}$ für alle $v \in \mathcal{X} \setminus \{0\}$ (wobei $\delta \|v\|_{\mathcal{X}}^{-1} v \in S_{\delta}(0) \subset \overline{B}_{\delta}(0)$ einging) und damit $\|L\|_{\mathcal{L}(\mathcal{X}, \mathcal{Y})} \leq \delta^{-1} < \infty$.

Eine einfache und in der Praxis fast immer ausreichende Möglichkeit zum Nachweis totaler Differenzierbarkeit ergibt sich mit dem nächsten Sachverhalt:

Hauptsatz (hinreichendes Kriterium für totale Differenzierbarkeit). *Seien \mathcal{Y} ein normierter Raum und $f: A \rightarrow \mathcal{Y}$ eine Funktion auf $A \subset \mathbb{R}^n$. Existieren die partiellen Ableitungen von f nach allen n Variablen in einer Umgebung von $a \in A$ und sind all diese **partiellen Ableitungen in a stetig**, so ist f in a total differenzierbar mit $f'(a)(v) = \sum_{i=1}^n v_i \partial_i f(a)$ für alle $v \in \mathbb{R}^n$.*

Ein Beweis des Hauptsatzes auf Basis des Schrankensatzes wird in der Vorlesung geführt.

Differenzierbare Funktionen mit stetiger Ableitung bezeichnet man wie in der Differentialrechnung einer Variablen auch allgemein als stetig differenzierbar oder als C^1 -Funktionen:

Definition (stetige Differenzierbarkeit, C^1 -Funktionen). *Seien \mathcal{X} und \mathcal{Y} normierte Räume sowie A eine offene Teilmenge von \mathcal{X} . Wir nennen $f: A \rightarrow \mathcal{Y}$ **stetig differenzierbar** auf A oder **Funktion der Klasse C^1** auf A , wenn die totale Ableitung $f': A \rightarrow \mathcal{L}(\mathcal{X}, \mathcal{Y})$ auf ganz A existiert und bezüglich der Operatornorm auf $\mathcal{L}(\mathcal{X}, \mathcal{Y})$ stetig ist. Den **Raum aller C^1 -Funktionen** $A \rightarrow \mathcal{Y}$ bezeichnen wir mit $C^1(A, \mathcal{Y})$.*

Aufgrund dieser Definition und des Kriteriums aus dem Hauptsatz können wir festhalten:

Korollar. *Für $A \subset \mathbb{R}^n$, einen normierten Raum \mathcal{Y} und $f: A \rightarrow \mathcal{Y}$ gilt:*

$$f \in C^1(A, \mathcal{Y}) \iff \partial_1 f, \partial_2 f, \dots, \partial_n f \text{ existieren und sind stetig auf } A.$$

Wie in der Vorlesung genauer erläutert wird, ergibt sich dabei „ \Leftarrow “ aus dem Hauptsatz und „ \Rightarrow “ aus Bemerkung (3) zu totalen Ableitungen.

Außerdem lässt sich an dieser Stelle folgende Version des Konstanzsatzes mit Voraussetzung nur an die partiellen Ableitungen erhalten.

Satz (Konstanzsatz). *Sei \mathcal{Y} ein normierter Raum, und sei $f: A \rightarrow \mathcal{Y}$ auf offenem und **zusammenhängendem** $A \subset \mathbb{R}^n$ nach allen n Variablen differenzierbar mit verschwindenden partiellen Ableitungen $\partial_j f \equiv \mathbf{0}$ auf A für alle $j \in \{1, 2, \dots, n\}$. Dann ist **f konstant** auf A (also $f \equiv y_0$ auf A für ein $y_0 \in \mathcal{Y}$).*

Gemäß dem Kriterium des vorausgehenden Hauptsatzes liegt in der Situation des Konstanzsatzes zunächst totale Differenzierbarkeit mit $f' \equiv 0$ auf A vor, und im Fall eines konvexen A folgt die behauptete Konstanz von f dann aus dem Schrankensatz mit der totalen Ableitung. Der Schritt zu Beweis des Satzes in der vollen Allgemeinheit gelingt dann mit einem typischen Zusammenhangs-Argument, das in der Vorlesung ausgeführt wird.

Wie für alle Ableitungsbegriffe gelten auch für totale Ableitungen naheliegende Summen- und Faktorregeln, die wir hier aber nicht im Detail erörtern. Von größerem Interesse ist die folgende allgemeine Kettenregel (auf die man andere Regeln tatsächlich zurückführen kann).

Satz (Kettenregel für totale Ableitungen). *Seien \mathcal{X} , \mathcal{Y} und \mathcal{Z} normierte Räume sowie $f: A \rightarrow \mathcal{Y}$ mit $A \subset \mathcal{X}$ und $g: B \rightarrow \mathcal{Z}$ mit $f(A) \subset B \subset \mathcal{Y}$ Funktionen (für die $g \circ f$ wohldefiniert ist). Existieren für $a \in A$ mit $f(a) \in B$ die totalen Ableitungen $f'(a) \in \mathcal{L}(\mathcal{X}, \mathcal{Y})$ und $g'(f(a)) \in \mathcal{L}(\mathcal{Y}, \mathcal{Z})$, so existiert auch $(g \circ f)'(a) \in \mathcal{L}(\mathcal{X}, \mathcal{Z})$ und ergibt sich als Verkettung linearer Abbildungen*

$$\boxed{(g \circ f)'(a) = g'(f(a)) \circ f'(a)}.$$

Der Beweis der Kettenregel wird (bis auf einen in den Übungen thematisierten Teilschritt) in der Vorlesung ausgeführt und benutzt im Wesentlichen nur die Definition der totalen Ableitung.

Folgerungen. Aus der Kettenregel für totale Ableitungen folgen weitere Versionen der Kettenregel — für f und g wie im vorigen Satz, mit kleinen Zusatzargumenten aber auch dann, wenn nur die jeweils auftretenden Ableitungen von f und die totale Ableitung $g'(f(a))$ existieren.

- (1) Aus $f'(a)(v) = \partial_v f(a)$ für $v \in \mathcal{X}$ gewinnt man die **Kettenregel für Richtungsableitungen**

$$\partial_v(g \circ f)(a) = \partial_w g(f(a)) \in \mathcal{Z} \quad \text{mit } w := \partial_v f(a) \in \mathcal{Y}$$

- (2) Im Fall $\mathcal{X} = \mathbb{R}^n$, $\mathcal{Y} = \mathbb{R}^m$, $\mathcal{Z} = \mathbb{R}^\ell$ ergibt sich, weil die totale Ableitung durch die Funktionalmatrix dargestellt wird, die **Kettenregel für Funktionalmatrizen**

$$\boxed{D(g \circ f)(a) = Dg(f(a))Df(a)} \in \mathbb{R}^{\ell \times n},$$

bei der rechts das Produkt der Matrizen $Dg(f(a)) \in \mathbb{R}^{\ell \times m}$ und $Df(a) \in \mathbb{R}^{m \times n}$ steht.

- (3) Ebenfalls im Fall $\mathcal{X} = \mathbb{R}^n$, $\mathcal{Y} = \mathbb{R}^m$, $\mathcal{Z} = \mathbb{R}^\ell$ führt Ausschreiben des Matrix-Produkts in der vorigen Regel zur **Kettenregel für partielle Ableitungen**

$$\boxed{\partial_k(g_i \circ f)(a) = \sum_{j=1}^m \partial_j g_i(f(a)) \partial_k f_j(a)} \in \mathbb{R} \quad \text{für } i \in \{1, 2, \dots, \ell\}, k \in \{1, 2, \dots, n\}$$

(wobei f_j , g_i und $g_i \circ f = (g \circ f)_i$ die Komponentenfunktionen von f , g und $g \circ f$ bezeichnen).

Beispiele (zur Anwendung von Kettenregeln).

- (1) Für beliebiges differenzierbares $g: \mathbb{R}^2 \rightarrow \mathbb{R}^\ell$ gibt die Kettenregel für partielle Ableitungen (mit $n = 1$, $m = 2$)

$$\frac{d}{dt}g(t^2, \arctan t) = \partial_1 g(t^2, \arctan t) 2t + \partial_2 g(t^2, \arctan t) \frac{1}{1+t^2} \in \mathbb{R}^\ell \quad \text{für } t \in \mathbb{R}.$$

- (2) Für beliebiges differenzierbares $g: \mathbb{R}^2 \rightarrow \mathbb{R}^\ell$ gibt die Kettenregel für partielle Ableitungen (mit $n = 3$, $m = 2$)

$$\begin{aligned} \frac{\partial}{\partial x}g(y^2 e^x, z + \sin y) &= \partial_1 g(y^2 e^x, z + \sin y) y^2 e^x && \in \mathbb{R}^\ell, \\ \frac{\partial}{\partial y}g(y^2 e^x, z + \sin y) &= \partial_1 g(y^2 e^x, z + \sin y) 2y e^x + \partial_2 g(y^2 e^x, z + \sin y) \cos y && \in \mathbb{R}^\ell, \\ \frac{\partial}{\partial z}g(y^2 e^x, z + \sin y) &= \partial_2 g(y^2 e^x, z + \sin y) && \in \mathbb{R}^\ell \end{aligned}$$

für $(x, y, z) \in \mathbb{R}^3$. Setzen wir $f(x, y, z) := (y^2 e^x, z + \sin y)$, so lässt sich dasselbe Ergebnis aber auch aus der Kettenregel für Funktionalmatrizen durch Matrixmultiplikation von

$$Dg(s, t) = \left(\partial_1 g(s, t) \mid \partial_2 g(s, t) \right) \in \mathbb{R}^{\ell \times 2} \quad \text{und} \quad Df(x, y, z) = \begin{pmatrix} y^2 e^x & 2y e^x & 0 \\ 0 & \cos y & 1 \end{pmatrix} \in \mathbb{R}^{2 \times 3}$$

gewinnen: Man erhält

$$\begin{aligned} D(g \circ f)(x, y, z) &= Dg(f(x, y, z))Df(x, y, z) \\ &= \left(\partial_1 g(\dots) y^2 e^x \mid \partial_1 g(\dots) 2y e^x + \partial_2 g(\dots) \cos y \mid \partial_2 g(\dots) \right) \in \mathbb{R}^{\ell \times 3}, \end{aligned}$$

wobei bei den Pünktchen jeweils $(y^2 e^x, z + \sin y)$ einzusetzen ist und sich in den drei Spalten der resultierenden Matrix dieselben Terme wie zuvor ergeben.

Schließlich führen wir in diesem Abschnitt einige weitere Ableitungsoperatoren auf Basis der schon behandelten ein. Die für diese Vorlesung mit Abstand wichtigste Bildung folgt direkt.

Definition (Gradient). Sei $f: A \rightarrow \mathbb{R}$ eine Funktion auf $A \subset \mathbb{R}^n$. Ist f an einer Stelle $a \in \overset{\circ}{A}$ nach allen n Variablen differenzierbar, so heißt der zur Funktionalmatrix $Df(a) \in \mathbb{R}^{1 \times n}$ transponierte Vektor

$$\nabla f(a) := \begin{pmatrix} \partial_1 f(a) \\ \partial_2 f(a) \\ \vdots \\ \partial_n f(a) \end{pmatrix} \in \mathbb{R}^n$$

der **Gradient** von f an der Stelle a /in a /bei a .

Beispiel (zum Gradienten). Für die Funktion $f: \mathbb{R}^3 \rightarrow \mathbb{R}$ mit

$$f(x, y, z) := x^2 y^4 e^z \quad \text{ergibt sich} \quad \nabla f(x, y, z) = \begin{pmatrix} 2xy^4 e^z \\ 4x^2 y^3 e^z \\ x^2 y^4 e^z \end{pmatrix}.$$

Bemerkungen (zum Gradienten). Sei $f: A \rightarrow \mathbb{R}$ eine Funktion auf $A \subset \mathbb{R}$.

- (1) Der Gradient ∇f wird **nur für R-wertige Funktionen** f , die man auch als skalare Funktionen bezeichnet, erklärt.
- (2) Ist f an einer Stelle $a \in \overset{\circ}{A}$ total differenzierbar, so existiert der Gradient $\nabla f(a)$ und hat die **zentrale Eigenschaft**

$$\partial_v f(a) = v \cdot \nabla f(a) \quad \text{für alle } v \in \mathbb{R}^n,$$

wobei \cdot das Standard-Skalarprodukt auf \mathbb{R}^n bezeichnet. Man kann also **aus dem Gradienten alle Richtungsableitungen berechnen**.

(Der Beweis der behaupteten Formel gelingt dabei mit den Bemerkungen (3) und (4) zur totalen Ableitung sowie durch Umschreiben der Matrixmultiplikation „Zeile mal Spalte“ als Skalarprodukt wie folgt: $\partial_v f(a) = f'(a)(v) = Df(a)v = Df(a)^T \cdot v = \nabla f(a) \cdot v = v \cdot \nabla f(a)$.)

- (3) Per Cauchy-Schwarz-Ungleichung ergibt sich aus der Formel der vorigen Bemerkung die Abschätzung $\partial_v f(a) \leq |v| |\nabla f(a)|$ mit Gleichheit genau dann, wenn v und $\nabla f(a)$ in die gleiche Richtung zeigen (oder präziser, wenn sie positiv linear abhängig sind, wenn also entweder $v = 0$ oder $\nabla f(a) = 0$ oder $\nabla f(a) = rv$ mit *positivem* $r \in (0, \infty)$ eintritt). Denkt man an Richtungen in \mathbb{R}^n als Einheitsvektoren $v \in \mathbb{R}^n$ mit $|v| = 1$ und wendet das Vorige auf diese an, so ergibt sich die **geometrische Interpretation des Gradienten**: $\nabla f(a)$ zeigt in die **Richtung des steilsten Anstiegs** des Graphen von f an der Stelle $(a, f(a))$, und $|\nabla f(a)|$ ist die zugehörige steilste Steigung.
- (4) Existiert ∇f überall auf offenem $A \subset \mathbb{R}^n$, so macht das Vektorfeld $\nabla f: A \rightarrow \mathbb{R}^n$ Sinn und wird als **Gradienten(vektor)feld** von f bezeichnet.

Als Verallgemeinerung von Richtungsableitungen reißen wir kurz an:

Definition (Ableitung entlang einer Kurve). Seien \mathcal{X} und \mathcal{Y} normierte Räume sowie $f: A \rightarrow \mathcal{Y}$ eine Funktion auf $A \subset \mathcal{X}$ und $c: I \rightarrow A$ eine Kurve in A mit Definitionsintervall $I \subset \mathbb{R}$. Für $t_0 \in I$ heißt dann

$$\frac{d}{dt} \Big|_{t=t_0} f(c(t)) \in \mathcal{Y},$$

falls existent, die **Ableitung** von f im Punkt $c(t_0)$ **entlang der Kurve** c .

Genau genommen ist es dabei nur korrekt, von der Ableitung *im Punkt* $c(t_0)$ entlang c zu sprechen, wenn der Parameterwert $t_0 \in I$ durch $c(t_0) \in A$ eindeutig bestimmt ist. Ist dies nicht der Fall, weil an der Stelle $c(t_0)$ ein Selbstschnitt der Kurve c vorliegt, so ist t_0 im Prinzip zusätzlich anzugeben, in der Praxis wird dies aber nur selten nötig.

Bemerkung (zur Ableitung entlang einer Kurve). Existieren, unter den Voraussetzungen der vorausgehenden Definition, $c'(t_0)$ und $f'(c(t_0))$ in einem inneren Punkt $c(t_0) \in \overset{\circ}{A}$, so existiert auch die Ableitung von f in a entlang c und für diese gilt

$$\frac{d}{dt} \Big|_{t=t_0} f(c(t)) = \partial_v f(c(t_0)) \quad \text{mit } v := c'(t_0) \in \mathcal{X}.$$

Diese Formel ergibt sich durch direkte Anwendung der Kettenregel, und sie besagt mit anderen Worten, dass man — wie geometrisch plausibel — die Ableitung von f in a entlang c als **Richtungsableitung in Richtung des Tangentialvektors** v an die Kurve c im Punkt $c(t_0)$ erhält.

Beispiel (zur Ableitung entlang einer Kurve). Für die Funktion

$$f: \mathbb{R}^2 \rightarrow \mathbb{R}, x \mapsto \frac{1}{2}(x_1^2 - x_2^2)$$

und die Parametrisierung der Einheitskreislinie $S_1 = \{x \in \mathbb{R}^2 \mid |x| = 1\}$ durch die Kurve

$$c: \mathbb{R} \rightarrow \mathbb{R}^2, t \mapsto \begin{pmatrix} \cos t \\ \sin t \end{pmatrix}$$

ergibt sich als Ableitung von f entlang c durch Einsetzen und direkte Berechnung

$$\frac{d}{dt} f(c(t)) = \frac{d}{dt} \frac{1}{2}(\cos^2 t - \sin^2 t) = -2(\sin t)(\cos t)$$

für $t \in \mathbb{R}$. Alternativ berechnet aufgrund von $\nabla f(x) = (x_1, -x_2)$ und $c'(t) = (-\sin t, \cos t)$ mit der Formel der Bemerkung

$$\frac{d}{dt} f(c(t)) = \partial_{c'(t)} f(c(t)) = c'(t) \cdot \nabla f(c(t)) = \begin{pmatrix} -\sin t \\ \cos t \end{pmatrix} \cdot \begin{pmatrix} \cos t \\ -\sin t \end{pmatrix} = -2(\sin t)(\cos t)$$

für $t \in \mathbb{R}$ und erhält natürlich dasselbe Ergebnis.

Übrigens entsprechen hier die Nullstellen der Ableitung an allen Stellen $t \in \mathbb{Z} \frac{\pi}{2}$ den vier Punkten $(\pm 1, 0) \in S_1$ und $(0, \pm 1) \in S_1$ der Form $c(z \frac{\pi}{2})$ mit $z \in \mathbb{Z}$, von denen sich $(\pm 1, 0)$ als Maximalstellen und $(0, \pm 1)$ als Minimalstellen von f auf S_1 bzw. entlang c erweisen. Dies gibt einen ersten, in der Tat aber sehr eingeschränkten Vorgeschmack auf den folgenden Abschnitt 2.2 zur Extremstellenbestimmung auf Teilmengen des \mathbb{R}^n .

Zwei weitere Bildungen, die in der Analysis an verschiedenen Stellen nützlich sind, folgen.

Definitionen (Divergenz und Rotation). Sei $V: A \rightarrow \mathbb{R}^n$ ein Vektorfeld auf $A \subset \mathbb{R}^n$, das in $a \in A$ nach allen n Variablen differenzierbar ist und für das somit $DV(a) \in \mathbb{R}^{n \times n}$ erklärt ist.

(1) Die **Divergenz** von V an der Stelle a /in a /bei a ist die Zahl

$$\operatorname{div} V(a) := \operatorname{Spur}(DV(a)) = \sum_{i=1}^n \partial_i V_i(a) \in \mathbb{R}.$$

(2) Die **Rotation** oder **Rotationsmatrix** von V an der Stelle a /in a /bei a ist der doppelte antisymmetrische Anteil von $DV(a)$, also die Matrix

$$\operatorname{Rot} V(a) := DV(a) - DV(a)^T = (\partial_j V_i(a) - \partial_i V_j(a))_{i,j=1,2,\dots,n} \in \mathbb{R}^{n \times n}.$$

(2') **Speziell für $n = 3$** bildet man mit den drei wesentlichen Einträgen

$$\operatorname{Rot} V(a) = \begin{pmatrix} 0 & \partial_2 V_1(a) - \partial_1 V_2(a) & \partial_3 V_1(a) - \partial_1 V_3(a) \\ \partial_1 V_2(a) - \partial_2 V_1(a) & 0 & \partial_3 V_2(a) - \partial_2 V_3(a) \\ \partial_1 V_3(a) - \partial_3 V_1(a) & \partial_2 V_3(a) - \partial_3 V_2(a) & 0 \end{pmatrix} \in \mathbb{R}^{3 \times 3}$$

den **Rotationsvektor** oder die (**klassische**) **Rotation**

$$\operatorname{rot} V(a) := \begin{pmatrix} \partial_2 V_3(a) - \partial_3 V_2(a) \\ \partial_3 V_1(a) - \partial_1 V_3(a) \\ \partial_1 V_2(a) - \partial_2 V_1(a) \end{pmatrix} \in \mathbb{R}^3,$$

speziell für $n = 2$ bezeichnet man den einen wesentlichen Eintrag von

$$\operatorname{Rot} V(a) = \begin{pmatrix} 0 & \partial_2 V_1(a) - \partial_1 V_2(a) \\ \partial_1 V_2(a) - \partial_2 V_1(a) & 0 \end{pmatrix} \in \mathbb{R}^{2 \times 2}$$

auch als den **Rotationsskalar**

$$\operatorname{rot} V(a) := \partial_1 V_2(a) - \partial_2 V_1(a) \in \mathbb{R}.$$

Bemerkungen (zu **Divergenz** und **Rotation**).

(1) Falls die jeweilige Bildung an allen Stellen eines offenen $A \subset \mathbb{R}^n$ existiert, kann $\operatorname{div} V$ als \mathbb{R} -wertige Funktion, $\operatorname{Rot} V$ als $\mathbb{R}^{n \times n}$ -wertiges Matrixfeld sowie $\operatorname{rot} V$ im Fall $n = 3$ bzw. $n = 2$ als \mathbb{R}^3 -wertiges Vektorfeld bzw. \mathbb{R} -wertige Funktion aufgefasst werden.

(2) Mit dem **Operator „Nabla“**

$$\nabla = (\partial_1, \partial_2, \dots, \partial_{n-1}, \partial_n) = \begin{pmatrix} \partial_1 \\ \partial_2 \\ \vdots \\ \partial_{n-1} \\ \partial_n \end{pmatrix},$$

den wir nicht als präzises Objekt einführen, sondern nur in folgenden Schreibweisen/Merkregeln formal verwenden, drückt man Gradient, Divergenz und klassische Rotation als

$$\nabla f, \quad \operatorname{div} V = \nabla \cdot V \quad \text{und} \quad \operatorname{rot} V = \nabla \times V$$

aus (wobei ersteres ja sowieso unsere Standard-Notation für den Gradienten und letzteres nur für $n = 3$ und Vektorfelder V auf Teilmengen des \mathbb{R}^3 gemeint ist). Mit dieser Notation, die auch in der Physik sehr verbreitet ist, kann man sich die Definitionen aller drei Bildungen mit Hilfe von ∇ in formaler Analogie zu Skalarmultiplikation des \mathbb{R}^n , Skalarprodukt \cdot des \mathbb{R}^n und Kreuzprodukt \times des \mathbb{R}^3 merken.

Das **Kreuzprodukt** ist dabei für Vektoren $v, w \in \mathbb{R}^3$ präzise erklärt als

$$v \times w := \begin{pmatrix} v_2 w_3 - v_3 w_2 \\ v_3 w_1 - v_1 w_3 \\ v_1 w_2 - v_2 w_1 \end{pmatrix} \in \mathbb{R}^3$$

und hängt von (v, w) bilinear sowie antikommutativ im Sinn $v \times w = -w \times v$ ab. Bei linearer Abhängigkeit von v und w gilt stets $v \times w = 0$, während man bei linearer Unabhängigkeit von v und w als $v \times w$ den zu v und w orthogonalen Vektor erhält, für den $v, w, v \times w$ positiv orientierte Basis des \mathbb{R}^3 ist und $|v \times w| = |v| |w| \sin \angle(v, w)$ gilt. Insbesondere folgt $|v \times w|^2 + |v \cdot w|^2 = |v|^2 |w|^2 (\sin^2 \angle(v, w) + \cos^2 \angle(v, w)) = |v|^2 |w|^2$.

Einige allgemeine Eigenschaften von Divergenz und Rotation werden in den Übungen hergeleitet. In den Lernwerkstätten wird zumindest ansatzweise die Interpretation der **Divergenz als Quell- bzw. Senkendichte** eines Vektorfelds sowie der **Rotation als Wirbeldichte** besprochen. Für eine Präzisierung dieser beiden Interpretationen durch den Satz von Gauß und den (klassischen) Satz von Stokes sowie weitere „Vektor-Analyse“ mit Divergenz und Rotation wird auf die Vorlesung zur höheren Analysis und aufbauende Vorlesungen verwiesen.

2.2 Extremstellenbestimmung

In diesem Abschnitt kommen wir zur Extremstellenbestimmung bei Funktionen mehrerer Variablen oder, mit anderen Worten, bei Funktionen auf Definitionsbereichen in \mathbb{R}^n . (Teils kann das Vorgehen dabei sogar auf Definitionsbereiche in ∞ -dimensionalen normierten Räumen übertragen werden, doch solche Verallgemeinerungen werden wir hier nicht diskutieren.)

Als Erstes erklären wir in weitgehender Analogie zum Fall einer Variablen einige Begriffe.

Definitionen (Extremstellen und kritische Punkte). Sei $f: A \rightarrow \mathbb{R}$ eine \mathbb{R} -wertige Funktion auf $A \subset \mathbb{R}^n$.

- Wir nennen $a \in A$ eine **absolute Minimalstelle** von f auf A , wenn $f(a) \leq f(x)$ für alle $x \in A$ gilt. Eine **absolute Maximalstelle** $a \in A$ von f auf A erklären wir analog mit der umgekehrten Ungleichung.
- Wir nennen $a \in A$ eine **lokale Minimalstelle** von f auf A , wenn es eine Umgebung U von a in A gibt, so dass $f(a) \leq f(x)$ für alle $x \in U$ gilt. Kann U sogar so gewählt werden, dass die strikte Ungleichung $f(a) < f(x)$ für alle $x \in U \setminus \{a\}$ eintritt, so sprechen wir von einer **strikten lokalen Minimalstelle** a von f auf A . Eine (**strikte**) **lokale Maximalstelle** $a \in A$ von f auf A erklären wir analog mit der umgekehrten Ungleichung.

- Eine (**strikte**) **lokale Extremstelle** ist entweder eine (**strikte**) lokale Minimalstelle oder eine (**strikte**) lokale Maximalstelle.
- Eine inneren Punkt $a \in \overset{\circ}{A}$ von A nennen wir einen **kritischen Punkt** von f , wenn f in a differenzierbar⁴ ist und $\nabla f(a) = 0_{\mathbb{R}^n}$ gilt, also der **Gradient** von f in a **verschwindet**.

Als einfaches, aber sehr grundlegendes Kriterium für Extremstellen bekommen wir:

Satz (notwendiges Kriterium für innere Extremstellen). Sei $A \subset \mathbb{R}^n$. Ist $a \in \overset{\circ}{A}$ lokale Extremstelle von $f: A \rightarrow \mathbb{R}$ auf A und ist f in a differenzierbar, so ist a **notwendig ein kritischer Punkt** von f .

Beweis. Für jedes $i \in \{1, 2, \dots, n\}$ ist die i -te Komponente $a_i \in \mathbb{R}$ von $a \in \mathbb{R}^n$ lokale Extremstelle von $s \mapsto f(a_1, \dots, a_{i-1}, s, a_{i+1}, \dots, a_n)$. Das notwendige Kriterium für Extremstellen in einer Variablen gibt daher $\partial_i f(a) = (f(a_1, \dots, a_{i-1}, \cdot, a_{i+1}, \dots, a_n))'(a_i) = 0$ für alle $i \in \{1, 2, \dots, n\}$, und es folgt $\nabla f(a) = (\partial_1 f(a), \partial_2 f(a), \dots, \partial_n f(a)) = 0$. \square

Bemerkung. Tatsächlich greift ein ähnliches Kriterium auch noch unter deutlich schwächeren Differenzierbarkeitsvoraussetzungen: Existiert in einer lokalen Minimalstelle a von $f: A \rightarrow \mathbb{R}$ auf $A \subset \mathbb{R}^n$ nur die einseitige Richtungsableitung $\frac{d}{dt}\Big|_{t=0+} f(a+tv) := \lim_{h \searrow 0} \frac{f(a+hw) - f(a)}{h}$ in Richtung eines Vektors $v \in \mathbb{R}^n$, so ergibt sich aus der Differentialrechnung einer Variablen, dass notwendig

$$\frac{d}{dt}\Big|_{t=0+} f(a+tv) \geq 0$$

gelten muss. Für lokale Maximalstelle gilt natürlich Analoges mit „ \leq “ anstelle von „ \geq “.

Auf den vorausgehenden Satz baut ein **schlagkräftiges Rechenverfahren** auf:

Verfahren (zur Bestimmung absoluter Extremstellen bei mehreren Variablen). Sei $f: A \rightarrow \mathbb{R}$ eine Funktion auf $A \subset \mathbb{R}^n$. Zur Bestimmung der Extremstellen von f auf einer Teilmenge A von D empfiehlt sich ein **Vorgehen in drei Schritten**:

(I) **Existenz sicherstellen:**

Ist A kompakt und ist f stetig auf A , so folgt die Existenz einer absoluten Minimal- und einer absoluten Maximalstelle von f auf A aus dem allgemeinen **Extremalsatz** des Abschnitts 1.3.

Andernfalls untersucht man das Verhalten von f bei Grenzübergängen gegen nicht zu A gehörige Randpunkte in $(\partial A) \setminus A$, im Unendlichen und/oder bei Unstetigkeitsstellen. Meist folgt dann die Existenz absoluter Extremstellen aus der Erweiterung des Extremalsatzes gemäß Abschnitt 1.3, oder ihre Nicht-Existenz wird offensichtlich.

(II) **Vollständige Kandidatenliste aufstellen:**

In diese Liste sind einerseits **kritische Punkte** und **Nichtdifferenzierbarkeitsstellen** von f **im Innern** $\overset{\circ}{A}$ von A aufzunehmen. Zur Berechnung der kritischen Punkte $x \in \overset{\circ}{A}$ geht man dabei vom Gleichungssystem $\nabla f(x) = 0$ aus, das ausgeschrieben aus den n im Allgemeinen nicht-linearen Gleichungen $\partial_1 f(x) = 0, \partial_2 f(x) = 0, \dots, \partial_n f(x) = 0$ besteht. Dieses Gleichungssystem für die n Variablen $x_1, x_2, \dots, x_n \in \mathbb{R}$ kann keine, eine, mehrere oder gar unendlich viele Lösungen besitzen.

⁴Soweit nicht anders spezifiziert, ist Differenzierbarkeit fortan immer als totale Differenzierbarkeit zu verstehen.

Des Weiteren sind in die Kandidatenliste **gewisse zu A gehörige Randpunkte** aus $A \cap \partial A$ aufzunehmen. In glücklichen Fällen (beispielsweise, wenn A offen und somit $A \cap \partial A = \emptyset$ ist, wenn nur endlich viele Randpunkte zu A gehören, oder, wenn die Menge der zugehörigen Funktionswerte überschaubar bleibt) kann man einfach alle solchen Punkte in die Kandidatenliste aufnehmen und zum letzten Schritt des Verfahrens übergehen. Im Allgemeinen braucht man aber **Kriterien zur Auswahl der Randkandidaten**, die erst im weiteren Verlauf dieses Abschnitts behandelt werden.

(III) **Wertevergleich durchführen:**

Man berechnet und vergleicht die **Funktionswerte in allen auf die Kandidatenliste genommenen Punkten**. Die Punkte mit den kleinsten und größten Funktionswerten sind damit — so die Existenz gesichert und die Liste korrekt aufgestellt wurde — als **absolute Extremstellen identifiziert**.

Beispiele (zur Bestimmung absoluter Extremstellen bei mehreren Variablen). Wir geben Beispiele für die Durchführung des zuvor erläuterten Verfahrens:

(1) Zunächst betrachten wir die Funktion $f: \mathbb{R}^3 \rightarrow \mathbb{R}$ mit

$$f(x, y, z) := x + y + yz - y^2 - z^2 - 2x^4 \quad \text{für } (x, y, z) \in A = \mathbb{R}^3.$$

(I) Da \mathbb{R}^3 nicht kompakt ist, ist der Extremalsatz nicht direkt anwendbar. Aufgrund der Stetigkeit von f und von⁵ $\lim_{|(x,y,z)| \rightarrow \infty} f(x, y, z) = -\infty$ ist aber klar, dass es keine absolute Minimalstelle von f auf \mathbb{R}^3 gibt, während eine absolute Maximalstelle von f auf \mathbb{R}^3 gemäß der Erweiterung des Extremalsatzes existiert.

(II) Man berechnet $\nabla f(x, y, z) = (1-8x^3, 1+z-2y, y-2z)$ und erhält das Gleichungssystem $1-8x^3 = 0$, $1+z-2y = 0$, $y-2z = 0$ für innere kritische Punkte. Als einzige Lösung dieses Gleichungssystems ergibt sich $(x, y, z) = (\frac{1}{2}, \frac{2}{3}, \frac{1}{3})$. Nichtdifferenzierbarkeitsstellen oder Randkandidaten gibt es hier nicht, da f differenzierbar und \mathbb{R}^3 offen ist.

(III) Da es nur einen Kandidaten gibt, ist kein Wertevergleich nötig.

Insgesamt ist nachgewiesen, dass es keine absolute Minimalstelle von f auf \mathbb{R}^3 gibt, während $(\frac{1}{2}, \frac{2}{3}, \frac{1}{3})$ die einzige absolute Maximalstelle von f auf \mathbb{R}^3 ist.

(2) Unser zweites Beispiel auf der abgeschlossenen Einheitskreisscheibe $\overline{B}_1^2(0, 0) \subset \mathbb{R}^2$ ist die Funktion $f: \overline{B}_1^2(0, 0) \rightarrow \mathbb{R}$ mit

$$f(x, y) := x(1 - x^2 - y^2) \quad \text{für } (x, y) \in A = \overline{B}_1^2(0, 0).$$

(I) Da $\overline{B}_1^2(0, 0)$ kompakt ist, garantiert der Extremalsatz die Existenz einer absoluten Minimal- und einer absoluten Maximalstelle der stetigen Funktion f auf $\overline{B}_1^2(0, 0)$.

(II) Man berechnet $\nabla f(x, y) = (1-3x^2-y^2, -2xy)$ und erhält das Gleichungssystem $1-3x^2-y^2 = 0$, $xy = 0$ für innere kritische Punkte. Es kommen nur Lösungen mit $x = 0$ oder $y = 0$ in Frage. Der erste Fall führt formal auf die beiden Lösungen $(x, y) = (0, \pm 1)$, die jedoch nicht im Innern $B_1^2(0, 0)$ von $\overline{B}_1^2(0, 0)$ liegen. Der zweite Fall führt auf die beiden Kandidaten $(\pm \frac{1}{3}\sqrt{3}, 0) \in B_1^2(0, 0)$. Zudem sind (mangels behandelte Randkriterien) alle Punkte in $\partial B_1^2(0, 0)$ auf die Kandidatenliste zu nehmen.

⁵Zum präzisen Nachweis dieser Konvergenz schätzt man mit Hilfe der Youngschen Ungleichung $yz \leq \frac{1}{2}y^2 + \frac{1}{2}z^2$ und $-2x^4 \leq -\frac{1}{4}x^4 \leq -\frac{1}{2}x^2 + \frac{1}{4}$ ab und erhält $f(x, y, z) \leq 2|(x, y, z)| - \frac{1}{2}|(x, y, z)|^2 + \frac{1}{4} \rightarrow -\infty$ bei $|(x, y, z)| \rightarrow \infty$.

(III) Es reicht hier, $f(\frac{1}{3}\sqrt{3}, 0) > 0 > f(-\frac{1}{3}\sqrt{3}, 0)$ und den in diesem Beispiel glücklichen Umstand, dass f auf $\partial B_1^2(0, 0)$ konstant gleich 0 ist, zu bemerken.

Damit sind $(-\frac{1}{3}\sqrt{3}, 0)$ als einzige absolute Minimalstelle von f auf $\bar{B}_1^2(0, 0)$ und $(\frac{1}{3}\sqrt{3}, 0)$ als einzige absolute Maximalstelle von f auf $\bar{B}_1^2(0, 0)$ identifiziert.

(3) Als drittes Beispiel betrachten wir auf der rechten Halbebene $(0, \infty) \times \mathbb{R} \subset \mathbb{R}^2$ die Funktion $f: (0, \infty) \times \mathbb{R} \rightarrow \mathbb{R}$ mit

$$f(x, y) := x^2 y^2 + 2xy + x + \frac{1}{x} \quad \text{für } (x, y) \in A = (0, \infty) \times \mathbb{R}.$$

(I) Da $(0, \infty) \times \mathbb{R}$ nicht kompakt ist, ist der Extreimalsatz nicht direkt anwendbar. Aufgrund der Stetigkeit von f und von $\lim_{A \ni (x, y) \rightarrow (0, y_0)} f(x, y) = \infty$ für alle $y_0 \in \mathbb{R}$ samt⁶ $\lim_{|(x, y)| \rightarrow \infty} f(x, y) = \infty$ ist aber einerseits klar, dass es keine absolute Maximalstelle von f auf $(0, \infty) \times \mathbb{R}$ gibt, während andererseits die Existenz einer absoluten Minimalstelle von f auf $(0, \infty) \times \mathbb{R}$ gemäß der Erweiterung des Extreimalsatzes folgt.

(II) Man berechnet $\nabla f(x, y) = (2xy^2 + 2y + 1 - \frac{1}{x^2}, 2x^2 y + 2x)$ und erhält das Gleichungssystem $2xy^2 + 2y + 1 - \frac{1}{x^2} = 0$, $x(xy + 1) = 0$ für innere kritische Punkte. Da x im Definitionsbereich positiv ist, gibt die zweite Gleichung $x = -\frac{1}{y}$ mit negativem y . Einsetzen in die erste Gleichung führt auf $1 - y^2 = 0$, so dass $(1, -1)$ als einziger innerer Kandidat verbleibt. Nichtdifferenzierbarkeitsstellen oder Randkandidaten gibt es auch hier nicht, da f differenzierbar und $(0, \infty) \times \mathbb{R}$ offen ist.

(III) Da es nur einen Kandidaten gibt, ist kein Wertevergleich nötig.

Insgesamt ist nachgewiesen, dass es keine absolute Maximalstelle von f auf $(0, \infty) \times \mathbb{R}$ gibt und $(1, -1)$ die einzige absolute Minimalstelle von f auf $(0, \infty) \times \mathbb{R}$ ist.

Als Nächstes wenden wir uns Kriterien für **nicht-innere Extremstelle** und **Extremstellen auf nieder-dimensionalen Mengen** zu, und werden damit insbesondere Möglichkeiten zur Bestimmung von **Randkandidaten für Extremstellen** erhalten. Das Vorgehen hierbei unterscheidet sich stark, je nachdem ob von einer **parametrischen Darstellung** oder **impliziten Darstellung** der relevanten Menge im Definitionsbereich ausgeht. Wir beginnen mit der Diskussion der parametrischen Situation:

Bemerkung (zur **Bestimmung von Extremstellen bei parametrischer Darstellung**).

Sei $f: D \rightarrow \mathbb{R}$ eine \mathbb{R} -wertige Funktion auf offenem $D \subset \mathbb{R}^n$, und sei

$$S = D \cap u(\Sigma) = \{x \in D \mid \exists p \in \Sigma: u(p) = x\}$$

in **parametrischer Darstellung** mit offenem Parameterbereich $\Sigma \subset \mathbb{R}^\ell$ für $\ell \in \{1, 2, \dots, n-1\}$ und Parametrisierung $u: \Sigma \rightarrow \mathbb{R}^n$ gegeben. In der Regel ist dann S eine „ **ℓ -dimensionale**“⁷

⁶Zum präzisen Nachweis von $\lim_{|(x, y)| \rightarrow \infty} f(x, y) = \infty$ betrachtet man eine beliebige Folge $(x_k, y_k)_{k \in \mathbb{N}}$ in A mit $\lim_{k \rightarrow \infty} |(x_k, y_k)| = \infty$, für die ohne Einschränkung Existenz von $x_* := \lim_{k \rightarrow \infty} x_k \in [0, \infty]$ angenommen werden kann. Nach Umschreiben $x_k^2 y_k^2 + 2x_k y_k = (x_k y_k + 1)^2 - 1$ erhält man dann $\lim_{k \rightarrow \infty} f(x_k, y_k) = \infty$ einerseits im Fall $x_* = \infty$ aus der Abschätzung $f(x_k, y_k) \geq x_k - 1$, andererseits im Fall $0 < x_* < \infty$ aus $f(x_k, y_k) \geq (x_k y_k + 1)^2 - 1$ samt der Beobachtung, dass dann $\lim_{k \rightarrow \infty} |y_k| = \infty$ gilt, und schließlich im Fall $x_* = 0$ aus $f(x_k, y_k) \geq \frac{1}{x_k} - 1$.

⁷Der Dimensionsbegriff der linearen Algebra greift hier nicht, da S normalerweise kein (Unter-)Vektorraum ist. Vielmehr verstehen wir die Dimension von S in einem naiven Sinn, in dem eine „normale“ Kurve 1-dimensional, eine „normale“ Fläche 2-dimensional, der „normale“ Raum 3-dimensional ist.

Fläche“ mit leerem Innern $\overset{\circ}{S} = \emptyset$ in \mathbb{R}^n , so dass die bisher behandelten Kriterien für innere Extremstellen zur Untersuchung von Extremstellen *auf* S *nicht* greifen. Dennoch stellt dieser Fall (prinzipiell) kein Problem dar: Wenn u zumindest stetig und an den meisten Stellen differenzierbar ist, so ist nämlich $u^{-1}(D)$ offen in \mathbb{R}^ℓ , man kann die **Extremstellen der Komposition** $f \circ u$ auf $u^{-1}(D)$ als innere Extremstellen **„normal“ bestimmen**, und als die Bilder dieser Extremstellen unter u erhält man dann genau die gesuchten Extremstellen von f auf S .

Beispiel (zur Bestimmung von Extremstellen bei parametrischer Darstellung). Im Fall $n = 2$, $D = \mathbb{R}^2$ und $\ell = 1$, $\Sigma = \mathbb{R}$ der vorausgehenden Bemerkung betrachten wir die Funktion $f: \mathbb{R}^2 \rightarrow \mathbb{R}$ und die Kurve $u: \mathbb{R} \rightarrow \mathbb{R}^2$ mit

$$f(x, y) := 4x - 3(y^2 + 1)^2 \quad \text{für } (x, y) \in \mathbb{R}^2 \quad \text{und} \quad u(t) := (2t^3, t) \quad \text{für } t \in \mathbb{R},$$

und wir suchen die Extremstellen von f auf der Spur $S = u(\mathbb{R})$ der Kurve u . Durch Einsetzen erhalten wir für $f \circ u: \mathbb{R} \rightarrow \mathbb{R}$ erst $(f \circ u)(t) = 8t^3 - 3(t^2 + 1)^2$ für $t \in \mathbb{R}$, dann (nach Umformung/Rechnung) $(f \circ u)'(t) = -12t(t-1)^2$ für $t \in \mathbb{R}$. Also sind 0 und 1 die einzigen kritischen Punkte von $f \circ u$. Aus $(f \circ u)(0) = -3$ und $(f \circ u)(1) = -4$ zusammen mit $\lim_{|t| \rightarrow \infty} (f \circ u)(t) = -\infty$ folgt, dass 0 Maximalstelle von $f \circ u$ ist und keine Minimalstelle von $f \circ u$ existiert. Folglich ist $u(0) = (0, 0)$ Maximalstelle von f auf $S = u(\mathbb{R})$, und es gibt keine Minimalstelle von f auf S .

Als Nächstes behandeln wir den schon angekündigten Fall einer **impliziten Darstellung** und formulieren ein für diesen Fall geeignetes Kriterium wie folgt.

Satz (notwendiges Kriterium für Extremstellen bei Nebenbedingungen). Für offenes $D \subset \mathbb{R}^n$ sei

$$S = \{x \in D \mid g_i(x) = 0 \text{ für } i = 1, 2, \dots, m\}$$

in **impliziter Darstellung** durch $m \in \mathbb{N}$ Gleichungen $g_i(x) = 0$, genannt **Nebenbedingungen**, mit Funktionen $g_i: D \rightarrow \mathbb{R}$ beschrieben, und zudem sei $f: D \rightarrow \mathbb{R}$ gegeben. Ist dann $a \in S$ eine lokale Extremstelle von f auf S , ist f in a differenzierbar, und sind g_1, g_2, \dots, g_m nahe a von der Klasse C^1 , so sind die Gradienten $\nabla f(a), \nabla g_1(a), \nabla g_2(a), \dots, \nabla g_m(a)$ **linear unabhängig**.

Ein Beweis des Satzes (mit Vorgriff auf den Satz über implizite Funktionen) wurde in der Vorlesung am Ende des Abschnitts ausgeführt.

Bemerkungen (zur Extremstellenbestimmung bei Nebenbedingungen und am Rand).

- (1) Von Interesse ist der Satz im Fall $m \leq n$ (höchstens so viele Nebenbedingungen wie Variablen), denn für $\ell := n - m \in \{0, 1, 2, \dots, n - 1\}$ ist dann S in der Regel wieder eine **„ ℓ -dimensionale Fläche**“ mit leerem Innern $\overset{\circ}{S} = \emptyset$ in \mathbb{R}^n . Im wichtigsten Fall sind zudem für $x \in S$ die m Gradienten $\nabla g_1(x), \nabla g_2(x), \dots, \nabla g_m(x)$ **linear unabhängig**. Man kann sich dann vorstellen, dass diese m Vektoren an der Stelle x senkrecht auf der Niveaumenge S von (g_1, g_2, \dots, g_m) stehen und einen m -dimensionalen Untervektorraum V_x von \mathbb{R}^n aufspannen, so dass V_x (oder eigentlich der affine Unterraum $x + V_x$) an der Stelle x senkrecht auf S steht. Für lokale Extremstellen x von f auf S gilt neben $x \in S$ nach dem Kriterium des Satzes dann $\nabla f(x) \in V_x$, was ausgeschrieben dem **System von $n + m$ im Allgemeinen nicht-linearen Gleichungen**

$$\boxed{\begin{aligned} \nabla f(x) &= \lambda_1 \nabla g_1(x) + \lambda_2 \nabla g_2(x) + \dots + \lambda_m \nabla g_m(x), \\ g_1(x) &= 0, \quad g_2(x) = 0, \quad \dots, \quad g_m(x) = 0 \end{aligned}} \quad (*)$$

für insgesamt $(n+m)$ Variablen $x = (x_1, x_2, \dots, x_n) \in D \subset \mathbb{R}^n$ und $\lambda_1, \lambda_2, \dots, \lambda_m \in \mathbb{R}$ entspricht. In der Praxis möchte man **aus diesem System** durch sukzessive Elimination von Variablen die **Kandidatenstellen** x für Extremstellen von f auf S **berechnen**, während die sogenannten **Lagrange-Multiplikatoren** $\lambda_1, \lambda_2, \dots, \lambda_m$ weniger relevant sind und beim Lösen des Systems in der Regel als Erste eliminiert werden sollten. Prinzipiell kann man als Lösungen (und damit Kandidaten für Extremstellen) übrigens eine beliebige Anzahl von Punkten $x \in S$ erhalten (keinen, genau einen, endlich oder unendlich viele), und anders als bei den linearen Gleichungssystemen der linearen Algebra bilden die Lösungen hier im Allgemeinen auch keinen (affinen) Unterraum.

Zusätzlich zu den Lösungen des Gleichungssystems sind in die Liste der Kandidatenstellen auch **Nicht- C^1 -Stellen** von f und den g_i auf S sowie **Stellen linearer Abhängigkeit von $\nabla g_1, \nabla g_2, \dots, \nabla g_m$** auf S aufzunehmen. In diese Kategorie fallen unter anderem Ecken, Kanten und andere geometrisch ausgezeichnete Stellen von S .

- (2) Insbesondere nützen der Satz und das aufbauende, gerade erläuterte Rechenverfahren im **Fall $m = 1$** einer einzigen Nebenbedingung oft zur **Bestimmung von Randkandidaten für Extremstellen**. Genauer greift dieses Vorgehen bei $A \subset \mathbb{R}^n$ für $(n-1)$ -dimensionale Teile des Rands ∂A , die $\partial A \subset A$ erfüllen und eine implizite Darstellung

$$D \cap \partial A = \{x \in D \mid g(x) = 0\}$$

mit einer Funktion $g: D \rightarrow \mathbb{R}$ auf offenem $D \subset \mathbb{R}^n$ haben (wie sie sich typischerweise zugleich mit $D \cap A = \{x \in D \mid g(x) \geq 0\}$ ergibt). Es folgt dann, dass C^1 -Stellen $x \in D \cap \partial A$ nur als lokale Extremstellen von f auf A in Frage kommen, wenn $\nabla f(x)$ und $\nabla g(x)$ linear abhängig sind. Mit anderen Worten gilt es in die Kandidatenliste als **Randkandidaten** im Randteil $D \cap \partial A$ neben eventuellen Nicht- C^1 -Stellen alle Nullstellen von ∇g in $D \cap \partial A$ und alle **Lösungen $x \in D$ des Systems von $n+1$ Gleichungen**

$$\nabla f(x) = \lambda \nabla g(x), \quad g(x) = 0$$

für die n Variablen $x \in D$ und einen einzelnen Lagrange-Multiplikator $\lambda \in \mathbb{R}$ aufzunehmen.

Beispiel (für eine **Extremstellenbestimmung am Rand**). Als einfaches Beispiel zur Bestimmung von Randkandidaten für Extremstellen untersuchen wir

$$\text{auf dem Dreieck } A := \{(x, y) \in \mathbb{R}^2 \mid x + y \geq 0, \max\{x, y\} \leq 1\}$$

die Funktion $f: \mathbb{R}^2 \rightarrow \mathbb{R}$ mit

$$f(x, y) := x^3 - y^2 \quad \text{für } (x, y) \in \mathbb{R}^2$$

anhand der bisher entwickelten Verfahrensweisen:

- (I) Da A kompakt ist, wird die Existenz einer absoluten Minimal- und einer absoluten Maximalstelle durch den Extremalsatz sichergestellt
- (II) Der Gradient $\nabla f(x, y) = (3x^2, -2y)$ verschwindet nur für $(x, y) = (0, 0)$. Da $(0, 0)$ nicht im Innern, sondern auf dem Rand von A liegt, gibt es keine inneren Kandidaten für Extremstellen.

Als Nächstes fassen wir das Randstück $\partial_{(1)}A := \{(x, y) \in (-1, 1)^2 \mid x + y = 0\}$ ins Auge. Eine Vorgehensweise besteht darin, $\partial_{(1)}A$ durch $u(t) := (t, -t)$ über $\Sigma = (-1, 1)$ zu parametrisieren. Dann zeigt Differentialrechnung einer Variablen, dass die kritischen Punkte für $f(u(t)) = t^3 - t^2$ genau bei $t = 0$ und $t = \frac{2}{3}$ liegen, und wir erhalten für Randextremstellen von f in $\partial_{(1)}A$ die Kandidaten $u(0) = (0, 0)$ und $u(\frac{2}{3}) = (\frac{2}{3}, -\frac{2}{3})$. Alternativ kann man $\partial_{(1)}A$ implizit durch die Nebenbedingung $g(x, y) := x + y = 0$ darstellen. Dann ist $\nabla g \equiv (1, 1)$, und $\nabla f(x, y) = (3x^2, -2y)$ ist genau dann linear abhängig von $(1, 1)$, wenn die beiden Einträge von $\nabla f(x, y)$ übereinstimmen. Aus dem resultierenden Gleichungssystem $3x^2 = -2y$, $x + y = 0$ erhält man wiederum die Kandidaten $(x, y) = (0, 0)$ und $(x, y) = (\frac{2}{3}, -\frac{2}{3})$.

Die Kandidaten auf den Randstücken $\partial_{(2)}A := (-1, 1) \times \{1\}$ und $\partial_{(3)}A := \{1\} \times (-1, 1)$ kann man analog berechnen. Da es sich um achsenparallele Strecken handelt, laufen alle Vorgehensweise einfach darauf hinaus, die Nullstellen von $\partial_1 f$ auf $\partial_{(2)}A$ und von $\partial_2 f$ auf $\partial_{(3)}A$ zu bestimmen. Man erhält die Kandidaten $(0, 1) \in \partial_{(2)}A$ und $(1, 0) \in \partial_{(3)}A$.

Neben den Randstücken $\partial_{(1)}A$, $\partial_{(2)}A$, $\partial_{(3)}A$ gehören zum Rand ∂A noch die Ecken $(-1, 1)$, $(1, -1)$, $(1, 1)$ des Dreiecks, die zusätzlich in die Kandidatenliste aufzunehmen sind.

(III) Schließlich vergleichen wir die Funktionswerte in den sieben ermittelten Kandidaten:

$$\begin{aligned} f(0, 0) = 0, \quad f\left(\frac{2}{3}, -\frac{2}{3}\right) = -\frac{4}{27}, \quad f(0, 1) = -1, \quad f(1, 0) = 1, \\ f(-1, 1) = -2, \quad f(1, -1) = 0, \quad f(1, 1) = 0. \end{aligned}$$

Wir lesen ab, dass $(-1, 1)$ die einzige absolute Minimalstelle und $(1, 0)$ die einzige absolute Maximalstelle von f auf dem Dreieck A ist.

2.3 Der Umkehrsatz und der Satz über implizite Funktionen

In diesem Abschnitt verallgemeinern wir zunächst die Regel für die Ableitung der Umkehrfunktion zum allgemeinen Umkehrsatz für Funktionen mehrerer Variablen.

Hauptsatz (Umkehrsatz). Für $D \subset \mathbb{R}^n$ sei $f: D \rightarrow \mathbb{R}^n$ zumindest auf einer Umgebung von $a \in \overset{\circ}{D}$ eine C^1 -Funktion. Ist die totale Ableitung $f'(a)$ in $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^n)$ invertierbar, so gibt es offene Umgebungen U von a und V von $f(a)$ in \mathbb{R}^n , so dass $f|_U: U \rightarrow V$ bijektiv und C^1 ist und eine **C^1 -Umkehrfunktion** $f^{-1}: V \rightarrow U$ hat. Zudem gilt dann die **Regel für die Ableitung der Umkehrfunktion**

$$(f^{-1})'(y) = [f'(f^{-1}(y))]^{-1} \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^n) \quad \text{für alle } y \in V.$$

Bemerkungen (zum Umkehrsatz).

(1) Der Satz kann auch mit Funktionalmatrizen anstelle totaler Ableitungen formuliert werden, und die Ableitungsregel lautet dann

$$D(f^{-1})(y) = [Df(f^{-1}(y))]^{-1} \in \mathbb{R}^{n \times n} \quad \text{für alle } y \in V.$$

Die -1 an der eckigen Klammer steht dabei für die Bildung der inversen linearen Abbildung (im Satz) bzw. der inversen Matrix (in der gerade angegebenen Variante) und wirkt jeweils „punktweise“ nach Einsetzen eines festen y . Die -1 an der eckigen Klammer spielt also eine etwas andere Rolle als die -1 bei der Umkehrfunktion f^{-1} und reduziert sich im Fall $n = 1$ einfach auf die Reziproken-Bildung in der bekannten Regel für Funktionen einer Variablen.

- (2) Die **vorausgesetzte Invertierbarkeit** der linearen Abbildung $f'(a): \mathbb{R}^n \rightarrow \mathbb{R}^n$ oder äquivalent der quadratischen Matrix $Df(a) \in \mathbb{R}^{n \times n}$ **ist wesentlich**. Ein **notwendiges und hinreichendes Kriterium** hierfür ist bekanntlich

$$\boxed{\det(Df(a)) \neq 0}.$$

- (3) Der Satz gilt analog für Banach-Räume \mathcal{X}, \mathcal{Y} und Funktionen $f: D \rightarrow \mathcal{Y}$ auf $D \subset \mathcal{X}$ unter der Voraussetzung, dass $f'(a) \in \mathcal{L}(\mathcal{X}, \mathcal{Y})$ eine Inverse $[f'(a)]^{-1} \in \mathcal{L}(\mathcal{Y}, \mathcal{X})$ besitzt.
- (4) Sind A, B offene Teilmengen normierter Räume, so nennt man eine bijektive C^1 -Abbildung $f: A \rightarrow B$ mit C^1 -Umkehrabbildung $f^{-1}: B \rightarrow A$ einen **(C^1 -)Diffeomorphismus** von A auf B . In dieser Terminologie ist die Aussage des Umkehrsatzes, dass $f|_U$ für (ausreichend kleine) Umgebungen U von a und V von $f(a)$ ein C^1 -Diffeomorphismus von U auf V ist. Etwas verkürzt spricht man auch davon, dass f *lokal* bei a ein C^1 -Diffeomorphismus ist.
- (5) Ist $f: D \rightarrow \mathbb{R}^n$ lokal bei jedem Punkt eines offenen $D \subset \mathbb{R}^n$ ein C^1 -Diffeomorphismus, so muss f noch nicht unbedingt ein C^1 -Diffeomorphismus auf ganz D sein. Dies zeigt das Beispiel der komplexen Exponentialfunktion (in reellen Koordinaten)

$$\exp: \mathbb{R}^2 \rightarrow \mathbb{R}^2, (x, y) \mapsto (e^x \cos y, e^x \sin y),$$

die lokaler Diffeomorphismus von \mathbb{R}^2 auf $\mathbb{R}^2 \setminus \{0\}$, aber nicht global injektiv ist (denn es gilt $\det(D \exp(x, y)) = e^x \neq 0$ für alle $(x, y) \in \mathbb{R}^2$, aber zum Beispiel $\exp(0, 0) = 1 = \exp(0, 2\pi)$). Es gibt Kriterien für globale statt nur lokaler Umkehrbarkeit, die man aber eher selten braucht und die wir im Rahmen dieser Vorlesung nicht behandeln.

Ein Beweis des Umkehrsatzes, der auf einer Anwendung des Banachschen Fixpunktsatzes basiert, wurde in der Vorlesung behandelt.

Als Nächstes kommen wir zu einer Erweiterung des Umkehrsatzes, auf die bereits im vorigen Abschnitt zurückgegriffen wurde.

Hauptsatz (Satz über implizite Funktionen). Sei $D \subset \mathbb{R}^n = \mathbb{R}^\ell \times \mathbb{R}^m$ mit $n = \ell + m$ für $\ell, m \in \{1, 2, \dots, n-2, n-1\}$, sei $a = (a', a'') \in \overset{\circ}{D} \subset \mathbb{R}^\ell \times \mathbb{R}^m$, und sei $f: D \rightarrow \mathbb{R}^m$ zumindest auf einer Umgebung von a eine C^1 -Funktion mit einer Nullstelle in a , also mit $f(a) = 0$. Ist die „partielle totale Ableitung“ $\frac{\partial f}{\partial x''}(a)$ nach den letzten m Variablen $x'' \in \mathbb{R}^m$ (gemeint ist damit die totale Ableitung von $\mathbb{R}^m \rightarrow \mathbb{R}^m, x'' \mapsto f(a', x'')$ an der Stelle a'') in $\mathcal{L}(\mathbb{R}^m, \mathbb{R}^m)$ **invertierbar**, so gibt es offene Umgebungen U' von a' in \mathbb{R}^ℓ und U'' von a'' in \mathbb{R}^m sowie eine **C^1 -Funktion** $g: U' \rightarrow U''$, so dass für alle $(x', x'') \in D$ mit $x' \in U'$ gilt:

$$\boxed{x'' \in U'', f(x', x'') = 0 \iff x'' = g(x')}.$$

Bemerkungen (zum Satz über implizite Funktionen).

- (1) Der Satz besagt, dass man das **unterbestimmte System** $f(x', x'') = 0$ **von m** im Allgemeinen **nicht-linearen (Komponenten-)Gleichungen** für die $n = \ell + m$ Variablen $x = (x', x'')$ **lokal eindeutig nach** den m Variablen x'' **auflösen** kann. Ähnlich wie bei unterbestimmten *linearen* Gleichungssystemen kann man also einen Teil der Variablen, hier die ℓ Variablen x' , vorgeben und daraus die anderen Variablen, hier die m Variablen x'' , berechnen. Anders

als im linearen Fall kann man für nicht-lineare Gleichungssysteme aber auch bei gegebenem x' keine globale Eindeutigkeit von x'' , sondern nur lokale Eindeutigkeit innerhalb der tendenziell kleinen Umgebung U'' erwarten. Dies ist der Grund, warum obige Äquivalenz nur mit der einschränkenden Bedingung $x'' \in U''$ auf der linken Seite allgemein gelten kann.

- (2) Die Funktion g im Satz bezeichnet man als (lokal) **auffösende Funktion** des Gleichungssystems $f(x', x'') = 0$ oder **implizit** durch $f(x', x'') = 0$ **gegebene** Funktion. Sie erfüllt insbesondere $g(a') = a''$ und

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

- (3) Die **vorausgesetzte Invertierbarkeit** der partiellen totalen Ableitung $\frac{\partial f}{\partial x''}(a): \mathbb{R}^m \rightarrow \mathbb{R}^m$ oder äquivalent Invertierbarkeit der partiellen Funktionalmatrix $D_{x''}f(a) \in \mathbb{R}^{m \times m}$ (die aus den x'' -Spalten von $Df(a) \in \mathbb{R}^{m \times n}$ besteht) **ist wesentlich**. Als **notwendiges und hinreichendes Kriterium** hierfür ergibt sich

$$\boxed{\det(D_{x''}f(a)) \neq 0}.$$

- (4) Durch Ableiten von $f(x', g(x')) = 0$ für $x' \in U'$ mit der Kettenregel und Auflösen erhält man die Formel $g'(x') = -\left[\frac{\partial f}{\partial x''}(x', g(x'))\right]^{-1} \frac{\partial f}{\partial x'}(x', g(x'))$ für die totale Ableitung g' beziehungsweise $Dg(x') = -\left[D_{x''}f(x', g(x'))\right]^{-1} D_{x'}f(x', g(x'))$ für die Funktionalmatrix Dg .

- (5) Ist $\frac{\partial f}{\partial x''}(a)$ nicht invertierbar, aber hat $f'(a) \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ beziehungsweise $Df(a) \in \mathbb{R}^{m \times n}$ den **maximal möglichen Rang m** , so kann man stets m linear unabhängige Spalten von $Df(a)$ finden. Analog zum Satz kann man dann statt nach den letzten m Variablen x'' **nach einem anderen Set von m Variablen auflösen** (und tatsächlich nach *jedem* Set von m Variablen, das zu m linear unabhängigen Spalten gehört. Formal ergibt sich dies durch Anwendung des Satzes nach geeigneter Permutation der Variablen.

- (6) **Geometrisch** besagt der Satz, dass die **Nullstellenmenge** $\{(x', x'') \in D \mid f(x', x'') = 0\}$ der C^1 -Funktion $f: D \rightarrow \mathbb{R}^m$ zumindest lokal (d.h. geschnitten mit $U' \times U''$) **als Graph** $\{(x', g(x')) \mid x' \in U'\}$ der C^1 -Funktion $g: U' \rightarrow \mathbb{R}^m$ dargestellt ist. Mit anderen Worten ist die Graphenabbildung $u: U' \rightarrow \mathbb{R}^n$, $x' \mapsto (x', u(x'))$ eine lokale C^1 -Parametrisierung der Nullstellenmenge mit maximalem Rang; sie erfüllt $u(a') = a$ sowie $f(u(x')) = 0$ und $\text{Rang}(Du(x')) = \ell$ für alle $x' \in U'$, und dies ging tatsächlich beim Beweis des Kriteriums für Extremstellen bei Nebenbedingungen in Abschnitt 2.2 bereits ein.

- (7) Unter den gleichen Voraussetzungen an f (abgesehen nur davon, dass $f(a) = 0$ nicht mehr benötigt wird) kann man auch die **Gleichung $f(x', x'') = y$ mit allgemeiner rechter Seite** $y \in \mathbb{R}^m$ lokal zu $x'' = \tilde{g}(x', y)$ auflösen, wobei $\tilde{g}: U' \times V \rightarrow U''$ eine C^1 -Funktion ist, U' und U'' dieselben Umgebungen wie zuvor bezeichnen und V eine jetzt zusätzlich benötigte Umgebung von $f(a)$ in \mathbb{R}^m ist. Insbesondere gelten damit $\tilde{g}(a', f(a)) = a''$ und $f(x', \tilde{g}(x', y)) = y$ für alle $(x', y) \in U' \times V$, und geometrisch hat man damit nicht nur die Nullstellenmenge, sondern eine ganze Schar von Niveaumengen $\{x \in D \mid f(x) = y\}$ mit $|y - f(a)| \ll 1$ als Graphen von C^1 -Funktionen dargestellt.

Diese Variante des Satzes, die C^1 -Abhängigkeit der Lösung und der Graphendarstellung von der rechten Seite y beinhaltet, gewinnt man durch Anwendung der obigen Version auf $\tilde{f}((x', y), x'') := f(x', x'') - y$ (oder kann sie alternativ auch direkter aus dem Beweis ablesen).

Ein Vorteil dieser Betrachtungsweise besteht darin, dass man den Umkehrsatz als genaues Analogon des Satzes über implizite Funktionen im Fall $\ell = 0$, $m = n$ erkennt: Streicht man nämlich alle Auftreten von x' , so entspricht \tilde{g} gerade einer lokalen Umkehrfunktion zu f .

- (8) Der Satz gilt allgemeiner für Banach-Räume \mathcal{X}' , \mathcal{X}'' , \mathcal{Y} und Funktionen $f: D \rightarrow \mathcal{Y}$ auf $D \subset \mathcal{X}' \times \mathcal{X}''$. Wie beim Umkehrsatz muss man dann voraussetzen, dass $\frac{\partial f}{\partial x''}(a) \in \mathcal{L}(\mathcal{X}'', \mathcal{Y})$ eine Inverse in $\mathcal{L}(\mathcal{Y}, \mathcal{X}'')$ besitzt.

Beispiele (zum Satz über implizite Funktionen). Wir betrachten nur den Fall $\ell = n-1$, $m = 1$, in dem f eine skalare Funktion ist, $\frac{\partial f}{\partial x''}(a) \in \mathcal{L}(\mathbb{R}, \mathbb{R})$ bzw. $D_{x''}f(a) \in \mathbb{R}^{1 \times 1}$ einer Zahl $\frac{\partial f}{\partial x''}(a) \in \mathbb{R}$ entspricht und sich die Invertierbarkeitsvoraussetzung zu $\frac{\partial f}{\partial x''}(a) \neq 0$ reduziert.

- (1) Die Gleichung

$$f(x) := (x'')^2 - |x'|^3 = 0 \quad \text{für } x \in \mathbb{R}^n$$

kann wegen $\frac{\partial f}{\partial x''}(x) = 2x''$ nahe allen Nullstellen $(x', x'') \in \mathbb{R}^{n-1} \times \mathbb{R}$ von f mit $x'' \neq 0$ lokal durch eine C^1 -Funktion nach x'' aufgelöst werden. Die aufgelöste Gleichung lautet $x'' = \pm |x'|^{3/2} := g(x')$, wobei für $x'' > 0$ das Vorzeichen Plus, für $x'' < 0$ das Vorzeichen Minus zu wählen ist. Global gibt es also zu $x' \neq 0$ stets zwei Lösungen x'' , nur lokal ist x'' eindeutig. Nahe der Nullstelle $0_{\mathbb{R}^n}$ von f und f' dagegen ist kein Auflösen möglich (auch nicht nach einer der x' -Variablen). Dies kann man für $n = 2$ auch gut an einer Skizze der Nullstellenmenge erkennen, die sich im Ursprung verzweigt und dort eben nicht die lokale Struktur eines Graphen hat.

- (2) Die Gleichung

$$f(x) := |x'| - x'' \log |x''| = 0 \quad \text{für } x \in D$$

betrachten wir auf dem Definitionsbereich $D := \{(x', x'') \in \mathbb{R}^{n-1} \times \mathbb{R} \mid x' \neq 0, x'' \neq 0\}$ betrachtet, auf dem eine C^1 -Funktion f vorliegt. Wegen $\frac{\partial f}{\partial x''}(x) = -1 - \log |x''|$ ist nahe allen Nullstellen $(x', x'') \in D$ von f mit $|x''| \neq \frac{1}{e}$ das lokale Auflösen zu $x'' = g(x')$ möglich, die auflösende Funktion g lässt sich aber nicht explizit angeben. Als Nullstellen von f in D mit $|x''| = \frac{1}{e}$, bei denen man nicht nach x'' auflösen kann, verbleiben $(\frac{1}{e}, -\frac{1}{e})$ und $(-\frac{1}{e}, -\frac{1}{e})$. Ersatzweise ist bei diesen Stellen aber ein lokales Auflösen nach jeder der x' -Variablen möglich, denn $\frac{\partial f}{\partial x'_1}, \frac{\partial f}{\partial x'_2}, \dots, \frac{\partial f}{\partial x'_{n-1}}$ sind dort alle ungleich Null. Auch bei diesem Beispiel ist eine Skizze der Nullstellenmenge instruktiv.

Ein Beweis des Satzes über implizite Funktionen, bei dem man triviale Komponentenfunktionen ergänzt, auf diese Weise ein voll bestimmtes Gleichungssystem herstellt und die Aussage dann auf den Umkehrsatz zurückführt, wurde in der Vorlesung ausgeführt.

2.4 Ableitungen zweiter und höherer Ordnung

In diesem Abschnitt erweitern wir auch Ableitungen zweiter und höherer Ordnung auf den Fall mehrerer Variablen.

Naheliegender ist zunächst die Bildung iterierter partieller Ableitungen:

Definitionen (partielle Ableitungen höherer Ordnung, Hesse-Matrizen). Sei $f: A \rightarrow \mathcal{Y}$ eine Funktion von $A \subset \mathbb{R}^n$ in einen normierten Raum \mathcal{Y} , und sei $a \in A$.

(I) Existiert für $\ell \in \mathbb{N}$ und $i_1, i_2, \dots, i_{\ell-1}, i_\ell \in \{1, 2, \dots, n\}$ die iterierte partielle Ableitung

$$\partial_{i_\ell} \partial_{i_{\ell-1}} \dots \partial_{i_2} \partial_{i_1} f(a) := \partial_{i_\ell} (\partial_{i_{\ell-1}} \dots \partial_{i_2} (\partial_{i_1} f) \dots)(a) \in \mathcal{Y}$$

(was insbesondere die Existenz der Ableitungen $\partial_{i_1} f, \partial_{i_2} \partial_{i_1} f, \dots, \partial_{i_{\ell-1}} \dots \partial_{i_2} \partial_{i_1} f$ bis Ordnung $\ell-1$ auf einer Umgebung von a erfordert), so spricht man von einer **partiellen Ableitung ℓ -ter Ordnung** von f an der Stelle a /in a /bei a . Speziell heißt

$$\partial_i^\ell f(a) := \underbrace{\partial_i \partial_i \dots \partial_i}_{\ell\text{-mal}} f(a)$$

mit $\ell \in \mathbb{N}$, $i \in \{1, 2, \dots, n\}$ die **reine partielle Ableitung ℓ -ter Ordnung** von f an der Stelle a /in a /bei a nach der i -ten Variablen.

(II) Existieren speziell im \mathbb{R} -wertigen Fall $\mathcal{Y} = \mathbb{R}$ alle partiellen Ableitungen zweiter Ordnung von f in a , so heißt

$$Hf(a) := D(\nabla f)(a) = (\partial_j \partial_i f(a))_{i,j=1,2,\dots,n} \in \mathbb{R}^{n \times n}$$

die **Hesse-Matrix** von f an der Stelle a /in a /bei a . Ausgeschrieben bedeutet dies

$$Hf(a) = \begin{pmatrix} \partial_1^2 f(a) & \partial_2 \partial_1 f(a) & \cdots & \partial_{n-1} \partial_1 f(a) & \partial_n \partial_1 f(a) \\ \partial_1 \partial_2 f(a) & \partial_2^2 f(a) & \cdots & \partial_{n-1} \partial_2 f(a) & \partial_n \partial_2 f(a) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \partial_1 \partial_{n-1} f(a) & \partial_2 \partial_{n-1} f(a) & \cdots & \partial_{n-1}^2 f(a) & \partial_n \partial_{n-1} f(a) \\ \partial_1 \partial_n f(a) & \partial_2 \partial_n f(a) & \cdots & \partial_{n-1} \partial_n f(a) & \partial_n^2 f(a) \end{pmatrix}.$$

Gemäß dem nächsten Satz ist die Reihenfolge, in der die einzelnen Ableitungsoperatoren ∂_{i_k} in der vorausgehenden Definition angewandt werden, „meistens“ irrelevant.

Satz (über Vertauschbarkeit partieller Ableitungen). Seien $f: A \rightarrow \mathcal{Y}$ eine Funktion von $A \subset \mathbb{R}^n$ in einen normierten Raum \mathcal{Y} und $a \in \overset{\circ}{A}$. Existieren für $i, j \in \{1, 2, \dots, n\}$ die partiellen Ableitungen $\partial_i f$ und $\partial_j f$ auf einer Umgebung von a , so gilt

$$\boxed{\partial_i \partial_j f(a) = \partial_j \partial_i f(a)},$$

vorausgesetzt eine der folgenden schwachen Zusatzvoraussetzungen ist erfüllt:

- (a) Die partielle Ableitung $\partial_j \partial_i f$ existiert auf einer Umgebung von a und ist in a stetig.
- (b) Die beiden totalen Ableitungen $(\partial_v f)'(a)$ und $(\partial_w f)'(a)$ existieren.

Unter Voraussetzung (a) ist der Satz auch als **Satz von Schwarz** (nach H.A. Schwarz) bekannt. Ein Beweis (für beide angegebenen Fälle) auf Basis der Definition der (totalen) Ableitung und des Schrankensatzes wurde in der Vorlesung geführt.

Mit anderen Worten drückt der Satz genau folgende Eigenschaft der Hesse-Matrix aus.

Korollar (Symmetrie der Hesse-Matrix). Sei $f: A \rightarrow \mathcal{Y}$ eine Funktion von $A \subset \mathbb{R}^n$ in einen normierten Raum \mathcal{Y} . Wenn alle partiellen Ableitungen $\partial_i \partial_j f$ mit $i, j \in \{1, 2, \dots, n\}$ auf einer Umgebung eines $a \in \overset{\circ}{A}$ existieren und in a stetig sind, so gilt

$$\boxed{Hf(a)^T = Hf(a)}.$$

Bemerkung (zur **Rotationsfreiheit von Gradientenfeldern**). Wie in den Übungen schon früher gezeigt, ergibt sich in der Situation des Korollars auch

$$\text{Rot}(\nabla f)(a) = D(\nabla f)(a) - D(\nabla f)(a)^T = Hf(a) - Hf(a)^T = 0.$$

Somit erweist sich $\text{Rot } V \equiv 0$ als grundlegende notwendige Bedingung dafür, dass ein C^1 -Vektorfeld V ein Gradientenfeld sein kann.

Genau wie partielle Ableitungen lassen sich auch Richtungsableitungen iterieren:

Definition (Richtungsableitungen höherer Ordnung). Seien \mathcal{X}, \mathcal{Y} normierte Räume, $f: A \rightarrow \mathcal{Y}$ eine Funktion auf $A \subset \mathcal{X}$ und $a \in \mathring{A}$. Existiert für $\ell \in \mathbb{N}$ und $v_1, v_2, \dots, v_{\ell-1}, v_\ell \in \mathcal{X}$ die iterierte Richtungsableitung

$$\partial_{v_\ell} \partial_{v_{\ell-1}} \dots \partial_{v_2} \partial_{v_1} f(a) := \partial_{v_\ell} (\partial_{v_{\ell-1}} \dots \partial_{v_2} (\partial_{v_1} f) \dots)(a) \in \mathcal{Y}$$

(was insbesondere die Existenz der Ableitungen $\partial_{v_1} f, \partial_{v_2} \partial_{v_1} f, \dots, \partial_{v_{\ell-1}} \dots \partial_{v_2} \partial_{v_1} f$ bis Ordnung $\ell-1$ auf einer Umgebung von a erfordert), so spricht man von einer **Richtungsableitung ℓ -ter Ordnung** von f an der Stelle a /in a /bei a . Speziell heißt

$$\partial_v^\ell f(a) := \underbrace{\partial_v \partial_v \dots \partial_v \partial_v f(a)}_{\ell\text{-mal}}$$

mit $\ell \in \mathbb{N}$, $v \in \mathcal{X}$ die **reine Richtungsableitung ℓ -ter Ordnung** von f an der Stelle a /in a /bei a in Richtung v .

Bemerkungen (zu **Richtungsableitungen höherer Ordnung**). Seien $f: A \rightarrow \mathcal{Y}$ eine Funktion von $A \subset \mathbb{R}^n$ in einen normierten Raum \mathcal{Y} und $a \in \mathring{A}$.

(1) Auch bei beliebiger Ordnung $\ell \in \mathbb{N}$ sind partielle Ableitungen ein Spezialfall von Richtungsableitungen. Genauer gilt $\partial_{i_\ell} \partial_{i_{\ell-1}} \dots \partial_{i_2} \partial_{i_1} f(a) = \partial_{e_{i_\ell}} \partial_{e_{i_{\ell-1}}} \dots \partial_{e_{i_2}} \partial_{e_{i_1}} f(a)$ mit den kanonischen Basisvektoren $e_1, e_2, \dots, e_n \in \mathbb{R}^n$.

(2) Auch bei Richtungsableitungen ist die **Vertauschung**

$$\partial_v \partial_w f(a) = \partial_w \partial_v f(a)$$

unter zum Fall partieller Ableitungen **analogen schwachen Voraussetzungen gerechtfertigt**. Dies folgt durch Anwendung des Vertauschungs-Resultats für partielle Ableitungen auf $f(s, t) := f(a+sv+tw)$ mit $\partial_1 \partial_2 f(0, 0) = \partial_v \partial_w f(a)$ und $\partial_2 \partial_1 f(0, 0) = \partial_w \partial_v f(a)$.

Ein weiteren Ableitungsbegriff höherer Ordnung werden wir in dieser Vorlesung nur anreißen:

Bemerkung (zur **totalen Ableitung höherer Ordnung**). Für normierte Räume \mathcal{X}, \mathcal{Y} , eine Funktion $f: A \rightarrow \mathcal{Y}$ auf $A \subset \mathcal{X}$ und $a \in \mathring{A}$ kann die totale Ableitung $f^{(\ell)}(a)$ der Ordnung $\ell \in \mathbb{N}_{\geq 2}$ von f an der Stelle a iterativ als eine stetige multi-lineare Abbildung $\mathcal{X}^\ell \rightarrow \mathcal{Y}$ definiert werden. Wir gehen auf Details dieser Definition (wie etwa Stetigkeit multi-linearer Abbildungen, Norm auf dem Raum stetiger multi-linearer Abbildungen) hier nicht ein, sondern illustrieren das Konzept nur dadurch, dass bei Existenz von $f^{(\ell)}(a)$ die Formel

$$f^{(\ell)}(a)(v_1, v_2, \dots, v_\ell) = \partial_{v_\ell} \partial_{v_{\ell-1}} \dots \partial_{v_2} \partial_{v_1} f(a)$$

für alle $v_1, v_2, \dots, v_n \in \mathcal{X}$ gilt, dass also die ℓ Argumente von $f^{(\ell)}(a)$ im Wesentlichen ℓ Ableitungsrichtungen spezifizieren. Speziell für $\mathcal{X} = \mathbb{R}^n, \mathcal{Y} = \mathbb{R}$, $\ell = 2$ ist $f^{(2)}(a)$ eine Bilinearform auf \mathbb{R}^n und wird als solche durch die Hesse-Matrix $Hf(a) \in \mathbb{R}^{n \times n}$ dargestellt.

Im Folgenden ziehen wir es aber tatsächlich vor, ohne totale Ableitungen höherer Ordnung auszukommen und stattdessen weiter mit partiellen Ableitungen zu arbeiten. An einzelnen Stellen macht dies leicht stärkere Voraussetzungen erforderlich, reicht aber für so gut wie alle Zwecke völlig aus.

Die Definition von C^ℓ -Funktionen, die oft per Stetigkeitsforderung an $f^{(\ell)}$ erfolgt, formulieren wir mit dem Kriterium für C^1 -Funktionen aus Abschnitt 2.1 im Hinterkopf daher äquivalent wie folgt.

Definition (C^ℓ -Funktionen). Seien A offen in \mathbb{R}^n und \mathcal{Y} ein normierter Raum. Für $\ell \in \mathbb{N}_{\geq 2}$ nennen wir $f: A \rightarrow \mathcal{Y}$ eine **C^ℓ -Funktion** auf A , wenn alle partiellen Ableitungen ℓ -ter Ordnung $\partial_{i_\ell} \partial_{i_{\ell-1}} \dots \partial_{i_2} \partial_{i_1} f$ mit $i_1, i_2, \dots, i_\ell \in \{1, 2, \dots, n\}$ auf ganz A existieren und stetig sind. Den **Raum der C^ℓ -Funktionen** $A \rightarrow \mathcal{Y}$ bezeichnen wir als $C^\ell(A, \mathcal{Y})$ und speziell im Fall $\mathcal{Y} = \mathbb{R}$ auch als $C^\ell(A) := C^\ell(A, \mathbb{R})$. Ist die C^ℓ -Eigenschaft für alle $\ell \in \mathbb{N}$ erfüllt, so sprechen wir von einer **C^∞ -Funktion** und vom Raum $C^\infty(A, \mathcal{Y})$ beziehungsweise $C^\infty(A)$.

Bemerkung (zu C^ℓ -Funktionen). Für eine C^ℓ -Funktion sind nicht nur die durch die Definition erfassten partiellen Ableitungen der Ordnung $= \ell$, sondern tatsächlich sogar **alle partiellen Ableitungen der Ordnung $\leq \ell$ stetig**.

Zum Beweis dieser Tatsache argumentiert man, dass zunächst die partiellen Ableitungen der Ordnung $= \ell - 1$ (noch ein Weiteres mal) stetig partiell differenzierbar sind, nach Abschnitt 2.1 also auch total differenzierbar und daher selbst stetig. Iterativ ergibt sich dann die Stetigkeit aller partiellen Ableitungen der Ordnung $\leq \ell$.

Als sehr praktisch für den Umgang mit partiellen Ableitungen höherer Ordnung erweist sich folgende Konvention.

Definition (Multiindex-Notation). Für $A \subset \mathbb{R}^n$, einen normierten Raum \mathcal{Y} , eine Funktion $f: A \rightarrow \mathcal{Y}$, die C^ℓ nahe $a \in A$ ist, und einen **Multiindex** $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_n) \in (\mathbb{N}_0)^n$ der **Ordnung** $|\alpha| := \alpha_1 + \alpha_2 + \dots + \alpha_n \leq \ell$ vereinbaren wir die Schreibweise

$$\partial^\alpha f(a) := \partial_n^{\alpha_n} \dots \partial_2^{\alpha_2} \partial_1^{\alpha_1} f(a)$$

für eine partielle Ableitung $|\alpha|$ -ter Ordnung.

Beispiel (zur Multiindex-Notation). Für $n = 5$ und $\ell \geq 7$ ist

$$\partial_3 \partial_1 \partial_2 \partial_3 \partial_1 \partial_3 \partial_5 f(a) = \partial^\alpha f(a) \quad \text{mit } \alpha := (2, 1, 3, 0, 1) \in (\mathbb{N}_0)^5, |\alpha| = 7.$$

Bemerkungen (zur Multiindex-Notation).

- (1) **Jede partielle Ableitung** (der Ordnung $\leq \ell$) lässt sich **eindeutig** als ∂^α -Ableitung **mit einem Multiindex α darstellen**. Dies liegt einfach daran, dass die verschiedenen Operatoren ∂_i mit $i \in \{1, 2, \dots, n\}$ gemäß dem Satz über Vertauschbarkeit partieller Ableitungen in eine beliebige Reihenfolge „umsortiert“ werden können.
- (2) Für Monome wurde bereits in Abschnitt 2.1 die weitgehend analoge Multiindex-Notation

$$x^\alpha := x_1^{\alpha_1} x_2^{\alpha_2} \cdot \dots \cdot x_n^{\alpha_n} \quad \text{für } x \in \mathbb{R}^n, \alpha \in (\mathbb{N}_0)^n$$

eingeführt.

(3) Des Weiteren treffen wir für Fakultäten die Festlegung

$$\alpha! := (\alpha_1!) (\alpha_2!) \cdot \dots \cdot (\alpha_n!) \quad \text{für } \alpha \in (\mathbb{N}_0)^n.$$

Als Nächstes reißen wir kurz die Erweiterung des Konzepts der **Taylor-Entwicklung** auf den Fall mehrerer Variablen an.

Definition (Taylor-Polynome). Seien $f: A \rightarrow \mathbb{R}^m$ eine Funktion auf $A \subset \mathbb{R}^n$ und $a \in \overset{\circ}{A}$. Wenn für $\ell \in \mathbb{N}_0$ alle partiellen Ableitungen zu Ordnungen $\leq \ell$ von f an der Stelle a existieren, so heißt die Polynomfunktion $T_a^\ell f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ mit

$$T_a^\ell f(x) := \sum_{|\alpha| \leq \ell} \frac{1}{\alpha!} \partial^\alpha f(a) (x-a)^\alpha \quad \text{für } x \in \mathbb{R}^n$$

das ℓ -te **Taylor-Polynom** von f an der Stelle a /bei a . (Dabei läuft die Summe über alle Multiindizes $\alpha \in (\mathbb{N}_0)^n$ der Ordnung $\leq \ell$. Analog sind ähnliche Summen im Folgenden zu verstehen.)

Beispiel (zu Taylor-Polynomen). Speziell für $n = 2$ Variablen, beliebiges $m \in \mathbb{N}$, Ordnung $\ell = 3$ und Entwicklungspunkt $a = \mathbf{0} = (0, 0) \in \mathbb{R}^2$ lässt sich das Taylor-Polynom $T_{\mathbf{0}}^3 f: \mathbb{R}^2 \rightarrow \mathbb{R}^m$ als Funktion der Variablen $(x, y) \in \mathbb{R}^2$ in der Form

$$\begin{aligned} T_{\mathbf{0}}^3 f(x, y) &= f(\mathbf{0}) && \text{(Ordnung 0)} \\ &+ \partial_1 f(\mathbf{0})x + \partial_2 f(\mathbf{0})y && \text{(Ordnung 1)} \\ &+ \frac{1}{2} \partial_1^2 f(\mathbf{0})x^2 + \partial_1 \partial_2 f(\mathbf{0})xy + \frac{1}{2} \partial_2^2 f(\mathbf{0})y^2 && \text{(Ordnung 2)} \\ &+ \frac{1}{6} \partial_1^3 f(\mathbf{0})x^3 + \frac{1}{2} \partial_1^2 \partial_2 f(\mathbf{0})x^2 y + \frac{1}{2} \partial_1 \partial_2^2 f(\mathbf{0})xy^2 + \frac{1}{6} \partial_2^3 f(\mathbf{0})y^3 && \text{(Ordnung 3)} \end{aligned}$$

ausschreiben.

Bemerkung (zu Taylor-Polynomen). Allgemein handelt es sich bei $T_a^\ell f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ um eine Polynomfunktion in n Variablen $x = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$ mit Koeffizienten $\frac{1}{\alpha!} \partial^\alpha f(a)$ in \mathbb{R}^m , den sogenannten Taylor-Koeffizienten. Die entscheidende Eigenschaft besteht darin, dass $T_a^\ell f$ an der Stelle a bis Ordnung ℓ dieselben Ableitungen wie die Funktion f hat und damit naturgemäß ein guter Kandidat für eine Approximation von f nahe a ist. Genauer gilt für partielle Ableitungen

$$\partial^\beta (T_a^\ell f)(a) = \partial^\beta f(a) \quad \text{für alle } \beta \in (\mathbb{N}_0)^n \text{ mit } |\beta| \leq \ell,$$

wobei die Herleitung dieser Formel in der Vorlesung im Zusammenhang mit dem Beweis des nun folgenden Hauptsatzes diskutiert wurde und im Wesentlichen auf einer in den Übungen behandelten Aufgabe beruht.

Als Hauptresultat sammeln wir nun eine Mischung qualitativer und quantitativer Aussagen in weitgehender Analogie zum Fall einer einzelnen Variablen:

Hauptsatz (Satz von Taylor). Seien A offen in \mathbb{R}^n , $f: A \rightarrow \mathbb{R}^m$ eine Funktion, $a \in \overset{\circ}{A}$ und $\ell \in \mathbb{N}_0$. Dann gelten:

(I) Ist $f \in C^\ell(A, \mathbb{R}^m)$, so ist $T_a^\ell f$ die eindeutige Polynomfunktion $\mathbb{R}^n \rightarrow \mathbb{R}^m$ vom Grad $\leq \ell$ mit der **Taylor-Eigenschaft**

$$f(x) = T_a^\ell f(x) + o(|x-a|^\ell) \quad \text{bei } A \ni x \rightarrow a.$$

(II) Ist $f \in C^{\ell+1}(A, \mathbb{R}^m)$, so gilt sogar die **verschärfte Taylor-Formel**

$$f(x) = T_a^\ell f(x) + \mathcal{O}(|x-a|^{\ell+1}) \quad \text{bei } A \ni x \rightarrow a.$$

(III) Ist $f \in C^{\ell+1}(A, \mathbb{R}^m)$ und ist für $a \neq x \in A$ die Strecke $[a, x] := \{a+s(x-a) \mid s \in [0, 1]\}$ ganz in A enthalten, so gelten mit dem Vektor $v := \frac{x-a}{|x-a|}$ die **Integral-Restglied-Formel**

$$\begin{aligned} f(x) - T_a^\ell f(x) &= \frac{1}{\ell!} |x-a|^{\ell+1} \int_0^1 (1-t)^\ell \partial_v^{\ell+1} f(a+t(x-a)) dt \\ &= \sum_{|\alpha|=\ell+1} \frac{\ell+1}{\alpha!} \int_0^1 (1-t)^\ell \partial^\alpha f(a+t(x-a)) dt (x-a)^\alpha \end{aligned}$$

und die daraus folgende **Lagrange-Restglied-Abschätzung**

$$\begin{aligned} |f(x) - T_a^\ell f(x)| &\leq \frac{1}{(\ell+1)!} |x-a|^{\ell+1} \sup_{\xi \in [a, x]} |\partial_v^{\ell+1} f(\xi)| \\ &= \sup_{\xi \in [a, x]} \left| \sum_{|\alpha|=\ell+1} \frac{1}{\alpha!} \partial^\alpha f(\xi) (x-a)^\alpha \right|. \end{aligned}$$

Bemerkung (zum **Satz von Taylor**). Für die Teile (I) und (II) des Satzes reicht, dass f nur auf einer (eventuell kleinen) offenen Umgebung von a eine C^ℓ - bzw. $C^{\ell+1}$ -Funktion ist. Für Teil (III) reicht analog, dass f nur auf einer offenen Umgebung von $[a, x]$ eine $C^{\ell+1}$ -Funktion ist.

In der Vorlesung wurde die Taylor-Eigenschaft (I) mit Hilfe des Schrankensatzes bewiesen sowie für die verschärfte Taylor-Formel (II) und die Restglied-Formel/-Abschätzung (III) die Zurückführung auf den Fall einer Variablen diskutiert.

Als letzten Themenkomplex dieses Kapitels erwähnen wir Kriterien für lokale Extremstellen auf Basis zweiter Ableitungen.

Satz (Zweiter-Ordnung-Kriterien für lokale Extremstellen). Sei $f: A \rightarrow \mathbb{R}$ eine \mathbb{R} -wertige Funktion auf $A \subset \mathbb{R}^n$, die C^2 auf einer offenen Umgebung eines Punktes $a \in \overset{\circ}{A}$ ist. Dann gelten mit dem Gradienten $\nabla f(a) \in \mathbb{R}^n$ und der Hesse-Matrix $Hf(a) \in \mathbb{R}^{n \times n}$ von f an der Stelle a einerseits die **notwendigen Kriterien** für lokale Extremstellen

$$\begin{aligned} a \text{ lokale Minimalstelle von } f &\implies \nabla f(a) = 0, Hf(a) \text{ positiv semidefinit,} \\ a \text{ lokale Maximalstelle von } f &\implies \nabla f(a) = 0, Hf(a) \text{ negativ semidefinit,} \end{aligned}$$

andererseits die **hinreichenden Kriterien** für lokale Extremstellen

$$\begin{aligned} \nabla f(a) = 0, Hf(a) \text{ positiv definit} &\implies a \text{ strikte lokale Minimalstelle von } f, \\ \nabla f(a) = 0, Hf(a) \text{ negativ definit} &\implies a \text{ strikte lokale Maximalstelle von } f, \end{aligned}$$

wobei positive/negative (Semi-)Definitheit der symmetrischen Matrix $Hf(a)$ oft auch in Form der Ungleichungen $Hf(a) \geq 0$, $Hf(a) \leq 0$, $Hf(a) > 0$ bzw. $Hf(a) < 0$ notiert wird.

Der Beweis des Kriterien erfolgt durch eine Standard-Anwendung von Taylor-Entwicklung. Wie man im Einzelnen die Kriterien an der Taylor-Entwicklung zweiter Ordnung abliest, wurde dabei in der Vorlesung erläutert.

Bemerkungen (zu den **Zweiter-Ordnung-Kriterien**).

- (1) Als weiteres hinreichendes Kriterien, aber nicht für Extremstellen, sondern für Sattelpunkte (wie man kritische Punkte, die keine lokalen Extremstellen sind, nennt), folgt unter denselben Voraussetzungen wie im Satz aus den beiden notwendigen Kriterien:

$$\nabla f(a) = 0, Hf(a) \text{ indefinit} \implies a \text{ Sattelpunkt von } f.$$

- (2) Standard-Verfahren zum Nachweis der Definitheit, Semidefinitheit oder Indefinitheit einer Matrix sind — wenn nicht gerade Diagonalstruktur oder ein anderer besonders günstiger Fall vorliegt — die Anwendung des sogenannten Hauptminorenkriteriums und notfalls die Berechnung aller Eigenwerte über das charakteristische Polynom. Näheres hierzu ist Thema der Übungen. Da die Rechnungen aufwändig sein können und die Extremstellenfrage nicht einmal unbedingt final entscheiden (etwa dann nicht, wenn sich die Matrix als semidefinit, aber nicht definit erweist), lohnt die Anwendung der Zweiter-Ordnung-Kriterien bei praktischen Extremwertproblemen allerdings eher selten.

Literaturverzeichnis

Die Themen der Vorlesung werden in einer Vielzahl von Büchern behandelt, von denen hier einige aufgelistet werden. Abgesehen vom fachdidaktischen Buch [3] und hauptsächlich für das letzte Vorlesungskapitel relevanten Quelle [4] handelt es sich dabei um bekannte fachwissenschaftliche Standardwerke der Analysis:

- [1] H. AMANN, J. ESCHER: *Analysis II*. Birkhäuser, 2006.
- [2] O. FORSTER: *Analysis 2*. Springer, 2010.
- [3] G. GREEFRATH, R. OLDENBURG, H.-S. SILLER, V. ULM, H.-G. WEIGAND: *Didaktik der Analysis*. Springer, 2016.
- [4] N. HENZE, G. LAST: *Mathematik für Wirtschaftsingenieure und naturwissenschaftlich-technische Studiengänge. Band 2*. Vieweg+Teubner, 2010.
- [5] H. HEUSER: *Lehrbuch der Analysis. Teil 2*. Vieweg+Teubner, 2008.
- [6] S. HILDEBRANDT: *Analysis 2*. Springer, 2003.
- [7] K. KÖNIGSBERGER: *Analysis 2*. Springer, 2002.
- [8] W. WALTER: *Analysis 2*. Springer, 2002.