Analysis II - Notizen[1]

Marcel Schmidt

[1]Die vorliegenden Notizen sind eine von Matthias Keller und mir leicht überarbeitete Version der Notizen von Daniel Lenz zur Analysis II im Sommersemester 2014. Sie sind kein Skript zur Vorlesung.

Inhaltsverzeichnis

Vorbemerkung: Zwei kleine Rechnungen

Eine Rechnung

Der Kurs wird mit 8 Leistungspunkten (LP) gewertet. Jeder Leistungspunkt entspricht 30h Arbeit. Damit geht es um

$$240h \text{ Arbeit}$$

Davon gehen ab:

$$-90h \quad (6 \text{ h Vorlesung und Übung in 15 Wochen})$$

$$-40h \ \ (1 \text{ Woche Prüfungsvorbereitung}).$$

Damit verbleiben noch 110 h Arbeit. Auf 15 Wochen verteilt bedeutet dies ca.

$$7 \text{ h Arbeit/ Woche}$$

also

$$1 \text{ h Arbeit / Tag.}$$

Eine andere Rechnung

Der Stoff der ersten drei Semester wurde beginnend mit Newton und Leibniz um 1670 bis etwa 1920 entwickelt. Es handelt sich also um

$$250 \text{ Jahre Entwicklung.}$$

Bei 45 Wochen für die ersten drei Semester, wird also in einer Woche Vorlesung etwa

$$5 \text{ Jahre Entwicklung} \sim 260 \text{ Wochen}$$

behandelt.

Folgerung

Es muss gearbeitet werden!

KAPITEL 1

Metrische Räume und topologische Grundbegriffe

Bisher haben wir Analysis (Grenzwert, Konvergenz, Stetigkeit, Differenzierbarkeit) auf $\mathbb{R}$ betrieben. Tatsächlich interessiert man sich oft für allgemeinere Situationen. Dabei geht es dann zum Beispiel um (Teilmengen) von

- $\mathbb{K}^N = \{(x_1, \ldots, x_N) \mid x_j \in \mathbb{K}\} = \{x : \{1, \ldots, N\} \to \mathbb{K}\}$ wobei $\mathbb{K} = \mathbb{R}, \mathbb{C}$
- $M^{N \times N} =$ N$\times$ N-Matrizen über $\mathbb{K}$.
- $\ell^2 = \{x : \mathbb{N} \longrightarrow \mathbb{K} \mid \sum_j |x_j|^2 < \infty\}$.

Will man auf diesen Mengen Grenzwerte zu untersuchen bzw. Analysis betreiben, so erweist sich das Konzept der Metrik als grundlegend.

1. Metrische Räume

In diesem Abschnitt geht es um Metrik. Das Konzept der Metrik gibt einen quantitativen Abstandsbegriff. Mit diesem Abstandsbegriff kann dann Konvergenz und damit Stetigkeit gefasst werden.
Der Abstand zwischen zwei Punkten sollte folgende Eigenschaften haben: **Zeichnung.**

- Hin- und Rückweg sind gleich lang,
- Umwege verlängern den Weg,
- Verschiedene Punkte haben eine positive Entfernung.

Eine präzise Fassung gibt folgende Definition.

DEFINITION. Sei X eine Menge. Eine *Metrik* auf X ist eine Abbildung $d : X \times X \longrightarrow [0, \infty)$ für die gilt:

(M1) $d(x, y) = d(y, x)$ für alle $x, y \in X$ (Symmetrie)
(M2) $d(x, y) \leq d(x, z) + d(z, y)$ für alle $x, y, z \in X$. (Dreiecksungleichung)
(M3) $d(x, y) = 0$ genau dann, wenn $x = y$ (Nichtausgeartet)

Ein Paar (X, d) bestehend aus einer Menge X und einer Metrik d auf X heißt *metrischer Raum.*

BEISPIEL. Ist (X, d) ein metrischer Raum und Y eine Teilmenge von X, so ist auch (Y, d) ein metrischer Raum.

BEISPIEL.
- Auf $\mathbb{K}$ definiert $d(x, y) = |x - y|$ eine Metrik.

- Auf $\mathbb{Z}^N$ ist die Manhatten Metrik / Blockmetrik gegeben durch
$$d(x,y) = \sum_{j=1}^{N} |x_j - y_j|.$$
(Zeichnung mit Blöcken)
- Auf $\mathbb{K}^N$ ist die ℓ^1 Metrik gegeben durch
$$d_1(x,y) := \sum_{j=1}^{n} |x_j - y_j|.$$
(Das verallgemeinert die Manhatten Metrik.)
- Auf $\mathbb{K}^N$ ist die ℓ^∞ Metrik gegeben durch
$$d_\infty(x,y) := \max\{|x_j - y_j| \mid j = 1, \dots, N\}.$$
Die ℓ^p-Metrik
$$d_p(x,y) := \left(\sum_{j=1}^{n} |x_j - y_j|^p\right)^{1/p}.$$
Die Dreiecksungleichung zu zeigen ist nicht trivial, wir zeigen sie später für $p = 2$. In der Übung zeigen wir $d_p(x,y) \to d_\infty(x,y)$.
- Sei M eine Menge. Auf
$$B(M) = \{f : M \to \mathbb{K} \mid f(M) \text{ beschränkt in } \mathbb{K}\}$$
ist durch
$$d_\infty(f,g) := \sup_{x \in M} |f(x) - g(x)|$$
eine Metrik gegeben. Das verallgemeintert die ℓ^∞ Metrik auf $\mathbb{K}$ (setze $M = \{1, \dots, N\}$).
- Auf beliebigem X ist die diskrete Metrik gegeben durch
$$d_D(x,y) = \begin{cases} 1 & \text{falls } x \neq y \\ 0 & \text{falls } x = y \end{cases}.$$
- Sei $\mathcal{A}$ eine endliche Menge (Alphabet). Sei
$$\mathcal{W}_\mathcal{A} = \{x : \mathbb{N} \longrightarrow \mathcal{A}\} = \{\text{Folgen mit Werten in } \mathcal{A}\}.$$
  - Ist $\mathcal{A}$ das übliche Alphabet mit Satzzeichen, so handelt es sich bei $\mathcal{W}_\mathcal{A}$ um die Menge aller 'Bücher'.
  - Ist $\mathcal{A} = \{U, C, A, G\}$, so handelt es sich bei $\mathcal{W}_\mathcal{A}$ um die Menge aller (unendlichen) Gensequenzen.
  - Ist $\mathcal{A} = \{0, 1\}$, so handelt es sich bei $\mathcal{W}_\mathcal{A}$ um die Menge aller (unendlichen) Computerprogramme.

Es definiert
$$d(v,w) := \sum_{j=1}^{\infty} \frac{d_D(v(j),w(j))}{2^j}$$
eine Metrik auf $\mathcal{W}$. Damit kann man dann den Abstand zwischen zwei Genen oder zwischen zwei Büchern ;-) berechnen.

- X seien die Städte in Europa
  - $d(x,y) =$ Autokilometer zwischen x und y ist eine Metrik
  - $d(x,y) =$ Preis für Direktflug von x und y ist in der Regel keine Metrik.
- X seien die Personen auf Facebook und $d(x,y)$ die Anzahl der minimale Kontakte über die x mit y befreundet ist.

PROPOSITION (Umgekehrte Dreiecksungleichung). *Sei (X,d) ein metrischer Raum. Dann gilt*
$$|d(x,y) - d(x,z)| \leq d(y,z)$$
für alle $x,y,z \in X$.

Beweis. Es gilt
$$d(x,y) \leq d(x,z) + d(z,y)$$
sowie nach Vertauschen von y und z
$$d(x,z) \leq d(x,y) + d(y,z).$$
Damit folgt die Aussage. □

Auf Vektorräumen werden Metriken meist von Normen induziert. Eine Norm liefert ein Konzept von Länge eines Vektors. Normen stellen also eine Verallgemeinerung der Betragsfunktion auf mehrdimensionale Situationen dar. Der Abstand zwischen zwei Vektoren ist dann die Länge ihrer Differenz. **Zeichnung.**

DEFINITION. Ein *Vektorraum* über dem Körper $\mathbb{K}$ (d.h. ein reeller oder komplexer Vektorraum) ist eine Menge V mit zwei Operationen $+ : V \times V \to V$, $(v,w) \mapsto v + w$ und $\cdot : \mathbb{K} \times V \to V$, $(\alpha, v) \mapsto \alpha v$, so dass $(V,+)$ eine abelsche Gruppe ist (d.h. Addition ist kommutativ, assoziativ, ein neutrales Element und für jedes Element ein inverses Element) und die Skalarmultiplikation ist assoziativ und distributiv bezüglich der Addition und die Addition ist distributiv bezüglich der Skalarmultiplikation.

BEISPIEL.
- $\mathbb{K}^N$ ist ein Vektorraum mit
$$x + y = (x_1 + y_1, \ldots, x_N + y_N)$$
und
$$\alpha x = (\alpha x_1, \ldots, x_N),$$
für $x = (x_1, \ldots, x_N)$, $y = (y_1, \ldots, y_N) \in \mathbb{K}^N$ und $\alpha \in \mathbb{K}$.

- $\ell^2 := \{x : \mathbb{N} \longrightarrow \mathbb{K} \mid \sum_j |x_j|^2 < \infty\}$ ist ein Vektorraum mit
$$x + y = (x_j + y_j)_{j\in\mathbb{N}} \text{ und } \alpha x = (\alpha x_j)_{j\in\mathbb{N}}$$
für $x = (x_j), y = (y_j) \in \ell^2$ und $\alpha \in \mathbb{K}$. (Zeige $x + y \in \ell^2$ in Übung.)
- Für eine beliebige Menge M ist die Menge der Funktion
$$F(M) = \{f : M \to \mathbb{K}\} = M^{\mathbb{K}}$$
ein Vektorraum mit $f + g$ und αg gegeben durch
$$(f + g)(m) = f(m) + g(m) \text{ und } (\alpha f)(m) = \alpha f(m), \quad m \in M$$
für $f, g \in F(M)$, $\alpha \in \mathbb{K}$. Der Raum der beschränkten Funktionen $B(M)$ ist ein Unterraum von $F(M)$. (Übung)

Ende 1. Vorlesung

Definition. Sei V ein Vektorraum über dem Körper $\mathbb{K}$. Eine Abbildung $\|\cdot\| : V \longrightarrow [0, \infty)$ heißt *Norm*, wenn die folgenden Aussagen gelten.

(N1) $\|\alpha x\| = |\alpha|\|x\|$ für alle $\alpha \in \mathbb{K}$ und $x \in V$. (Homogenität)
(N2) $\|x + y\| \leq \|x\| + \|y\|$ für alle $x, y \in V$. (Dreiecksungleichung)
(N3) $\|x\| = 0$ genau dann, wenn $x = 0$. (Nichtausgeartet)

Definition. Sei V eine Vektorraum mit Norm $\|\cdot\|$. Dann heißt
$$d_{\|\cdot\|} : V \times V \longrightarrow [0, \infty), (v, w) \mapsto d_{\|\cdot\|}(v, w) := \|v - w\|$$
die durch $\|\cdot\|$ *induzierte Metrik.*

Beispiel.
- Auf $V = \mathbb{K}$ wird durch $|\cdot|$ eine Norm gegeben.
- Auf $V = \mathbb{K}^N$ wird durch $\|x\|_1 := \sum_{j=1}^N |x_j|$ eine Norm, die sogenannte ℓ^1-Norm definiert. Die induzierte Metrik ist die ℓ^1 Metrik.
- Auf $V = \mathbb{K}^N$ wird durch $\|x\|_\infty := \max\{|x_j| : j = 1, \ldots, N\}$ eine Norm definiert, die sogenannte ℓ^∞-Norm. Die induzierte Metrik ist die ℓ^∞ Metrik.

Proposition (Umgekehrte Dreiecksungleichung). *Ist $\|\cdot\|$ eine Norm auf V, so gilt die umgekehrte Dreiecksungleichung*
$$|\|x\| - \|y\|| \leq \|x - y\|.$$

Beweis. Das folgt unmittelbar aus $\|x\| = d_{\|\cdot\|}(x, 0)$ und der entsprechenden Aussage für Metriken. □

Jedes Skalarprodukt liefert eine Norm (und damit eine Metrik). Dafür brauchen wir die Cauchy-Schwarz Ungleichung.

Definition. Sei V eine Vektorraum. Eine Abbildung $\langle\cdot,\cdot\rangle : V \times V \to \mathbb{K}$ heißt *Skalarprodukt*, falls

(S1) $\langle x, y\rangle = \overline{\langle y, x\rangle}$ für alle $x, y \in V$. (Antisymmetrie)

(S2) $\langle \alpha x + \beta y, z\rangle = \alpha\langle x, y\rangle + \beta\langle x, z\rangle$ und $\langle x, \alpha y + \beta z\rangle = \overline{\alpha}\langle x, y\rangle + \overline{\beta}\langle x, z\rangle$ für alle $\alpha, \beta \in \mathbb{K}$ und $x, y, z \in V$. (Sesquilinearität)
(S3) $\langle x, x\rangle \geq 0$ und $\langle x, x\rangle$ genau dann, wenn $x = 0$. (Positiv definit)

Lemma (Cauchy-Schwarz Ungleichung). *Sei V eine Vektorraum mit Skalarprodukt $\langle\cdot,\cdot\rangle$. Dann gilt*

$$|\langle x, y\rangle| \leq \langle x, x\rangle^{\frac{1}{2}}\langle y, y\rangle^{\frac{1}{2}}$$

für alle $x, y \in V$.

Beweis. Seien $x, y \in V$. Es gilt

$$0 \leq \langle x + \alpha y, x + \alpha y\rangle = \langle x, x\rangle + 2\mathrm{Re}\alpha\langle y, x\rangle + |\alpha|^2\langle y, y\rangle, \qquad (*)$$

für alle $\alpha \in \mathbb{K}$.
Falls $\langle y, y\rangle = 0$, so gilt $y = 0$ nach (S3) und somit $\langle x, y\rangle = 0$ nach (S2). Somit ist die Aussage klar.
Nehmen wir also an $\langle y, y\rangle > 0$, dividieren in (*) und erhalten

$$0 \leq \frac{\langle x, x\rangle}{\langle y, y\rangle} + 2\frac{\mathrm{Re}\alpha\langle y, x\rangle}{\langle y, y\rangle} + \alpha^2.$$

Quadratisches Ergänzen ergibt dann

$$0 \leq \left|\alpha + \frac{\langle y, x\rangle}{\langle y, y\rangle}\right|^2 + \frac{\langle x, x\rangle}{\langle y, y\rangle} - \frac{|\langle y, x\rangle|^2}{|\langle y, y\rangle|^2}.$$

Da dies für alle α gelten soll, folgt mit $\alpha = -\frac{\langle y, x\rangle}{\langle y, y\rangle}$

$$0 \leq \frac{\langle x, x\rangle}{\langle y, y\rangle} - \frac{|\langle y, x\rangle|^2}{|\langle y, y\rangle|^2}.$$

Damit folgt die gewünschte Ungleichung. □

Proposition. *Sei V eine Vektorraum mit Skalarprodukt $\langle\cdot,\cdot\rangle$. Dann ist*

$$\|\cdot\| : V \longrightarrow [0, \infty), \|x\| := \langle x, x\rangle^{1/2}$$

eine Norm.

Beweis. Wir zeigen die Dreiecksungleichung. Die übrigen Aussagen folgen einfach. Es gilt nach Cauchy-Schwarz Ungleichung:

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

Das beendet den Beweis. □

BEISPIEL. **Bekannteste Metrik und Norm.** Die *euklidische Metrik* auf $\mathbb{R}^N$ bzw. $\mathbb{C}^N$ ist gegeben durch

$$d_2(x,y) = \left(\sum_{j=1}^{N} |x_j - y_j|^2\right)^{1/2}.$$

Sie wird durch das Standardskalarprodukt

$$\langle x,y\rangle := \sum_{j=1}^{N} x_j\overline{y_j}$$

bzw. die dazugehörige Norm

$$\|x\|_2 := \left(\sum_{j=1}^{N} |x_j|^2\right)^{1/2}$$

induziert.

BEISPIEL. **Der Vektorraum ℓ^2.** Wir betrachten die Menge

$$\ell^2 = \left\{x : \mathbb{N} \longrightarrow \mathbb{K} \,\middle|\, \sum |x_j|^2 < \infty\right\}.$$

Dann ist ℓ^2 ein Untervektorraum des Raumes aller reell bzw. komplexwertigen Funktionen auf $\mathbb{N}$. Für beliebige $x, y \in \ell^2$ ist

$$\sum_{j=1}^{\infty} x_j\overline{y_j}$$

absolut konverent und durch

$$\langle\cdot,\cdot\rangle : \ell^2 \to \mathbb{C}, \quad \langle x,y\rangle := \sum_{j=1}^{\infty} x_j\overline{y_j}$$

wird ein Skalarprodukt auf ℓ^2 definiert (siehe Übung). Damit ist

$$\|x\|_2 = \left(\sum_{j=1}^{\infty} |x_j|^2\right)^{1/2}, \qquad x \in \ell^2$$

eine Norm und

$$d_2(x,y) = \left(\sum_{j=1}^{\infty} |x_j - y_j|^2\right)^{1/2}, \qquad x,y \in \ell^2$$

die durch $\|\cdot\|_2$ induzierte Metrik.

BEISPIEL. Sei $C([a,b])$ der Raum der stetigen Funktionen von $[a,b]$ nach $\mathbb{R}$. Dann wird durch

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

ein Skalarprodukt definiert, wobei das Integral, das Riemannintegral ist. (Übung)

Metriken erlauben es, den Begriff der Konvergenz von Folgen zu definieren.

Definition (Konvergenz im metrischen Raum). Sei (X, d) ein metrischer Raum. Eine Folge (x_n) in X heißt *konvergent* gegen $x \in X$, wenn gilt

$$d(x_n, x) \to 0, \quad n \to \infty.$$

Wir schreiben dann $x_n \to x$ oder $\lim_{n\to\infty} x_n = x$ und nennen x *Grenzwert* der Folge (x_n).

Lemma (Eindeutigkeit Grenzwert). *Sei (X, d) ein metrischer Raum und (x_n) eine konvergente Folge in X. Dann ist der Grenzwert von (x_n) eindeutig.*

Beweis. Seien $x, y \in X$ Grenzwerte von (x_n). Dann gilt nach (M2) $d(x, y) \leq d(x, x_n) + d(x_n, y) \to 0$. Damit folgt $d(x, y) = 0$ und somit $x = y$ wegen (M3). □

Ende 2. Vorlesung

Proposition (Stetigkeit von Metrik und Norm). *Sei (X, d) ein metrischer Raum. Gilt $x_n \to x$ und $y_n \to y$, so folgt $d(x_n, y_n) \to d(x, y)$. Sei V ein Vektorraum mit Norm $\|\cdot\|$ und zugehöriger Metrik d. Gilt $x_n \to x$ so folgt $\|x_n\| \to \|x\|$.*

Beweis. Das folgt jeweils aus der umgekehrten Dreiecksungleichung: Zum ersten Punkt:

$$\begin{aligned} |d(x_n, y_n) - d(x, y)| &= |d(x_n, y_n) - d(x_n, y) + d(x_n, y) - d(x, y)| \\ &\leq |d(x_n, y_n) - d(x_n, y)| + |d(x_n, y) - d(x, y)| \\ \text{(Umgek. } \triangle\text{-Ugl)} \quad &\leq d(y_n, y) + d(x_n, x). \end{aligned}$$

Zum zweiten Punkt: Nach der umgekehrten Dreiecksungleichung gilt

$$|\|x_n\| - \|x\|| \leq \|x_n - x\|.$$

Damit folgt die Aussage sofort. □

Ein nützliches Konzept zur weiteren Untersuchung (und zur Visualisierung von Konverenz) sind Kugeln und Umgebungen bzw. offene Mengen.

Definition (Kugel). Sei (X, d) metrischer Raum und $x \in X$. Dann heißt für $r \geq 0$

$$B_r(x) := \{y \in X \mid d(x, y) \leq r\}$$

abgeschlossene Kugel um x mit Radius r und

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

offene Kugel um x mit Radius r.

Beispiel.
- Sei $\mathbb{R}$ mit der Metrik $d(x,y) = |x-y|$ versehen. Dann ist $x \in \mathbb{R}$ und $r > 0$ die Kugel $U_r(x) = (x-r, x+r)$. Entsprechend ist $B_r(x) = [x-r, x+r]$. (Zeichnung)
- Sei $\mathbb{C}$ mit der Metrik $d(x,y) = |x-y|$ versehen. Dann ist für $z \in \mathbb{C}$ und $r > 0$ die Kugel $U_r(z)$ gegeben durch den (offenen) Kreis um z mit Radius r. Entsprechend ist $B_r(z)$ der volle Kreis um z mit Radius r. (Zeichnung)
- Sei $X = \mathbb{R}^2$ mit der Euklidischen Metrik d_2 versehen. Dann ist für $x = 0$ und $r > 0$ die Kugel $U_r(x)$ ($B_r(x)$ gegeben durch den offenen (vollen) Kreis um 0 mit Radius r. (Vom Standpunkt des metrischen Raumes aus, sind also $\mathbb{R}^2$ und $\mathbb{C}$ nicht zu unterscheiden.)
- Sei $X = \mathbb{R}^2$ mit $d = d_\infty$. Dann ist für $x = 0$ und $r > 0$ die offenen Kugel $U_r(0)$ gegeben durch das Quadrat $(-r,r)^2$.
- Sei $X = \mathbb{R}^2$ mit $d = d_1$. Dann ist für $x = 0$ und $r > 0$ die offenen Kugel $U_r(x)$ gegeben durch ... (Zeichnung)
- Sei X eine beliebige Menge und $d = d_D$. Dann ist für $r = 1/2$ die offene und die abgeschlossene Kugel gegeben durch $\{0\}$.

Lemma. *Sei (X,d) ein metrischer Raum und (x_n) eine Folge in X und $x \in X$. Dann gilt $x_n \to x$ genau dann, wenn es zu jeder offenen nichtleeren Kugel K um x ein $n_K \in \mathbb{N}$ gibt mit $x_n \in K$ für alle $n \geq n_K$.*

Beweis. Eine offene nichtleere Kugel um x hat die Form $K = U_r(x)$ mit $r > 0$. Damit folgt die Aussage sofort. □

Wir untersuchen nun Konvergenz in den oben gegebenen Beispielen.

Beispiel. **Konvergenz in $\mathcal{A}$.** Sei $\mathcal{A}$ eine beliebige Menge mit der diskreten Metrik d_D. Dann gilt $x_n \to x$ genau dann, wenn ein N existiert mit $x_n = x$ für alle $n \geq N$.
Konvergenz in $\mathcal{W}_\mathcal{A}$: Sei $\mathcal{W}_\mathcal{A} = \{x : \mathbb{N} \to \mathcal{A}\}$ wie oben mit der Metrik

$$d(x,y) = \sum_{k=1}^{\infty} \frac{d_D(x(k), y(k))}{2^k}.$$

Dann gilt $x_n \to x$ genau dann, wenn für jedes $k \in \mathbb{N}$ gilt $x_n(k) \to x(k)$ bezüglich d_D auf $\mathcal{A}$ (d.h. wenn für jedes $k \in \mathbb{N}$ schließlich $x_n(k) = x(k)$ gilt).
Bew. $\Longrightarrow$: Sei $k \in \mathbb{N}$ beliebig. Offenbar gilt

$$d_D(x_n(k), x(k)) \leq 2^k d(x_n, x).$$

Damit folgt die Aussage sofort.
$\Longleftarrow$: Betrachte $L \in \mathbb{N}$ beliebig. Dann gibt es ein n_L mit $x_n(k) = x(k)$ für alle $k = 1, \ldots, L$, falls $n \geq N_L$. Damit gilt dann für solche n

$$d(x_n, x) = \sum_{k=1}^{\infty} \frac{d_D(x_n(k), x(k)}{2^k} = \sum_{k=L+1}^{\infty} \frac{d_D(x_n(k), x(k)}{2^k} \leq \sum_{k=L+1} \frac{1}{2^k} = \frac{1}{2^L}.$$

Da L beliebig war, folgt die gewünschte Konvergenz.

Konvergenz in $\mathbb{K}^N$. Sei $x^{(n)} = (x_1^{(n)}, \ldots, x_N^{(n)})$ eine Folge in $\mathbb{R}^N$ und $x = (x_1, \ldots, x_N) \in \mathbb{R}^N$. Dann gilt:
$x^{(n)} \stackrel{d_1}{\to} x \Longleftrightarrow x_j^{(n)} \to x_j$ für jedes $j = 1, \ldots, N$. (Bew: ...)
$x^{(n)} \stackrel{d_\infty}{\to} x \Longleftrightarrow x_j^{(n)} \to x_j$ für jedes $j = 1, \ldots, N$. (Bew: ...)
$x^{(n)} \stackrel{d_2}{\to} x \Longleftrightarrow x_j^{(n)} \to x_j$ für jedes $j = 1, \ldots, N$. (Bew: ...)
Das ist kein Zufall. Es gilt vielmehr folgender bemerkenswerter Satz.

THEOREM (Äquivalenz aller Normen im $\mathbb{K}^N$). *Ist $\|\cdot\| : \mathbb{K}^N \longrightarrow [0,\infty)$ eine Norm, so gibt es $c, C \geq 0$ mit*

$$\|x\| \leq C\|x\|_1 \;\; \textit{und} \;\; \|x\|_1 \leq c\|x\|$$

für alle $x \in \mathbb{K}^N$.

BEMERKUNG. Es sind damit alle Normen im $\mathbb{K}^N$ äquivalent in dem Sinne, dass zu beliebigen Normen $\|\cdot\|_\clubsuit$ und $\|\cdot\|_\diamond$ Konstanten $\lambda, \mu > 0$ existieren mit

$$\|\cdot\|_\clubsuit \leq \lambda\|\cdot\|_\diamond$$

und

$$\|\cdot\|_\diamond \leq \mu\|\cdot\|_\clubsuit.$$

Beweis. Wir zeigen zunächst die erste Ungleichung: Sei (e_j) die Standardbasis im $\mathbb{K}^N$. Dann gilt also

$$x = (x_1, \ldots, x_n) = \sum_{j=1}^N x_j e_j.$$

Dann folgt

$$\|x\| = \|\sum_{j=1}^N x_j e_j\| \leq \sum_{j=1}^N |x_j|\|e_j\| \leq C\sum_{j=1}^N |x_j| = C\|x\|_1$$

mit

$$C := \max\{\|e_j\| \mid j = 1, \ldots, N\}.$$

Ende 3. Vorlesung

Wir zeigen nun die zweite Ungleichung

$$\|x\|_1 \leq c\|x\|.$$

Wir nehmen an, dass sie nicht gilt. Dann existiert also zu jedem $n \in \mathbb{N}$ ein $x^{(n)}$ mit

$$\|x^{(n)}\|_1 > n\|x^{(n)}\|.$$

Ohne Einschränkung gelte $\|x_n\|_1 = 1$ für alle $n \in \mathbb{N}$. Dann folgt also

$$\|x_n\| \leq \frac{1}{n}.$$

Damit folgt sofort

$$x^{(n)} \to 0 \text{ bezüglich } \|\cdot\|.$$

Weiterhin sind wegen

$$1 = \|x^{(n)}\|_1 = \sum_{j=1}^{N} |x_j^{(n)}|$$

die Folgen $(x_j^{(n)})_n$ für jedes $j \in \{1, \ldots, N\}$ beschränkt. Daher können wir nach Bolzano Weierstrass (BW) ohne Einschränkung annehmen, dass x_n bezüglich $\|\cdot\|_1$ konvergiert gegen ein $x = (x_1, \ldots, x_N)$. (Genauer: $(x_1^{(n)})_n$ beschränkt, also gibt es nach BW eine konvergente Teilfolge. Ohne Einschränkung sei die Folge $(x_1^{(n)})_n$ selber konvergent. Betrachte nun $(x_2^{(n)})_n$. Diese Folge ist beschränkt. Also gibt es nach nach BW eine konvergente Teilfolge. Ohne Einschränkung sei die Folge $(x_2^{(n)})_n$ selber konvergent ...)
Dann gilt

$$\|x\|_1 = \sum_{j=1}^{N} |x_j| = \sum_{j=1}^{N} \lim_{n\to\infty} |x_j^{(n)}| = \lim_{n\to\infty} \|x^{(n)}\|_1 = 1.$$

Insbesondere gilt $x \neq 0$. Aus Stetigkeit der Norm, $\|x^{(n)}\| \to 0$, der schon bewiesenen Ungleichung $\|\cdot\| \leq C\|\cdot\|_1$ und $\|x^{(n)} - x\|_1 \to 0$ folgt nun

$$0 \neq \|x\| = \lim_{n\to\infty} \|x^{(n)} - x\| \leq C \lim_{n\to\infty} \|x^{(n)} - x\|_1 = 0.$$

Das ist ein Widerspruch. Somit folgt $\|\cdot\|_1 \leq c\|\cdot\|$ für ein gewisses c. □

Wir kommen nun noch zu einem fundamentalen Konzept, das später von großer Bedeutung sein wird.

Definition (Cauchy-Folge). Sei (X, d) ein metrischer Raum. Eine Folge (x_n) in X heißt *Cauchy-Folge*, wenn zu jedem $\varepsilon > 0$ ein $n_\varepsilon > 0$ existiert mit

$$d(x_n, x_m) \leq \varepsilon$$

für alle $n, m \geq n_\varepsilon$.

Proposition. *Sei (X, d) ein metrischer Raum. Dann ist jede konvergente Folge eine Cauchy-Folge.*

Beweis. Sei die Folge (x_n) in X konverent gegen x. Es gilt

$$d(x_n, x_m) \leq d(x_n, x) + d(x, x_m).$$

Damit folgt die Aussage leicht. □

Definition (Vollständigkeit). Ein metrischer Raum heißt *vollständig*, wenn jede Cauchy-Folge in ihm einen Grenzwert besitzt.

Bemerkung. Vollständigkeit hängt von der Metrik ab und nicht nur von der Topologie (siehe Übung).

Beispiel. **Vollständige Räume.**

- Der Raum $\mathbb{R}$ mit der (euklidischen Metrik) $d(x, y) = |x - y|$ ist vollständig. Bew. Analysis I.
- Die Raum $\mathbb{C}$ mit der (euklidischen Metrik) $d(x, y) = |x - y|$ ist vollständig. Bew. Analysis I.
- Der Raum $\mathbb{R}^N$ mit der Euklidischen Metrik ist vollständig. Bew: Sei $(x^{(n)})$ eine Cauchy-Folge bezüglich d_2 und $x^{(n)} = (x_1^{(n)}, \ldots, x_N^{(n)})$. Wegen

$$|x_j^{(n)} - x_j^{(m)}| \leq \left(\sum_{k=1}^{N} |x_k^{(n)} - x_k^{(m)}|^2 \right)^{1/2} = d_2(x^{(n)}, x^{(m)})$$

ist dann $(x_j^{(n)})$ eine Cauchy-Folge in $\mathbb{R}$ für jedes $j = 1, \ldots, N$. Damit existiert also nach dem im ersten Beispiel diskutieren für jedes $j = 1, \ldots, N$ ein $x_j \in \mathbb{R}$ mit $x_j^{(n)} \to x_j \in \mathbb{R}$ d.h.

$$|x_j^{(n)} - x_j| \to 0, n \to \infty.$$

Für $x = (x_1, \ldots, x_N)$ gilt dann also

$$d_2(x^{(n)}, x) = \left(\sum_{k=1}^{N} |x_k^{(n)} - x_k^{(m)}|^2 \right)^{1/2} \to 0, n \to \infty.$$

Damit konvergiert also (x_n) gegen x.

Wichtig. Tatsächlich haben auf dem $\mathbb{R}^N$ aufgrund des Satzes von der Äquivalenz aller Normen alle von Normen induzierten Metriken dieselben Cauchy-Folgen und dieselben konvergenten Folgen. Insbesondere ist also der Raum $\mathbb{R}^N$ mit jeder von einer Norm induzierten Metrik vollständig.
- Der Raum $\mathbb{C}^N$ mit der Euklidischen Metrik ist vollständig. Bew. Wie für $\mathbb{R}^N$.
- Der Raum $M^{N \times N}$ der $N \times N$ Matrizen über $\mathbb{R}$ (über $\mathbb{C}$) ist vollständig bezüglich der durch die Norm (Übung)

$$\|A\| := \sup\{\|Ax\|_2 \mid \|x\|_2 \leq 1\}$$

induzierten Metrik.
Beweis. Es handelt sich um den Vektorraum $\mathbb{R}^{N^2}$ auf dem alle Normen äquivalent sind. Es reicht also zu zeigen, dass $\|\cdot\|$ eine Norm ist. Das folgt aus den Betrachtungen der Übung. □

- Sei X eine beliebige Menge und d_D die diskrete Metrik auf X. Dann ist (X, d_D) vollständig. *Beweis.* Sei (x_n) eine Cauchy-Folge. Dann gilt für alle genügend großen n und m $d_D(x_n, x_m) \leq 1/2$. Damit gilt für alle genügend großen n und m also $x_n = x_m$. Es gibt also ein $x \in X$ mit $x_n = x$ für alle genügend großen n. Damit folgt die Konvergenz von (x_n) gegen x. □

- Sei $\mathcal{A}$ endlich und $\mathcal{W}$ der Raum der Wörter über $\mathcal{A}$ mit der Metrik $d(x,y) = \sum_{j=1}^{\infty} \frac{d_D(x(j),y(j))}{2^j}$. Dann ist $\mathcal{W}$ vollständig.
  Beweis. Sei (x_n) eine Cauchy-Folge. Wegen
  $$d_D(x_n(j), x_m(j)) \le 2^j d(x_n, x_m)$$
  ist dann $(x_n(j))_n$ für jedes feste $j \in \mathbb{N}$ eine Cauchy-Folge bezüglich d_D. Also existiert nach dem schon bewiesenen zu jedem $j \in \mathbb{N}$ ein $x(j)$ mit
  $$x_n(j) \to x(j).$$
  Dann gilt also $x_n \to x$ für $x : \mathbb{N} \longrightarrow \mathcal{A}$, $j \mapsto x(j)$. □
- ℓ^2 ist vollständig (Zusatzaufgabe in Übung). Das bedeutet ℓ^2 ist ein sogenannter Hilbertraum.
- Sei M eine Menge. Dann ist der Raum $B(M)$ der beschränkten Funktionen von M nach $\mathbb{K}$ mit der Supremumsmetrik
  $$d_\infty(f,g) = \sup_{m \in M} |f(m) - g(m)|$$
  ist vollständig.
  Beweis: Sei (f_n) eine Cauchy-Folge. Offensichtlich konvergiert f_n punktweise gegen ein $f : M \to \mathbb{K}$. (Für jedes $x \in M$ konvergiert $f_n(x)$ gegen ein $f(x)$, da $(f_n(x))$ eine Cauchy-Folge in $\mathbb{K}$ ist wegen $|f_n(x) - f_m(x)| \le d_\infty(f_n, f_m)$.) Weiterhin ist f beschränkt: Sei N, so dass $d_\infty(f_n, f_m) \le 1$, $n,m \ge N$ und setze $C = \sup_{x \in M} |f_N(x)|$. Für $x \in X$ wähle jetzt $n = n_x \ge N$, so dass $|f_n(x) - f(x)| \le 1$. Dann gilt für alle $x \in M$
  $$|f(x)| \le |f(x) - f_n(x)| + |f_n(x) - f_N(x)| + |f_N(x)| \le 2 + C.$$
  Daraus folgt $f \in B(M)$. Es bleibt zu zeigen, dass $d_\infty(f_n, f) \to 0$, $n \to \infty$. Für $\varepsilon > 0$ wähle nun N, so dass $d_\infty(f_n, f_m) \le \varepsilon$ und für $x \in M$ sei $m_x \ge N$, so dass $|f_{m_x}(x) - f(x)| \le \varepsilon$. Dann gilt für alle $x \in M$
  $$|f(x) - f_n(x)| \le |f(x) - f_{m_x}(x)| + |f_{m_x}(x) - f_n(x)| \le 2\varepsilon.$$
- Der Raum der stetigen Funktionen $C([a,b])$ von $[a,b]$ nach $\mathbb{R}$ mit der durch die Norm $\|f\|_\infty = \sup_{x \in [a,b]} |f(x)|$ induzierten Metrik ist vollständig (Übung).

Beispiel. **Unvollständige Räume**

- Das Interval $(0,1)$ mit der Euklidischen Metrik ist nicht vollständig. (Übung: Finden Sie eine Metrik d auf $(0,1)$, so dass $(0,1)$ vollständig bzgl d ist und eine bezüglich der Euklidischen Metrik in $(0,1)$ konvergente Folge auch bzgl d konvergiert und zwar gegen denselben Grenzwert.)
- Die Menge $\mathbb{Q}$ mit der Euklidischen Metrik ist nicht vollständig.

- Der Raum der stetigen Funktionen $C([a,b])$ von $[a,b]$ nach $\mathbb{R}$ mit der durch das Skalarprodukt $\langle f,g\rangle = \int_a^b f(x)g(x)dx$ induzierten Metrik ist nicht vollständig (Übung).

Ende 4. Vorlesung

2. Etwas Topologie metrischer Räume

Neben dem Konzept der Kugel um einen Punkt gibt es auch noch das Konzept der Umgebung eines Punktes, um die Nähe zweier Punkte zu untersuchen. Der Vorteil der Umgebungen über Kugeln liegt darin, dass sie mit dem Anwenden von Funktionen besser verträglich sind. Es gibt nun Mengen, die Umgebung jedes ihrer Punkte sind. Solche Mengen spielen eine besondere Rolle. Das untersuchen wir nun.

Definition (Umgebung). Sei (X,d) metrischer Raum und $x \in X$. Dann heißt eine Menge V *Umgebung* von x, wenn ein $r > 0$ existiert mit $U_r(x) \subseteq V$ (oder äquivalent, wenn ein $r' > 0$ existiert mit $B_{r'}(x) \subseteq V$.)

Bemerkung. Es ist also V Umgebung von x, wenn es eine Obermenge einer Kugel um x ist, d.h. wenn um x herum noch etwas "Platz in V ist„. Die Elemente von V sind dann in gewisser Weise "nahe„ an x.

Beispiel. In einem metrischen Raum sind $U_r(x)$ und $B_r(x)$ Umgebungen von x für jedes $r > 0$.

Definition (Offene und abgeschlossene Mengen). Sei (X,d) ein metrischer Raum. Dann heißt eine Teilmmenge U von X *offen*, wenn sie Umgebung jedes ihrer Punkte ist d.h. wenn zu jedem $x \in U$ ein $r_x > 0$ existiert mit $U_{r_x}(x) \subseteq U$. Eine Teilmenge A von X heißt *abgeschlossen*, wenn $X \setminus A$ offen ist.

Bemerkung. Es ist NICHT so, daß jede Teilmenge eines metrischen Raumes genau eine der beiden Eigenschafen Offenheit bzw. Abgeschlossenheit haben muss.

- Es kann eine Teilmenge offen und abgeschlossen sein.
- Es kann eine Teilmenge weder offen noch abgeschlossen sein.

Lemma. *Jede offene Kugel ist in einem metrischen Raum ist offen. Jede abgeschlossene Kugel ist abgeschlossen.*

Beweis. Zur Offenheit von $U_r(x)$: Diese folgt aus der Dreiecksungleichung: Ist $y \in U_r(x)$, so gilt $d(y,x) < r$. Damit folgt nach Dreiecksungleichung/Zeichnung mit $s := r - d(x,y) > 0$ also

$$U_s(y) \subseteq U_r(x).$$

(Denn für $v \in U_s(y)$ gilt $d(v,x) \leq d(v,y) + d(y,x) < s + d(y,x) = r$.)
Zur Abgeschlossenheit von $B_r(x)$: Es ist die Offenheit des Komplementes zu zeigen. Diese folgt wieder aus der Dreiecksungleichung: Sei

$y \notin B_r(x)$. Dann gilt also $d(x,y) > r$. Damit gilt nach Dreiecksungleichung / Zeichnung für $s := d(x,y) - r$ also

$$U_s(y) \subseteq X \setminus B_r(x).$$

(Denn für $v \in U_s(y)$ gilt $d(v,x) \geq d(y,x) - d(v,y) > d(y,x) - s = r$.) Da $y \in X \setminus B_r(x)$ beliebig war folgt die Behauptung. □

Lemma (Charakterisierung Konvergenz). *Sei (X,d) metrischer Raum, (x_n) Folge in X, $x \in X$. Dann sind äquivalent:*

(i) *Es konvergiert (x_n) gegen x (d.h. $\lim_{n\to\infty} d(x_n, x) = 0$).*
(ii) *Für alle $\varepsilon > 0$ existiert ein $n_\varepsilon \in \mathbb{N}$ mit $x_n \in U_\varepsilon(x)$ für alle $n \geq n_\varepsilon$*
(iii) *Für jede Umgebung U von x existiert ein $n_U \in \mathbb{N}$ mit $x_n \in U$ für alle $n \geq n_U$.*
(iv) *Für jede offene Menge V mit $x \in V$ existiert ein $n_V \in \mathbb{N}$ mit $x_n \in V$ für alle $n \geq n_U$.*

Bemerkung.
- Alle vier Aussagen geben eine präzise Version davon, dass die (x_n) dem Punkt x beliebig nahe kommen, wenn n beliebig groß wird.
- Natürlich könnte man in (ii) auch U_ε durch B_ε ersetzen, da $U_\varepsilon \subseteq B_\varepsilon$ und $B_{\varepsilon/2} \subseteq U_\varepsilon$ gilt.

Beweis. (iv) $\Longrightarrow$ (iii): Jede Umgebung von x enthält eine offene Menge V mit $x \in V$.
(iii) $\Longrightarrow$ (ii): Es ist $U_\varepsilon(x)$ eine Umgebung von x.
(ii) $\Longrightarrow$ (i): Das ist klar nach Definition von $d(x, x_n) \to 0$ und $U_\varepsilon(x)$.
(i) $\Longrightarrow$ (iv): Sei V eine offene Menge mit $x \in V$. Dann existiert ein $\varepsilon > 0$ mit $U_\varepsilon(x) \subseteq V$. Wegen (i) gibt es ein $n_\varepsilon > 0$ mit $d(x_n, x) < \varepsilon$ für $n \geq n_\varepsilon$. Dann gilt für solche n also

$$x_n \in U_\varepsilon(x) \subseteq V.$$

Das beendet den Beweis. □

Definition. Sei (X,d) ein metrischer Raum. Dann heißt die Familie aller offenen Mengen auf X die durch d *induzierte Topologie* und wird mit $\mathcal{T} = \mathcal{T}_d$ bezeichnet.

Das System aller offenen Mengen in einem metrischen Raum hat einige bemerkenswerte Eigenschaften.

Proposition. *Sei (X,d) ein metrischer Raum mit von d erzeugter Topologie $\mathcal{T}$. Dann gilt:*

(T1) *Es gehören $\emptyset$ und X zu $\mathcal{T}$.*
(T2) *Gehören $V_1, \ldots, V_n$ zu $\mathcal{T}$, so auch $V = \bigcap_{j=1}^n V_j$. ("Endliche Schnitte offener Mengen sind offen.„)*
(T3) *Ist I eine Menge und gehören V_α, $\alpha \in I$, zu $\mathcal{T}$, so auch $\bigcup_{\alpha \in I} V_\alpha$. ("Beliebige Vereinigungen offener Menge sind offen.„)*

Beweis. Die Eigenschaften (T1) und (T3) sind einfach nachzuweisen. Nun zu (T2): Seien $V_1, \ldots, V_n \in \mathcal{T}$ und $V := \cap_{j=1}^n V_j$. Für $x \in V$ gilt $x \in V_j$ für alle $j = 1, \ldots, n$. Wegen $V_j \in \mathcal{T}$ existieren also $r_j > 0$ mit $U_{r_j}(x) \subseteq V_j$. Sei

$$r := \min\{r_j \mid j = 1, \ldots, n\} > 0.$$

Dann gilt

$$U_r(x) \subseteq U_{r_j}(x) \subseteq V_j$$

für alle $j = 1, \ldots, n$ und daher

$$U_r(x) \subseteq V.$$

Das beweist (T2). □

Bemerkung.
- Ist X eine beliebige Menge und $\mathcal{T}$ ein Familie von Teilmengen von X, die die Eigenschaften (T1), (T2)und (T3) hat, so heißt $\mathcal{T}$ eine Topologie und $(X, \mathcal{T})$ ein topologischer Raum.
- Auf topologischen Räumen kann man Konvergenz, Umgebung, Stetigkeit etc. definieren. Im allgemeinen wird man aber nicht mehr alles auf Folgen zurückführen können, sondern wird sogenannte Netze brauchen. Alle in dieser Vorlesung gegebenen Formulierungen mit Folgen nutzen also die spezielle Struktur eines metrischen Raumes. Die Formulierungen mit offenen Mengen und Umgebungen sind auch im allgemeinen gültig. Daher rührt auch ihre Relevanz.
- Im allgemeinen ist der Schnitt von abzählbar vielen offenen Mengen nicht mehr offen. Bsp. Sei $U_n = U_{1/n}(0)$ in $\mathbb{R}$ mit der Euklidischen Metrik. Dann gilt $\cap U_n = \{0\}$.

Folgerung. *Ist (X, d) ein metrischer Raum, so sind beliebige Schnitte von abgeschlossenen Mengen abgeschlossen.*

Beweis. Das folgt durch Komplementbildung □

Folgerung. *Sei (X, d) ein metrischer Raum und $A \subseteq X$ abgeschlossen und $U \subseteq X$ offen. Dann ist $U \setminus A$ offen und $A \setminus U$ abgeschlossen.*

Beweis. Es gilt $U \setminus A = U \cap (X \setminus A)$ und $A \setminus U = A \cap (X \setminus U)$. Damit folgen die Aussagen dann aus dem schon bewiesenen. □

Eine Teilmenge eines metrischen Raumes ist im allgemeinen weder offen noch abgeschlossen. Man kann aus einer Menge aber jeweils abgeschlossene und offene Menge “machen„.

Definition. (X, d) ein metrischer Raum. Dann heißt

$$M^\circ := \bigcup_{U \subseteq M, U \text{offen}} U$$

das *Innere* (oder der *offene Kern*) von M und

$$\overline{M} := \bigcap_{M\subseteq A, A \text{abgeschlossen}} A$$

der *Abschluss* (oder die *abgeschlossene Hülle*) von M und

$$\partial M := \overline{M} \setminus M^\circ$$

der *Rand* von M.

Bemerkung. Es ist M° die größte offene Teilmenge, die in M enthalten ist (da Vereinigung von offenen Mengen offen sind.) Insbesondere stimmt jede offene Menge mit ihrerem Inneren überein.

Bemerkung. Es ist $\overline{M}$ die kleinste abgeschlossene Menge, die M enthält. Insbesondere stimmt jede abgeschlossene Menge mit ihrer Hülle überein.

Ende 5. Vorlesung

Bemerkung. Es gilt (siehe Übung):

$$\begin{aligned} M^\circ &= \{x \in X \mid M \text{ ist Umgebung von } x\} \\ &= \{x \in X \mid \text{es existiert } r_x > 0 \text{ mit } U_{r_x}(x) \subseteq M\} \end{aligned}$$

$$\begin{aligned} \overline{M} &= \{x \in X \mid \text{jede Umgebung von } x \text{ enthält Punkt von } M\} \\ &= \{x \in X \mid \text{ es gibt } (x_n) \text{ in } M \text{ mit } x_n \to x\}. \end{aligned}$$

$$\begin{aligned} \partial M = \{x \in X \mid &\text{für jede Umgebung } U \text{ von } x \text{ gilt } U \cap M \neq \emptyset \\ &\text{und } U \cap X \setminus M \neq \emptyset\}. \end{aligned}$$

Beispiel. Betrachte $\mathbb{K}^m$ mit der Euklidischen Metrik und der induzieren Topologie. Dann ist das Innere von $B_r(x)$ gerade $U_r(x)$ und der Abschluss von $U_r(x)$ ist $B_r(x)$. (Übung.)

Beispiel. Für jeden metrischen Raum gilt,

$$U_r(x) \subseteq B_r(x)^\circ,$$

(da $U_r(x)$ offen ist) und,

$$\overline{U_r(x)} \subseteq B_r(x),$$

(da $B_r(x)$ abgeschlossen ist). Diese Inklusionen können strikt sein! Beispiele:

- $X =$ abgeschlossene Einheitskugel in $\mathbb{R}^2$. Dann ist $X = B_1(0)$ offen, also sein Inneres.
- $X =$ abgeschlossene Einheitskugel in $\mathbb{R}^2$ ohne eine punktierte Umgebung von $(1,0)$. Dann enthält der Abschluss von $U_1(0)$ nicht den Punkt $(1,0)$. Er gehört aber zu $B_1(0)$.

Sei $\mathbb{R}$ mit der durch die Euklidische Metrik erzeugten Topologie versehen. Dann ist der Rand von $\mathbb{Q}$ gerade $\partial\mathbb{Q} = \overline{\mathbb{Q}} \setminus \mathbb{Q}^\circ = \mathbb{R} \setminus \emptyset = \mathbb{R}$ (Vgl. Übung).

Zum Abschluss noch eine praktische Charakterisierung von Abgeschlossenheit.

Proposition. *Sei (X, d) ein metrischer Raum. Eine Teilmenge A von X ist genau dann abgeschlossen, wenn der Grenzwert jeder (in X) konvergenten Folge aus A wieder in A liegt.*

Beweis. Sei A abgeschlossen und (x_n) eine Folge in A mit $x_n \to x$. Zu zeigen: $x \in A$. Wäre x nicht in A, so gehörte es zu der offenen Menge $X \setminus A$. Damit wäre also x_n für alle großen n ebenfalls in $X \setminus A$. Widerspruch.
Es habe A die angegebene Eigenschaft. Wir müssen zeigen, daß $X \setminus A$ offen ist. Sei $x \in X \setminus A$ beliebig. Zu zeigen: ein $\delta > 0$ existiert mit $U_\delta(x) \cap A = \emptyset$. Angenommen: Für jedes $\delta > 0$ gilt $U_\delta(x) \cap A \neq \emptyset$. Dann existiert also ein Folge (x_n) mit $x_n \in U_{1/n}(x) \cap A$. Damit ist (x_n) eine Folge in A mit $x_n \to x$. Aus der angegebenen Eigenschaft folgt $x \in A$. Das ist ein Widerspruch. $\square$

3. Konvergenz und Stetigkeit

Mithilfe des Begriffes der Konvergenz können wir nun die Stetigkeit von Abbildungen fassen. Darum geht es in diesem Abschnitt.

Definition. Seien (X_1, d_1), (X_2, d_2) metrische Räume und $f : X_1 \longrightarrow X_2$ eine Abbildung. Es heißt f *stetig in* $x \in X_1$, wenn für jede Folge (x_n) in X_1 mit $x_n \to x$ bezüglich d_1 gilt $f(x_n) \to f(x)$ bezüglich d_2.

Kurzfassung: Es ist f stetig in x genau dann, wenn gilt: $x_n \to x \Longrightarrow f(x_n) \to f(x)$.

Lemma. *Seien (X_1, d_1), (X_2, d_2) metrische Räume und $f : X_1 \longrightarrow X_2$ eine Abbildung und $x \in X_1$. Dann sind äquivalent:* ***Zeichnung.***

(i) *Für jede Folge (x_n) in X_1 mit $x_n \to x$ gilt $f(x_n) \to f(x)$.*
(ii) *Für alle $\varepsilon > 0$ existiert ein $\delta > 0$ mit $f(U_\delta(x)) \subseteq U_\varepsilon(f(x))$. (Beachte: Es handelt sich um Kugeln in verschiedenen Räumen!)*
(iii) *Für jede Umgebung V von $f(x)$ existiert eine Umgebung U von x mit $f(U) \subseteq V$.*
(iv) *Für jede Umgebung V von $f(x)$ ist auch $f^{-1}(V)$ eine Umgebung von x. (Erinnerung: Das Urbild $f^{-1}(V)$ von V unter f ist gegeben als $f^{-1}(V) = \{x \in X_1 \mid f(x) \in V\}$)*

Bemerkung.
- Alle vier Aussagen geben eine präzise Version davon, dass Punkte nahe an x auf Punkte nahe an $f(x)$ abgebildet werden.

- In (ii) kann man natürlich statt mit offenen Kugeln auch mit abgeschlossenen Kugeln argumentieren.
- Die Formulierung in (iii) und (iv) nehmen keinen direkten Bezug auf eine Metrik.
- Die Formulierung (iv) zeigt die Nützlichkeit des Umgebungsbegriffes im Vergleich mit dem Begriff der Kugel: Im allgemeinen werden die (Ur)bilder von Kugeln auch für “schöne„ Funktionen keine Kugeln sein. So sind schon bei linearen Funktionen die Urbilder von Kugeln im allgemeinen Ellipsen, die keine Kugeln sind.

Beweis. (iv) $\Longrightarrow$ (iii): Man kann $U = f^{-1}(V)$ setzen.
(iii) $\Longrightarrow$ (ii): Es ist $U_\varepsilon(f(x))$ eine Umgebung von $f(x)$. Daher existiert nach (iii) eine Umgebung U von x mit $f(U) \subseteq B_\varepsilon(f(x))$. Da U Umgebung von x ist, existiert also ein $\delta > 0$ mit $U_\delta(x) \subseteq U$. Damit gilt also

$$f(U_\delta(x)) \subseteq f(U) \subseteq U_\varepsilon(f(x)).$$

(ii) $\Longrightarrow$ (i): Sei $\varepsilon > 0$. (Z.z. $d_2(f(x_n), f(x)) \leq \varepsilon$ für große n.) Sei zu diesem ε ein $\delta > 0$ gemäß (ii) gewählt. Wegen $x_n \to x$ existiert ein $n_\delta \in \mathbb{N}$ mit $d_1(x_n, x) \leq \delta$ für alle $n \geq n_\delta$. Damit gilt dann nach (ii) $d_2(f(x_n), f(x)) \leq \varepsilon$ für $n \geq n_\delta$.
(i) $\Longrightarrow$ (iv): Sei V eine Umgebung von $f(x)$. Angenommen: Es ist $f^{-1}(V)$ keine Umgebung von x. Dann gilt also für kein $\delta > 0$ die Inklusion $U_\delta(x) \subseteq f^{-1}(V)$. Daher gibt es also (mit $\delta = 1/n$) zu jedem $n \in \mathbb{N}$ eine $x_n \in U_{1/n}(x)$ mit $f(x_n) \notin V$. Dann gilt aber $x_n \to x$ und $f(x_n)$ konvergiert nicht gegen $f(x)$. Das ist ein Widerspruch. □

Stetigkeit ist verträglich mit den “üblichen„ Operationen.

PROPOSITION. *Seien (X_j, d_j), $j = 1, 2, 3$, metrische Räume und $f : X_1 \longrightarrow X_2$, $g : X_2 \longrightarrow X_3$ gegeben. Ist f stetig in $x \in X_1$ und g stetig in $f(x)$, so ist $g \circ f : X_1 \longrightarrow X_3$ stetig in x.*

Beweis. Das folgt aus der Folgencharakterisierung der Stetigkeit wie folgt: $x_n \to x$ impliziert, $f(x_n) \to f(x)$ (da f stetig in x) und das impliziert $g(f(x_n)) \to g(f(x))$ (da g stetig in $f(x)$). □

PROPOSITION. *Sei (X, d) ein metrischer Raum und $f, g : X \longrightarrow \mathbb{C}$ seien stetig in $x \in X$. Dann sind auch $f + g : X \longrightarrow \mathbb{C}$ und $fg : X \longrightarrow \mathbb{C}$ stetig in x*

Beweis. Das folgt sofort aus der Folgencharakterisierung der Stetigkeit.
□

PROPOSITION. *Ist (X, d) ein metrischer Raum und sind $f, g : X \longrightarrow \mathbb{R}$ stetig in $x \in X$, so sind auch* $\max\{f, g\}$, $\min\{f, g\}$ *und* $|f|$ *stetig in $x \in X$.*

Beweis. (Übung) □

DEFINITION. Seien (X_1, d_1), (X_2, d_2) metrische Räume und $f : X_1 \longrightarrow X_2$ eine Abbildung. Dann heißt f *stetig*, wenn es in jedem Punkt von X_1 stetig ist.

Ende der 6. Vorlesung

THEOREM (Charakterisierung Stetigkeit). *Seien (X_1, d_1), (X_2, d_2) metrische Räume und $f : X_1 \longrightarrow X_2$ eine Abbildung. Dann sind äquivalent:*

(i) *f ist stetig.*
(ii) *Die Urbilder offener Mengen unter f sind offen, d.h. für jedes $V \in \mathcal{T}_2$ gehört $f^{-1}(V)$ zu $\mathcal{T}_1$.*

BEMERKUNG. Dieser Satz allein rechtfertigt schon die Einführung des Konzeptes der offenen Menge.

Beweis. (i) $\Longrightarrow$ (ii): Sei $V \in \mathcal{T}_2$ beliebig und $U := f^{-1}(V)$. Zu zeigen: U ist Umgebung von jedem $x \in U$. Sei also $x \in U$ beliebig. Dann folgt $f(x) \in V$. Da V offen ist, ist es eine Umgebung von $f(x)$ und damit folgt nach (i) (und (iii) aus Lemma oben) die Existenz einer Umgebung U_x von x mit

$$f(U_x) \subseteq V.$$

Damit folgt also $x \in U_x \subseteq U$ (nach Definition von U). Damit ist U Umgebung von x.

(ii) $\Longrightarrow$ (i): Zu zeigen: f ist stetig in jedem $x \in X_1$. Sei also $x \in X_1$ beliebig. Sei V eine Umgebung von $f(x)$. Dann existiert also eine offene Menge W mit

$$f(x) \in W \subseteq V.$$

Nach (ii) ist dann $U := f^{-1}(W)$ offen. Weiterhin gilt nach Definition $x \in U$. Damit ist also U eine Umgebung von x mit $f(U) \subseteq W \subseteq V$. Also ist f stetig in x nach (iii) aus Lemma oben. □

BEISPIEL. Alle linearen Funktionen $A : \mathbb{R}^N \longrightarrow \mathbb{R}^M$ sind stetig. Bew. Übung.

Zum Abschluss des Abschnittes disktuieren wir noch eine Anwendung:
Stetigkeit und Produkträume.
Seien (X_j, d_j), $j = 1, \ldots, N$ metrische Räume und sei

$$X := X_1 \times \ldots \times X_N = \{(x_1, \ldots, x_N) \mid x_j \in X_j, j = 1, \ldots, N\}$$

ihr Produkt. Ähnlich wie man $\mathbb{R}^N$ mittels der Euklidischen Metrik von $\mathbb{R}$ metrisiert, kann man dann auch den Raum X metrisieren. So ist zum Beispiel

$$d(x_1, \ldots, x_N), (y_1, \ldots, y_N)) := \sum_{j=1}^{N} d_j(x_j, y_j)$$

eine Metrik auf X. Es gilt dann offenbar

$$x^{(n)} = (x_1^{(n)}, \dots, x_N^{(n)}) \to x = (x_1, \dots, x_N)$$

bezüglich d genau dann, wenn $x_j^{(n)} \to x_j$ bzgl d_j für jedes $j = 1, \dots, N$. Offenbar gilt für d die Gleichheit

$$d((x_1, \dots, x_N), (y_1, \dots, y_N)) = \|((d_1(x_1, y_1), \dots, d_N(x_N, y_N))\|_1.$$

Allgemeiner kann man zu jeder Norm $\|\cdot\|$ auf $\mathbb{R}^N$ eine Metrik $e_{\|\cdot\|}$ auf X einführen durch

$$e_{\|\cdot\|}(x_1, \dots, x_N), (y_1, \dots, y_N)) = \|((d_1(x_1, y_1), \dots, d_N(x_N, y_N))\|.$$

Aufgrund der Äquivalenz aller Normen im $\mathbb{R}^N$, gibt es dann für jede Norm $\|\cdot\|$ Konstanten $c, C > 0$ mit

$$cd \le e_{\|\cdot\|} \le Cd.$$

Darum spielt es für die Untersuchungen von konvergenter Folgen bzw Cauchy-Folgen keine Rolle, welche dieser Metrik wir tatsächlich verwenden.

Es ist (Übung) (X, d) genau dann vollständig, wenn (X_j, d_j) vollständig ist für jedes $j = 1, \dots, N$.

Für Stetigkeit von Funktionen erhalten wir im Zusammenhang mit Produkträumen folgendes:

Sei (Z, e) ein metrischer Raum und (X, d) der obige Produktraum. Dann ist $f : X \longrightarrow Z$ stetig genau dann, wenn aus $x_j^{(n)} \to x_j$, $j = 1, \dots, N$, folgt $f((x_1^{(n)}, \dots, x_N^{(n)})) \to f(x_1, \dots x_N))$.

Es ist $f : Z \longrightarrow X, f(z) = (f_1(z), \dots, f_N(z))$ genau dann stetig, wenn jede einzelne Komponente $f_j : Z \longrightarrow X_j$ stetig ist.

Beispiel. Ist (Z, e) ein beliebiger metrischer Raum, so ist $e : Z \times Z \longrightarrow \mathbb{R}$ stetig. (Bew. Die entsprechende Aussage für Folgenkonvergenz wurde oben schon gezeigt.)

Beispiel. Es sind $+ : \mathbb{K} \times \mathbb{K} \longrightarrow \mathbb{K}$ und $\cdot : \mathbb{K} \times \mathbb{K} \longrightarrow \mathbb{K}$ stetig. (Bew. In der Folgencharakterisierung ist das schon bekannt).

Zum Ende des Kapitels führen wir noch eine Notation ein, nämlich Grenzwerte.

Seien (X, d) und (Y, e) metrische Räume und $f : X \longrightarrow Y$ eine Funktion. Sei $p \in X$ und $q \in Y$. Dann hat f bei p den Grenzwert q geschrieben als

$$\lim_{x \to p} f(x) = q$$

genau dann, wenn gilt

- $f(x_n) \to q$ für jede Folge (x_n) in X mit $x_n \to p$.

bzw. äquivalent, wenn gilt

- zu jedem $\varepsilon > 0$ ein $\delta > 0$ existiert mit $e(f(x), q) \le \varepsilon$ für alle $x \in U$ mit $d(x, p) \le \delta$.

Offenbar ist dann also f stetig in p, genau dann wenn der Grenzwert von f in p existiert.
Unter Umständen hat man es mit Funktionen zu tun, die in p nicht definiert sind. Dann hat f auf in p den Grenzwert q, geschrieben

$$\lim_{x \to p, x \neq p} f(x) = q,$$

falls

- $f(x_n) \to q$ für jede Folge (x_n) in $X \setminus \{p\}$ mit $x_n \to p$.

bzw. äquivalent, wenn

- zu jedem $\varepsilon > 0$ ein $\delta > 0$ existiert mit $e(f(x), q) \leq \varepsilon$ für alle $x \in X \setminus \{p\}$ mit $d(x, p) \leq \delta$.

4. Kompaktheit

In diesem Abschnitt lernen wir ein fundamentales Konzept der Topologie kennen nämlich das Konzept der Kompaktheit. Kompaktheit ist eine ausgesprochen nützliche Eigenschaft (verlangt aber eine etwas abstrakte Definition).

Definition. Sei (X, d) ein metrischer Raum. Eine Teilmenge $K \subseteq M$ heißt *kompakt*, wenn jede Folge (x_n) in K ein Teilfolge (x_{n_k}) besitzt, die gegen einen Grenzwert aus K konvergiert.

Bemerkung. Das so definierte Konzept ist auch als Folgenkompaktheit bekannt. In metrischen Räumen stimmt es mit anderen Konzepten von Kompaktheit überein (s.u.). Da uns nur metrische Räume interessieren, werden wir nicht zwischen Folgenkompaktheit und Kompaktheit unterscheiden.

Bemerkung. Der Fall $K = X$ in obiger Definition ist natürlich möglich. In "Anwendungen„ hat man es aber oft mit der Situation zu tun, dass X nicht kompakt ist, man sich aber für die kompakten Mengen in X interessiert.

Beispiel. Sei $I = [a, b]$ ein abgeschlossenes beschränktes Intevall in $\mathbb{R}$. Dann ist I kompakt. (Bew. Nach Bolzano/Weierstrass hat jede Folge in I eine (in $\mathbb{R}$) konvergente Teilfolge. Da das Intervall abgeschlossen ist, liegt der Grenzwert der Teilfolge in I).

Beispiel. Sei A eine abgeschlossene bezüglich der Euklidischen Metrik beschränkte Menge in $\mathbb{K}^N$. Dann ist A kompakt. Das werden wir unten genauer diskutieren. Hier weisen wir schon darauf hin, daß wir das eigentlich kürzlich mitbewiesen haben, als es um die Äquivalenz aller Normen im $\mathbb{K}^N$ ging.

Kompaktheit ist eine nützliche Eigenschaft, wie die folgenden Resultate zeigen. Diese Resultate haben wir für stetige Funktionen auf abgeschlossenen beschränkten Intervallen in $\mathbb{R}$ schon untersucht. Die dort gegebenen Beweise überträgen sich wortwörtlich. Weil sie so schön sind, geben wir sie hier noch einmal.

THEOREM. *Sei* (K, d) *kompakter metrischer Raum und* $f : K \longrightarrow \mathbb{R}$ *stetig. Dann nimmt* f *Minimum und Maximum an.*

Beweis. Wir behandeln nur das Maximum. (Das Minimum kann analog behandelt werden.) Wähle eine Folge (x_n) in K mit

$$\lim_{n\to\infty} f(x_n) = \sup_{x\in X} f(x).$$

(Hier ist apriori auch die bestimmte Divergenz gegen ∞ möglich.) Da K kompakt ist, hat (x_n) eine in K konvergente Teilfolge, d.h. es gibt $(x_{n_k})_k$ und $x \in K$ mit $x_{n_k} \to x$. Da f stetig ist, folgt dann

$$f(x) = \lim_{k\to\infty} f(x_{n_k}) = \sup_{x\in K} f(x).$$

Das beendet den Beweis. □

Ende 7. Vorlesung

THEOREM. *Sei* (K, d) *ein kompakter metrischer Raum und* (Y, e) *ein metrischer Raum und* $f : K \longrightarrow Y$ *stetig. Dann ist* f *gleichmäßig stetig (d.h. für alle* $\varepsilon > 0$ *existiert ein* $\delta > 0$*, so dass für alle* $x, y \in K$ *mit* $d(x, y) < \delta$ *gilt* $e(f(x), f(y)) < \varepsilon$*.)*

Beweis. Angenommen f ist nicht gleichmäßig stetig und sei $\varepsilon > 0$. Dann existieren also zu jedem $n \in \mathbb{N}$ Elemente $x_n, y_n \in K$ mit $d(x_n, y_n) \leq \frac{1}{n}$ und $e(f(x_n), f(y_n)) \geq \varepsilon$. Da K kompakt ist, haben (x_n), (y_n) eine in K konvergente Teilfolge. Ohne Einschränkung sei (x_n) konvergent gegen $x \in K$. Dann konvergiert auch (y_n) gegen x. Damit folgt dann

$$0 = e(f(x), f(x)) = \lim_{n\to\infty} e(f(x_n), f(y_n)) \geq \varepsilon.$$

Das ist ein Widerspruch. □

THEOREM. *Sei* (K, d) *ein kompakter metrischer Raum und* (Y, e) *ein metrischer Raum und* $f : K \longrightarrow Y$ *stetig. Dann ist* $f(K)$ *kompakt.*

Beweis. Sei (y_n) eine Folge in $f(K)$. Dann existieren x_n so dass $f(x_n) = y_n$, $n \geq 1$. Da K kompakt ist existiert eine konvergente Teilfolge (x_{n_k}). Da f stetig ist, konvergiert $(y_{n_k}) = (f(x_{n_k})$. Das heißt, $f(K)$ ist kompakt. □

BEMERKUNG. (a) Das ist einer der beiden einzigen globalen Sätze über stetige Funktionen. (b) Anders als bei den Betrachtungen zur Stetigkeit, geht es hier um die Bilder (und nicht die Urbilder) von f.

PROPOSITION. *Sei* (X, d) *metrischer Raum. Ist* $K \subseteq X$ *kompakt und* $A \subseteq K$ *abgeschlossen, so ist* A *kompakt.*

Beweis. Sei (x_n) eine Folge in A. Dann ist (x_n) eine Folge in K. Daher hat also (x_n) eine in K konvergente Teilfolge. Da A abgeschlossen ist, liegt der Grenzwert der Teilfolge sogar in A. $\square$

PROPOSITION. *Sei (X, d) metrischer Raum und $K \subseteq X$ kompakt. Dann ist K abgeschlossen.*

Beweis. Sei (x_n) eine Folge in K, die gegen $x \in X$ konvergiert. Aufgrund der Kompaktheit enthält (x_n) eine Teilfolge, die gegen ein $y \in K$ konvergiert. Dann muss aber gelten $x = y \in K$. $\square$

BEMERKUNG. Als Konsequenz der beiden vorangehenden Propositionen ist der Schnitt einer kompakten Menge mit einer abgeschlossenen Menge wieder kompakt.

THEOREM (Automatische Stetitgkeit der Inversen). *Sei (K, d) ein kompakter metrischer Raum und (L, e) ein metrischer Raum und $f : K \longrightarrow L$ sei bijektiv und stetig. Dann ist die inverse Abbildung $g = f^{-1} : L \longrightarrow K$ stetig.*

Beweis. Zu zeigen $g^{-1}(U)$ offen für jedes offene $U \subseteq K$. Es reicht zu zeigen: $g^{-1}(A)$ abgeschlossen für jedes abgeschlossene $A \subseteq K$. Ist $A \subseteq K$ abgeschlossen, so ist A kompakt. Daher ist aufgrund der Stetigkeit von f auch

$$g^{-1}(A) = f(A)$$

kompakt, also nach vorheriger Proposition abgeschlossen. $\square$

Wir kommen nun zur konzeptuellen Untersuchung und Charakterisierung von Kompaktheit.

DEFINITION. Sei (X, d) ein metrischer Raum. Eine Teilmenge $B \subseteq X$ (z.B. $B = X$) heißt *total beschränkt*, wenn zu jedem $\varepsilon > 0$ ein $N \in \mathbb{N}$ und $x_1, \ldots, x_N \in X$ (äquivalent $x_1, \ldots, x_N \in B$) existieren mit

$$B \subseteq \bigcup_{k=1}^{N} B_\varepsilon(x_k).$$

BEMERKUNG.
- Statt mit abgeschlossenen Kugeln könnte man auch offenen Kugeln arbeiten.
- Wir wissen schon, dass jede Metrik d zu einer Metrik e äqivalent ist, in der der ganze Raum eine beschränkte Menge ist (Z.B. $e = \frac{d}{1+d}$). Daher ist Beschränktheit einer Menge keine besonders kanonische Eigenschaft. Das Konzept der total Beschränktheit ersetzt das Konzept der Beschränktheit. Ist ein Raum bgzl. zweier äquivalenter Metriken vollständig, so haben beide Metriken dieselben total beschränkten Mengen (s.u.).

BEISPIEL. **Total beschränkte Mengen im $\mathbb{K}^N$.** Sei $\mathbb{K}^N$ mit der Euklidischen Metrik versehen. Dann ist ein $B \subseteq \mathbb{K}^N$ genau dann total beschränkt, wenn es beschränkt ist.
Beweis. $\Longrightarrow$: Sei B beschränkt. Betrachte zu $\varrho > 0$ das Gitter $\Gamma_\varrho := (\varrho\mathbb{Z})^N$. Dann schneidet B aufgrund der Beschränktheit (und des archimedischen Axioms) nur endllich viele Maschen des Gitters. Jede dieser Maschen kann mit einer Kugel vom Radius $N\rho$ überdeckt werden. Damit folgt die gewünschte Implikation (da $\varrho > 0$ beliebig klein gemacht werden kann).
$\Longleftarrow$: Das ist klar. □

Ende 8. Vorlesung

BEMERKUNG. Eine total beschränkte Menge muss nicht vollständig sein. Bsp: $(0,1)$ mit der üblichen Euklidischen Metrik.

Für die folgenden Betrachtungen werden noch zwei Stücke Notation von Nutzen sein.

NOTATION. Sei (X, d) ein metrischer Raum.

- Ist (x_n) eine Folge in X und B eine Teilmenge von X, so sagen wir, dass in B *unendlich viele Folgeglieder liegen*, wenn die Menge $\{n \in \mathbb{N} \mid x_n \in B\}$ unendlich viele Elemente hat.
- Ist B eine Teilmenge von X und sind U_α, $\alpha \in A$, Teilmengen von X so sagen wir, dass die U_α die Menge B *überdecken*, wenn $B \subseteq \cup_{\alpha\in A} U_\alpha$ gilt.

Wir erinnern an den Satz von Bolzano/Weierstrass. Dieser besagt, dass eine Menge I in $\mathbb{R}$ genau dann beschränkt ist, wenn jede Folge eine konvergente Teilfolge hat. (Hier ist die schwere Richtung die eigentliche Aussage des Satzes von Bolzano/Weierstrass und die andere Richtung ist klar). Die total beschränkten Menge lassen sich mit folgender Variante des Satzes von Bolzano/Weierstrass charakterisieren.

LEMMA (Charakterisierung total beschränkt). *Sei (M, d) ein metrischer Raum und $B \subseteq M$. Dann sind äquivalent:*

(i) *Es ist B total beschränkt.*
(ii) *Jede Folge in B hat eine Teilfolge, die eine Cauchy-Folge ist.*

Beweis. (i) $\Longrightarrow$ (ii): Sei (x_n) eine Folge in B. Wir überdecken nun B durch endlich viele Kugeln vom Radius 1. In einer von diesen Kugeln müssen dann unendlich viele Folgeglieder von (x_n) liegen.
Bezeichne den Schnitt dieser Kugel mit B als B_1. Wir überdecken nun B_1 (das als Teilmenge des total beschränkten B ebenfalls total beschränkt ist) durch endlich viele Kugeln vom Radius $1/2$. Der Schnitt einer dieser Kugeln mit B_1 muss dann unendlich viele Folgeglieder von (x_n) enthalten. Nenne diesen Schnitt B_2. Induktiv können wir so eine Folge von Mengen B_n, $n \in \mathbb{N}$, konstruieren mit

- $B \supseteq B_1 \supseteq \ldots \supseteq B_{n-1} \supseteq B_n$
- $\sup_{x,y\in B_n} d(x,y) =: \mathrm{diam} B_n \leq \frac{1}{2^n}$,
- Jedes B_n enthält unendlich viele Folgeglieder.

(Vgl. Intervallhalbierungsverfahren zum Beweis des Satzes von Bolzano/ Weierstrass im ersten Semester). Aufgrund der dritten Eigenschaft können wir dann eine Teilfolge (x_{n_k}) von (x_n) konstruieren mit $x_{n_k} \in B_k$ für alle $k \in \mathbb{N}$. Aufgrund der ersten beiden Eigenschaften handelt es sich um eine Cauchy-Folge.
(ii) $\Longrightarrow$ (i): Angenommen B ist nicht total beschränkt. Dann existiert also ein $\varepsilon > 0$ so dass B nicht mit endlich vielen Kugeln mit dem Radius ε überdeckt werden kann. Dann können wir induktiv eine Folge (x_n), $n \in \mathbb{N}$, in B von Kugelmittelpunkten konstuieren mit

$$x_{n+1} \notin \bigcup_{k=1}^{n} B_\varepsilon(x_n).$$

$n = 1$: Wähle x_1 beliebig.
$n \Longrightarrow n+1$: Es wird B nicht von $\cup_{k=1}^{n} B_\varepsilon(x_n)$ überdeckt nach Annahme. Die Elemente dieser Folge erfüllen

$$d(x_n, x_k) \geq \varepsilon$$

für alle $n, k \in \mathbb{N}$ mit $n \neq k$. Daher kann diese Folge keine Teilfolge haben, die eine Cauchy-Folge ist. Widerspruch zu (ii). □

Damit können wir nun Kompaktheit charakterisieren.

Theorem (Charakterisierung Kompaktheit). *Sei (X, d) ein metrischer Raum und $K \subseteq X$. Dann sind äquivalent:*

(i) *Es ist K total beschränkt und vollständig.*
(ii) *Jede Folge in K hat eine Teilfolge, die in K konvergiert.*

Beweis. Das folgt leicht aus dem vorherigen Lemma.
(i) $\Longrightarrow$ (ii): Da K total beschränkt ist, hat nach dem vorherigen Lemma jede Folge in K eine Teilfolge, die Cauchy-Folge ist. Aufgrund der Vollständigkeit von K konvergiert diese dann in K.
(ii) $\Longrightarrow$ (i): Nach (ii) hat insbesondere jede Folge in K eine Teilfolge, die Cauchy-Folge ist. Daher ist nach dem Lemma also die Menge K total beschränkt. Noch zu zeigen: K vollständig. (Hier kommt das einzig neue im Beweis, das nicht schon aus dem vorherigen Lemma folgt.) Sei (x_n) eine Cauchy-Folge in K. Dann enthält (x_n) nach (ii) eine gegen ein $x \in K$ konvergente Teilfolge (x_{n_k}). Damit ist (nach Standardschlüssen) die Folge (x_n) selber konvergent gegen x. Denn es gilt

$$d(x, x_n) \leq d(x, x_{n_k}) + d(x_{n_k}, x_n).$$

Es konvergiert nun $(d(x, x_{n_k}))$ gegen 0, da $x_{n_k} \to x$ und es wird $d(x_{n_k}, x_n)$ beliebig klein für k, n groß, da es sich um eine Cauchy-Folge handelt.
□

Wir erinnern im Zusammenhang mit Vollständigkeit von Teilmengen noch an folgenden Zusammenhang.

LEMMA. *Sei* (X, d) *ein vollständiger metrischer Raum. Sei* A *eine Teilmenge von* X*. Dann ist* A *abgeschlossen (in* X*) genau dann, wenn der metrische Raum* (A, d) *vollständig ist.*

Beweis. Sei A abgeschlossen. Sei (x_n) eine Cauchy-Folge in (A, d). Dann ist (x_n) eine Cauchy-Folge in X. Damit konvergiert (x_n) in X gegen ein x. Da A abgeschlossen ist, gehört x zu A.
Sei (A, d) vollständig. Sei (x_n) eine gegen $x \in X$ konvergente Folge in A. Zu zeigen $x \in A$: Es ist (x_n) eine Cauchy-Folge in X und damit auch in A. Daher hat (x_n) einen Grenzwert in A. Dieser muss dann mit x übereinstimmen.

□

BEISPIEL. **Kompaktheit in $\mathbb{K}^N$.** In $\mathbb{K}^N$ ist eine Menge total beschränkt genau dann, wenn sie beschränkt ist. Sie ist vollständig, genau dann, wenn sie abgeschlossen ist. Damit erhalten wir folgendes Resultat: Ein $K \subseteq \mathbb{K}^N$ ist kompakt genau dann, wenn es abgeschlossen und beschränkt ist.

Eine Variante der Betrachtungen von $\mathbb{R}^N$ führt auf folgende Charakterisierung von Totaler Beschränktheit mittels Kompaktheit.

FOLGERUNG. *Sei* (M, d) *ein vollständiger metrischer Raum. Sei* $A \subseteq M$*. Dann sind äquivalent:*

(i) A *ist total beschränkt.*
(ii) *Es ist* $\overline{A}$ *kompakt.*

Beweis. Man sieht leicht, dass eine Menge A von X total beschränkt ist genau dann wenn $\overline{A}$ total beschränkt ist. Damit können wir wie folgt weiterschließen:
Ist $\overline{A}$ kompakt, so ist es also total beschränkt. Dann ist auch A total beschränkt.
Sei umgekehrt A total beschränkt. Dann ist auch $\overline{A}$ total beschränkt (einfach) und vollständig (als abgeschlossene Menge eines vollständigen Raumes). Daher ist $\overline{A}$ kompakt. □

DEFINITION. Sei (X, d) ein metrischer Raum. Eine Menge A in X heißt relativ kompakt, wenn ihr Abschluss kompakt ist.

Wir kommen nun zu einer weiteren Charakterisierung von Kompaktheit. Diese lässt sich auch für allgemeine topologische Räume geben.

THEOREM. *Sei* (M, d) *ein metrischer Raum und* $K \subseteq M$*. Dann sind äquivalent:*

(i) *Jede Folge in* K *hat eine in* K *konvergente Teilfolge.*

(ii) *Jede offenen Überdeckung von K hat eine endliche Teilüberdeckung. (Das heißt: Zu allen U_α, $\alpha \in A$, offen mit $K \subseteq \cup U_\alpha$ existiert $N \in \mathbb{N}$ und $\alpha_1, \ldots, \alpha_N$ mit $K \subseteq \cup_{j=1}^N U_{\alpha_j}$.) „K ist überdeckungskompakt".*

Bemerkung. Die Eigenschaft (ii) kann auch als Definition von Kompaktheit genommen werden. Man spricht dann von Überdeckungskompaktheit (oder der Heine-Borel Eigenschaft). Der Satz besagt dann, dass in metrischen Räumen Überdeckungskompaktheit äquivalent zu Folgenkompaktheit ist.

Bemerkung. In der Eigenschaft (ii) wird kein Bezug auf die Metrik genommen sondern nur auf die erzeugte Topologie. Insbesondere haben also zwei Metriken, die dieselbe Topologie erzeugen, dieselben kompakten Mengen.

Beweis. (ii) $\Longrightarrow$ (i): Sei (x_n) eine Folge in K. Angenommen (x_n) enthält keine konvergente Teilfolge. Dann hat (x_n) also keinen Häufungspunkt. Dann existiert zu jedem $x \in K$ also ein $\delta_x > 0$ so dass in $U_{\delta_x}(x)$ nur endlich viele Folgeglieder liegen. Dann ist $U_{\delta_x}(x)$ eine offene Überdeckung von K und hat also eine endliche Teilüberdeckung

$$U_{\delta_{p_j}}(p_j), \;\; j = 1, \ldots, N.$$

Da jedes $U_{\delta_{p_j}}$ nur endlich viele Folgeglieder enthält, gibt es also insgesamt nur endlich viele Folgeglieder. Das ist ein Widerspruch.
(i) $\Longrightarrow$ (ii): Sei U_α, $\alpha \in A$, eine offene Überdeckung von K. Wir gehen in drei Schritten vor.
Schritt 1: Es gibt ein $\delta > 0$, so dass für jedes $x \in K$ ein $\alpha_x \in A$ existiert mit $x \subseteq U_\delta(x) \subseteq U_{\alpha_x}$. 'Jede δ-Kugel liegt in einem U_α'.
Bew. Angenommen nein! Dann gibt es also zu jedem $n \in \mathbb{N}$ (für $\delta = 1/n$) ein $x_n \in K$ sodass $U_{1/n}(x_n)$ nicht in U_α enthalten ist für alle $\alpha \in A$. Aufgrund von (i) hat (x_n) eine konvergente Teilfolge. Ohne Einschränkung sei $x_n \to x \in K$. Dann gibt es ein $\alpha \in A$ mit $x \in U_\alpha$. Da U_α offen ist, ist dann auch $U_r(x) \subseteq U_\alpha$ für ein geeignetes $r > 0$. Für hinreichend große n ist dann aber $d(x_n, x) < r/2$ und $1/n < r/2$ also

$$U_{1/n}(x_n) \subseteq U_{r/2}(x_n) \subseteq U_r(x) \subseteq U_\alpha.$$

Das ist ein Widerspruch zur Konstruktion der (x_n).
Schritt 2: Zu $\delta > 0$ aus Schritt 1, gibt es ein $N \in \mathbb{N}$ und $x_1, \ldots, x_N \in K$ mit $K \subseteq \cup_{j=1}^N U_\delta(x_j)$.
Bew. Nach (i) und der Charakterisierung von Kompaktheit, ist K total beschränkt. Damit folgt Schritt 2 (für jedes $\delta > 0$).
Schritt 3: Es gilt (ii).
Bew. Seien $\delta > 0$ und $x_1, \ldots, x_N$ aus Schritt 2 gegeben. Dann gilt also

$$K \subseteq \bigcup_{j=1}^N U_\delta(x_j).$$

Nach Schritt 1 ist jedes $U_\delta(x_j)$ in einem U_{α_j} enthalten und (ii) folgt. □

Ende 9. Vorlesung

Proposition. *Seien (X, d) und (Y, e) metrische Räume und $K \subseteq X$ und $L \subseteq Y$ kompakt. Dann ist $K \times L \subseteq X \times Y$ kompakt.*

Beweis. Sei (x_n, y_n) eine Folge in $K \times L$. Dann ist (x_n) eine Folge in K. Damit hat (x_n) eine in K gegen ein $x \in K$ konvergente Teilfolge (x_{n_k}). Dann ist (y_{n_k}) eine Teilfolge in L. Daher hat es eine gegen ein $y \in L$ konvergente Teilfolge (y_{m_k}). Dann gilt

$$x_{m_k} \to x \text{ und } y_{m_k} \to y.$$

Damit konvergiert (x_{m_k}, x_{n_k}) gegen $(x, y) \in K \times L$. □

Bemerkung. Alternativ hier noch ein Überdeckungsbeweis: Sei W_ι, $\iota \in I$, eine offene Überdeckung von $K \times L$.
Der Beweis wird nun in drei Schritten geführt. (Zeichnung)
Schritt 1: Es gibt zu jedem $(x, y) \in K \times L$ ein $\iota(x, y) \in A$ und Umgebungen $U_{(x,y)}$ von x und $V_{(x,y)}$ von y mit

$$(x, y) \in U_{(x,y)} \times V_{(x,y)} \subseteq W_{\iota(x,y)}.$$

Bew. Das folgt aus der Überdeckungseigenschaft unter Definition der Produkttopologie.
Schritt 2: Für festes $x \in K$ ist $(V_{(x,y)})_{y\in L}$ eine offene Überdeckung von Y. Also gibt es $y_1(x), \ldots, y_{N(x)}(x)$ mit

$$L \subseteq V_{(x,y_1(x))} \cup \ldots \cup V_{(x,y_{N(x)}(x))}.$$

Setze nun

$$U_x := U_{(x,y_1(x))} \cap \ldots \cap U_{(x,y_{N(x)}(x))}.$$

Schritt 3: Aufgrund der Kompaktheit von K gibt $x_1, \ldots, x_M$ in K mit

$$K \subseteq U_{x_1} \cup \ldots \cup U_{x_M}.$$

Damit überdecken dann die endlich vielen

$$W_{\iota(x_j,y_k(x_j))},\ j = 1, \ldots, M, k = 1, \ldots, N(x_j),$$

die Menge $K \times L$.

5. Zusammenhang

In diesem Abschnitt diskutieren wir ein weiteres topologisches Konzept, nämlich Zusammenhang.

Definition. Ein metrischer Raum (X, d) heißt *zusammenhängend*, wenn er nicht in zwei offene disjunkte nichtleere Teilmengen zerlegt werden kann (d.h. wenn aus $X = U \cup V$ mit U, V offen, $U \cap V = \emptyset$ folgt $U = X$ oder $V = X$). Eine Teilmenge A eines metrischen Raumes heißt *zusammenhängend*, wenn (A, d) als metrischer Raum zusammenhängend ist.

Zeichnung.

Bemerkung (Erinnerung offene Mengen in Teilräumen). Sei (X, d) ein metrischer Raum und $A \subseteq X$ eine Teilmenge. Ist dann U eine Teilmenge des metrischen Raumes (A, d) so ist U offen (im metrischen Raum (A, d)) genau dann, wenn es eine offene Menge U' in (X, d) gibt mit $U = A \cap U'$. **Zeichnung.**
(Bew. Es gilt offenbar

$$U_r^A(x) := \{a \in A \mid d(a, x) < r\} = U_r^X(x) \cap A.$$

Damit folgt die Aussage leicht.)

Hier ist das Hauptergebnis über stetige Funktionen auf zusammenhängenden Mengen.

Theorem. *Seien (X, d) und (Y, e) metrische Räume und (X, d) zusammenhängend und $f : X \longrightarrow Y$ stetig, dann ist $f(X)$ zusammenhängend.*

Beweis. Seien U, V disjunkte offene Mengen in $(f(X), d)$ mit $U \cup V = f(X)$. Dann sind aufgrund der Stetigkeit von f also

$$U_X := f^{-1}(U), \;\; V_X := f^{-1}(V)$$

offene disjunkte Mengen in X mit $X = U_X \cup V_X$. Da X zusammenhängend ist, folgt $U_X = \emptyset$ oder $V_X = \emptyset$. Das liefert die Behauptung. $\square$

Bemerkung. Wie wir schon erwähnt hatten, gibt es eigentlich nur zwei globale Sätze zu stetigen Funktionen. Diese besagen:

- Stetige Bilder kompakter Mengen sind kompakt.
- Stetige Bilder zusammenhängender Mengen sind zusammenhängend.

Beispiel. **Zusammenhängende Mengen in $\mathbb{R}$.** Wir untersuchen nun noch die Struktur der zusammenhängenden Mengen in $\mathbb{R}$. Das wird ein strukturelles Verständnis des Zwischenwertsatzes liefern.

Theorem. *Die zusammenhängenden Mengen in $\mathbb{R}$ sind gerade die Intervalle.*

Beweis. Sei $I \subseteq \mathbb{R}$ zusammenhängend. Zu zeigen: Mit $x, y \in I$ gehört auch jeder Punkt zwischen x und y zu I.
Sei ohne Einschränkung $x < y$. Betrachte p mit $x < p < y$. Angenommen $p \notin I$. Dann bilden $U := (-\infty, p) \cap I$ und $V : (p, \infty) \cap I$ eine offene disjunkte Zerlegung von I in nichtleere Mengen. Widerspruch.
Sei nun $I \subseteq \mathbb{R}$ ein Intervall. Sei $I = U \cup V$ eine Zerlegung von I in offene disjunkte Mengen. Betrachte die charakteristische Funktion

$$1_U : I \longrightarrow \mathbb{R},$$

mit $1_U(x) = 1$ falls $x \in U$ und $1_U(x) = 0$ sonst. Dann sind die Urbilder von beliebigen Mengen in $\mathbb{R}$ unter 1_U offenbar entweder $\emptyset$ oder U oder V oder I. Damit ist 1_U stetig. Da 1_U nur Werte in $\{0, 1\}$ annimmt, folgt nach dem Zwischenwertsatz, dass 1_U konstant sein muss. $\square$

Definition (Wegzusammenhängend). Sei (X,d) ein metrischer Raum. Dann heißt (X,d) *wegzusammenhängend*, wenn zu beliebigen $x, y \in X$ eine stetige Abbildung $\gamma : [0,1] \longrightarrow X$ existiert mit $\gamma(0) = x$, $\gamma(1) = y$. Ein solches γ heißt dann Weg von x nach y.

Beispiel. Eine $C \subseteq \mathbb{R}^N$ heißt konvex, wenn mit $x, y \in C$ auch die Verbindungsstrecke $V = \{tx + (1+t)y \mid 0 \leq t \leq 1\}$ zwischen x und y zu C gehört. Jede konvexe Menge in $\mathbb{R}^N$ ist offenbar wegzusammenhängend (Klar!?). Insbesondere ist jede Kugel in $\mathbb{R}^N$ wegzusammenhängend.

Ende 10. Vorlesung

Folgerung. *Ist (X,d) wegzusammenhängend, so ist (X,d) zusammenhängend.*

Beweis. Sei $X = U \cup V$ mit U, V offen und disjunkt. Angenommen $U \neq \emptyset$ und $V \neq \emptyset$. Dann gibt es also $x \in U$ und $y \in V$. Da (X,d) wegzusammenhängend ist, gibt es einen Weg γ von x nach y. Dann ist $I := [0,1] = \gamma^{-1}(U) \cup \gamma^{-1}(V)$ eine Zerlegung in offene (in I) disjunkte nichtleere Teilmengen. Das ist ein Widerspruch. $\square$

Bemerkung. Im allgemeinen gilt die Umkehrung der vorigen Folgerung nicht. Es ist

$$Y = \{(y, \sin y) \mid y > 0\} \cup \{(0,0)\} \subseteq \mathbb{R}^2$$

versehen mit der euklidischen Metrik zusammenhängend aber nicht wegzusammenhängend (Übung).

In speziellen Fällen aber gilt eine Umkehrung. Das wird im folgenden Theorem behandelt.

Theorem. *Sei $U \subseteq \mathbb{K}^N$ offen und zusammenhängend. Dann ist U auch wegzusammenhängend.*

Beweis. Sei $p \in U$ beliebig. Sei C_p die Menge der $x \in U$ für die eine stetige Kurve $\gamma : [0,1] \longrightarrow U$ existiert mit $\gamma(0) = p$ und $\gamma(1) = x$. Dann gilt: Sei $R_p := U \setminus C_p$.

- C_p ist offen, da U offen ist und jede Kugel wegzusammenhängend. **Zeichnung**
- R_p ist offen, da U offen ist und jede Kugel wegzusammenhängend. **Zeichnung**

Da U zusammenhängend ist und $C_p \neq \emptyset$ gilt, folgt nun $R_p = \emptyset$. $\square$

Proposition. *Ist (X,d) ein wegzusammenhängender metrischer Raum, (Y,e) ein metrischer Raum und $f : X \longrightarrow Y$ stetig, so ist $(f(X),e)$ wegzusammenhängend.*

Beweis. Das ist klar. **Zeichnung.** (Ein Weg zwischen $f(x)$ und $f(y)$ wird gerade durch $f \circ \gamma$ gegeben, wobei γ ein Weg von x nach y ist). $\square$

6. Anwendungen - Der Banachsche Fixpunktsatz (Bonusmaterial)

In diesem Abschnitt lernen wir den Banachschen Fixpunktsatz kennen, der an vielen Stellen nützlich sein kann. Wir skizzieren kurz eine Anwendung (Erzeugung von Fraktalen), bei der kompakte Mengen und Metriken eine große Rolle spielen.

THEOREM (Banachsche Fixpunktsatz). *Sei (X, d) ein vollständiger metrischer Raum und $f : X \longrightarrow X$ gegeben. Gibt es eine ein $0 < c < 1$ mit*

$$d(f(x), f(y)) \leq c d(x, y)$$

für alle $x, y \in X$, so existiert ein eindeutiges $p \in X$ mit $f(p) = p$. Ist $x \in X$ beliebig und (x_n) induktiv definiert durch

$$x_0 := x, \text{ und } x_{n+1} := f(x_n)$$

(d.h. $x_n = f \circ \ldots \circ f(x)$), so konvergergiert (x_n) gegen p, und es gilt die apriori Abschätzung

$$d(p, x_n) \leq c^n \frac{d(x_0, x_1)}{1 - c}.$$

Bemerkung Eine Abbildung f wie im Theorem heißt Kontraktion mit Kontraktionsfaktor c. Ein p wie im Theorem heißt Fixpunkt von f.
Beweis. Wir zeigen Eindeutigkeit, Existenz und die Apriori-Abschätzung.
Eindeutigkeit. Seien $x, y \in X$ mit $f(x) = x$ und $f(y) = y$ geben. Dann gilt

$$d(x, y) = d(f(x), f(y)) \leq c d(x, y).$$

Wegen $c < 1$ und $d \geq 0$ folgt $d(x, y) = 0$.
Existenz. Sei $x \in X$ beliebig und x_n wie in der Aussage des Theorem. Dann gilt (Induktion)

$$d(x_j, x_{j+1}) \leq c^j d(x_0, x_1)$$

für alle $j \in \mathbb{N}$ und damit für $n < m$

$$\begin{aligned} d(x_n, x_m) &\leq \sum_{j=n}^{m-1} d(x_j, x_{j+1}) \\ &\leq \sum_{j=n}^{m-1} c^j d(x_0, x_1) \\ &\leq c^n d(x_0, x_1) \sum_{j=0}^{m-n-1} c^j \\ &\leq c^n d(x_0, x_1) \frac{1}{1 - c}. \end{aligned}$$

Damit ist also (x_n) eine Cauchy-Folge. Aufgrund der Vollständigkeit konvergiert dann (x_n) und für den Grenzwert p gilt

$$d(p, f(p)) = \lim_{n\to\infty} d(x_n, f(x_n)) = \lim_{n\to\infty} d(x_n, x_{n+1}) = 0.$$

Damit ist p ein Fixpunkt.
Abschätzung. Schließlich gilt aufgrund des schon gezeigten

$$d(p, x_n) = \lim_{m\to\infty} d(x_m, x_n) \leq c^n d(x_0, x_1)\frac{1}{1-c}.$$

Das beendet den Beweis. □

Anwendung - Hausdorffmetrik. Sei (X, d) ein metrischer Raum. Wir werden eine Abstandsfunktion auf den (kompakten) Teilmengen von (X, d) einführen und damit die Menge der kompakten Teilmengen zu einem metrischen Raum machen.
Für eine Teilmenge A von X und $x \in X$ definiert man

$$d(x, A) := \inf\{d(x, a) \mid a \in A\}.$$

Aufgrund der Dreiecksungleichung gilt dann (einfach!)

$$d(x, A) - d(y, A) \leq d(x, y)$$

für alle $x, y \in X$. Damit ist also für gegebenes A die Abstandsfunktion

$$d_A := d(\cdot, A) : X \longrightarrow \mathbb{R}$$

stetig auf X. Ist A abgeschlossen, so gilt weiterhin

$$d(x, A) = 0 \Longleftrightarrow x \in A.$$

($\Longrightarrow$: $x \notin A$ impliziert $U_r(x) \in X \setminus A$ da A abgeschlosse ist. Damit folgt $d(x, A) \geq r$. $\Longleftarrow$: klar).
Ist A kompakt, so gibt es ein $a \in A$ mit $d(x, A) = d(x, a)$ und man kann das Infimum durch ein Minimum ersetzen (da die stetig Funktion $d(x, \cdot) : K \longrightarrow \mathbb{R}$ ihr Minimum annimmt). Sind K, L kompakte Teilmengen von X, so können wir nun d_K und d_L auf L bzw K untersuchen. Diese Funktionen sind stetig (s.o.) und nehmen also auf L bzw. K ihr Maxmimum.
- Es gilt $\max_K d_L := \max\{d(x, L) \mid x \in K\} \leq \varepsilon$ genau dann, wenn zu jedem $x \in K$ ein $y \in L$ existiert mit $d(x, y) \leq \varepsilon$. **Zeichnung.**
- Es gilt $\max_L d_K := \max\{d(K, y) \mid y \in L\} \leq \varepsilon$ genau dann, wenn zu jedem $y \in L$ ein $x \in K$ existiert mit $d(x, y) \leq \varepsilon$. **Zeichnung.**

Für kompakte Teilmengen K, L von X definiert man dann den Haussdorfabstand

$$d_H(K, L) := \max\{\max_K d_L, \max_L d_K\}.$$

Damit gilt $d_H(K, L) \leq \varepsilon$ genau dann, wenn zu jedem $x \in K$ ein $y_x \in L$ existiert mit $d(x, y_x) \leq \varepsilon$ und umgekehrt zu jedem $y \in L$ ein $x_y \in K$ existiert mit $d(x_y, y) \leq \varepsilon$.

THEOREM. *Ist (X, d) ein vollständiger metrischer Raum, so ist die Menge $\mathcal{K}(X)$ der kompakten Teilmengen von X mit der Hausdorffmetrik d_H ein vollständiger metrischer Raum.*

Beweis. Beweis nicht hier. (Evtl. in Übung). □

FOLGERUNG. *(Erzeugung von Fraktalen) Ist (X, d) ein vollständiger metrischer Raum und $F : \mathcal{K}(X) \longrightarrow \mathcal{K}$ eine Kontraktion. Dann gibt genau eine kompakte Menge K in X mit $F(K) = K$.*

Beweis. Das folgt sofort aus dem Banachschen Fixpunktsatz und dem vorangegangenen Theorem. □

Eine kleine Rechnung zeigt folgende interessante Eigenschaft der Haussdorffmetrik.

PROPOSITION.

$$d_H(A \cup B, C \cup D) \leq \max\{d_H(A, C), d_H(B, D)\}.$$

Anwendung. Sei $X = \mathbb{R}$ mit der Euklidischen Metrik und

$$F : \mathcal{K} \longrightarrow \mathcal{K}, \ F(K) = \frac{1}{3}K + (2/3 + \frac{1}{3}K).$$

Dann ist gilt (kleine Rechnung)

$$d_H(F(L), F(K)) \leq \frac{1}{3}d(K, L).$$

Damit hat also F nach dem Banachschen Fixpunktsatz einen Fixpunkt C und für C gilt

$$C = \lim_{n\to\infty} F \circ \ldots \circ F(K)$$

für jedes kompakte K in $\mathbb{R}$.
Wir 'bestimmen' den Fixpunkt auf zwei Arten, nämlich durch Wahl von $K_0 = [0, 1]$ und $K_0 = \{0\}$. **Zeichnung.**
Der Fixpunkt ist die uns aus dem ersten Semester schon vertraute Cantormenge.

KAPITEL 2

Differenzierbarkeit im Höherdimensionalen

In diesem Kapitel geht es um Funktionen von Teilmengen des $\mathbb{R}^N$ nach $\mathbb{R}^M$. Dabei werden $\mathbb{R}^N$ und $\mathbb{R}^M$ als metrische Räume mit der Euklidischen Metrik aufgefasst. Ein wesentlicher Inhalt dieses Kapitels ist die Untersuchung von Differenzierbarkeit solcher Funktionen. Anschaulich bedeutet die Differenzierbarkeit von f, dass sich f gut durch eine lineare Abbbildung, seine Ableitung, approximieren lässt.

1. Zum Aufwärmen: Funktionen von $\mathbb{R}^N$ nach $\mathbb{R}^M$

Es geht um Funktionen von Teilmengen U des $\mathbb{R}^N$ nach $\mathbb{R}^M$ (mit geeigneten N und M). Den Fall $N = M = 1$ haben wir im vergangenen Semester schon untersucht.
Eine Funktion $f : U \longrightarrow \mathbb{R}^M$ lässt sich offenbar schreiben als $f(x) = (f_1(x), \ldots, f_M(x))$ mit $f_j : U \longrightarrow \mathbb{R}$. Es ist f stetig, genau dann, wenn jedes f_j stetig ist (siehe Übung). Einige spezielle Situationen haben eigene Namen:

- Ist $U \subseteq \mathbb{R}$ ein Intervall, so heißt $\gamma : U \longrightarrow \mathbb{R}^M$ eine *Kurve.* **Zeichnung.**
  Beispiel: Es ist $\mathbb{R} \longrightarrow \mathbb{R}^M, t \mapsto p + tv$ eine stetige Kurve (für gegebenes $p \in \mathbb{R}^M$ und $v \in \mathbb{R}^M$).
- Ist $U \subseteq \mathbb{R}^N$, so heißt $f : U \longrightarrow \mathbb{R}$ eine *skalare Funktion.* **Zeichnung.**
  Beispiel: Es sind die Koordinatenfunktionen $\pi_j : \mathbb{R}^N \longrightarrow \mathbb{R}, (\xi_1, \ldots, \xi_N) \mapsto \xi_j$ stetige skalare Funktionen.
- Ist $U \subseteq \mathbb{R}^N$, so heißt $f : U \longrightarrow \mathbb{R}^N$ ein *Vektorfeld* auf U. **Zeichnung.**
  Beispiel: Es liefert die Identität $id : U \longrightarrow \mathbb{R}^N, x \mapsto x$, ein stetiges Vektorfeld.

Natürlich sind Funktionen von einer Variable besonders leicht zu handhaben. Die folgende Prozedur erlaubt es, aus Funktionen mehrerer Variablen, mehrere Funktionen einer Variablen zu machen. Damit lassen sich viele (aber nicht alle) Betrachtungen im Mehrdimensionalen auf Betrachtungen im eindimensionalen zurückführen: Hält man für $f : U \subseteq \mathbb{R}^N \longrightarrow \mathbb{R}^M$ alle Variablen bis auf eine - z.B. die j-te fest, so erhält man die *partiellen Funktionen*

$$t \mapsto f(\xi_1, \ldots, \xi_{j-1}, \xi_j + t, \xi_{j+1}, \ldots, \xi_N).$$

Genauer definiert man mit der Standard Normalbasis (e_j) im $\mathbb{R}^N$ zu $p \in U$ und jedem $j \in \{1, \dots, N\}$ die sogenannte partielle Funktion

$$f_{p,j} : \{t \in \mathbb{R} : p + te_j \in U\} \longrightarrow \mathbb{R}^M, \ f_{p,j}(t) := f(p + te_j).$$

Dann heißt f *partiell stetig in* p, wenn $f_{p,j}$ stetig in 0 ist für alle $j = 1, \dots, N$. Gilt das für alle $p \in U$, so heißt f *partiell stetig*.

Ende 11. Vorlesung

Allgemeiner kann man zu jedem Einheitsvektor $e \in \mathbb{R}^N$ und jedem $p \in U$ die Funktion

$$f_{p,e} : \{t \in \mathbb{R} \mid p + te \in U\} \longrightarrow \mathbb{R}^M, f_{p,e}(t) := f(p + te),$$

definieren. Offenbar gilt dann also $f_{p,e_j} = f_j$, $j = 1, \dots, N$. Es heißt f *richtungsstetig in* p, wenn für alle Einheitsvektoren $e \in \mathbb{R}^N$, die Funktion $f_{p,e}$ stetig in 0 ist. Gilt dies für alle $p \in U$, so heißt f *richtungsstetig*. Offenbar gilt folgende Proposition. (Übung)

PROPOSITION. *f stetig (in p) $\Longrightarrow$ f richtungsstetig (in p) $\Longrightarrow$ f partiell stetig (in p).*

Beweis. Die Funktionen $f_{p,e}$ können als Verknüpfung von f mit geeigneten Kurven γ dargestellt werden. □

BEMERKUNG. Die Umkehrung ist aber nicht wahr. (Übung)

Besonders einfache zu verstehende Funktion sind lineare Abbildungen. Sie werden durch Matrizen beschrieben an welchen man ihre wesentlichen Eigenschaften 'ablesen' kann. Detailliert werden sie in der Vorlesung Lineare Algebra 1 besprochen. Wir geben hier nur einen kleinen Überblick.

DEFINITION (Lineare Abbildung). Eine Abildung $A : \mathbb{R}^N \to \mathbb{R}^M$ heißt linear, falls für alle $\lambda, \mu \in \mathbb{R}$ und alle $x, y \in \mathbb{R}^N$ gilt

$$A(\lambda x + \mu y) = \lambda Ax + \mu Ay.$$

Wir haben hier die übliche Schreibweise Ax statt $A(x)$ verwendet.

PROPOSITION. *Sei $A : \mathbb{R}^N \to \mathbb{R}^M$ linear. Dann ist A stetig.*

Beweis. Siehe Übung oder nutze, dass lineare Abbildungen differenzierbar sind (später). □

Sei (e_j) die Standardbasis von $\mathbb{R}^N$, das heißt

$$e_j = (0, \dots, 0, \underset{\substack{\uparrow \\ j\text{-te Stelle}}}{1}, 0, \dots, 0).$$

Jeder Vektor $x = (x_1, \dots, x_N) \in \mathbb{R}^N$ hat also die Darstellung

$$x = (x_1, \dots, x_N) = \sum_{j=1}^N x_j e_j.$$

Für $1 \leq j \leq N$ definieren wir $A_j := Ae_j = ((Ae_j)_1, \ldots, (Ae_j)_M) \in \mathbb{R}^M$ und für $1 \leq i \leq M$ sei $A_{ij} := (Ae_j)_i \in \mathbb{R}$.

Proposition. *Sei $A : \mathbb{R}^N \to \mathbb{R}^M$ linear. Für $x = (x_1, \ldots, x_N)$ gilt*

$$Ax = \sum_{j=1}^{N} x_j A_j = \sum_{i=1}^{M} \left(\sum_{j=1}^{N} A_{ij} x_j \right) e_i.$$

Insbesondere ist die i-te Koordinate von Ax gegeben durch

$$(Ax)_i = \sum_{j=1}^{N} A_{ij} x_j.$$

Beweis. Das folgt unmittelbar aus den Definitionen und der Rechnung

$$\begin{aligned} Ax &= A \left(\sum_{j=1}^{N} x_j e_j \right) \\ (A \text{ linear}) &= \sum_{j=1}^{N} x_j A e_j \\ &= \sum_{j=1}^{N} x_j \left(\sum_{i=1}^{M} (Ae_j)_i \, e_i \right) \\ &= \sum_{i=1}^{M} \left(\sum_{j=1}^{N} (Ae_j)_i \, x_j \right) e_i. \end{aligned}$$

□

Bemerkung.

- Die vorherige Proposition zeigt, dass die Abbildung A eindeutig durch die Zahlen A_{ij} bestimmt ist. Statt mit der Standardbasis (e_j) hätte man auch mit anderen Basen arbeiten können.
- In unserer Notation entspricht Ax gerade dem Transponierten vom Ergebnis der Matrixmultiplikation

$$\begin{pmatrix} A_{11} & \cdots & A_{1N} \\ \vdots & \ddots & \vdots \\ A_{M1} & \cdots & A_{MN} \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_N \end{pmatrix}.$$

  In diesem Sinne sprechen wir von der Matrixdarstellung der Abbildung A bezüglich der Standardbasis.

Proposition (Injektivität linearer Abbildungen). *Sei $A : \mathbb{R}^N \to \mathbb{R}^M$ linear. Die folgenden Aussagen sind äquivalent.*

(i) *A ist injektiv.*
(ii) *$Ax = 0$ impliziert $x = 0$.*
(iii) *$A_1, \ldots, A_N$ sind linear unabhängig.*

Beweis. (i) $\Rightarrow$ (ii): Da A linear ist gilt $A0 = 0$. Damit folgt Aussage (ii) aus der Injektivität von A.
(ii) $\Rightarrow$ (i): Es gelte $Ax = Ay$. Aufgrund der Linearität von A folgt $A(x-y) = 0$. Aussage (ii) impliziert $x - y = 0$.
(ii) $\Leftrightarrow$ (iii): Es gilt

$$Ax = \sum_{j=1}^{N} x_j A_j.$$

Damit ist (ii) offensichtlich äquivalent zu (iii). $\square$

BEMERKUNG. Ist $A : \mathbb{R}^N \to \mathbb{R}^M$ linear und injektiv, so sind $A_1, \ldots, A_N$ N linear unabhängige Vektoren im $\mathbb{R}^M$. Damit muss aus Dimensionsgründen $N \leq M$ gelten.

PROPOSITION (Surjektivität linearer Abbildungen). *Eine lineare Abbildung $A : \mathbb{R}^N \to \mathbb{R}^M$ ist genau dann surjektiv, wenn*

$$\operatorname{span}\{A_1, \ldots, A_N\} = \mathbb{R}^M.$$

Beweis. Das folgt unmittelbar aus der Darstellung

$$Ax = \sum_{j=1}^{N} x_j A_j.$$

$\square$

BEMERKUNG. Ist $A : \mathbb{R}^N \to \mathbb{R}^M$ eine lineare und surjektive Abbildung, so gilt $\operatorname{span}\{A_1, \ldots, A_N\} = \mathbb{R}^M$. Damit muss aus Dimensionsgründen $M \geq N$ gelten.

PROPOSITION. *Seien $A : \mathbb{R}^N \to \mathbb{R}^M$ und $B : \mathbb{R}^M \to \mathbb{R}^L$ lineare Abbildungen. Dann ist $BA := B \circ A$ linear und für $1 \leq i \leq L$ und $1 \leq j \leq N$ gilt*

$$(BA)_{ij} = \sum_{k=1}^{M} B_{ik} A_{kj}.$$

Beweis. Die Linearität von BA ist offensichtlich. Wir berechnen

$$(BA)_{ij} = (BAe_j)_i = \sum_{k=1}^{M} B_{ik}(Ae_j)_k = \sum_{k=1}^{M} B_{ik} A_{kj}.$$

$\square$

BEMERKUNG. Die vorherige Proposition besagt, dass die BA darstellende Matrix in der Standardbasis durch das Ergebniss der Matrixmultiplikation

$$\begin{pmatrix} B_{11} & \cdots & B_{1M} \\ \vdots & \ddots & \vdots \\ B_{L1} & \cdots & B_{LM} \end{pmatrix} \cdot \begin{pmatrix} A_{11} & \cdots & A_{1N} \\ \vdots & \ddots & \vdots \\ A_{M1} & \cdots & A_{MN} \end{pmatrix}$$

gegeben ist.

Schließlich erinnern wir zum Abschluss noch an das Konzept des Grenzwertes (vgl. Analysis I, bzw. Betrachtungen zu metrischen Räumen): Sei $U \subseteq \mathbb{R}^N$ offen und $f : U \longrightarrow \mathbb{R}^M$ gegeben. Sei $p \in \overline{U}$ (d.h. p ist Grenzwert einer Folge in U). Dann hat f bei p den Grenzwert $c \in \mathbb{R}^M$ geschrieben als

$$\lim_{x \to p} f(x) = c$$

genau dann, wenn gilt

- $f(x_n) \to c$ für jede Folge (x_n) in U mit $x_n \to p$.

bzw. äquivalent, wenn gilt

- zu jedem $\varepsilon > 0$ ein $\delta > 0$ existiert mit $|f(x) - c| \leq \varepsilon$ für alle $x \in U$ mit $|x - p| \leq \delta$.

Entsprechend schreibt man $\lim_{x \to p, x \neq p} f(x) = c$, falls für die Einschränkung $f|_{U \setminus \{p\}}$ von f auf $U \setminus \{p\}$ gilt

$$\lim_{x \to p} f|_{U \setminus \{p\}}(x) = c.$$

Ende 12. Vorlesung

2. Definition der Ableitung und einfache Eigenschaften

Es geht um lineare Approximationen. Die Ableitung ist die lineare Approximation.
Aus dem ersten Semester wissen wir, dass für $f : (a, b) \longrightarrow \mathbb{R}$ und $p \in (a, b)$ äquivalent sind:

- $f'(p) := \lim_{x \to p}(f(x) - f(p))/(x - p)$ existiert.
- Es gilt $f(x) = f(p) + c(x - p) + \varphi(x)$ mit $\varphi(x)/|x - p| \to 0$, $x \to p$.

In diesem Fall heißt f differenzierbar in p. Wir werden das nun auf höherdimensionale Situationen verallgemeinern. Dabei gehen wir von der im zweiten Punkt präsentierten Variante aus (da die Division durch $x - p$ im höherdimensionalen keinen Sinn hat).

Definition. Sei $U \subseteq \mathbb{R}^N$ offen und $f : U \to \mathbb{R}^M$ gegeben. Sei $p \in U$. Dann heißt f in p *differenzierbar*, wenn eine lineare Abbildung $L : \mathbb{R}^N \to \mathbb{R}^M$ existiert mit

$$f(x) = f(p) + L(x - p) + \varphi_p(x),$$

sodass für den Fehler

$$\varphi_p(x) = f(x) - f(p) - L(x - p)$$

gilt

$$\frac{\varphi_p(x)}{|x - p|} \to 0, x \to p.$$

In diesem Fall heißt L Ableitung von f in p. Ist f in jedem $p \in U$ differenzierbar, so heißt f auf ganz U differenzierbar.

Bemerkung.

- Ist f differenzierbar in p, so wird es also “sehr gut„ durch die linear (affine) Funktion
$$f(p) + L(\cdot - p)$$
approximiert.
- Die Offenheit von U kommt ins Spiel, da wir f von allen “Seiten„ approximieren können wollen. Das werden wir später im Konzept der Richtungsableitung noch genauer machen (s.u.).
- Im Falle einer Funktion von $U \subseteq \mathbb{R} \longrightarrow \mathbb{R}$ ist die lineare Abbildung L gerade durch eine Zahl gegeben (lineare Abbildung von $\mathbb{R}$ nach $\mathbb{R}$). Diese Zahl ist die uns schon aus dem ersten Semester bekannte Ableitung.

Proposition. *In der Situation der Definition ist L eindeutig.*

Beweis. Seien L_1 und L_2 mit
$$f(x) = f(p) + L_1 \cdot (x - p) + \varphi_1(x)$$
$$f(x) = f(p) + L_2 \cdot (x - p) + \varphi_2(x)$$
und $\varphi_j(x)/|x - p| \to 0$, $x \to p$, $j = 1, 2$. Dann gilt
$$(L_1 - L_2)(x - p) = \varphi_1(x) - \varphi_2(x) = \frac{\varphi_1(x) - \varphi_2(x)}{|x - p|}|x - p|$$
Betrachtet man $x = p + he$ mit $h \in \mathbb{R}$, $h \neq 0$ und einem beliebigen Einheitsvektor $e \in \mathbb{R}^N$, so gilt also $x - p = he$ und
$$(L_1 - L_2)(he) = \frac{\varphi_1(x) - \varphi_2(x)}{|x - p|}|he|$$
und damit konvergiert die rechte Seite von
$$(L_1 - L_2)e = \frac{\varphi_1(x) - \varphi_2(x)}{|x - p|}$$
für $h \to 0$ gegen 0. Es folgt also
$$(L_1 - L_2)e = 0.$$
Da e ein beliebiger Einheitsvektor ist, folgt $L_1 = L_2$. □

Notation. Die vorherigen Propostion erlaubt es uns, von DER *Ableitung* von f in p zu sprechen. Wir bezeichnen diese dann mit $Df(p)$.

Beispiel (Lineare Funktion). Sei $A : \mathbb{R}^N \longrightarrow \mathbb{R}^M$ linear, $a \in R^M$ und $q \in \mathbb{R}^N$ gegeben. Dann ist die Funktion
$$f : \mathbb{R}^N \longrightarrow \mathbb{R}^M, f(x) = a + A(x - q)$$
differenzierbar und es gilt $Df(x) = A$ für alle $x \in \mathbb{R}^N$.
Bew. Es gilt $f(x) - f(p) = A(x - p)$, also
$$f(x) = f(p) + A(x - p) + 0.$$
Damit folgt die Differenzierbarkeit (mit $L = A$ und $\varphi = 0$).

Theorem (Differenzierbarkeit impliziert Stetigkeit). *Sei $U \subseteq \mathbb{R}^N$ offen und $f : U \to \mathbb{R}^M$ gegeben. Ist f in p differenzierbar, so ist es dort auch stetig.*

Beweis. Es gilt (für $x \neq p$)

$$f(x) = f(p) + L(x-p) + \varphi(x) = f(p) + L(x-p) + \frac{\varphi(x)}{|x-p|}|x-p|.$$

Damit folgt sofort, dass die rechte Seite gegen $f(p)$ konvergiert für $x \to p$. □

Wir haben nun die Ableitung L *charakterisiert*. Als Nächstes soll es darum gehen, wie man diese lineare Funktion *berechnet*.

Theorem (Berechnen der Ableitung). *Sei $U \subseteq \mathbb{R}^N$ offen und $f : U \to \mathbb{R}^M$ differenzierbar in $p \in U$. Dann existieren die partiellen Ableitungen*

$$D_j f_i(p) := \partial_j f_i(p) := \frac{\partial f_i}{\partial x_j}(p) := \lim_{h\to 0, h\neq 0} \frac{1}{h}(f_i(p+he_j) - f_i(p))$$

für $i = 1, \ldots, M$, $j = 1, \ldots, N$ und die Ableitung L wird bezüglich der Standardbasen in $\mathbb{R}^N$ bzw. $\mathbb{R}^M$ durch die Matrix

$$Df(p) = (\partial_j f_i(p))_{i,j}\,.$$

gegeben.

Bemerkung.
- Die partielle Ableitung von f_i nach x_j in p ist also gerade die "gewöhnliche„ Ableitung der i-ten Komponente der oben betrachteten partiellen Funktion $f_{p,j}$

  $$t \mapsto f_i(p + te_j)$$

  an der Stelle 0.
- Die partiellen Ableitungen werden gebildet, indem man alle Variablen bis auf eine festhält und dann nach dieser einen so ableitet, wie wir es im ersten Semester gelernt haben.
- Der Satz beschreibt, wie man die Ableitung von f ausrechnet, wenn man schon weiss, dass f differenzierbar ist. Der Satz besagt NICHT, dass f differenzierbar ist, wenn die partiellen Ableitungen existieren (und das ist im allgemeinen auch falsch (siehe Übung)).

Beweis. Nach Voraussetzung gilt

$$f(x) = f(p) + L(x-p) + \varphi(x)$$

mit $\frac{\varphi(x)}{|x-p|} \to 0$ und $L = (L_{i,j})_{i,j}$. Wir setzen $x = p + he_j$ und betrachten die i-te Komponente. Für diese gilt

$$f_i(p + he_j) = f_i(p) + (L(he_j))_i + \varphi_i(x).$$

Das impliziert

$$f_i(p + he_j) - f_i(p) = h(Le_j)_i + \varphi_i(x).$$

Damit folgt nach Division durch h also

$$\frac{f_i(p+he_j)-f_i(p)}{h}=L_{i,j}+\frac{\varphi_i(p+he_j)}{h}$$

Wegen $\frac{|\varphi_i(p+he_j)|}{|h|}=\frac{|\varphi_i(p+he_j)|}{|x-p|}\to 0,\ h\to 0$, folgt

$$\frac{f_i(p+he_j)-f_i(p)}{h}\to L_{i,j},\ h\to 0.$$

Das beendet den Beweis. $\square$

Ende 13. Vorlesung

DEFINITION (Partielle Differenzierbarkeit). Sei $U\subseteq\mathbb{R}^N$ offen. Eine Funktion $f:U\longrightarrow\mathbb{R}^M$ heißt in $p\in U$ *partiell differenzierbar*, wenn die im vorherigen Theorem definierten partiellen Ableitungen in p existieren. In diesem Fall heißt $(\partial_j f_i(p))_{i,j}$ die *Jacobi-Matrix* oder auch *Differentialmatrix* von f in p.

NOTATION (Gradient). Existieren für $f:U\subseteq\mathbb{R}^N\longrightarrow\mathbb{R}$ die partiellen Ableitungen in $p\in U$, so nennt man den Vektor $(\partial_1 f(p),\ldots,\partial_N f(p))^\top$ den *Gradienten* von f in p und schreibt ihn als $\nabla f(p)$. Ist f differenzierbar, so ist also $Df(p)=\nabla f(p)^\top$.

BEISPIEL. (a) $f:U\subseteq\mathbb{R}^N\to\mathbb{R}$ differenzierbar in p. Dann ist:

$$\begin{aligned}f(x)&=f(p)+Df(p)(x-p)+\varphi(x)\\&=f(p)+\langle\nabla f(p),x-p\rangle+\varphi(x)\end{aligned}$$

mit $\varphi(x)/|x-p|\to 0,\ x\to p$.
(b) Sei $\gamma:(a,b)\to\mathbb{R}^N$ gegeben. Dann ist γ differenzierbar in $t_0\in(a,b)$ genau dann, wenn

$$\gamma'(t_0):=\lim_{t\to t_0}\frac{1}{t-t_0}(\gamma(t)-\gamma(t_0))$$

existiert (Bew. Übung). Dann gilt

$$\gamma(t)=\gamma(t_0)+\gamma'(t_0)(t-t_0)+\varphi(t)$$

mit $\varphi(t)/(t-t_0)\to 0,\ t\to t_0$. Insbesondere haben wir für $\gamma(t)=t_0+vt$,

$$\gamma'(t)=v.$$

Tatsächlich kann man in diesem Fall auch $\gamma:[a,b]\longrightarrow U$ betrachten und die Ableitung von γ in a bzw. b durch den obigen Grenzwert definieren (wenn der Grenzwert existiert) (vgl. Übung).
(c) Ist $f:\mathbb{R}^N\setminus\{0\}\to\mathbb{R}$, $f(x)=|x|$, so gilt (Übung)

$$\partial_j f(x)=\tfrac{x_j}{|x|},\ \text{also}\ \nabla f(x)=\tfrac{x}{|x|}$$

Tatsächlich (s.u.) ist f in diesem Fall differenzierbar.

Wir lernen als nächstes einige Rechenregeln für die Ableitung kennen.

Theorem (Kettenregel). *Sei $U \subseteq \mathbb{R}^N$, $V \subseteq \mathbb{R}^M$ offen und $f : U \to V$, $g : V \to \mathbb{R}^L$ gegeben. Sind f in p und g in $f(p)$ differenzierbar, so ist $g \circ f : U \to \mathbb{R}^L$ differenzierbar in p, und es gilt*

$$D(g \circ f)(p) = Dg(f(p))Df(p).$$

Insbesondere gilt

$$D(g \circ f)_{ij}(p) = \partial_j (g \circ f)_i = \sum_{k=1}^{M} \partial_k g_i(f(p))\partial_j f_k(p).$$

Bemerkung. Die Ableitung der Komposition ist also die Komposition der Ableitungen. Diese Komposition wird auf der Ebene von Matrizen durch das Produkt der Matrizen gegeben. Die Reihenfolge der Matrizen ist dabei natürlich wichtig.

Beweis. Es reicht die erste Aussage zu zeigen. Das "Insbesondere„ folgt dann durch Einsetzen von e_j. Mit $q = f(p)$ gilt nach Voraussetzung:

$$f(x) = f(p) + Df(p)(x - p) + \varphi_f(x)$$

mit $\frac{\varphi_f(x)}{|x-p|} \to 0$, $x \to p$ und

$$g(y) = g(q) + Dg(q)(y - q) + \varphi_g(x)$$

mit $\frac{\varphi_g(y)}{|y-q|} \to 0$, $y \to q$

also gilt für $h(x) = g(f(x))$ mit $y = f(x)$ und weiterhin $f(p) = q$ dann

$$\begin{aligned} h(x) &= \underbrace{g(f(p))} + Dg(f(p))(f(x) - f(p)) + \varphi_g(f(x)) \\ &= h(p) + Dg(f(p))\Big(Df(p)(x-p) + \varphi_f(x)\Big) + \varphi_g(f(x)) \\ &= h(p) + Dg(f(p))Df(p)(x-p) + Dg(f(p))\varphi_f(x) + \varphi_g(f(x)) \\ &= h(p) + Dg(f(p))Df(p)(x-p) + \varphi(x) \end{aligned}$$

mit

$$\varphi(x) = Dg(f(p))\varphi_f(x) + \varphi_g(f(x)) =: T(x) + S(x).$$

Es bleibt zu zeigen, dass gilt

$$\frac{\varphi(x)}{|x-p|} \to 0, x \to p.$$

Das machen wir nun: Es gilt

$$\frac{T(x)}{|x-p|} = \frac{1}{|x-p|} Dg(f(p))\varphi_f(x) = \underbrace{Dg(f(p))}_{\text{beschränkt}} \underbrace{\frac{\varphi_f(x)}{|x-p|}}_{\to 0} \to 0.$$

Für den zweiten Term gilt

$$\frac{S(x)}{|x-p|} = \frac{1}{|x-p|}\varphi_g(f(x)) = \frac{\varphi_g(f(x))}{|f(x)-f(p)|}\frac{|f(x)-f(p)|}{|x-p|}$$

falls $f(x) \neq f(p)$. (Falls $f(x) = f(p)$ gilt $\varphi_g(f(x)) = \varphi_g(f(p)) = 0$ und es ist nichts zu zeigen.) Nun gilt aber

$$\frac{\varphi_g(f(x))}{|f(x) - f(p)|} \to 0, x \to p \ \text{ (da } f(x) \to f(p) \text{ für } x \to p)$$

und es bleibt

$$\frac{|f(x) - f(p)|}{|x - p|} = \frac{|Df(p)(x-p) + \varphi_f(x)|}{|x-p|} \leq \frac{|Df(p)(x-p)|}{|x-p|} + \frac{|\varphi_f(x)|}{|x-p|}$$

beschränkt für $x \to p$. Damit folgt

$$\frac{|T(x) + S(x)|}{|x-p|} \to 0, x \to p.$$

Das beendet den Beweis. □

Beispiel (Verknüpfung von Funktion mit Kurve). Sei $U \subseteq \mathbb{R}^N$ offen und $\gamma : (a,b) \longrightarrow U$ differenzierbar in t_0 und $f : U \subseteq \mathbb{R}^N \longrightarrow \mathbb{R}$ differenzierbar in $\gamma(t_0)$. Dann ist $f \circ \gamma$ differenzierbar und es gilt

$$(f \circ \gamma)'(t) = \sum_{k=1}^{N} \partial_k f(\gamma(t_0))\gamma_k'(t_0) = \langle \nabla f(\gamma(t_0), \gamma'(t_0)\rangle.$$

Für $\gamma : [a,b] \longrightarrow U$ gilt eine entsprechende Formel auch in a bzw. b, wenn γ in a bzw. b differenzierbar ist (s.o.).

Beispiel. $\gamma : (a,b) \longrightarrow \mathbb{R}^{N+1}$, $\gamma(t) = (\gamma_1(t), \ldots, \gamma_N(t), t)$.
$F : \mathbb{R}^{N+1} \longrightarrow \mathbb{R}$, $(x,t) \mapsto F(x,t)$, differenzierbar. Dann gilt

$$(F \circ \gamma)'(t) = \sum_{k=1}^{N} \partial_k F(\gamma(t_0))\gamma_k'(t_0) + \partial_{N+1} F(\gamma(t_0)).$$

(Das wird manchmal unter dem Begriff der totalen Ableitung von F gefasst.)

Beispiel ($g(|x|)$). $g : \mathbb{R} \longrightarrow \mathbb{R}$ differenzierbar, $f : \mathbb{R}^N \longrightarrow \mathbb{R}$, $f(x) = g(|x|)$. Dann ist f in $x \neq 0$ differenzierbar, und es gilt $Df(p) = g'(|p|)\frac{1}{|p|}p$. (Hier haben wir die weiter unten bewiesene Differenzierbarkeit der Betragsfunktion in $p \neq 0$ benutzt.)

Theorem (Produktregel). *Sei $U \subseteq \mathbb{R}^N$ offen und $f, g : U \longrightarrow \mathbb{R}$ differenzierbar in $p \in U$. Dann ist auch $fg : U \longrightarrow \mathbb{R}$ differenzierbar in p, und es gilt*

$$D(fg)(p) : \mathbb{R}^N \longrightarrow \mathbb{R},\ y \mapsto g(p)\langle \nabla f(p), y\rangle + f(p)\langle \nabla g(p), y\rangle.$$

Insbesondere ist also

$$\partial_j(gf)(p) = f(p)(\partial_j g)(p) + g(p)\partial_j f(p)$$

für alle $j = 1, \ldots, N$.

Beweis. Es reicht die erste Aussage zu beweisen. Die "Insbesondere"-Aussage folgt dann durch Einsetzen von e_j.
Nach Voraussetzung gilt

$$f(x) = f(p) + Df(p)(x-p) + \varphi_f(x)$$

mit $\varphi_f(x)/|x-p| \to 0$, $x \to p$, und

$$g(x) = g(p) + Dg(p)(x-p) + \varphi_g(x)$$

mit $\varphi_g(x)/|x-p| \to 0$, $x \to p$. Damit folgt

$$g(x)f(x) = g(p)f(p) + (Dg(p)(x-p))f(p) + g(p)Df(p)(x-p) + \varphi,$$

wobei φ aus Termen besteht, die mindestens den Faktor $\varphi_f(x)$ bzw. $\varphi_g(x)$ oder zweimal den Faktor $(x-p)$ enthalten. Also gilt

$$\frac{\varphi(x)}{|x-p|} \to 0, x \to p$$

und die Aussage folgt. □

Ende 14. Vorlesung

Wir gehen nun noch auf eine geometrische Deutung des Gradienten ein. Dazu benötigen wir noch ein Konzept.

Definition (Richtungsableitung). Sei $U \subseteq \mathbb{R}^N$ offen und $f : U \longrightarrow \mathbb{R}$ gegeben. Sei $p \in U$ und e ein beliebiger Einheitsvektor in $\mathbb{R}^N$. Sei $\delta > 0$ mit $p + te \in U$ für $t \in (-\delta, \delta)$. Sei $g : (-\delta, \delta) \longrightarrow \mathbb{R}$, $g(t) = f(p + te)$. Ist g differenzierbar in 0, so heißt der Grenzwert

$$\partial_e f(p) := g'(0) = \lim_{t\to 0} \frac{1}{t}(g(t) - g(0)) = \lim_{t\to 0} \frac{1}{t}(f(p+te) - f(p))$$

die *Richtungsableitung* von f in p in Richtung e. **Zeichnung.**

Lemma (Geometrische Deutung Gradient). *Sei $f : U \longrightarrow \mathbb{R}$ differenzierbar in p. Dann gilt:*

- *Es ist $\nabla f(p)$ senkrecht auf der Konstanzfläche $M := \{y \mid f(y) = f(p)\}$ (d.h. ist $\gamma : I \longrightarrow M$ stetig differenzierbar, mit $\gamma(0) = p$, so gilt $\langle \gamma'(0), \nabla f(p)\rangle = 0$.* ***Zeichnung.*** *(siehe unten)*
- *Es gilt $\partial_e f(p) = \langle \nabla f(p), e\rangle$ für alle Einheitsvektoren $e \in \mathbb{R}^N$.*
- *Ist $\nabla f(p) \neq 0$, so zeigt $\nabla f(p)$ in die Richtung in der f am stärksten wächst (d.h. die Richtungsableitung von f in der Richtung von $\nabla f(p)$ ist maximal unter den Richtungsableitungen).*

Bemerkung. Da f differenzierbar ist, gilt

$$f(q) = f(p) + \langle \nabla f(p), q-p\rangle + \text{kleiner Fehler}.$$

Für die Funktion $x \mapsto f(p) + \langle \nabla f(p), (x-p)\rangle$ sind die entsprechenden Aussagen des Lemma klar.

Beweis. Nach der Kettenregel gilt für jedes differenzierbare $\gamma : (-\delta, \delta) \longrightarrow U$ mit $\gamma(0) = p$

$$\frac{d}{dt}(f \circ \gamma)(0) = \langle \nabla f(p), \gamma'(0) \rangle.$$

Ist γ eine Kurve mit Werten in M so gilt also $f \circ \gamma = constant$ und es folgt (mit obigem)

$$0 = \frac{d}{dt}(f \circ \gamma)(t) = \langle \nabla f(p), \gamma'(0) \rangle.$$

Das beweist den ersten Punkt.
Andererseits mit $\gamma(t) = p + te$ folgt also

$$\partial_e f(p) = \frac{d}{dt} f \circ \gamma(t_0) = \langle \nabla f(p), e \rangle.$$

Das beweist den zweiten Punkt.
Für jeden Einheitsvektor e gilt nach der Cauchy-Schwarz Ungleichung

$$|\partial_e f(p)| = |\langle \nabla f(p), e \rangle| \leq |\nabla f(p)|$$

und für $e = \frac{1}{|\nabla f(p)|} \nabla f(p)$ gilt dann

$$|\partial_e f(p)| = |\langle \nabla f(p), \frac{1}{|\nabla f(p)|} \nabla f(p) \rangle| = \frac{1}{|\nabla f(p)|} |\nabla f(p)|^2 = |\nabla f(p)|.$$

Damit folgt die letzte Aussage. □

Beispiel. Sei $f : \mathbb{R}^N \setminus \{0\}$, $f(x) = |x|$. Dann gilt $Df(x) = \nabla f(x) = \frac{1}{|x|} x$. Das ist in der Tat der Einheitsvektor in Richtung des stärksten Anstieges von f. **Zeichnung.** (Konstanzflächen sind Sphären).

Bisher haben wir unter der Annahme der Differenzierbarkeit von f gearbeitet. Hier lernen wir nun ein (d a s) Kriterium kennen, das die Differenzierbarkeit von f liefert.

Theorem (Stetigkeit der partiellen Ableitung impliziert Differenzierbarkeit). *Sei $U \subseteq \mathbb{R}^N$ offen und $f : U \longrightarrow \mathbb{R}^M$ gegeben. Existieren die partiellen Ableitungen*

$$\partial_j f_i(q) = \frac{\partial f_i}{\partial x_j}(q) = \lim_{h \to 0, h \neq 0} \frac{1}{h}(f_i(q + he_j) - f_i(q))$$

für alle $q \in U$, $i = 1, \ldots, M$, $j = 1, \ldots, N$ und sind stetig in $p \in U$, so ist f differenzierbar in p.

Bemerkung. In der Situation des Theorem ist die Ableitung natürlich durch die Jacobimatrix/Differentialmatrix $\left(\frac{\partial f_i}{\partial x_j}(p)\right)_{i,j}$ gegeben.

Beweis. Sei L die durch die Matrix $\left(\frac{\partial f_i}{\partial x_j}(p)\right)_{i,j}$ induzierte Abbildung. Zu zeigen:

$$\frac{\varphi(x)}{|x-p|} = \frac{1}{|x-p|} |f(x) - f(p) - L(x-p)| \to 0, x \to p.$$

Wir zeigen: Zu jedem $\varepsilon > 0$ existiert ein $\delta > 0$ mit

$$\frac{1}{|x-p|}|f(x) - f(p) - L(x-p)| \leq \varepsilon$$

für alle $|x-p| \leq \delta$. Sei also $\varepsilon > 0$ gegeben.
Ziel. Wir setzen

$$t = (\tau_1, \dots, \tau_N) := x - p.$$

Für $i = 1, \dots, M$ ist dann die i-te Komponente von $f(x)-f(p)-L(x-p)$ gegeben durch

$$(f(x) - f(p) - L(x-p))_i = f_i(x) - f_i(p) - \sum_{j=1}^{N} \partial_j f_i(p)\tau_j.$$

Es gilt also diesen Term abzuschätzen.
Aufgrund der Stetigkeit der $\partial_j f_i$ existiert ein $\delta > 0$ mit

$$|\partial_j f_i(x) - \partial_j f_i(p)| \leq \frac{\varepsilon}{\sqrt{M} \cdot N} \quad (*)$$

für alle $x \in U$ mit $|x-p| < \delta$. Sei nun $x \in U_\delta(p)$ und

$$t = (\tau_1, \dots, \tau_N) := x - p$$

und

$$t_j := (\tau_1, \dots, \tau_j, 0, \dots, 0) = \sum_{l=1}^{j} \tau_l e_l, \ t_0 = 0$$

für $j = 0, \dots, N$. Insbesondere ist also $t_N = t = x - p$. Dann gilt (Teleskopsumme)

$$f_i(x) - f_i(p) = \sum_{j=1}^{N} \left(f_i(p+t_j) - f_i(p+t_{j-1})\right). \ (T)$$

Zeichnung. Nach dem 1-dimensionalen Mittelwertsatz der Differentialrechnung angewendet auf die Funktion

$$s \mapsto f_i(x + t_{j-1} + se_j)$$

exisistieren $\sigma_{j,i}$ zwischen 0 und τ_j mit

$$f_i(p+t_j) - f_i(p+t_{j-1}) = f_i(p+t_{j-1}+\tau_j e_j) - f_i(p+t_{j-1}) = \tau_j \partial_j f_i(p+t_{j-1}+\sigma_{j,i}e_j)$$

für $j = 1, \dots, N$, $i = 1, \dots, M$. Mit $(*)$ folgt dann

$$\begin{aligned}|f_i(p+t_j) - f_i(p+t_{j-1}) - \tau_j \partial_j f_i(p)| &\leq |\tau_j||\partial_j f_i(p+t_{j-1}+\sigma_{j,i}e_j) - \partial_j f_i(p)| \\ &\overset{(*)}{\leq} |x-p|\frac{\varepsilon}{\sqrt{M}N}.\end{aligned}$$

Mit (T) und dieser Abschätzung folgt

$$|f_i(x) - f_i(p) - \sum_{j=1}^{N} \partial_j f_i(p)\tau_j| \leq \sum_{j=1}^{N} |f_i(p+t_j) - f_i(p+t_{j-1}) - \tau_j \partial_j f_i(p)|$$
$$\leq |x-p|\frac{\varepsilon}{\sqrt{M}}.$$

Damit ergibt sich für x mit $|x-p| < \delta$

$$\begin{aligned}
|f(x) - f(p) - L(x-p)|^2 &= \sum_{i=1}^{M} |f_i(x) - f_i(p) - \sum_{j=1}^{N} \partial_j f_i(p)\tau_j|^2 \\
&\leq \sum_{i=1}^{M} \left| |x-p| \frac{\varepsilon}{\sqrt{M}} \right|^2 \\
&\leq |x-p|^2 \varepsilon^2.
\end{aligned}$$

Damit folgt

$$\frac{1}{|x-p|}|f(x) - f(p) - L(x-p)| \leq \varepsilon.$$

Das beendet den Beweis. $\square$

Beispiel. Die Funktion $f : \mathbb{R}^N \setminus \{0\} \longrightarrow \mathbb{R}$, $x \mapsto |x|$ ist differenzierbar.

Ende 15. Vorlesung

Definition (Stetig differenzierbar). Sei $U \subseteq \mathbb{R}^N$ offen und $f : U \longrightarrow \mathbb{R}^M$ gegeben. Existieren für jedes $p \in U$ die partiellen Ableitungen

$$\partial_j f_i(p) = \frac{\partial f_i}{\partial x_j}(p) := \lim_{h \to 0, h \neq 0} \frac{1}{h}(f_i(p + he_j) - f_i(p))$$

für $i = 1, \ldots, M$, $j = 1, \ldots, N$ und sind stetig in $p \in U$, so heißt f stetig differenzierbar (auf U).

Bemerkung. Nach dem vorherigen Theorem ist jede stetig differenzierbare Funktion auch differenzierbar (wie es ja auch - rein sprachlich - sein sollte ;-)

Wir schließen diesen Abschnitt mit einer Diskussion von Mittelwertsätzen ab.

Theorem (Mittelwertsatz). *Sei $U \subseteq \mathbb{R}^N$ offen, $f : U \to \mathbb{R}$ differenzierbar, und seien $x, y \in U$ mit $x + t(y-x) \in U$ für $t \in [0,1]$ gegeben. Dann existiert ein $\vartheta \in (0,1)$, sodass*

$$f(y) - f(x) = Df(x + \vartheta(y-x))(y-x).$$

Bemerkung.
- Ist U konvex, so gilt $x + t(y-x) \in U$ für alle $x, y \in U$ (und umgekehrt). Da alle Kugeln konvex sind, gilt der obige Satz also insbesondere lokal (d.h. jedes $x \in U$ hat eine Umgebung $U_r(x)$, in der der Satz gilt).

- Es ist wesentlich (siehe Übung), dass der Wertebereich von f im eindimensionalen $\mathbb{R}$ enthalten ist.

Beweis. Sei $\gamma : [0,1] \to U,\ \gamma(t) = x + t(y-x)$ und

$$h : [0,1] \to \mathbb{R},\ h = f \circ \gamma.$$

Dann ist h stetig auf $[0,1]$ und nach Kettenregel differenzierbar auf $(0,1)$ mit

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

Der eindimensionale Mittelwertsatz liefert dann

$$\begin{array}{rcl} f(y) - f(x) & = & h(1) - h(0) \\ (1 - dim\ MWS) & = & h'(\vartheta)(1-0) \\ (\gamma'(t) = y - x) & = & \langle \nabla f(x + \vartheta(y-x)), y - x \rangle \\ & = & Df(x + \vartheta(y-x))(y-x). \end{array}$$

$\square$

Der Satz hat Konsequenzen für Funktionen $f : U \longrightarrow \mathbb{R}^M$.

FOLGERUNG. *Sei $U \subseteq \mathbb{R}^N$ offen und $f : U \longrightarrow \mathbb{R}^M$ differenzierbar. Dann gilt:*

(a) *Gilt $|\partial_k f_i(z)| \leq C$ für alle $z \in U$ und $1 \leq k \leq N$, $1 \leq i \leq M$, so gilt*

$$|f(y) - f(x)| \leq C\sqrt{MN}|y - x|$$

für alle x, y, deren Verbindungstrecke ganz in U liegt.

(b) *Für alle $x, y \in U$, deren Verbindungstrecke ganz in U liegt existiert ein ξ der Form $x + \vartheta(y - x)$ für $\vartheta \in (0,1)$ mit*

$$|f(y) - f(x)| \leq \|Df(\xi)\||y - x|.$$

(c) *Ist U zusammenhängend und $Df(x) = 0$ für alle $x \in U$, so folgt $f \equiv constant$ auf U.*

Erinnerung: $\|L\| := \sup\{|Lx| \mid |x| \leq 1\}$. Es gilt $|Lx| \leq \|L\||x|$.)

Beweis. (a) Nach dem vorangehenden Mittelwertsatz gibt es $\vartheta_i \in (0,1)$, $i = 1, \ldots, d$ mit

$$|f_i(y) - f_i(x)| = |\langle \nabla f_i(x + \vartheta_i(y-x), (y-x) \rangle| \leq \sum_{j=1}^{N} C|y_j - x_j| \leq C\sqrt{N}|y - x|.$$

(Hier wurde Cauchy-Schwarz in der letzten Abschätzung benutzt). Damit folgt dann sofort

$$|f(y) - f(x)| = \left(\sum_{i=1}^{M} 1|f_i(y) - f_i(x)|^2 \right)^{1/2} \leq C\sqrt{NM}|y - x|.$$

(b) *Spezialfall:* Es gilt $f_i(x) = f_i(y)$ für alle $i \neq 1$. Dann folgt aus obigem Mittelwertsatz

$$|f(y) - f(x)| = |f_1(y) - f_1(x)| = |\langle \nabla f_1(x + \vartheta(y-x)), (y-x) \rangle| \leq \|Df(\xi)\||y - x|.$$

Allgemeiner Fall: Sei U eine Drehung, sodass $Uf(y)$ und $Uf(x)$ sich höchstens in der ersten Komponente unterscheiden (drehe zunächst so, dass die durch $0, f(x), f(y)$ gebildete Ebene gerade die durch die ersten beiden Koordinatenachsen gegebene Ebene ist. Drehe anschließend weiter.) Dann gilt nach dem Spezialfall

$$|f(y)-f(x)| = |Uf(y)-Uf(x)| \leq \|DUf(\xi)\||y-x| = \|Df(\xi)\||y-x|.$$

(c) Da U offen und zusammenhängend ist, ist es wegzusammenhängend. Sei nun $\gamma : [0,1] \longrightarrow U$ mit $\gamma(0) = x$, $\gamma(1) = y$ gegeben. Da U offen ist und γ stetig, können wir ohne Einschränkung annehmen, dass γ stückweise linear ist, d.h. es gibt $p = x, x_1, \ldots, x_n = y \in U$, so dass γ durch die Verbindungsstrecken zwischen x_i und x_{i+1} beschrieben ist. Dann gilt nach (a) oder (b) $f(x_i) = f(x_{i+1})$. Damit folgt die Behauptung. □

Ende 16. Vorlesung

3. Höhere Ableitungen und der Satz von Taylor

In diesem Abschnitt untersuchen wir, wie Funktionen gut durch Polynome approximiert werden können, wenn sie genügend oft differenzierbar sind.
Wie schon bekannt, heißt f stetig differenzierbar, wenn die partiellen Ableitungen aller f_i, $i = 1, \ldots, m$ existieren und stetig sind. Wir können das Iterieren: Ist f einmal stetig differenzierbar und existieren alle ersten Ableitungen der partiellen Ableitungen $D_j f_i$ und sind stetig, so heißt f zweimal stetig differenzierbar und man nennt die $D_i D_j f_k$ die zweiten partiellen Ableitungen von f. Induktiv definiert man dann die $(q+1)$-ten partiellen Ableitungen von f als die ersten Ableitungen der p-ten Ableitungen $D_{i_1} \ldots D_{i_q} f$ (falls existent).

Definition. Sei $U \subseteq \mathbb{R}^N$ offen. Die Funktion $f : U \longrightarrow \mathbb{R}^M$ heißt *q mal stetig differenzierbar*, wenn f $(q-1)$ mal stetig differenzierbar ist und die q-ten partiellen Ableitungen $D_{i_q} \ldots D_{i_1} f$, $i_1, \ldots, i_{q-1} \in \{1, \ldots, N\}$, existieren und stetig sind. In diesem Fall nennt man die entstehenden partiellen Ableitungen $D_{i_q}(D_{i_{q-1}} \ldots D_{i_1} f)$ die *q-ten partiellen Ableitungen* von f. Der Vektorraum der q-mal stetig differenziebaren Funktionen wird mit $C^q(U, \mathbb{R}^M)$ bezeichnet. Die Elemente aus

$$C^\infty(U, \mathbb{R}^M) := \bigcap_{q=1}^{\infty} C^q(U, \mathbb{R}^M)$$

heißen *beliebig oft differenzierbare Funktionen.*

Es zeigt sich, dass die Reihenfolge der partiellen Ableitungen keine Rolle spielt (wenn f genügend oft s t e t i g differenzierbar ist).

Theorem (Satz von Schwarz). *Sei $U \subseteq \mathbb{R}^N$ offen und $f : U \to \mathbb{R}^M$. Ist die Funktion f zweimal stetig differenzierbar, so gilt für jedes $p \in U$,*

$k \in \{1, \ldots, M\}$ und $i, j \in \{1, \ldots, N\}$

$$D_i D_j f_k(p) = D_j D_i f_k(p).$$

Beweis. Wir können die Situation ohne Einschränkung etwas vereinfachen:

- Es genügt eine Komponente $k \in \{1, \ldots, n\}$ zu betrachten, d.h. ohne Einschränkung setzen wir $M = 1$ voraus.
- Für $i = j$ gibt es nichts zu beweisen
- Wir setzen ohne Einschränkung voraus $N = 2$ und setzen $i = 1,\ j = 2$ und $N = 2$. (Ansonsten umnummerieren.)
- Betrachte o. B. d. A. $p = 0$

Es ist somit zu zeigen: $D_1 D_2 f(0) = D_2 D_1 f(0)$.
Wir zeigen dies im wesentlichen, indem wir den entsprechenden Differenzenquotienten zweiter Ordnung auf zwei verschiedene Weisen ausrechnen und anschließend einen Grenzübergang durchführen.
Seien $t, s > 0$ beliebig. Wir betrachten nun

$$\begin{aligned}\Delta(s,t) &:= (f(s,t) - f(s,0)) - (f(0,t) - f(0,0)) \\ &= (f(s,t) - f(0,t)) - (f(s,0) - f(0,0)).\end{aligned}$$

Dann gilt für diesen Ausdruck also

$$\begin{aligned}\Delta(s,t) &= (f(s,t) - f(s,0)) - (f(0,t) - f(0,0)) \\ (g(r) := f(r,t) - f(r,0)) \quad &= g(s) - g(0) \\ (MWS) \quad &= g'(s_1)(s-0) \qquad \text{für } s_1 \in (0,s) \\ (g' = ...) \quad &= s(D_1 f(s_1,t) - D_1 f(s_1,0)) \\ (h(r) := D_1 f(s_1, r)) \quad &= s(h(t) - h(0)) \\ &= sth'(t_1) \qquad \text{mit } t_1 \in (0,t) \\ &= stD_2 D_1 f(s_1, t_1).\end{aligned}$$

Die andere Darstellung von Δ liefert nach analoger Rechnung dann

$$\begin{aligned}\Delta(s,t) &= (f(s,t) - f(0,t)) - (f(s,0) - f(0,0)) \\ &= \ldots \\ &= stD_1 D_2 f(s_2, t_2) \qquad \text{mit } s_2 \in (0,s),\ t_2 \in (0,t).\end{aligned}$$

Für $s, t \neq 0$ kann man dividieren und erhält

$$D_2 D_2 f(s_1, t_1) = D_1 D_2 f(s_2, t_2).$$

Betrachtet man nun den Grenzwert $(s,t) \to (0,0)$ so folgt aus der Stetigkeit der zweiten Ableitungen also

$$D_1 D_2 f(0,0) = D_2 D_1 f(0,0).$$

Das beendet den Beweis. □

FOLGERUNG (Vertauschbarkeit der partiellen Ableitungen). *Sei $U \subseteq \mathbb{R}^N$ offen und $f : U \to \mathbb{R}^M$ p-mal stetig differenzierbar, so können in*

$$D_{j_1}, \dots, D_{j_p} f$$

die partiellen Ableitungen beliebig vertauscht werden.

Beweis. Die Behauptung folgt aus dem Satz von Schwarz und Induktion. □

Für die weiteren Betrachtungen brauchen wir nun etwas Notation. Dabei geht es im wesentlichen um gute „Buchhaltung“:
Für $U \subseteq \mathbb{R}^N$ offen und $f : U \longrightarrow \mathbb{R}^M$ q-mal stetig differenzierbar. Dann definiert man für $r \leq q$ und $j = 1, \dots, N$

$$D_j^r f = \underbrace{D_j \dots D_j}_{r\text{-mal}} f \qquad (D_j^0 f := f)$$

Für $(j_1, \dots, j_q) \in \{1, \dots, N\}^q$ bezeichne α_i die Anzahl des Auftretens der Zahl i unter den j_k. Dann gilt nach dem Satz von Schwarz bzw. seinem Korollar

$$D_{j_1} \dots D_{j_q} f = D_1^{\alpha_1} \dots D_N^{\alpha_N} f.$$

Für *Multiindizes* $\alpha = (\alpha_1, \dots, \alpha_N) \in \mathbb{N}_0^N$ sei im folgenden

$$|\alpha| := \sum_{j=1}^N \alpha_j \qquad \text{der Betrag/die Länge von } \alpha,$$

$$\alpha! := \prod_{j=1}^N \alpha_j! \qquad \text{die Fakultät von } \alpha,$$

$$x^\alpha := x_1^{\alpha_1} \cdot \ldots \cdot x_N^{\alpha_N} = \prod_{j=1}^N x_j^{\alpha_j} \qquad \text{mit } x = (x_1, \dots, x_N) \in \mathbb{R}^N$$

Sei weiterhin für ein q mal stetige differenziebares f und einen Multiindex α mit $|\alpha| \leq q$

$$D^\alpha f := D_1^{\alpha_1} \dots D_N^{\alpha_N} f = (\prod_{j=1}^N D_j^{\alpha_j}) f.$$

Anwendung. (Übung) Die Multiindexschreibweise kann anhand des *Polynomischen Satzes* verdeutlicht werden (auf einen Beweis wird an dieser Stelle verzichtet):

$$\left(\sum_{j=1}^N x_j \right)^k = \sum_{\alpha : |\alpha| = k} \frac{k!}{\alpha!} x^\alpha$$

Wendet man diese Formel zum Beispiel auf $(a+b)^3$ an, so ergibt sich:

$$\begin{aligned}(a+b)^3 &= \sum_{\alpha:|\alpha|=3} \frac{3!}{\alpha!}x^\alpha = \frac{3!}{(3,0)!}x^{(3,0)} + \frac{3!}{(2,1)!}x^{(2,1)} + \frac{3!}{(1,2)!}x^{(1,2)} + \frac{3!}{(0,3)!}x^{(0,3)} \\ &= \frac{3!}{3!}a^3b^0 + \frac{3!}{2!}a^2b^1 + \frac{3!}{2!}a^1b^2 + \frac{3!}{3!}a^0b^3 = a^3+3a^2b+3ab^2+b^3\end{aligned}$$

Mit den vorherigen Resultaten und den obigen Beziehungen sind wir nun in der Lage den Taylor'schen Satz auf Funktionen von N Variablen zu übertragen.

Theorem (Taylor). *Sei $U \subseteq \mathbb{R}^N$ offen, $q \geq 1$, $f \in C^{q+1}(U,\mathbb{R})$ und $p \in U$. Dann gilt für alle $x \in U$, für die die Verbindungsstrecke von p und x in U liegt:*

$$f(x) = \sum_{|\alpha|\leq q} \frac{1}{\alpha!} D^\alpha f(p)(x-p)^\alpha + R_q(x)$$

mit

$$R_q(x) = \sum_{|\alpha|=q+1} \frac{1}{\alpha!} D^\alpha f(p+\vartheta(x-p))(x-p)^\alpha \qquad \textit{mit } \vartheta \in (0,1)$$

Bemerkung. Zur Verdeutlichung geben wir dieses Resultat für $q = 0,1,2$ explizit an. Sei dazu $x = (x_1,\ldots,x_N)$ und $p = (p_1,\ldots,p_N)$. Dann gilt:

$\underline{q=0}$:
$$\begin{aligned} f(x) =& f(p) + \sum_{j=1}^N D_j f(p+\vartheta(x-p))(x_j-p_j) \\ =& f(p) + \langle \nabla f(p+\vartheta(x-p)), x-p\rangle \end{aligned}$$

$\underline{q=1}$:
$$\begin{aligned} f(x) =& f(p) + \sum_{j=1}^N D_j f(p)(x_j-p_j) \\ &+ \frac{1}{2}\sum_{j,k=1}^N D_jD_k f(p+\vartheta(x-p))(x_j-p_j)(x_k-p_k) \\ =& f(p) + \langle \nabla f(p), (x-p)\rangle + \frac{1}{2}\langle H(p+\theta(x-p))(x-p),(x-p)\rangle \end{aligned}$$

$\underline{q=2}$:
$$\begin{aligned} f(x) =& f(p) + ... \\ =& f(p) + \langle \nabla f(p), (x-p)\rangle + \frac{1}{2}\langle H(p)(x-p),(x-p)\rangle + R_2(x) \end{aligned}$$

mit der *Hesse-Matrix*

$$H(f)(\cdot) = H(\cdot) = (D_jD_kf(\cdot))_{j,k=1}^N$$

Ende 17. Vorlesung

Beweis. Bonusmaterial: Der Satz wird bewiesen durch Rückführung auf den eindimensionalen Fall. Dazu sei

$$\begin{gathered} y = (y_1, \dots, y_m) := (x - p) \\ g : \ [-\varepsilon, 1 + \varepsilon] \to \mathbb{R} \\ g(t) = f(p + ty) = (f \circ h)(t), \text{ mit } h(t) = p + ty \end{gathered}$$

wobei $\varepsilon > 0$ so klein gewählt sei, dass das Gesamtstück

$$\{p + ty \mid t \in [-\varepsilon, 1 + \varepsilon]\}$$

ganz in U liegt (möglich, da U offen).
Offensichtlich ist g stetig differenzierbar und mit der Kettenregel folgt

$$g'(t) = Df(h(t))h'(t) = \sum_{j=1}^{N} D_j f(h(t)) \underbrace{h_j'(t)}_{y_j} = \sum_{j=1}^{N} y_j D_j f(p + ty)$$

Ist $q \geq 1$, so ist auch g' stetig differenzierbar und es gilt:

$$\begin{aligned} g,,(t) &= \sum_{j=1}^{N} y_j \frac{d}{dt} D_j f(p + ty) \\ &= \sum_{j=1}^{N} y_j \sum_{i=1}^{N} D_i D_j f(p + ty) y_i \\ &= \left(\sum_{j=1}^{N} y_j D_j \right)^2 f(p + ty) \end{aligned}$$

wobei der Term $(\sum y_j D_j)^2$ formal auszumultiplizieren ist. Allgemeiner erhält man durch induktive Anwendung des Verfahrens für $k \leq q + 1$

$$g^{(k)}(t) = \left(\sum_{j=1}^{N} y_j D_j \right)^2 f(p + ty)$$

wobei auch hier die k-te Potenz formal auszurechnen ist.
Nach dem Polynomischen Satz (siehe oben) gilt für beliebige $z = (z_1, \dots, z_N) \in \mathbb{R}^N$ und $k \in \mathbb{N}$

$$\left(\sum_{j=1}^{N} \zeta_j \right)^k = \sum_{\alpha : |\alpha| = k} \frac{k!}{\alpha!} z^\alpha$$

In dieser Formel ersetzen wir $(z_1, \dots, z_N)$ durch $(y_1 D_1, \dots, y_N D_N)$. Damit erhalten wir

$$\begin{aligned}
\frac{1}{k!} g^{(k)}(t) &= \frac{1}{k!} \left(\sum_{j=1}^{N} y_j D_j \right)^k f(p + ty) \\
&= \left(\sum_{|\alpha|=k} \frac{1}{\alpha!} (y_1 D_1, \dots, y_N D_N)^\alpha \right) f(p + ty) \\
&= \sum_{|\alpha|=k} \frac{y^\alpha}{\alpha!} D^\alpha f(p + ty)
\end{aligned}$$

Die Taylorsche Formel liefert für $t \in [0, 1]$

$$g(t) = \sum_{k=1}^{q} \frac{1}{k!} g^{(k)}(0) t^k + \widetilde{R_q}(t)$$

mit $\widetilde{R_q}(t) = \frac{1}{(q+1)!} t^q g^{(q)}(\vartheta t)$ mit $0 < \vartheta < 1$. Damit erhalten wir:

$$\begin{aligned}
f(x) = g(1) &= \sum_{k=0}^{q} \frac{1}{k!} g^{(k)}(0) + \widetilde{R_q}(1) \\
&= \sum_{k=0}^{q} \sum_{|\alpha|=k} \frac{(x-p)^\alpha}{\alpha!} D^\alpha f(p) + \widetilde{R_q}(1) \\
&= \sum_{|\alpha| \le q} \frac{1}{\alpha!} D^\alpha f(p) (x-p)^\alpha + R_q(x)
\end{aligned}$$

mit

$$\begin{aligned}
R_q(x) = \widetilde{R_q}(1) &= \frac{1}{(q+1)!} g^{(q+1)}(\vartheta) \\
&= \sum_{|\alpha|=q+1} \frac{y^\alpha}{\alpha!} D^\alpha f(p + \vartheta y) \\
&= \sum_{|\alpha|=q+1} \frac{1}{\alpha!} D^\alpha f(p + \vartheta(x-p))(x-p)^\alpha
\end{aligned}$$

Das liefert die gewünschte Aussage. □

4. Extrema von Funktionen

Wie im eindimensionalen Fall werden wir Kriterien für (lokale) Extrema von Funktionen mehrerer Veränderlicher (im Inneren eines Gebietes U) angeben.

Definition (Extrema im mehrdimensionalen Fall). Sei $U \subseteq \mathbb{R}^N$ gegeben, $f : U \to \mathbb{R}$ und $p \in U$. Dann hat f in p ein *lokales* $\substack{\textit{Maximum}\\\textit{Minimum}}$, falls ein $\delta > 0$ existiert, so dass gilt

$$f(x) \substack{\leq \\ \geq} f(p)$$

für alle $x \in U$ mit $|x-p| < \delta$. Es hat f in p in *strenges lokales* $\substack{\textit{Maximum}\\\textit{Minimum}}$, falls ein $\delta > 0$ existiert, so dass gilt

$$f(x) \substack{< \\ >} f(p)$$

für alle $x \in U$ mit $0 < |x-p| < \delta$. Ein $p \in U$ ist ein *Extremum* von f, wenn es ein Minimum oder ein Maximum ist.

Bemerkung. Gelten die entsprechenden Ungleichungen für alle $x \in U$, so handelt es sich um globale Extrema.

Theorem (Notwendige Kriterien für Extrema). *Sei $U \subseteq \mathbb{R}^N$ offen, $p \in U$ und die Funktion $f : U \to \mathbb{R}$ sei differenzierbar in p. Hat f in p ein lokales Maximum bzw. Minimum, so gilt*

$$\nabla f(p) = 0$$

Beweis. Da f bei p differenzierbar ist, sind die Funktionen

$$g_j : (-\varepsilon, \varepsilon) \to \mathbb{R}, \; g_j(t) = f(p + te_j)$$

in $t = 0$ differenzierbar für jedes $j = 1, \ldots, N$. Außerdem haben sie bei 0 ein Extremum. Damit folgt

$$g_j'(0) = 0 = \frac{\partial}{\partial x_j} f(p).$$

Das beendet den Beweis. □

Bemerkung. Ist $M \subseteq \mathbb{R}^N$ beliebig und $f : M \longrightarrow \mathbb{R}$ differenzierbar im Inneren von M, so liefet dieses Theorem eine notwendige Bedingung für Extremwerte im Inneren von M. Das Verhalten von f auf dem Rand muss dann natürlich noch getrennt untersucht werden.

Erinnerung. Eine symmetrische $N \times N$ Matrix A heißt *positiv definit*, falls für alle $x \in \mathbb{R}^N \setminus \{0\}$ gilt

$$\langle Ax, x \rangle > 0.$$

Sie heißt *positiv semidefinit*, wenn für alle $x \in \mathbb{R}^N \setminus \{0\}$ gilt

$$\langle Ax, x \rangle \geq 0.$$

Durch Diagonalisieren von A sieht man leicht, dass gilt

$$\langle Ax, x \rangle \geq \lambda_m \|x\|^2$$

mit $\lambda_m :=$ kleinste Eigenwert von A, wobei Gleichheit gilt für x aus dem Eigenraum zu λ_m. Damit folgt, dass A positiv (semi)definit ist, genau dann, wenn die Eigenwerte von A strikt positiv bzw. nichtnegativ sind. Entsprechend definiert man negative (semi)definite Matrizen.

Theorem (Kriterien für Extrema). *Sei $U \subseteq \mathbb{R}^N$ offen, $f \in C^2(U, \mathbb{R})$ und $p \in U$. Dann gilt:*

- *(Notwendiges Kriterium) Hat f bei p ein (lokales) $\genfrac{}{}{0pt}{}{\textit{Maximum}}{\textit{Minimum}}$, so gilt $\nabla f(p) = 0$ und die Hesse-Matrix $(D_k D_j f(p))_{j,k=1}^N$ ist $\genfrac{}{}{0pt}{}{\textit{negativ}}{\textit{positiv}}$ semidefinit.*
- *(Hinreichendes Kriterium) Gilt $\nabla f(p) = 0$ und die Hesse-Matrix ist $\genfrac{}{}{0pt}{}{\textit{negativ}}{\textit{positiv}}$ definit in p, so hat f ein striktes lokales $\genfrac{}{}{0pt}{}{\textit{Maximum}}{\textit{Minimum}}$ in p.*

Bemerkung. Obiges Theorem liefert im eindimensionalen Fall gerade die uns schon bekannte Charakterisierung.

Beweis. **Idee.** Es gilt nach Taylor und dem vorherigen Satz

$$\begin{aligned} f(x) &= f(p) + \langle \nabla f(p), (x-p) \rangle + \frac{1}{2} \langle H(f)(x)(x-p), (x-p) \rangle \\ &= f(p) + \frac{1}{2} \langle H(f)(\xi)(x-p), (x-p) \rangle \end{aligned}$$

mit einer Zwischenstelle ξ zwischen x und p. Damit wird also 'alles' durch das Verhalten von $H(f)(\xi)$ entschieden. Dieses Verhalten ist aber (im wesentlichen) das Verhalten von $H(f)(p)$.

Ende 18. Vorlesung

Wir zeigen zunächst die notwendige Bedingung: Nach obigen Satz gilt $\nabla f(p) = 0$. Angenommen $H(p)$ ist nicht negativ semidefinit, dann existiert ein $y \in \mathbb{R}^N$, $|y| = 1$ und $\langle H(p)y, y \rangle > 0$.
Da $f \in C^2(U, \mathbb{R})$ gilt, existiert ein $\delta > 0$ mit

$$\langle H(x)y, y \rangle > 0$$

für alle $x \in U$ mit $|x - p| < \delta$.
Dann folgt mit dem Satz von Taylor für alle $t \in (0, \delta]$

$$\begin{aligned} f(p+ty) &= f(p) + 0 + \sum_{|\alpha|=2} \frac{1}{\alpha!} D^\alpha f(p + \vartheta t y) t^{|\alpha|} y^\alpha \\ &= f(p) + \frac{1}{2} \sum_{i,j=1}^N t^2 D_i D_j f(p + \vartheta t y) y_i y_j \\ &= f(p) + \frac{1}{2} t^2 \langle H(p + \vartheta t y) y, y \rangle \\ &> f(p) \end{aligned}$$

Wir zeigen nun die hinreichende Bedingung:
Sei $Df(p) = 0$ und $(D_i D_j f(p))_{i,j}$ negativ definit, dann gilt also mit $y = (y_1, \ldots, y_m)$

$$0 > \sum_{i,j=1}^N D_i D_j f(p) y_i y_j = \langle H(f)(p) y, y \rangle$$

für alle y mit $y \neq 0$.

Da die Abbildung

$$S^N = \{y \in \mathbb{R}^N \ \mid \ |y| = 1\} \to \mathbb{R}, y \mapsto \langle H(f)(p)y, y\rangle$$

stetig ist, nimmt sie auf dem kompakten S^N ihr Maximum an. Es gibt also ein $\lambda > 0$ mit $\langle H(f)(p)y, y\rangle \leq -\lambda$ für alle y mit $|y| = 1$. Da alle $D_i D_j f$ stetig sind, existiert ein $\delta > 0$ mit

$$|D_i D_j f(x) - D_i D_j f(p)| \leq \frac{\lambda}{2N^2}$$

für alle $x \in U$ mit $|x - p| < \delta$. Damit folgt

$$\begin{aligned}
\langle H(f)(x)y, y\rangle &= \sum_{i,j=1}^{N} D_i D_j f(x) y_i y_j \\
&= \sum_{i,j=1}^{N} (D_i D_j f(x) - D_i D_j f(p)) y_i y_j + \langle H(f)(p)y, y\rangle \\
&\leq \frac{\lambda}{2} - \lambda \\
&= -\frac{\lambda}{2}
\end{aligned}$$

für alle y mit $|y| = 1$. Damit folgt dann sofort

$$\langle H(f)(x)y, y\rangle \leq -\frac{\lambda}{2}|y|^2 < 0$$

für alle $|y| \neq 0$, und x mit $|x - p| < \delta$. Nachg dem Satz von Taylor gilt aber

$$f(x) = f(p) + Df(p)(x - p) + \frac{1}{2}\langle H(f)(p + \vartheta(x - p))(x - p), (x - p)\rangle$$

für ein $\vartheta \in (0, 1)$. Mit der vorherigen Abschätzung und $Df(p) = 0$ ergibt sich dann sofort

$$f(x) - f(p) < 0$$

für $|x - p| < \delta$, $x \neq p$. Damit hat f also in p ein striktes Maximum. Die Aussage über Minima kann analog bewiesen werden. Alternativ kann man auch die schon bewiesene Aussage auf $-f$ anwenden. □

Bemerkung.

- Der Beweis nutzt: A nahe $B \Rightarrow \langle Ay, y\rangle$ nahe $\langle By, y\rangle$
- Der Beweis zeigt: A negativ definit, größter Eigenwert ist $-\lambda$, B nahe A, B symmetrisch $; \Rightarrow$ größter Eigenwert von B ist nahe $anischerAsylbewerberinnenh - \lambda$,

KAPITEL 3

Implizite Funktionen und lokale Invertierbarkeit

In diesem Abschnitt untersuchen wir (lokale) Lösbarkeit von Gleichungen der Form

$$F(x,y) = 0$$

nach x bzw y bei gegebenen y bzw. x.
Wir beginnen mit einer Diskussion von lokaler Invertierbarkeit. Zur Einstimmung betrachten wir zwei Spezialfälle.
Beispiel. Sei I ein Interval in $\mathbb{R}$ und $f : I \longrightarrow \mathbb{R}$ stetig. Dann hat f eine Umkehrfunktion genau dann wenn f streng monoton ist. Das ist eine gute Charakterisierung aber unter Umständen nicht einfach zu prüfen. Ist f stetig differenzierbar, so wird die Lage einfacher: Ist $f'(p) \neq 0$ so ist f lokal invertierbar. (Denn dann hat f' in der Nähe von p festes Vorzeichen und ist damit nach dem Mittelwertsatz streng monoton.) Die (lokale) Inverse g ist stetig differenzierbar mit

$$g'(y) = \frac{1}{f(x)}$$

mit $f(x) = y$.
Beispiel. Sei L eine lineare Abbidlung von $\mathbb{R}^N$ nach $\mathbb{R}^N$ und $f : \mathbb{R}^N \longrightarrow \mathbb{R}^N, f(x) = a + Lx$. Dann ist f invertierbar genau dann, wenn L invertierbar ist. In diesem Fall gilt für die Umkehrabbildung g also

$$g : \mathbb{R}^N \longrightarrow \mathbb{R}^N, g(y) = -L^{-1}a + L^{-1}y.$$

Insbesondere ist g stetig differenzierbar mit

$$Dg(y) = Df(x)^{-1}.$$

In beiden Fällen gilt also: Ist f stetig differenzierbar und $Df(x)$ invertierbar, so ist f lokal invertierbar und seine Umkehrabbildung ist stetig differenzierbar mit $Dg(y) = Df(x)^{-1}$. Tatsächlich gilt folgender allgemeiner Satz.

Theorem (Satz über die Umkehrfunktion). *Sei $U \subseteq \mathbb{R}^N$ offen, $f : U \longrightarrow \mathbb{R}^N$ stetig differenzierbar, $p \in U$, $q := f(p)$ und $Df(p)$ invertierbar. Dann gibt es eine offene Umgebung V von p eine offene Umgebung W von q so dass folgendes gilt:*

- *$f|_V : V \longrightarrow W$ ist bijektiv,*
- *die Umkehrabbildung $g := (f|_V)^{-1} : W \longrightarrow V$ ist stetig differenzierbar,*

- *für alle $x \in V$ ist $Df(x)$ invertierbar und es gilt $Dg(y) = Df(x)^{-1}$ für $y = f(x)$.*

Zeichnung.

Bemerkung.

- Ist die Umkehrfunktion g von f differenzierbar, so folgt mit $g \circ f = id$ aus der Kettenregel, dass
$$1 = Dg(y)Df(x)$$
für $y = f(x)$. Die Invertierbarkeit von $Df(x)$ und die Formel für $Dg(y)$ sind also Konsequenzen der Diffbarkeit von g.
- In einer Dimension gilt: Ist $f : \mathbb{R} \longrightarrow \mathbb{R}$ stetig differenzierbar mit $Df(x) = f'(x) \neq 0$, so hat also f' festes Vorzeichen, f ist streng monoton und damit auf ganz $\mathbb{R}$ invertierbar. Wir zeigen nun an zwei Beispielen, dass in beliebigen Dimensionen globalen Invertierbarkeit i. a. nicht gilt, auch wenn Df überall invertierbar ist.

  Beispiel. Sei $f : \mathbb{R}^2 \setminus \{0\} \longrightarrow \mathbb{R}^2 \setminus \{0\}$, $f(\xi_1, \xi_2) = (\xi_1^2 - \xi_2^2, 2\xi_1\xi_2)$. Dann gilt
$$Df(\xi_1, \xi_2) = (2\xi_1 \;\; -2\xi_2, 2\xi_2 \;\; 2\xi_1).$$
  Also $\det Df(\xi_1, \xi_2) = 4(\xi_1^2 + \xi_2^2) \neq 0$. Damit ist Df überall invertierbar. Aber es ist f nicht injektiv, denn es gilt $f(-\xi_1, -\xi_2) = f(\xi_1, \xi_2)$. (Setzt man $z = \xi_1 + i\xi_2$, so ist $f(\xi_1, \xi_2) = (\Re z^2, \Im z^2)$. Es handelt sich also bei f um die komplexe Quadratfunktion und bei Df umd die Multiplikation mit $2z$.)

  Beispiel. Sei $f : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$, $f(\xi_1, \xi_2) = e^{\xi_1}(\cos\xi_2, \sin\xi_2)$. Dann ist f nicht injektiv. Aber die Ableitung ist überall invertierbar: Es ist $Df(\xi_1, \xi_2) = e^{\xi_1}(\cos\xi_2 \;\; -\sin\xi_2, \sin\xi_2 \;\; \cos\xi_2)$. Also ist $\det Df = e^{\xi_1} \neq 0$. Aber es ist $f(\xi_1, \xi_2) = f(\xi_1, \xi_2 + 2\pi)$. (Setzt man $z = \xi_1 + i\xi_2$ so ist $f(z) = (\Re e^z, \Im e^z)$. Es handelt sich also um die komplexe Exponentialfunktion und ihre Ableitung.)

Ende 19. Vorlesung

Zur Vorbereitung auf den Beweis des Satzes beweisen wir zwei Resultate.

LEMMA. *Sei $A : \mathbb{R}^N \longrightarrow \mathbb{R}^N$ linear. Dann sind äquivalent:*

(i) *Es ist A invertierbar.*
(ii) *Es gibt ein $C > 0$ mit $|Ax| \geq C|x|$ für alle $x \in \mathbb{R}^N$*
(iii) *Es ist 0 nicht Eigenwert von A.*

In diesem Fall kann $C = \frac{1}{\|A^{-1}\|}$ gewählt werden.

Beweis. (i) $\Longrightarrow$ (ii): Sei $y = Ax$. Dann gilt nach (i) also $x = A^{-1}y$. Damit folgt
$$|x| \leq \|A^{-1}\||y|.$$

Das liefert dann
$$|Ax| = |y| \geq \frac{1}{\|A^{-1}\|}|x|.$$
(ii) $\Longrightarrow$ (iii): Das ist klar.
(iii) $\Longrightarrow$ (i): Nach (iii) ist A injektiv. Damit ist es bijektiv. $\square$

PROPOSITION. *Sei $h : B_R(0) \subseteq \mathbb{R}^N \longrightarrow \mathbb{R}^N$ stetig und differenzierbar in $U_R(0)$ und*
$$\psi : U_R(0) \longrightarrow \mathbb{R}, \psi(x) = |h(x)|^2.$$
Hat ψ ein Extremum in p in U_R, so gilt $Dh(p)^\top h(p) = 0$. Insbesondere verschwindet also h in p, wenn Dh auf ganz $U_R(0)$ invertierbar ist.

Beweis. Das folgt direkt durch Bilden des Gradienten:
$$0 = \partial_j\psi(p) = \partial_j \sum_{k=1}^{N} h_k(p)^2 = 2\sum_{k=1}^{N} \partial_j h_k(p) \cdot h_k(p).$$
$\square$

BEMERKUNG. Die Aussage ist bemerkenswert, weil auf den Funktionswert im Extremum geschlossen werden kann (und nicht nur auf den Gradienten).

BEMERKUNG. (Übung) Ist $h : \mathbb{R}^N \longrightarrow \mathbb{R}$ stetig differenzierbar mit
- $Dh(x)$ invertierbar für alle $x \in \mathbb{R}^N$,
- $|h| \to \infty$ für $x \to \infty$,

so gibt es ein $x \in \mathbb{R}^N$ mit $h(x) = 0$.

Wir kommen nun zum Beweis des Satzes über die Umkehrfunktion.
Beweis. Wir setzen
$$\lambda := \frac{1}{2}\|Df(p)^{-1}\|^{-1} > 0.$$
Es ist λ die einzige relevante Größe in unserem Beweis!
Da f stetig differenzierbar ist, existiert ein $\widetilde{\delta} > 0$ mit
$$\|Df(x) - Df(p)\| \leq \lambda$$
für alle $x \in U$ mit $|x - p| \leq \widetilde{\delta}$. Da U offen ist, gibt es ein $\delta > 0$ mit $\delta < \widetilde{\delta}$ und $U_\delta(p) \subseteq U$. Setze
$$V := U_\delta(p) \subseteq U.$$
Dann gilt also
$$\|Df(x) - Df(p)\| \leq \lambda \ \ (*)$$
für alle $x \in V$.
Zwischenbehauptung: Für alle $x, x' \in V$ gilt $|f(x) - f(x')| \geq \lambda|x - x'|$.
Bew. Sei $\varphi(x) = f(x) - Df(p)(x - p) - f(p)$, also
$$f(x) = f(p) + Df(p)(x - p) + \varphi(x).$$

Dann gilt:

$$\begin{array}{rcl} |f(x) - f(x'| & = & |Df(p)(x - x') + \varphi(x) - \varphi(x')| \\ (\Delta - Ugl) & \geq & |Df(p)(x - x')| - |\varphi(x) - \varphi(x')| \\ (Lemma) & \geq & 2\lambda|x - x'| - |\varphi(x) - \varphi(x')| \\ (!) & \geq & 2\lambda|x - x'| - \lambda|x - x'| \\ & = & \lambda|x - x'|. \end{array}$$

(!: Das folgt aus dem Mittelwertsatz:

$$\begin{array}{rcl} |\varphi(x) - \varphi(x')| & \leq & \|D\varphi(\xi)\||x - x'| \\ & = & \|Df(\xi) - Df(p)\||x - x'| \\ (\xi \in V,\ Def\ V) & \leq & \lambda|x - x'|.) \end{array}$$

Nach diesen Vorbereitungen können wir nun die Aussagen des Satzes (in anderer Reihenfolge) beweisen.
Behauptung. Für alle $x \in V$ ist $Df(x)$ invertierbar.
Bew. Es gilt

$$\begin{array}{rcl} |Df(x)y| \geq |Df(p)y| - |(Df(x) - Df(p))y| & & \\ (Lemma, \text{Def. } V, \lambda) & \geq & 2\lambda|y| - \lambda|y| \\ & = & \lambda|y| \end{array}$$

Damit folgt die Aussage nach dem Lemma zur Invertierbarkeit.
Behauptung. $W := f(V)$ ist offen.
Bew. Sei $y' := f(x') \in W$ beliebig. Sei $\delta' := \delta - |x' - p|, \varepsilon := \frac{\lambda}{2}\delta'$. Wir zeigen $U_\varepsilon(y') \subseteq W$. **Zeichnung.**
Sei also $z \in U_\varepsilon(y')$ beliebig. Damit folgt also

$$z \in U_{\frac{\lambda r}{2}}(y') \ (*)$$

für ein $0 < r < \delta'$. Wir betrachten

$$\psi : B_r(x') \longrightarrow [0, \infty), \psi(x) = |f(x) - z|^2.$$

Da ψ (offenbar) stetig ist, nimmt es auf dem kompakten $B_r(x')$ sein Minimum an. (Wir müssen zeigen, dass dieses Minimum 0 ist.) Wir zeigen nun, dass das Minimum im Inneren angenommen wird: Für x auf dem Rand von $B_r(x')$ gilt $|x - x'| = r$ und damit

$$\begin{array}{rcl} \psi(x) = |f(x) - z|^2 & \geq & (|f(x) - f(x')| - |y' - z|)^2 \\ (ZB, (*)) & \geq & (\lambda r - \frac{1}{2}\lambda r)^2 \\ & = & (\frac{1}{2}\lambda r)^2. \end{array}$$

Für den Mittelpunkt x' von $B_r(x')$ gilt aber

$$\psi(x') = |f(x') - z|^2 = |y' - z|^2 \overset{(*)}{<} (\frac{1}{2}\lambda r)^2.$$

Damit wird also das Minimum im Inneren von $B_r(x')$ angenommen. Dann muss aber nach dem vorherigen Lemma (mit $h = f - z$) für den entsprechenden Punkt x_m gelten

$$0 = Df(x_m)^t(f(x_m) - z).$$

Da $Df(x_m)$ invertierbar ist (nach vorherigen Behauptung) folgt also $f(x_m) - z = 0$. Das liefert die gewünschte Offenheit.
Behauptung. $f : V \longrightarrow W$ ist bijektiv.
Bew. Nach Definition von W ist f surjektiv. Nach ZB ist f injektiv.

Ende 20. Vorlesung

Behauptung. Es ist $g = (f|_V)^{-1}$ differenzierbar mit $Dg(y) = Df(x)^{-1}$ für $f(x) = y$.
Bew. Wir betrachten nur den Punkt $y = q = f(p)$. (In den anderen Punkten von V können analoge Betrachtungen gemacht werden, da Differenzierbarkeit eine lokale Eigenschaft ist.) Sei $\varepsilon = \frac{\lambda}{2}\delta$ wie oben zu $x' = p$ gewählt. Dann gilt also $U_\varepsilon(q) \subseteq W$, $U_\delta(p) \subseteq V$. **Zeichnung.** Für $y \in \mathbb{R}^N$ mit $|y| < \varepsilon$ gilt dann

$$g(q + y) = p + x$$

mit $|x| < \delta$. Insgesamt hat mal also

$$g(q + y) = p + x, \; g(q) = p, \quad f(p + x) = q + y, \; f(p) = q.$$

Wir können nun abschatzen

$$|y| = |(y + q) - q| = |f(p + x) - f(p)| \overset{ZB}{\geq} \lambda|x| \;\; (**).$$

Weiterhin gilt

$$f(p + x) = f(p) + Df(p)x + \varphi_f(p + x)$$

also

$$q + y = q + Df(p)x + \varphi_f(p + x) \;\; (***)$$

mit

$$\frac{\varphi(p + x)}{|x|} \to 0, x \to 0.$$

Damit erhält man insgesamt:

$$\begin{aligned}
\frac{\varphi_g(q+y)}{|y|} &= \frac{1}{|y|}|g(q+y) - g(q) - Df(p)^{-1}y| \\
&= \frac{1}{|y|}|x - Df(p)^{-1}y| \\
&= \frac{1}{|y|}|Df(p)^{-1}(Df(p)x - y)| \\
(***) \quad &= \frac{1}{|y|}|Df(p)^{-1}\varphi_f(p+x)| \\
&\leq \frac{1}{|y|}\|Df(p)^{-1}\||\varphi_f(p+x)| \\
(**) \quad &\leq \frac{\|Df(p)^{-1}\|}{\lambda|x|}|\varphi_f(p+x)| \\
&= \frac{\|Df(p)^{-1}\|}{\lambda}\frac{|\varphi_f(p+x)|}{|x|} \\
&\to 0, |x| \to 0.
\end{aligned}$$

Für $|y| \to 0$ folgt aber aus $(**)$ sofort $|x| \to 0$. Damit ergibt sich aus der vorherigen Abschätzung

$$\frac{\varphi_g(q+y)}{|y|} \to 0, \ |y| \to 0.$$

Das beendet den Beweis. $\square$

In der bisherigen Diskussion haben wir versucht, die Gleichung

$$g(x,y) = f(x) - y = 0$$

bei gegebenen y nach x aufzulösen, d.h. x als Graph einer Funktion von y darzustellen. Wir fragen nun, wieweit das für allgemeine Funktionen g möglich ist. Dazu vertauschen wir die Rollen von x und y und betrachten $g : \mathbb{R}^N \times \mathbb{R}^M \longrightarrow \mathbb{R}^M$ und seine Nullstellenmenge

$$\mathcal{N} = \mathcal{N}(g) := \{(x,y) \mid g(x,y) = 0\}.$$

Die Frage lautet, ob $\mathcal{N}$ lokal als Graph darstellbar ist d.h. ob man y als Funktion von x schreiben kann $y = g(x)$. Es geht also um das Auflösen der Gleichung $g(x,y) = 0$ nach y als Funktion von x.

Beispiel (Höhenlinien einer Landkarte)). Sei g : Landkarte $\longrightarrow \mathbb{R}$, $g(p)$ = Höhe von p über dem Meeresspiegel. Sei für $c \in \mathbb{R}$

$$\mathcal{N}_c := \{p \in \text{Landkarte} : g(p) - c = 0\}.$$

Dann ist $\mathcal{N}_c$ gerade die Höhenlinie zu c. **Zeichnung**

Beispiel (Einheitskreislinie). Sei $g : \mathbb{R}\times\mathbb{R} \longrightarrow \mathbb{R}$, $g(x,y) = x^2+y^2-1$.

$$\mathcal{N} := \{(x,y) \in \mathbb{R}^2 \mid x^2 + y^2 = 1\}.$$

Dann gilt: **Zeichnung.**
$\mathcal{N} \cap \{y > 0\}$: Oberer Halbkreis $y = \sqrt{1 - x^2}$.
$\mathcal{N} \cap \{x < 0\}$: Linker Halbkreis $x = -\sqrt{1 - y^2}$.
$\mathcal{N} \cap \{y < 0\}$: Unterer Halbkreis $y = -\sqrt{1 - x^2}$.
$\mathcal{N} \cap \{x > 0\}$: Rechter Halbkreis. $x = \sqrt{1 - y^2}$.
Die Gleichung kann also lokal nach x bezw y aufgelöst werden. Das Auflösen nach y ist dabei in Bereichen mit $\partial_y g(x, y) = 2y \neq 0$ möglich und das Auflösn nach x in Bereichen mit $\partial_x g(x, y) = 2x \neq 0$.

NOTATION. Für $z \in \mathbb{R}^N \times \mathbb{R}^M$ schreiben wir $z = (x, y)$ mit $x \in \mathbb{R}^N$ und $y \in \mathbb{R}^M$. Ist $W \subseteq \mathbb{R}^N \times \mathbb{R}^M$ offen und $g : W \longrightarrow \mathbb{R}^M$ differenzierbar, so setzen wir

$$D_x g(z) := (\partial_j g_i(z))_{i=1,\dots,M, j=1,\dots,N}$$

$$D_y g(z) := (\partial_{j+N} g_i(z))_{i=1,\dots,M, j=1,\dots,M}.$$

Damit gilt dann also

$$Dg(z) = (D_x g(z)\ D_y g(z)) = \text{Matrix malen.}$$

Wichtig. $D_y g(z))$ ist eine quadratische Matrix ($M \times M$ Matrix).

THEOREM (Satz über implizite Funktionen). *Sei $W \subseteq \mathbb{R}^N \times \mathbb{R}^M$ offen und $g : W \longrightarrow \mathbb{R}^M$ stetig differenzierbar. Sei*

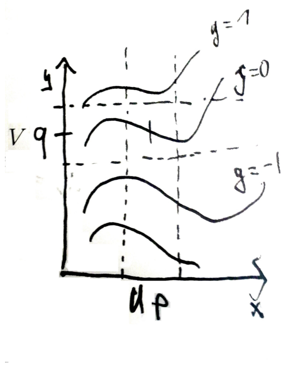
$$\mathcal{N} := \{(x, y) \in W \mid g(x, y) = 0\}.$$

Ist $z_0 := (p, q) \in \mathcal{N}$ und $D_y g(z_0)$ invertierbar, so existieren offene Umgebungen U von p und V von q und ein stetig differenzierbares $f : U \longrightarrow V$ mit

$$\mathcal{N} \cap (U \times V) = \textit{Graph von } f = \{(x, f(x)) \mid x \in U\}$$

sowie $Df(x) = -D_y g(x, f(x))^{-1} D_x g(x, f(x))$.

Kurz: Ist $g(p, q) = 0$ und $D_y g(p, q)$ invertierbar, so kann man die Lösung von $g(x, y) = 0$ in einer Umgebung von p als Graph darstellen.



Ende 21. Vorlesung

Beweis. Wir zeigen zunächst die Existenz von f.

Idee (vgl. Zeichnung). Nach dem Satz zur Umkehrfunktion ist um (p, q) herum ist jeder Punkt eindeutig durch seine x Koordinate und den Wert von g beschrieben. Daher können wir (x, g) als Koordinatensystem verwenden. Das liefert eine Zerlegung der Umgebung von (p, g) in Konstanzflächen von g. Diese sind nach Konstruktion (Bilder von) Graphen.
Hier sind die Details: Sei $G : W \longrightarrow \mathbb{R}^N \times \mathbb{R}^M$, $(x, y) \mapsto (x, g(x, y))$. Dann gilt

$$G(p, q) = (p, 0)$$

sowie

$$DG(p, q) \equiv \begin{pmatrix} 1_N & 0_{NM} \\ D_x g(p, q) & D_y g(p, q) \end{pmatrix}.$$

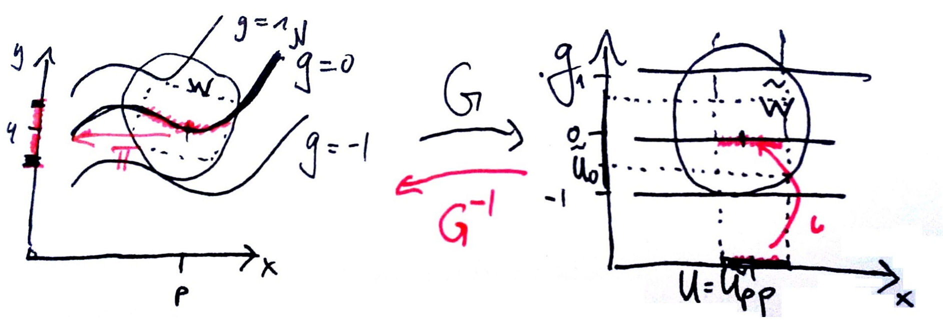
Damit folgt

$$\det DG(p, q) = \det D_y g(p, q) \neq 0$$

nach unserer Vorausetzung an die Invertierbarkeit von $D_y g(p, q)$. Nach dem Satz über die Umkehrfunktion gibt es also offenen Umgebungen W_0 von (p, q) und $\tilde{W}_0$ von $G(p, q) = (p, 0)$, so dass $G : W_0 \longrightarrow \tilde{W}_0$ bijektiv mit stetig differenzierbarer Inverser ist. Ohne Einschränkung (sonst Verkleinern) können wir

$$\tilde{W}_0 = \tilde{U}_p \times \tilde{U}_0$$

mit Umgebungen $\tilde{U}_p$ von p und $\tilde{U}_0$ von 0 annehmen.
Weiteres Vorgehen in Worten: Wir können $\tilde{W}_0$ offenbar in Konstanzflächen (der zweiten Koordinate) zerlegen. Zurückziehen mit G liefert dann eine Zerlegung von $\mathcal{N} \cap W_0$ in Konstanzflächen von g.



Nach Konstruktion sind diese Konstanzflächen aber gerade Graphen.
Hier sind die Details:
Sei

$$\pi : \mathbb{R}^N \times \mathbb{R}^M \longrightarrow \mathbb{R}^M, \begin{pmatrix} x \\ y \end{pmatrix} \mapsto y = (0, 1_M) \begin{pmatrix} x \\ y \end{pmatrix},$$

$$\iota : \mathbb{R}^N \longrightarrow \mathbb{R}^N \times \mathbb{R}^M, x \mapsto \begin{pmatrix} x \\ 0 \end{pmatrix} = \begin{pmatrix} 1_N \\ 0 \end{pmatrix} x.$$

Mit $U = \tilde{U}_p$ und $V = \pi \circ G^{-1} \circ \iota(U)$ hat

$$f = \pi \circ G^{-1} \circ \iota : U \longrightarrow V$$

die gewünschten Eigenschaften.

- ι bildet U auf $U \times \{0\}$ ab, $x \mapsto (x, 0)$.
- G^{-1} bildet $U \times \{0\}$ ab nach $U \times V$, $(x, 0) \mapsto (x, y)$ mit $g(x, y) = 0$.
- π bildet $U \times V$ nach V ab $(x, y) \mapsto y$, wobei für die in Frage kommenden (x, y) gilt $g(x, y) = 0$.

Es bleibt die Ableitung zu berechnen. Nach Kettenregel ist f differenzierbar und es gilt

$$Df(x) = D\pi(G^{-1}(\iota(x))DG^{-1}(\iota(x))D\iota(x).$$

Es ist (beachte π, ι linear):

$$D\iota = \iota, D\pi = \pi.$$

Weiterhin gilt

$$DG(x,y) \equiv \begin{pmatrix} 1_N & 0_{NM} \\ D_x g(x,y) & D_y g(x,y) \end{pmatrix} =: \begin{pmatrix} 1_N & 0_{NM} \\ L_1 & L_2 \end{pmatrix}$$

mit $L_1 = D_x g(x, y)$ und $L_2 = D_y g(x, y)$. Damit ergibt sich das Inverse (Probe!) zu

$$DG^{-1} = \begin{pmatrix} 1_N & 0_{NM} \\ -L_2^{-1} L_1 & L_2^{-1} \end{pmatrix}.$$

Insgesamt folgt

$$Df(x) = (0, 1_M) \begin{pmatrix} 1_N & 0_{NM} \\ -L_2^{-1} L_1 & L_2^{-1} \end{pmatrix} \begin{pmatrix} 1_N \\ 0 \end{pmatrix} = -L_2^{-1} L_1$$

und das ist die Behauptung. □

Eine Reihe von Kommentaren sind angemessen.

Bemerkung.

- Sind f, g differenzierbar mit $g(x, f(x)) \equiv 0$ und $D_y g(x, f(x))$ invertierbar, so folgt aus der Kettenregel

$$0 \equiv Dg(x, f(x)) \begin{pmatrix} 1_M \\ Df(x) \end{pmatrix} = D_x g(x, f(x)) + D_y g(x, f(x)) Df(x).$$

  Damit ergibt sich die Formel für die Ableitung von f

$$Df(x) = -D_y g(x, f(x))^{-1} D_x g(x, f(x)).$$

  Auch hier folgt also die Formel für die Ableitung von f automatisch aus der Differenzierbarkeit von f und der Kettenregel.
- Im Allgemeinen werden wir die Funktion f aus dem Satz nicht explizit bestimmen können. Es ist bemerkenswert, dass wir trotzdem die Ableitung von f berechnen können.

- Ist $D_y g(p,q)$ nicht invertierbar, so ist i.a. eindeutiges Auflösen von y nach x in Umgebung von p nicht möglich:

  Beispiel (kein eindeutiges Auflösen möglich):
  $$g : \mathbb{R}^2 \longrightarrow \mathbb{R}, g(x,y) = x^2 + y^2 - 1, (p,q) = (1,0).$$
  Beispiel (gar kein Auflösen möglich)
  $$h : \mathbb{R}^2 \longrightarrow \mathbb{R}, g(x,y) = x^2 + y^2, (p,q) = (0,0).$$
- Statt $g \equiv 0$ kann man auch $g \equiv c$ betrachten. (Dazu ersetzt man einfach g durch $\tilde{g} = g - c$...)
- Ist $W \subseteq \mathbb{R}^N \times \mathbb{R}^M$ offen und $g : W \longrightarrow \mathbb{R}^M$ stetig differenzierbar mit $g(z_0) = 0$ und $\mathrm{Rang} Dg(z_0) = M$, so kann man den Satz nach Umsortieren der Koordinaten anwenden. (Vgl. Beispiel mit Einheitskreis.)
- Wir haben den Satz über implizite Funktionen mittels des Satzes über die Umkehrfunktion gezeigt. Man kann auch umgekehrt vorgehen. Dazu betrachtet man $g(x,y) = f(x) - y$ in einer Umgebung von (p,q) mit $f(p) = q$ d.h. $g(p,q) = 0$. Ist $D_x f(p)$ invertierbar, so ist $D_x g(p,q)$ invertierbar und damit nach dem Satz über implizite Funktionen also x nahe an p als Funktion von y nahe q darstellbar. (Details: Übung).

KAPITEL 4

Untermannigfaltigkeiten, bedingte Extrema und all das

Untermannigfaltigkeiten sind die glatten Gebilde im Raum. Sie sind lokal Nullstellenmengen von glatten Funktionen oder äquivalent lokale Graphen.
Zeichnungen. Kreis, Sphäre, Torus, Parabel.

NOTATION. Die Nullstellenmenge einer Funktion g bezeichnen wir weiterhin mit $\mathcal{N}(g)$, d.h. zu $g : W \subseteq \mathbb{R}^N \longrightarrow \mathbb{R}^L$ definieren wir

$$\mathcal{N}(g) := \{z \in W \mid g(z) = 0\} = \cap_{j=1}^{L} \mathcal{N}(g_j).$$

Den Graphen einer Funktion f bezeichen wir mit $\mathcal{G}(f)$ d.h. zu $f : U \subseteq \mathbb{R}^N \longrightarrow \mathbb{R}^M$ definieren wir

$$\mathcal{G}(f) := \{(x, f(x)) \mid x \in U\} \subseteq \mathbb{R}^N \times \mathbb{R}^M.$$

DEFINITION. Eine stetig differenzierbare Funktion $f : U \subseteq \mathbb{R}^N \longrightarrow \mathbb{R}^M$ heißt *regulär in* x, wenn $Df(x)$ maximalen Rang hat. Ist die Funktion in jedem $x \in U$ regulär, so heißt sie *regulär.*

BEMERKUNG. Ist $f : U \subseteq \mathbb{R}^N \longrightarrow \mathbb{R}^M$ regulär, so gilt also

- falls $M \leq N$: Rang$Df(x) = M$ d.h. Zeilen von $Df(x)$ sind linear unabhängig.
- falls $N \leq M$: Rang$Df(x) = N$ d.h. Spalten von $Df(x)$ sind linear unabhängig.
- falls $N = M$: $Df(x)$ invertierbar.

Wie man am letzten Punkt sieht, ist Regularität eine Verallgemeinerung von Invertierbarkeit.

BEISPIEL.
- Ist $g : W \longrightarrow \mathbb{R}$ stetig differenzierbar, so ist g genau dann regulär, wenn $\nabla g(x) \neq 0$ in jedem $x \in W$.
- $U \subseteq \mathbb{R}^N$ offen, $f : U \longrightarrow \mathbb{R}^M$ stetig differenzierbar. Dann ist $F : U \longrightarrow \mathbb{R}^N \times \mathbb{R}^M$, $F(x) = (x, f(x))$ regulär. (Denn $DF(x) = (1, Df(x))^t$ und damit Rang$DF = N$.)

Ende 22. Vorlesung

LEMMA. *Sei $X \subseteq \mathbb{R}^N$, $p \in X$ und $L \leq N$. Dann sind äquivalent:*

(i) *Es existiert eine offene Umgebung W von p in $\mathbb{R}^N$ und reguläres $g : W \longrightarrow \mathbb{R}^{N-L}$ mit $X \cap W = \mathcal{N}(g)$. („$X$ ist lokal eine Nullstellenmenge.“)*

(ii) *Es existiert eine offenen Umgebung W von $p \in \mathbb{R}^N$ und eine offenen Menge $U \subseteq \mathbb{R}^L$ und eine stetig differenzierbare Funktion $f : U \longrightarrow \mathbb{R}^{N-L}$ und eine Permutation $P : \mathbb{R}^N \longrightarrow \mathbb{R}^N$ mit*
$$X \cap W = P\mathcal{G}(f).$$
(„X ist lokal Graph.“)

Beweis. Die Richtung (i) $\Longrightarrow$ (ii) folgt aus dem Satz über implizite Funktionen. Die Implikation (ii) $\Longrightarrow$ (i) folgt einfach durch Betrachten von $g(x,y) = y - f(x)$ (wenn man $P = Id$ annimmt). □

Definition (Untermannigfaltigkeit). Eine Teilmenge X von $\mathbb{R}^N$ heißt *Untermannigfaltigkeit der Dimension L*, wenn sie in jedem Punkt $p \in X$ eine der Bedingungen des vorangehenden Lemma erfüllt.

Bemerkung.
- Untermannigfaltigkeiten der Dimension $N-1$ heißen Hyperflächen. Sie sind also (lokal) Nullstellengebilde von reellwertigen Funktionen.
- Alle Untermannigfaltigkeiten sind lokal Schnitte von Hyperflächen (da $\mathcal{N}(g) = \cap_k^{N-L}\mathcal{N}(g_k)$).

Beispiel. Offenbar sind alle Graphen und alle Nullstellengebilde regulärer Funktionen also Untermannigfaltigkeiten. Insbesondere gehören dazu:
- Kreis ($g(x,y) = x^2 + y^2 - 1 = 0$)
- Die Sphäre $S^{N-1} := \{x \in \mathbb{R}^N \mid |x|^2 - 1 = 0\}$ d.h. $g(\xi_1, \ldots, \xi_N) = \sum_{j=1}^N x_j^2 - 1 = 0$.
- Torus (Übung)
- Parabel (Graph).

Wir führen nun die Tangentialflächen und Normalen an Untermannigfaltigkeiten ein. **Zeichnung.**

Definition. Sei $X \subseteq \mathbb{R}^N$ eine Untermannigfaltigkeit und $p \in X$. Dann heißt
$$T_pX := \{\gamma'(0) \mid \gamma : (-1,1) \longrightarrow X \text{stetig differenzierbar},\ \gamma(0) = p,$$
der *Tangentialraum* von X in p. Der *Normalraum* N_pX von X in p ist definiert als
$$N_pX := T_pX^{\perp}.$$

Bemerkung. Man könnte auch Kurven γ auf $(-\varepsilon_\gamma, \varepsilon_\gamma)$ betrachten.

Theorem (Beschreibung Tangentialraum und Normalraum). *Sei $X \subseteq \mathbb{R}^N$ eine Untermannigfaltigkeit der Dimension L. Sei $p \in X$ und $W \subseteq \mathbb{R}^N$ offen mit*
$$p \in X \cap W = \mathcal{N}(g) = \mathcal{G}(f) = \mathrm{Bild}(F),$$

wobei $U \subseteq \mathbb{R}^L$ *offen,* $f : U \longrightarrow \mathbb{R}^{N-L}$ *stetig differenzierbar,* $F : U \longrightarrow \mathbb{R}^N$*,* $F(x) = (x, f(x))$*,* $g : W \longrightarrow \mathbb{R}^{N-L}$ *stetig differenzierbar und regulär. Dann gilt*

$$T_pX = \mathrm{Ker} Dg(p) = \mathrm{Bild} DF(p_0)$$

für $p = F(p_0) = (p_0, f(p_0))$*. Inbesondere gilt*

$$T_pX = \bigcap_{j=1}^{N-L} \{\nabla g_j(p)\}^\perp,$$
$$N_pX = \mathrm{lin}\{\nabla g_j(p) \mid j = 1, \dots, N-L\}$$

sowie $\dim T_pX = L$*,* $\dim N_pX = N - L$*.*

Bemerkung.
- Es zeigt ∇g_i in Richtung des stärksten Wachstum von g_i. In Richtung von T_pX gibt es umgekehrt kein Wachstum der g_i (da Nullstellenmenge). Daher ist der Normalraum gerade durch die Gradienten der g_i aufgespannt. **Zeichnung**.
- Der Satz zeigt, dass T_pX ein Unterraum ist.

Beweis. $T_pX \subseteq \mathrm{Ker} Dg$: Sei $v \in T_pX$ beliebig und γ zugehörige Kurve d.h $\gamma : (-1, 1) \longrightarrow X$, $\gamma(0) = p$, $\gamma'(0) = v$. Dann gilt $g \circ \gamma \equiv 0$ also

$$0 = (g \circ \gamma)'(0) = Dg(\gamma(0))\gamma'(0) = Dg(p)v.$$

$\mathrm{Bild} DF \subseteq T_pX$: Sei $w \in \mathbb{R}^L$ beliebig, $\gamma : (-1, 1) \longrightarrow X$, $\gamma(t) = F(p_0 + tw)$. Dann gilt

$$DF(p_0)w = \gamma'(0) \in T_pX.$$

Damit folgt die Aussage.
Insgesamt zeigt dies

$$\mathrm{Bild} DF(p_0) \subseteq T_pX \subseteq \mathrm{Ker} Dg(p).$$

Weiterhin gilt

$$\dim \mathrm{Ker} D(g) = L$$

(nach Dimensionsformel $N = \dim \mathrm{Ker} Dg + \dim \mathrm{Bild} Dg$ und g regulär also $\dim \mathrm{Bild} Dg = N - L$) und

$$\mathrm{Rang} \mathrm{Bild} DF(p_0) = L$$

(klar, betrachte Ableitung). Damit folgt dann

$$\mathrm{Bild} DF(p_0) = T_pX = \mathrm{Ker} Dg(p).$$

$T_pX = \bigcap_{j=1}^{N-L} \{\nabla g_j(p)\}^\perp$*:* Das folgt sofort aus $T_pX = \mathrm{Ker} Dg(p)$.
$N_pX = \mathrm{lin}\{\nabla g_j(p) \mid j = 1, \dots, N-L\}$*:* Es gilt

$$\mathrm{lin}\{\nabla g_j(p) \mid j = 1, \dots, N-L\}^\perp = \bigcap_{j=1}^{N-L} \{\nabla g_j(p)\}^\perp = T_pX.$$

Damit folgt dann durch Bilden des orthogonalen Komplementes die Aussage. □

Bemerkung (Achtung!). Die Gleichheit $T_pX = \mathrm{Ker}Dg = \mathrm{Bild}DF$ ist gerade eine linearisierte Version von $X = \mathcal{N}(g) = \mathrm{Bild}F$. Ähnlich ist $Df(x)$ eine linearisierte Version von f. Tatsächlich kann man für $f : U \subseteq \mathbb{R}^N \longrightarrow V \subseteq \mathbb{R}^M$ die Ableitung $Df(x)$ deuten als eine Abbildung des Tangentialraumes $\mathbb{R}^N = T_xU$ in den Tangentialraum $\mathbb{R}^M = T_{f(x)}V$.

Ende 23. Vorlesung

Betrachtung von Untermannigfaltigkeiten ist in vielerlei Hinsicht nützlich. Wir lernen jetzt eine Anwendung auf sogenannte bedingte Extrema kennen. Diese ist unter dem Namen „Methode der Lagrange Multiplikatoren“ bekannt.

Definition. Sei $U \subseteq \mathbb{R}^N$ offen, $X \subseteq U$, $f : U \longrightarrow \mathbb{R}$, $p \in X$. Es nimmt f in p ein *bedingtes lokales Maximum/Minimum* auf X an, wenn ein $\delta > 0$ existiert mit

$$f(x) \leq f(p)/f(x) \geq f(p)$$

für alle $x \in X$ mit $|x - p| < \delta$.

Da X i.a. nicht offen ist, können die bisher entwickelten Methoden zur Untersuchung von bedingten Extrema nicht angewendet werden. Wir untersuchen nun den Fall, dass X eine Nullstellenmenge ist.

Theorem. *Seien $K \leq N$ aus $\mathbb{N}$ gegeben. Sei $U \subseteq \mathbb{R}^N$ offen, $g : U \longrightarrow \mathbb{R}^K$ stetig differenzierbar, $L := N - K$,*

$$X := \mathcal{N}(g) = \bigcap_{i=1}^{K} \mathcal{N}(g_i).$$

Sei p ein regulärer Punkt von g also $\mathrm{Rang}Dg(p) = K$ *(d.h. $\nabla g_i(p)$, $i = 1, \ldots, K$, linear unabhängig). Sei $f : U \longrightarrow \mathbb{R}$ stetig differenzierbar. Hat f in p ein bedingtes lokales Extremum auf X, so gibt es reelle Zahlen $\lambda_1, \ldots, \lambda_K$ mit*

$$\nabla f(p) = \sum_{j=1}^{K} \lambda_j \nabla g_j(p).$$

Ist insbesondere X ein Untermannigfaltigkeit (d.h. jeder Punkt von X ist regulär), so gilt $\nabla f(p) \in N_pX = T_pX^\perp$.

Interpretation. Zeichnung. Hätte $\nabla f(p)$ eine Komponente in Richtung von T_pX so könnte man auf X in dieser Richtung e wandern und dabei den Wert von f wegen

$$\partial_e f(p) = \langle \nabla f(p), e \rangle$$

vergrößern und durch Wandern in Gegenrichtung verkleinern. In einem Extremum kann also $\nabla f(p)$ keine Komponente in Richtung von T_pX

haben, muss also in N_pX liegen. Es ist aber N_pX die lineare Hülle der Gradienten der g_i (nach dem vorherigen Satz).

Beweis. Ohne Einschränkung ist X UMK (sonst Verkleinern). Ist $\gamma : (-\varepsilon, +\varepsilon) \longrightarrow X$ eine stetig differenzierbare Kurve in X mit $\gamma(0) = p$, so hat

$$h : (-\varepsilon, \varepsilon) \longrightarrow \mathbb{R},\ h(t) = f(\gamma(t))$$

in 0 ein lokales Extremum. Daher muss gelten (**Zeichnung**.)

$$0 = h'(0) = \langle \nabla f(p), \gamma'(0) \rangle.$$

Da γ beliebige Kurve in X durch p war, folgt $\nabla f(p) \perp T_P X$, also $\nabla f(p) \in N_pX$. Nach dem vorangehenden Satz folgt die Aussage. □

Verfahren/Methode der Lagrange Multiplikatoren. Um die Punkte zu finden, in denen *möglicherweise* Extrema vorliegen, hat man das Gleichungssystem $(*)$

$$g_j(p) = 0, j = 1, \ldots, K,\ \ (K - \mathit{Gleichungen})$$

$$\nabla f(p) = \sum_{j=1}^{K} \lambda_j \nabla g_j(x)\ \ (N - \mathit{Gleichungen})$$

nach $p = (p_1, \ldots, p_N)$ und $(\lambda_1, \ldots, \lambda_K)$ aufzulösen. Das sind $K + N$ Gleichungen und $K + N$ Unbekannte.

Bemerkung. Man erhält das Gleichungssystem $(*)$ auch durch Betrachten von

$$F : U \times \mathbb{R}^K \longrightarrow \mathbb{R},\ \ F(x, \lambda_1, \ldots, \lambda_K) = f(x) - \sum_{j=1}^{K} \lambda_j g_j(x)$$

und Bilden des Gradienten

$$\nabla F(p, \lambda) = (\nabla f(p) - \sum_{j=1}^{L} \lambda_j \nabla g_j(p), g_1(p), \ldots, g_K(p))^\top$$

und Nullsetzen des Gradienten. Das ist unter dem Stichwort „Ankoppeln der Nebenbedingungen“ bekannt. Die sich ergebenenden λ_j heißen Lagrangemultiplikatoren. In Beispielen können sie eine konkrete Interpretation haben (Zwangskräfte, Schattenpreise....).

Bemerkung. Die gesuchte Menge der bedingten Extrema und die berechnete Menge, in der $(*)$ gilt, können sehr verschieden sein: Seien

$$M_1 := \{\text{Extrempunkt von } f \text{ auf } X\}.$$

$$M_2 := \{\text{Extrempunkt von } f \text{ auf } X, \text{ in denen } Dg \text{ vollen Rang hat}\}.$$

$$M_3 := \{x \in X \mid (*)\}.$$

Dann gilt

$$M_1 \supseteq M_2 \subseteq M_3.$$

Es gilt:

- M_1 nicht leer, falls X kompakt.
- Erste Inklusion ist Gleichheit, falls Dg überall auf X vollen Rang hat, also X eine Untermannigfaltigkeit ist.
- Es können in M_3 Punkte dazukommen.

Folgerung. *Sei $U \subseteq \mathbb{R}^N$ offen, $f : U \longrightarrow \mathbb{R}$ stetig differenzierbar und $X \subseteq U$ eine kompakte Untermannigfaltigkeit. Dann nimmt f sein Maximum/Minimum auf X an und zwar in denjenigen Punkten von*

$$\{p \in X \mid \nabla f(p) \in N_p X\}$$

in denen es maximalen/minimalen Funktionswert hat.

Beispiel. $U = \mathbb{R}^2$, $g(\xi_1, \xi_2) = \xi_1^2 + \xi_2^2 - 1$, $X = \mathcal{N}(g) = S$. Gesucht Extrema von $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$, $f(x) = \xi_1 + 2\xi_2$ auf X.
Da f stetig und X kompakt ist, nimmt f auf X auf jeden Fall Minimum und Maximum an. Wir berechnen nun Gradienten von f und g. Es gilt

$$\nabla f(x) = (1, 2)^t, \ \ \nabla g(x) = 2(\xi_1, \xi_2)^t.$$

Insbesondere gilt $\nabla g \neq 0$ auf ganz X und X ist eine UMK. Auflösen von $(*)$ liefert dann die Extrempunkte von f auf M (und möglicherweise noch mehr Punkte). Es wird $(*)$ zu

$$\xi_1^2 + \xi_2^2 - 1 = 0 \quad (1, 2) - \lambda 2(\xi_1, \xi_2) = 0.$$

Auflösen liefert (nach endlicher Rechnung):

$$\lambda = \pm\frac{\sqrt{5}}{2}, \ \ (\xi_1, \xi_2) = \pm(\frac{1}{\sqrt{5}}, \frac{2}{\sqrt{5}}).$$

Es gilt $f(\frac{1}{\sqrt{5}}, \frac{2}{\sqrt{5}}) = \sqrt{5}$ (Maximum) und $f(-\frac{1}{\sqrt{5}}, \frac{2}{\sqrt{5}}) = -\sqrt{5}$ (Minimum).

Ende 24. Vorlesung

KAPITEL 5

Etwas Maß- und Integrationstheorie

Das Konzept des Maßes stellt eine präzise (und sehr allgemeine) Fassung von Volumenmessung zur Verfügung. Die charakteristische Eigenschaft ist
$$\mu(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mu(A_n)$$
für paarweise disjunkte A_n, $n \in \mathbb{N}$. Basierend auf einem Maß können dann Funktionen integriert werden. Im allgemeinen kann nicht allen Mengen ein Maß zugeordnet werden, und es können nicht alle Funktionen integriert werden. Die „guten" Mengen bzw. Funktionen heißen *messbar*.

1. σ-Algebren und messbare Funktionen

In diesem Abschnitt lernen wir die „guten" Mengen und Funktionen kennen.

Definition (σ-Algebra). Sei X eine Menge und $\mathcal{A}$ eine Familie von Teilmengen von X. Dann heißt $\mathcal{A}$ eine *σ-Algebra*, wenn folgende Eigenschaften erfüllt sind:

- X gehört zu $\mathcal{A}$.
- Mit A gehört auch $X \setminus A$ zu $\mathcal{A}$.
- Mit A_n, $n \in \mathbb{N}$, gehört auch $\bigcup_n A_n$ zu $\mathcal{A}$.

Ist $\mathcal{A}$ eine σ-Algebra auf X, so heißt $(X, \mathcal{A})$ ein *messbarer Raum* und die Mengen von $\mathcal{A}$ werden *messbar* genannt.

Eine σ-Algebra ist also abgeschlossen unter Komplementbildung und abzählbaren Vereinigungen und enthält jedenfalls den ganzen Raum.

Bemerkung. Es ist $\{X, \emptyset\}$ eine σ-Algebra auf X und in jeder anderen enthalten.

Proposition (Einfache Eigenschaften). *Sei $(X, \mathcal{A})$ ein messbarer Raum. Dann gilt:*

(a) $A_n \in \mathcal{A}$, $n \in \mathbb{N}$, $\Longrightarrow \bigcap_{n=1}^{\infty} A_n \in \mathcal{A}$.
(b) $A, B \in \mathcal{A} \Longrightarrow A \setminus B \in \mathcal{A}$.

Beweis. (a) Das folgt durch doppelte Komplementbildung:
$$\bigcap_n A_n = X \setminus \Big(\bigcup_n (X \setminus A_n) \Big).$$

(b) Das folgt aus $A \setminus B = A \bigcap (X \setminus B)$. □

Bemerkung. (Nur mündlich) Zum besseren Verständnis mag ein Vergleich mit dem Konzept der Topologie dienen: Eine Topologie τ auf X ist eine Familie von Teilmengen von X mit folgenden Eigenschaften:

- $\emptyset, X \in \tau$.
- Mit $U_1, \ldots, U_n$ gehört auch $\bigcap_j U_j$ zu τ.
- Mit U_α, $\alpha \in A$, gehört auch $\bigcup_\alpha U_\alpha$ zu τ.

Man kann σ-Algebren aus einer gegebene Familie von Teilmengen erzeugen.

Theorem (Erzeugung von σ-Algebren durch $\mathcal{F}$). *Sei $\mathcal{F}$ eine Familie von Teilmengen von X. Dann existiert genau eine σ-Algebra $\mathcal{A}$ auf X mit folgenden beiden Eigenschaften:*

- *$\mathcal{F} \subseteq \mathcal{A}$.*
- *Gilt $\mathcal{F} \subseteq \mathcal{B}$ für eine σ-Algebra $\mathcal{B}$, so folgt $\mathcal{A} \subseteq \mathcal{B}$.*

Beweis. Eindeutigkeit. Das ist klar. (Seien $\mathcal{A}_i$, $i = 1, 2$, solche σ-Algebren. Wähle $\mathcal{A}_2 = \mathcal{B}$ und schließe $\mathcal{A}_1 \subseteq \mathcal{A}_2$. Vertausche nun die Rollen von $\mathcal{A}_1$ und $\mathcal{A}_2$.)

Existenz. Es existiert ein σ-Algebra, die $\mathcal{F}$ enthält (z.B. die Potenzmenge von X). Der Durchschnitt aller solcher σ-Algebren ist wieder ein σ-Algebra (Check!) und enthält $\mathcal{F}$ (Check!). Er hat nach Konstruktion die gewünschten Eigenschaften. □

Definition. In der Situation des vorherigen Theorems nennt man $\mathcal{A}$ die durch $\mathcal{F}$ *erzeugte σ-Algebra* und schreibt sie auch als $\sigma(\mathcal{F})$.

Beispiel. Sei $X = \mathbb{R}$ und seien die Familien $\mathcal{F}_i$ gegeben als:
$\mathcal{F}_0$: Intervalle der Form (a, ∞), $a \in \mathbb{R}$.
$\mathcal{F}_1$: Alle Intervalle.
$\mathcal{F}_2$: Alle offenen Intervalle.
$\mathcal{F}_3$: Alle Intervalle der Form $(-\infty, a)$, $a \in \mathbb{R}$.
$\mathcal{F}_4$: Intervalle der Form $[a, \infty)$, $a \in \mathbb{R}$.
$\mathcal{F}_5$: Intervalle der Form $(-\infty, a]$, $a \in \mathbb{R}$.
$\mathcal{F}_6$: Intervalle der Form (a, b), $a \in \mathbb{R}$.
Dann gilt $\sigma(\mathcal{F}_0) = \ldots = \sigma(\mathcal{F}_6)$.
Beweis. Übung. Nutze Formeln der folgenden Art:

$$[a, b) = \bigcap_n (a - 1/n, b),$$

$$(a, b) = \bigcup_n [a + 1/n, b),$$

$$(a, b) = (a, \infty) \setminus [b, \infty) \ldots$$

□

Definition (Borel-σ-Algebra). Ist (X, d) ein metrischer Raum, so heißt die von der Familie aller offener Mengen erzeugte σ-Algebra die *Borel-σ-Algebra* auf X.

Beispiel. (a) Ist $X = \mathbb{R}$, so ist die Borel-σ-Algebra gerade die oben schon diskutierte von den offenen Intervallen erzeugte σ-Algebra. (Da jede offene Menge eine Vereinigung von abzählbar vielen offenen Intervallen ist.)
(b) Ist $X = \mathbb{R} \cup \{-\infty\} \cup \{\infty\}$, so wird durch

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

mit $\arctan(\infty) = \pi/2$ und $\arctan(-\infty) = -\pi/2$ eine Metrik auf $\mathbb{R} \cup \{-\infty\} \cup \{\infty\}$ definiert. Der metrische Raum $(\mathbb{R} \cup \{-\infty\} \cup \{\infty\}, d)$ ist kompakt und heißt Zweipunktkompaktifizierung von $\mathbb{R}$. Die Borel-σ-Algebra auf $\mathbb{R} \cup \{-\infty\} \cup \{\infty\}$ wird erzeugt z.B. durch die Familie

$$(a, \infty], \quad a \in \mathbb{R}.$$

(c) Ist $X = \mathbb{R}^N$, so wird die Borel-σ-Algebra erzeugt durch die Rechtecke der Form

$$R = (a_1, b_1) \times \ldots \times (a_N, b_N)$$

mit $a_i, b_i \in \mathbb{R}$ mit $a_i \leq b_i$ für $i = 1, \ldots, N$.

Definition (Messbarkeit). Seien $(X_i, \mathcal{A}_i)$, $i = 1, 2$, messbare Räume. Eine Funktion $f : X_1 \longrightarrow X_2$ heißt *messbar*, wenn $f^{-1}(A) \in \mathcal{A}_1$ für alle A aus $\mathcal{A}_2$ gilt.

Bemerkung. (a) Vergleiche Charakterisierung der Stetigkeit.
(b) Offenbar sind konstante Funktionen messbar (da die auftretenden Urbilder entweder leer oder der gesamte Raum sind).

Aus der Definition folgt sofort, dass die Komposition messbarer Funktionen wieder messbar ist:

Proposition (Komposition messbarer Funktionen ist messbar). *Seien $(X_i, \mathcal{A}_i)$, $i = 1, 2, 3$ messbare Räume und $f : X_1 \longrightarrow X_2$ und $g : X_2 \longrightarrow X_3$ messbar. Dann ist auch $g \circ f : X_1 \longrightarrow X_3$ messbar.*

Wesentlich ist folgende Verträglichkeit von Funktionen mit den charakteristischen Eigenschaften einer σ-Algebra.

Proposition. *Sei $(X, \mathcal{A})$ ein messbarer Raum und $f : X \longrightarrow Y$ eine Funktion. Dann ist die Menge*

$$\{B \subseteq Y : f^{-1}(B) \in \mathcal{A}\}$$

eine σ-Algebra.

Beweis. Bezeichne die genannte Menge mit $\mathcal{A}_f$. Dann gilt:

- $Y \in \mathcal{A}_f$: Klar (da $f^{-1}(Y) = X \in \mathcal{A}$).
- $B \in \mathcal{A}_f \Longrightarrow B^c \in \mathcal{A}_f$: $f^{-1}(B^c) = X \setminus f^{-1}(B) \in \mathcal{A}$ (wegen $f^{-1}(B) \in \mathcal{A}$).

- $B_n \in \mathcal{A}_f \Longrightarrow \bigcup B_n \in \mathcal{A}_f$: $f^{-1}(\bigcup_n B_n) = \bigcup_n f^{-1}(B_n) \in \mathcal{A}$ (wegen $f^{-1}(B_n) \in \mathcal{A}$).

$\square$

Übung. Untersuche entsprechende Aussage für Topologien.

LEMMA (Reicht Urbilder des Erzeuger zu testen). *Seien $(X, \mathcal{A})$ und $(Y, \mathcal{B})$ messbare Räume und $\mathcal{B}$ sei von der Familie $\mathcal{F}$ erzeugt. Dann sind für $f : X \longrightarrow Y$ äquivalent:*

(i) *Es ist f messbar.*
(ii) *Es ist $f^{-1}(B)$ messbar für alle $B \in \mathcal{F}$.*

Beweis. (i) $\Longrightarrow$ (ii): Das ist klar.
(ii) $\Longrightarrow$ (i): Betrachte $\mathcal{A}_f := \{B \subseteq Y : f^{-1}(B) \in \mathcal{A}\}$. Dann gilt:

- $\mathcal{F} \subseteq \mathcal{A}_f$ (nach (ii)).
- $\mathcal{A}_f$ ist eine σ-Algebra (nach voriger Proposition).

Damit folgt dann nach Konstruktion der erzeugten σ-Algebra

$$\mathcal{B} = \sigma(\mathcal{F}) \subseteq \mathcal{A}_f.$$

Das beendet den Beweis. $\square$

Wir ziehen einige Folgerungen aus dem vorherigen Lemma.

FOLGERUNG (Stetige Funktionen sind messbar). *Seien (X_i, τ_i), $i = 1, 2$, topologische Räume mit den zugehörigen Borel - σ -Algebren $\mathcal{B}_i$. Ist $f : X_1 \longrightarrow X_2$ stetig, so ist es messbar.*

Beweis. Da $\mathcal{B}_2$ von τ_2 erzeugt wird, reicht es nach dem vorherigen Lemma zu zeigen, dass $f^{-1}(U)$ messbar ist für alle $U \in \tau_2$. Es gilt aber

$$f^{-1}(U) \in \tau_1 \subseteq \mathcal{B}_1,$$

da f stetig ist und $\mathcal{B}_1$ von τ_1 erzeugt ist. $\square$

Wir ziehen nun Folgerungen für messbare reellwertige Funktionen. Insbesondere werden wir sehen, dass die alle 'üblichen' Operationen Messbarkeit erhalten. Insbesondere bleibt Messbarkeit (ANDERS als Stetigkeit) unter punktweisen Grenzwerten erhalten. Ende 25. Vorlesung

Konvention. Wenn es um Messbarkeit von Funktionen mit Werten in $\mathbb{R}$ (bzw. $\mathbb{R} \cup \{-\infty\} \cup \{\infty\}$) geht, wird $\mathbb{R}$ (bzw. $\mathbb{R} \cup \{-\infty\} \cup \{\infty\}$) immer mit der Borel - σ - Algebra versehen.

Aus dem vorherigen Lemma und den verschiedenen Arten die Borel-σ-Algebra zu erzeugen, erhält man sofort:

FOLGERUNG. *Sei $(X, \mathcal{A})$ ein messbarer Raum und $f : X \longrightarrow \mathbb{R}$ (bzw. $f : X \longrightarrow \mathbb{R} \cup \{\infty\} \cup \{-\infty\}$) eine Funktion. Dann sind äquivalent:*

(i) *Es ist f Borel-messbar.*
(ii) *Es ist $f^{-1}(I)$ messbar für alle Intervalle I.*

(iii) *Es ist* $f^{-1}(a,\infty]$ *messbar für alle* $a \in \mathbb{R}$.
(iv) *Es ist* $f^{-1}[-\infty,a)$ *messbar für alle* $a \in \mathbb{R}$.

Bemerkung. Aus (ii) in der Folgerung folgt sofort, dass mit f auch $-f$ messbar ist.

Theorem. *Sei* $(X,\mathcal{A})$ *ein messbarer Raum und seien* $f_n : X \longrightarrow \mathbb{R}\bigcup\{-\infty\}\bigcup\{\infty\}$, $n \in \mathbb{N}$, *messbare Funktionen. Dann sind auch die Funktionen*

$$\sup_{n\geq 1} f_n,\quad \inf_{n\geq 1} f_n,\quad \limsup_{n\to\infty} f_n \quad \textit{und} \quad \liminf_{n\to\infty} f_n$$

messbar.

Beweis. Setze $g = \sup f_n$. Dann gilt

$$g^{-1}(a,\infty]) = \bigcup_{n=1}^{\infty} f_n^{-1}((a,\infty])$$

und dies ist messbar als abzählbare Vereinigung messbarer Mengen. Ähnlich kann man $\inf f_n$ behandeln. Damit folgen dann auch die Aussagen für

$$\limsup_{n\to\infty} f_n = \inf_{k\geq 1}\left(\sup_{n\geq k} f_n\right)$$

und

$$\liminf_{n\to\infty} f_n = \sup_{k\geq 1}\left(\inf_{n\geq k} f_n\right).$$

Das beendet den Beweis. □

Aus dem Theorem erhalten wir die folgende sehr bemerkenswerte Konsequenz (vgl. obige Bemerkung zur Stetigkeit):

Folgerung (Punktweise Grenzwerte sind messbar). *Sei* $(X,\mathcal{A})$ *ein messbarer Raum. Konvergiert die Folge der messbaren Funktionen*

$$f_n : X \longrightarrow \mathbb{R}\cup\{\infty\}\cup\{-\infty\}$$

punktweise gegen die Funktion f, *so ist* f *messbar.*

Aus dem Theorem erhält man auch sofort folgendes Ergebnis.

Folgerung. *Sei* $(X,\mathcal{A})$ *ein messbarer Raum. Sind* $f,g : X \longrightarrow \mathbb{R}$ *messbar, so sind auch* $\max\{f,g\}$ *und* $\min\{f,g\}$ *messbar. Insbesondere sind auch* $f_+ = \max\{f,0\}$ *und* $f_- = -\min\{f,0\} = \max\{-f,0\}$ *messbar.*

Theorem (Komponentenweise Messbarkeit impliziert Messbarkeit). *Sei* $(X,\mathcal{A})$ *ein messbarer Raum und* $f_1,\ldots,f_N : X \longrightarrow \mathbb{R}$ *messbar. Dann ist auch die Funktion*

$$F : X \longrightarrow \mathbb{R}^N, F(x) = (f_1(x),\ldots,f_N(x)),$$

messbar.

Beweis. Es reicht zu zeigen, daß die Urbilder von Rechtecken messbar sind. Sei $R = (a_1, b_1) \times \ldots \times (a_N, b_N)$ ein solches Rechteck. Dann gilt

$$F^{-1}(R) = \bigcap_{j=1}^{N} f_j^{-1}(a_j, b_j)$$

und das ist messbar als endlicher Schnitt messbarer Mengen. $\square$

FOLGERUNG. *Sei $(X, \mathcal{A})$ ein messbarer Raum und seien $f, g : X \longrightarrow \mathbb{R}$ messbar. Dann sind $f + g$ und fg messbar.*

Beweis. Wir betrachten nur fg. Der Fall $f + g$ kann analog behandelt werden. Sei

$$\phi : \mathbb{R}^2 \longrightarrow \mathbb{R}, \phi(u, v) = uv.$$

Dann ist ϕ stetig und damit messbar. Sei weiterhin

$$F : X \longrightarrow \mathbb{R}^2, F(x) = (f(x), g(x)).$$

Dann ist F nach dem vorangehenden Theorem messbar. Damit ist

$$fg = \phi \circ F$$

als Verknüpfung messbarer Funktionen wieder messbar. $\square$

Zum Abschluss dieses Abschnittes führen wir noch eine besonders schöne Klasse von messbaren Funktionen ein.

DEFINITION (Einfache Funktionen). Sei $(X, \mathcal{A})$ ein messbarer Raum. Eine messbare Funktion $s : X \longrightarrow \mathbb{R}$ heißt *einfach*, wenn ihr Wertebereich endlich ist.

BEMERKUNG. Sind $\alpha_1, \ldots, \alpha_n$ die verschiedenen Werte der einfachen Funktion s, so sind die Urbilder $A_j := \{x \in X \mid s(x) = \alpha_j\}$ (Niveaumengen) messbar und disjunkt mit Vereinigung X, und es gilt

$$s = \sum_{j=1}^{n} \alpha_j 1_{A_j}.$$

Sind umgekehrt $A_1, \ldots, A_n$ messbar (und nicht notwendig disjunkt) und $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$, so ist $\sum \alpha_j 1_{A_j}$ einfach. (Tatsächlich ist diese Funktion messbar als Summe messbarer Funktionen und nimmt (offenbar) nur endlich viele Werte an.)

Meßbare Funktionen können gut durch einfache Funktionen approximiert werden.

THEOREM (Approximation durch einfache Funktionen). *Sei $(X, \mathcal{A})$ ein messbarer Raum und $f : X \longrightarrow [0, \infty]$ messbar. Dann existieren einfache Funktionen s_n, $n \in \mathbb{N}$, mit folgenden Eigenschaften:*

- $0 \leq s_1 \leq \ldots s_n \leq s_{n+1} \leq \ldots \leq f$.
- $s_n(x) \to f(x)$, $n \to \infty$, *für alle* $x \in X$.

Bemerkung. (Übung) Ist $f : X \longrightarrow \mathbb{R}$ messbar und beschränkt, so lässt es sich gleichmäßig durch einfache Funktionen approximieren.

Beweis. Sei für $n \in \mathbb{N}$ die Funktion $\phi_n : [0,\infty] \longrightarrow [0,\infty)$ definiert durch

$$\phi_n(t) = n, \text{ falls } n \leq t$$

und

$$\phi_n(t) = \frac{k}{2^n} \text{ falls } k/2^n \leq t < \tfrac{(k+1)}{2^n} \text{ für ein } k \in \mathbb{N} \text{ mit } 0 \leq k < n2^n.$$

Dann nehmen die ϕ_n nur endlich viele Werte an und sind messbar, und es gilt:

- $\phi_n(x) \leq \phi_{n+1}(x) \leq x$ für alle $x \in [0,\infty]$.
- $\phi_n(x) \to x$ für alle $x \in [0,\infty]$.

Damit folgt leicht, dass $s_n := \phi_n \circ f$ die gewünschten Eigenschaften hat. □

2. Maße und Integration positiver Funktionen

In diesem Abschnitt lernen wir Maße kennen. Das Konzept des Maßes verallgemeinert die Idee der Volumenmessung.

Definition (Maß). Sei $(X,\mathcal{A})$ ein messbarer Raum. Eine Abbildung

$$\mu : \mathcal{A} \longrightarrow [0,\infty]$$

heißt *Maß*, wenn gilt:

- $\mu(\emptyset) = 0$.
- $\mu(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mu(A_n)$ für alle messbaren paarweise diskunkten A_n, $n \in \mathbb{N}$.

Ist μ ein Maß auf $(X,\mathcal{A})$, so nennt man das Tripel $(X,\mathcal{A},\mu)$ einen *Maßraum*. Das Maß μ heißt *endlich*, wenn $\mu(X) < \infty$ gilt.

Bemerkung.

- Die zweite Bedingung ist die entscheidenden Bedingung. Sie ist als *σ-Additivität* bekannt.
- Gegeben die zweite Bedingung, so ist die Bedingung $\mu(\emptyset) = 0$ äquivalent dazu, dass es eine Menge A in $\mathcal{A}$ gibt mit $\mu(A) < \infty$ (Übung). Sie dient also nur dazu den trivialen Fall auszuschließen.

Beispiel (Zählmaß). Sei X eine beliebige Menge und $\mathcal{A}$ die Potenzmenge von $\mathbb{N}$. Dann ist

$$\mu : \mathcal{A} \longrightarrow [0,\infty], \;\; \mu(E) = \sharp E$$

ein Maß (wobei $\sharp E$ die Anzahl der Element von E bezeichnet). Dieses Maß heißt das Zählmaß auf X. Auf abzählbaren Mengen (z.B. für $X = \mathbb{N}$) ist das Zählmaß ein sehr natürliches Maß. Alle Punkte haben dann das gleiche "Gewicht„.

Proposition (Einfache Eigenschaften). *Sei μ ein Maß auf $(X, \mathcal{A})$. Sind A, B messbar mit $A \subseteq B$, so gilt*

$$\mu(B) = \mu(A) + \mu(B \setminus A)$$

und insbesondere $\mu(A) \leq \mu(B)$.

Beweis. Mit $A_1 = A$ und $A_2 = B \setminus A$, $A_j = \emptyset$, $j \geq 2$ folgt die Aussage. □

Ende 26. Vorlesung

Proposition (Kleiner Grenzwertsatz). *Sei μ ein Maß auf $(X, \mathcal{A})$. Seien $A_1 \subseteq A_2 \subseteq \ldots$ messbar. Dann gilt für $A = \bigcup_n A_n$ die Beziehung*

$$(*) \quad \mu(A) = \lim_{n\to\infty} \mu(A_n).$$

Sind $A_1 \supseteq A_2 \supseteq \ldots$ messbar mit $\mu(A_1) < \infty$. Dann gilt für $A = \bigcap_n A_n$

$$\mu(A) = \lim_{n\to\infty} \mu(A_n).$$

Bemerkung. (a) Ist μ additiv, so ist σ-Additivität äquivalent zu $(*)$. In diesem Sinne is $(*)$ eine fundamentale Eigenschaft des Maßes.
(b) Setzt man $A = \lim A_n$ (was in beiden Fällen naheliegt), so besagt die Proposition gerade

$$\mu(\lim_{n\to\infty} A_n) = \lim_{n\to\infty} \mu(A_n).$$

In diesem Sinne handelt es sich um eine Stetigkeitseigenschaft von μ.
(c) Die Voraussetzung $\mu(A_1) < \infty$ in der zweiten Behauptung ist notwendig. Betrachte zum Beispiel $\mathbb{N}$ mit dem Zählmaß und den messbaren Mengen $A_n = [n, \infty)$, $n \in \mathbb{N}$. Dann gilt $\mu(A_n) = \infty$ aber $\mu(A) = 0$ (da $A = \emptyset$).

Beweis. Erste Aussage: Setzte $B_1 := A_1$ und $B_n := A_n \setminus A_{n-1}$ für $n \geq 2$. Dann gilt:

- $A_n = \bigcup_{k=1}^n B_k$ und damit auch $A = \bigcup_{k=1}^\infty B_k$.
- Die B_n sind paarweise disjunkt und messbar.

Damit folgt dann aus der σ-Additivität:

$$\begin{aligned}
\mu(A) &= \mu\Big(\bigcup_{k\geq 1} B_k\Big) \\
(\sigma\text{-Additivität}) \quad &= \sum_{k=1}^\infty \mu(B_k) \\
&= \lim_{n\to\infty} \sum_{k=1}^n \mu(B_k) \\
&= \lim_{n\to\infty} \mu(A_n).
\end{aligned}$$

Zweite Aussage: Das folgt aus der ersten Aussage nach Komplementbildung und Subtraktion des endlichen (!) $\mu(A_1)$: Setze $C_n := A_1 \setminus A_n$. Dann gilt $C_1 \subseteq C_2 \subseteq \ldots$. Sei $C = \bigcup C_n$. Dann gilt nach der ersten Aussage
$$\mu(C) = \lim_{n\to\infty} \mu(C_n).$$
Weiterhin gilt
$$\mu(C_n) = \mu(A_1 \setminus A_n) = \mu(A_1) - \mu(A_n)$$
sowie
$$\mu(C) = \mu\left(\bigcup_n C_n\right) = \mu\left(A_1 \setminus \bigcap_n A_n\right) = \mu(A_1) - \mu\left(\bigcap_n A_n\right).$$
Damit folgt die Behauptung. □

Definition (Integral einfacher Funktionen). Sei μ ein Maß auf $(X, \mathcal{A})$. Sei s eine einfache nichtnegative Funktion mit den verschiedenen Werten $\alpha_1, \ldots, \alpha_n$ und den zugehörigen Niveaumengen $A_j = \{x \in X \mid s(x) = \alpha_j\}$, also $s = \sum_{j=1}^n \alpha_j 1_{A_j}$. Dann definiert man
$$\int s d\mu = \int_X s d\mu = \sum_{j=1}^n \alpha_j \mu(A_j).$$

Bemerkung – Rechnen in $[0, \infty]$. In dieser Definition (wie auch an anderen Stellen) brauchen wir Arithmetik d.h. Addition und Multiplikation auf $[0, \infty]$. Diese werden so definiert:
$$a \cdot \infty = 0 \quad \text{für } a = 0, \;\; a \cdot \infty = \infty \text{ sonst.}$$
$$a + \infty = \infty \text{ für alle } a \in [0, \infty].$$
Für $a, b \in [0, \infty)$ werden $a + b, a \cdot b$ wie üblich definiert. Mit diesen Definitionen lässst sich dann problemlos addieren und multiplizieren. (Probleme treten erst dann auf, wenn man subtrahieren will. Zum Beispiel lässt sich aus $a + c = b + c$ im allgemeinen NICHT schließen $a = b$ usw.)

Proposition (Wohldefiniert). *Sei μ ein Maß auf $(X, \mathcal{A})$. Ist $s = \sum_{j=1}^m \beta_j 1_{B_j}$ eine einfache Funktion mit disjunkten (messbaren) B_j und nichtnegativen β_j, $j = 1, \ldots, m$. Dann gilt*
$$\int s d\mu = \sum_{j=1}^n \beta_j \mu(B_j).$$

Beweis. Ohne Einschränkung seien alle β_j von Null verschieden. (Ein Term mit $\beta_j = 0$ trägt zur gewünschten Summe sowieso nicht bei.) Seien $\alpha_1, \ldots, \alpha_n$ die von Null verschiedenen Werte von s und $A_i = \{x \in X \mid s(x) = \alpha_i\}$ die zugehörigen Niveaumengen. Dann setzen sich die A_i aus den B_j (disjunkt) zusammen, da die B_j disjunkt sind und das liefert die Behauptung. Hier ist die genaue Buchhaltung: Sei

$$S_k = \{j \mid \beta_j = \alpha_k\}$$

für $k = 1, \ldots, n$. Dann gilt (aufgrund der Disjunktheit der B_j)

$$A_k = \bigcup_{j \in S_k} B_j,$$

wobei die Vereinigung disjunkt ist. Damit folgt also

$$\mu(A_k) = \sum_{j \in B_k} \mu(B_j).$$

Das liefert insgesamt

$$\begin{aligned} \sum_k \alpha_k \mu(A_k) &= \sum_k \alpha_k \left(\sum_{j \in S_k} \mu(B_j) \right) \\ &= \sum_k \sum_{j \in S_k} \beta_j \mu(B_j) \\ &= \sum_j \beta_j \mu(B_j). \end{aligned}$$

Das beendet den Beweis. □

Proposition (Einfache Eigenschaften des Integrals einfacher Funktionen). *Sei μ ein Maß auf $(X, \mathcal{A})$. Seien s, t einfache nichtnegative Funktionen.*
(a) Gilt $s \leq t$ so folgt $\int s d\mu \leq \int t d\mu$.
(b) Mit $\lambda > 0$ gilt $\int (s + \lambda t) d\mu = \int s d\mu + \lambda \int t d\mu$.

Beweis. (a) Das ist klar.
(b) Sei $s = \sum_j^n \alpha_j 1_{A_j}$ und $t = \sum_{k=1}^m \beta_k 1_{B_k}$. Ohne Einschränkung $n = m$ und $A_j = B_j$ (sonst Betrachten von $A_j \bigcap B_k$ und nutzen der Proposition zur Wohldefiniertheit). Nun folgt die Aussage einfach. □

Definition (Integral nichtnegativer Funktionen). Sei μ ein Maß auf $(X, \mathcal{A})$. Für ein messbares $f : X \longrightarrow [0, \infty]$ definiert man

$$\int f d\mu := \int_X f d\mu := \sup_{0 \leq s \leq f, \text{ s :einfach}} \int s d\mu$$

und nennt dies das Integral von f über X bzgl. μ. (Hier ist der Wert ∞ für das Integral erlaubt.)

Bemerkung. Aus der Definition ergeben sich sofort zwei nützliche Beobachtungen:

- Gilt $\int f d\mu < \infty$, so ist $\mu(\{x \in X \mid f(x) = \infty\}) = 0$. (Andernfalls: Setze $A := \{x \in X \mid f(x) = \infty\}$ und betrachte $s = n 1_A$ für $n \in \mathbb{N}$...)

- Sei $S := \{x \in X \mid f(x) \neq 0\}$. Gilt $\mu(S) = 0$, so folgt $\int f d\mu = 0$. (Betrachte eine beliebige einfache Funktion s mit $0 \leq s \leq f$. Dann gilt $s = \sum_{j=1}^n \alpha_j 1_{A_j}$ mit disjunkten A_j. Jedes A_j mit $\alpha_j \neq 0$ ist nun eine Teilmenge von S und erfüllt damit $\mu(A_j) \leq \mu(S) = 0$. Damit folgt sofort $\int s d\mu = 0$. Das liefert die gewünschte Behauptung.)

Proposition (Monotonie des Integrals). *Sei μ ein Maß auf $(X, \mathcal{A})$. Seien $f, g : X \longrightarrow [0, \infty]$ messbar mit $f \leq g$. Dann gilt $\int f d\mu \leq \int g d\mu$.*

Beweis. Das ist klar. □

Wir lernen nun DAS THEOREM zur Konvergenz von Integralen nichtnegativer Funktionen kennen. Die meisten der anschließend diskutierten Aussagen sind Folgerungen aus diesem Theorem.

Theorem (Monotone Konvergenz/Satz von Beppo Levi). *Sei μ ein Maß auf $(X, \mathcal{A})$. Seien $f_n : X \longrightarrow [0, \infty]$ messbar mit $f_n \leq f_{n+1}$ für alle $n \in \mathbb{N}$. Sei f der punktweise Grenzwert der f_n. Dann ist f messbar und nichtnegativ und es gilt*

$$\int f d\mu = \lim_{n \to \infty} \int f_n d\mu.$$

Bemerkung. Aufgrund von $f_n \leq f_{n+1}$ existiert der punktweist Grenzwert der f_n.

Beweis. Es ist f messbar als punktweiser Grenzwert messbarer Funktionen. Aufgrund der Monotonie des Integrals gilt

$$\int f_n d\mu \leq \int f_{n+1} d\mu \leq \int f d\mu.$$

Insbesondere existiert also

$$\alpha = \lim_{n \to \infty} \int f_n d\mu$$

und es gilt

$$\alpha \leq \int f d\mu.$$

Noch zu zeigen $\alpha \geq \int f d\mu$: Sei s eine einfache Funktion mit $0 \leq s \leq f$ und $c \in (0, 1)$. Betrachte

$$E_n := \{x \in X \mid f_n(x) \geq cs(x)\}.$$

Dann gilt

$$E_1 \subseteq E_2 \subseteq \ldots \subseteq E_n \subseteq ...$$

und

$$X = \bigcup_n E_n.$$

($f(x) = 0$ liefert $x \in E_1$, $f(x) > 0$ liefert $x \in E_n$ für genügend großes n wegen $f_n(x) \to f(x)$.) Damit folgt aus der Monotonie des Integrals also

$$\begin{aligned}\int f_n d\mu &\geq \int f_n 1_{E_n} d\mu \\ &\geq c \int s 1_{E_n} d\mu \\ &\stackrel{!}{\to} c \int s d\mu.\end{aligned}$$

Nimmt man (für den Moment) die letzte Konvergenz als bewiesen an, so kann man schließen

$$\alpha = \lim_{n\to\infty} \int f_n d\mu \geq c \int s d\mu$$

für alle einfachen s mit $0 \leq s \leq f$ und alle $c \in (0,1)$. Damit folgt

$$\alpha \geq \int f d\mu.$$

Es bleibt die Aussage (!) zu beweisen: Wegen $X = \bigcup_n E_n$ und den schon bekannten Aussagen über Konvergenz von Maßen, reicht es zu zeigen, dass

$$\phi : \mathcal{A} \longrightarrow [0,\infty], \phi(E) = \int s 1_E d\mu$$

ein Maß ist. Das folgt durch direkte Rechnung für einfache Funktionen. □

Bemerkung (Philosophie). Weiter oben haben wir den “kleinen Grenzwertsatz„ kennengelernt:

$$\mu(A) = \lim_{n\to\infty} \mu(A_n)$$

für $A_1 \subseteq A_2 \subseteq ...$ mit $A = \bigcup A_n$. Dieser Satz ist natürlich ein Spezialfall des vorangehenden Theorems (mit $f_n = 1_{A_n}$ und $f = 1_A$). Umgekehrt liegt er (für das Maß ϕ statt μ) dem Beweis des Theorems zugrunde. In diesen Sinne ist das Theorem nicht anderes als eine (extrem nützliche) Umformulierung des kleinen Grenzwertsatzes. Dieser Satz wiederum ist, wie oben schon diskutiert, im wesentlichen äquivalent zur σ-Additivität. Damit sind σ-Additvität, kleiner Grenzwertsatz und vorangehendes Theorem im wesentlichen nichts anderes als verschiedene Perspektiven auf dasselbe Geschehen.

Als Folgerung erhalten wir die Linearität des Integrals und viel mehr.

Folgerung. *Sei μ ein Maß auf $(X, \mathcal{A})$. Seien $f_n : X \longrightarrow [0,\infty]$ messbar. Dann gilt*

$$\int \sum_{n=1}^{\infty} f_n d\mu = \sum_{n=1}^{\infty} \int f_n d\mu.$$

Bemerkung. Auch hier sind (natürlich) die auftretenden Summem im Sinne der Arithmetik in $[0,\infty]$ zu verstehen.

Beweis. Wir zeigen zunächst $\int (f+g)d\mu = \int f d\mu + \int g d\mu$ für messbare nichtnegative f, g. Nach dem Theorem zur Approximation messbarer Funktionen existieren monoton wachstende Folgen einfacher nichtnegativer Funktionen s_n und t_n mit $s_n \to f$ und $t_n \to g$. Dann konvergiet also $t_n + s_n$ monoton wachstend gegen $f + g$. Damit folgt mit zweimaliger Anwendung des vorherigen Theorem und der Linearität des Integrals für einfache Funktionen dann

$$\begin{aligned}\int (f+g)d\mu &= \lim_{n\to\infty} \int (s_n + t_n)d\mu \\ &= \lim_{n\to\infty} (\int s_n d\mu + \int t_n d\mu) \\ &= \int f d\mu + \int g d\mu.\end{aligned}$$

Damit können wir nun zur eigentlichen Aussage kommen: Wir setzen

$$g_N := \sum_{k=1}^{N} f_k.$$

Dann konvergieren die g_N monoton wachsend gegen $f = \sum_{k=1}^{\infty} f_k$. Damit folgt aus dem vorherigen Theorem und der Vorüberlegung dann

$$\int f d\mu = \lim_{N\to\infty} \int g_N d\mu = \lim_{N\to\infty} \sum_{k=1}^{N} \int f_k d\mu = \sum_{k=1}^{\infty} \int f_k d\mu.$$

Das beendet den Beweis. $\square$

Bemerkung (Anwendung). Betrachte $\mathbb{N}$ versehen mit dem Zählmaß und $a : \mathbb{N} \times \mathbb{N} \longrightarrow [0,\infty]$. Dann gilt

$$\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} a_{ij} = \sum_{j=1}^{\infty} \sum_{i=1}^{\infty} a_{ij}.$$

(Beweis: $f_i : \mathbb{N} \longrightarrow [0,\infty]$, $f_i(j) = a_{ij}$...)

Theorem (Lemma von Fatou). *Sei μ ein Maß auf $(X, \mathcal{A})$. Seien $f_n : X \longrightarrow [0,\infty]$ messbar. Dann gilt*

$$\int \liminf_{n\to\infty} f_n d\mu \le \liminf_{n\to\infty} \int f_n d\mu.$$

Bemerkung. Im allgemeinen ist die Ungleichung strikt. Betrachte zum Beispiel $\mathbb{N}$ mit dem Zählmaß und $f_n = 1_{\{k\}}$ (führt auf $0 \le 1$) oder $f_n = 1_{\{k \ge n\}}$ (führt auf $0 \le \infty$).

Beweis. Setzt $g_k := \inf_{n\geq k} f_n$. Dann ist g_k messbar und nichtnegative. Nach Konstruktion konvergiert (g_k) monoton wachsend gegen $\liminf_n f_n$. Damit folgt aus dem Theorem über monotone Konvergenz

$$\int \liminf_{n\to\infty} f_n d\mu = \lim_{k\to\infty} \int g_k d\mu$$

für alle $k \in \mathbb{N}$. Offenbar gilt $g_k \leq f_k$ und wir erhalten also

$$\int \liminf_n f_n d\mu \leq \int f_k d\mu$$

für alle $k \in \mathbb{N}$. Bildet man auf der rechten (und der linken ;-) Seite den $\liminf$ über k, so folgt die gewünschte Aussage. □

3. Integration komplexer Funktionen und $\mathcal{L}^1(X,\mu)$

Im vorangehenden Abschnitt haben wir die Integration nichtnegativer Funktionen kennengelernt. In diesem Abschnitt lernen wir die Integration von komplexwertigen Funktionen mit einer gewissen Beschränktheitseigenschaft kennen. Damit haben wir dann die beiden gängigen Varianten von Integrationstheorie kennengelernt. Eine ähnliche Situation ist uns auch in Analysis I schon begegnet bei der Summation von Folgen: Man hat eine überzeugende Theorie für Folgen mit nichtnegativen Gliedern und eine weitere überzeugende Theorie für absolut konvergente Folgen. Tatsächlich sind entsprechende Betrachtungen ein Spezialfall unserer Erwägungen hier (mit dem Raum der natürlichen Zahlen versehen mit dem Zählmaß).

Definition. Sei $(X,\mathcal{A},\mu)$ ein Maßraum. Dann definieren wir

$$\mathcal{L}^1(X,\mathcal{A},\mu) = \mathcal{L}^1(X,\mu) = \mathcal{L}^1(X)$$

als die Menge der messbaren Funktionen $f : X \longrightarrow \mathbb{C}$ mit

$$\int |f| d\mu < \infty.$$

Für ein solches f mit Realteil u und Imaginärteil v (also $f = u + iv$) definieren wir dann

$$\int f d\mu := \int u_+ d\mu - \int u_- d\mu + i \int v_+ d\mu - i \int f_- d\mu.$$

Bemerkung.

- Hier ist $\mathbb{C}$ mit der Borel-σ-Algebra versehen. Dann bedeutet Messbarkeit von $f : X \longrightarrow \mathbb{C}$ gerade, dass die entsprechende Funktion
$$\widetilde{f} : X \longrightarrow \mathbb{R}^2, \widetilde{f}(x) = (\Re f(x), \Im f(x))$$
messbar ist.
- f messbar $\Longrightarrow |f| = |\cdot| \circ f$ messbar.

- Es gilt (einfach): f messbar $\Longleftrightarrow \Re f, \Im f$ messbar. Damit sind insbesondere die auftretenden Terme $u_\pm$ und $v_\pm$ messbar und nichtnegativ. Daher existieren die entsprechenden Integrale.
- Alle Integrale auf der rechten Seite der angegebenen Gleichung sind endlich (aufgrund von $|u_\pm| \leq |u| \leq |f|$ und $|v_\pm| \leq |v| \leq |f|$).

THEOREM ($\mathcal{L}^1$ ist Vektorraum). *Sei $(X, \mathcal{A}, \mu)$ ein Maßraum. Seien $f, g \in \mathcal{L}^1(X, \mathcal{A}, \mu)$ und $\alpha, \beta \in \mathbb{C}$ gegeben. Dann gehört auch $\alpha f + \beta g$ zu $\mathcal{L}^1(X, \mathcal{A}, \mu)$ und es gilt*

$$\int (\alpha f + \beta g) d\mu = \alpha \int f d\mu + \beta \int g d\mu.$$

Beweis. $\alpha f + \beta g \in \mathcal{L}^1(X, \mathcal{A}, \mu)$: Es ist $\alpha f + \beta g$ messbar als Linearkombination messbarer Funktionen.
Zur Gleichung: Übung. □

PROPOSITION (Dreiecksungleichung). *Sei $(X, \mathcal{A}, \mu)$ ein Maßraum. Dann gilt*

$$\left| \int f d\mu \right| \leq \int |f| d\mu$$

für alle $f \in \mathcal{L}^1(X, \mathcal{A}, \mu)$.

Beweis. Wähle $\alpha \in \mathbb{C}$ mit $|\alpha| = 1$ und $|\int f d\mu| = \alpha \int f d\mu$. Dann gilt

$$\begin{aligned} \left| \int f d\mu \right| &= \alpha \int f d\mu \\ &= \int \alpha f d\mu \\ \text{(Linke Seite reell)} \quad &= \int \Re(\alpha f) d\mu \\ (\Re(\alpha f) \leq |\alpha f| = |f|) \quad &\leq \int |f| d\mu. \end{aligned}$$

Das beendet den Beweis. □

BEMERKUNG. Auf $\mathcal{L}^1(X, \mathcal{A}, \mu)$ wird durch

$$\|f\|_1 := \int |f| d\mu$$

eine Halbnorm definiert. Durch entsprechende Quotientenbildung (s.u.) entsteht dann ein vollständiger (!) normierter Raum.

Der wesentliche Grenzwertsatz zur Integration nichtnegativer Funktionen ist das Monotone-Konvergenz-Theorem. Der entsprechende Grenzwertsatz zur Integration in $\mathcal{L}^1$ ist das folgende Theorem. Es ist (natürlich ;-) eine Folge aus dem Theorem über monotone Konvergenz in Form des Lemma von Fatou.

THEOREM (Dominierte Konvergenz/Satz von Lebesgue). *Sei* $(X, \mathcal{A}, \mu)$ *ein Maßraum. Seien* $f_n : X \longrightarrow \mathbb{C}$, $n \in \mathbb{N}$, *und* $f : X \longrightarrow \mathbb{C}$ *messbare Funktionen mit* $f_n(x) \to f(x)$, $n \to \infty$, *für alle* $x \in X$. *Gibt es ein* $g \in \mathcal{L}^1(X, \mathcal{A}, \mu)$ *mit* $|f_n| \leq g$ *für alle* $n \in \mathbb{N}$, *so gehört auch* f *zu* $\mathcal{L}^1(X, \mathcal{A}, \mu)$ *und es gilt*

$$\int |f_n - f| d\mu \to 0, n \to \infty$$

und damit insbesondere

$$\int f d\mu = \lim_{n\to\infty} \int f_n d\mu.$$

Beweis. Es ist f messbar als Grenzwert messbarer Funktionen und es gilt $|f| \leq g$ (da $|f_n| \leq g$ und f der punktweise Grenzwert der f_n ist). Damit gehört f zu $\mathcal{L}^1$.

Wir zeigen nun die Konvergenz: Betrachte

$$h_n := 2g - |f_n - f| \geq 0.$$

Dann gilt $h_n \geq 0$ und $h_n \to 2g$ punktweise. Damit folgt aus dem Lemma von Fatou also

$$\begin{aligned} \int 2g d\mu &= \int \lim h_n d\mu \\ \text{(Fatou)} \quad &\leq \liminf_n \int h_n \\ &= \liminf_n \left(\int 2g d\mu - \int |f_n - f| d\mu \right) \\ \text{(Rechenregeln } \liminf) \quad &= \int 2g d\mu - \limsup_n \int |f_n - f| d\mu. \end{aligned}$$

Wegen $g \in \mathcal{L}^1$ ist $\int 2g d\mu$ endlich und wir können es auf beiden Seiten subtrahieren und erhalten

$$\limsup_n \int |f_n - f| d\mu \leq 0.$$

Da die auftretenden Integranden nichtnegativ sind, folgt

$$0 \leq \liminf_n \int |f_n - f| d\mu \leq \limsup_n \int |f_n - f| d\mu \leq 0$$

und wir erhalten

$$0 = \lim_{n\to\infty} \int |f_n - f| d\mu.$$

Das ist gerade die erste Aussage. Aus der Dreiecksungleichung folgt dann die letzte Aussage. □

Wir gehen nun auf die (verschwindende) Rolle von Nullmengen in der Theorie ein. Hier heißt eine Menge *Nullmenge*, wenn sie messbar ist und ihr Maß gerade Null ist.

Definition (Gültigkeit μ - fast überall). Sei $(X, \mathcal{A}, \mu)$ ein Maßraum und $\mathcal{P}$ eine Eigenschaft, die ein Element von X haben kann. Dann gilt $\mathcal{P}$ μ-fast-überall (μ-f.ü. oder auch nur f.ü.), wenn es ein N in $\mathcal{A}$ gibt mit folgenden Eigenschaften:

- $\mu(N) = 0$.
- Es gilt $\mathcal{P}$ für alle $x \in X \setminus N$.

Bemerkung.

- Natürlich hängt das Konzept der Nullmenge bzw. der Gültigkeit fast überall stark vom gegebenen Maß μ ab. Betrachtet man z. B. $X = \mathbb{N}$ mit dem Zählmaß, so gilt ein Eigenschaft fast-überall genau dann, wenn sie für alle Punkte gilt.
- Man beachte, daß nicht gefordert wird, daß die Menge

$$\{x \mid \ \mathcal{P} \text{ gilt nicht für } x\}$$

eine Nullmenge ist. Tatsächlich kann es sein, daß diese Menge gar nicht messbar ist. Es gibt jedoch einen Satz der besagt, dass man eine σ-Algebra bezüglich aller Nullmengen eines Maßes vervollständigen kann und das Maß dann darauf fortsetzen kann.

Beispiel (Eigenschaften fast überall).

- $f(x) > 0$ (für gegebenes f).
- $f(x) = g(x)$ (für gegebene f und g).
- $f_n(x)$ konvergiert bzw. $f_n(x)$ konvergiert nicht (für gegebenen Folge (f_n)).

Anwendung - Integration fast überall gleicher Funktionen. Sei $(X, \mathcal{A}, \mu)$ ein Maßraum. Dann wird durch

$$f \sim g \Longleftrightarrow \quad f = g \ \mu \text{ -f.ü.}$$

eine Äquivalenzrelation auf den mßbaren Funktionen definiert (wie man leicht sieht). Gilt $f \sim g$, so gehört f zu $\mathcal{L}^1$ genau dann, wenn g zu $\mathcal{L}^1$ gehört. In diesem Fall gilt dann

$$\int f 1_E d\mu = \int g 1_E d\mu$$

für alle messbaren E.
Bew. Sei N eine Nullmenge mit $f = g$ auf $X \setminus N$ (z.B. kann man $N = \{x \mid f(x) \neq g(x)\}$ wählen). Dann gilt $\mu(N \cap E) = 0$ für alle

messbaren E. Damit folgt

$$\begin{aligned}
\int f 1_E d\mu &= \int f(1_{E\setminus N} + 1_{E\cap N}) d\mu \\
&= \int f 1_{E\setminus N} d\mu + \int f 1_{E\cap N} d\mu \\
(\mu(E\cap N) = 0) \quad &= \int f 1_{E\setminus N} d\mu \\
(f = g \text{ auf } E\setminus N) \quad &= \int g 1_{E\setminus N} d\mu \\
(\mu(E\cap N) = 0) \quad &= \int g 1_{E\setminus N} d\mu + \int g 1_{E\cap N} d\mu \\
&= \int g(1_{E\setminus N} + 1_{E\cap N} d\mu \\
&= \int g 1_E d\mu.
\end{aligned}$$

Das beendet den Beweis.

Soweit es um Integration geht, können also f und g nicht unterschieden werden. Das legt es nahe, eine Integrationstheorie für die Klassen zu entwickeln. Man definiert

$$L^1(X) := L^1(X,\mu) := L^1(X,\mathcal{A},\mu) := \mathcal{L}^1(X,\mathcal{A},\mu)/\sim$$

und setzt

$$\int : L^1(X,\mu) \longrightarrow \mathbb{C}, \int [f] d\mu := \int f d\mu.$$

Nach den vorhergehenden Überlegungen ist das wohldefiniert. Dann wird $L^1(X,\mu)$ mit

$$\|[f]\|_1 := \int |f| d\mu$$

zu einem vollständigen normierten Raum. Wir werden in dieser Vorlesung diesen Raum nicht benötigen.

4. Das Lebesgue-Maß

In den vorangehenden Abschnitten haben wir ausgehend von einem gegebenen Maßraum Theorie der Integration für nichtnegative und für $\mathcal{L}^1$ Funktionen entwickelt. Wir haben jedoch (vom Zählmass abgesehen) keine Beispiele für Maße diskutiert. In diesem Abschnitt lernen wir DAS Maß auf dem Euklidischen Raum kennen. Wir werden diese Maß charakterisieren, aber seine Existenz hier (aus Zeitgründen) nicht beweisen.

Theorem (Lebesgue Maß). *Sei $\mathcal{B}$ die Borel-σ-Algebra auf $\mathbb{R}^N$. Dann gibt es ein eindeutiges Maß λ auf $\mathcal{B}$ mit*

$$\lambda(R) = |R|$$

für alle Rechtecke R. Dabei ist für ein Rechteck $R = (a_1, b_1) \times \ldots \times (a_N, b_N)$ das Volumen $|R|$ gegeben durch

$$|R| = \prod_{j=1}^{N} |b_j - a_j|.$$

Bemerkung. Der Satz besagt, dass das Volumen auf den Rechtecken (eindeutig) zu einem Maß fortgesetzt werden kann.

Bemerkung. Unter Rechteck verstehen wir hier (wie schon oben) immer offene Rechtecke.

Beweis. [Beweisidee] Wir können hier aus Zeitgründen den Beweis nicht geben. Deshalb sei kurz die Idee geschildert: *Existenz von λ:* Die Aussage ist eigentlich eine Forsetzungsaussage. Man kann sie zum Beispiel mittels des Caratheodoryschen Fortsetzungsatzes zeigen.
Eindeutigkeit von λ: Seien λ_1 und λ_2 Maße mit der angegbenen Eigenschaft. Dann ist

$$\{A \in \mathcal{B} \mid \lambda_1(A) = \lambda_2(A)\}$$

abgeschlossen unter Vereinigung aufsteigender Mengen und dem Schnitt von absteigenden Mengen (mit endlichem Maß). Weiterhin enthält das System die schnittstabile Menge aller Rechtecke. Damit gilt dann nach sogenannten Monotonen Klassen Argumenten $\lambda_1 = \lambda_2$ auf der (von den Rechtecken erzeugten) Borel-σ-Algebra. □

Definition. Es heißt λ das *Lebesguemaß* und das Integral bzgl. des Lebesguemaßes wird als *Lebesgue-Integral* bezeichnet.

Bemerkung (Lebesgue-Integral als Fortsetzung des Riemann-Integral). Es setzt das Lebegue-Integral das Riemann Integral fort in folgendem Sinne: Ist R ein (relativ kompaktes) Rechteck und $f : R \longrightarrow \mathbb{C}$ Riemann-integrierbar so gilt für die Fortsetzung $\widetilde{f}$ von f auf $\mathbb{R}^N$ durch 0

$$\int_R f dx = \int \widetilde{f} d\lambda.$$

(Beweis. Das ist klar für einfache Funktionen, deren Niveaumengen Rechtecke sind. Der allgemeine Fall folgt durch Grenzübergang.)

Das Lebesgue-Integral ist durch die Translationsinvarianz charakterisiert.

Folgerung. *Sei μ ein Maß auf $\mathbb{R}^N$. Dann sind äquivalent:*

(i) $\mu = \lambda$.

(ii) *Es ist μ translationsinvariant (d.h. es gilt $\mu(t+V) = \mu(V)$ für alle $t \in \mathbb{R}^N$ und alle messbaren V) mit $\mu((0,1)^N = 1$.*

Beweis. (i) $\Longrightarrow$ (ii). Wir müssen zeigen, dass λ die beiden angegebenen Eigenschaften hat: Es gilt $\lambda((0,1)^N) = 1$ wegen $|(0,1)^N| = 1$. Um die Translationsinvarianz zu zeigen, betrachten wir für $t \in \mathbb{R}^N$ das Maß λ_t mit $\lambda_t(A) := \lambda(t+A)$. Dann hat λ_t die charakteristischen Eigenschaften des Lebesguemaß und muss also mit diesem übereinstimmen aufgrund der schon diskutierten Eindeutigkeit.

(ii) $\Longrightarrow$ (i): Wir betrachten nur $N = 1$, d.h. $\mathbb{R}$. Es gilt $\mu(\{p\}) = 0$, $p \in \mathbb{R}$ (da sonst aufgrund der Translationsinvarianz $\mu((0,1)) = \infty$ gelten müsste.) Damit hängt also das Maß eines Interval nicht davon ab, ob die Randpunkte dazugehören oder nicht. Damit folgt aus der Tranlationsinvarianz also

$$\mu([0,1/n]) = \frac{1}{n}$$

für alle $n \in \mathbb{N}$, da

$$[0,1) = \bigcup_{k=0}^{n} (\frac{k}{n} + [0, \frac{1}{n})).$$

Damit hat dann (wieder nach Translationsinvarianz) jedes Intervall der Länge $1/n$ das Maß $1/n$ und dann jedes Intervall der Länge k/n das Maß k/n. Durch Ausschöpfen eines beliebigen Intervalles durch solche mit rationaler Länge erhält man dann, dass das Maß eines Intervalles gerade sein Länge ist. Damit folgt (aus dem vorangehenden Theorem) dann $\mu = \lambda$. $\square$

Beispiel (Nichtmessbare Menge). Betrachte $[0,1]$ mit der Einschränkung des Lebesguemaß λ und die Äquivalenzrelation

$$x \sim y :\Longleftrightarrow x - y \in \mathbb{Q}.$$

Wähle nun aus jeder Äquivalenzuklasse einen Repräsentanten und nenne die entstehende Menge E. (Dieser Schritt nutzt das Auswahlaxiom!) Wir zeigen, dass dieses E nicht messbar ist.
Definiere

$$E_r := \{x + r \mod 1 \mid x \in E\} \subset [0,1)$$

für $r \in \mathbb{Q}$. Wie man leicht sieht, gilt dann folgendes:

- $E_r \cap E_s = \emptyset$ für $r, s \in \mathbb{Q}$ mit $r \neq s \mod 1$.
- $[0,1) = \bigcup_{r \in \mathbb{Q} \cap [0,1)} E_r$.

Wäre E messbar, so wären auch alle E_r messbar und hätten das gleiche Lebesguemaß (aufgrund der Translationsinvarianz). Damit erhielte man in diesem Fall aus den beiden Punkten

$$1 = \lambda([0,1)] = \sum_{r \in \mathbb{Q} \cap [0,1)} \lambda(E_r) = \sum_{r \in \mathbb{Q} \cap [0,1)} \lambda(E).$$

Wir unterscheiden nun zwei Fälle:

$\lambda(E) = 0$*:* Dann gilt also $\lambda([0,1)) = 0$. Das ist ein Widerspruch.
$\lambda(E) > 0$*:* Dann gilt also $\lambda([0,1)) = \infty$. Das ist ein Widerspruch.

5. Der Satz von Fubini-Tonelli

In diesem Abschnitt lernen wir einen Satz kennen, der beim Ausrechnen von Integralen sehr nützlich ist.

Seien $(X, \mathcal{A}_X)$ und $(Y, \mathcal{A}_Y)$ messbare Räume. Dann erzeugen die Mengen der Form $A \times B$ mit $A \in \mathcal{A}_X$ und $B \in \mathcal{A}_Y$ eine σ-Algebra auf $X \times Y$, die wir als die Produkt-σ-Algebra $\mathcal{A}_X \otimes \mathcal{A}_Y$ bezeichnen. Sind weiterhin μ_X und μ_Y Maße auf $(X, \mathcal{A}_X)$ und $(Y, \mathcal{A}_Y)$, so wird durch

$$\mu\Big(\bigcup_{j=1}^{\infty}(A_j \times B_j)\Big) = \sum_{j=1}^{\infty} \mu_X(A_j)\mu_Y(B_j)$$

für $A_j \in \mathcal{A}_X$ und $B_j \in \mathcal{A}_Y$, $j = 1, \ldots, n$ mit $(A_j \times B_j) \cap (A_k \times B_k) = \emptyset$ für $j \neq k$ ein Maß μ auf $(X \times Y, \mathcal{A}_X \otimes \mathcal{A}_Y)$ definiert welches das *Produktmaß* von μ_X und μ_Y genannt wird. Das ist in der Tat ein nicht offensichtlicher Satz, auf dessen Beweis wir hier jedoch verzichten.

Theorem (Fubini-Tonelli). *Seien $(X, \mathcal{A}_X, \mu_X)$ und $(Y, \mathcal{A}_Y, \mu_Y)$ Maßräume. Sei ein μ-messbares $f : X \times Y \longrightarrow \mathbb{C}$ gegeben. Dann sind äquivalent:*

(i) $f \in \mathcal{L}^1(X \times Y, \mu)$.
(ii) *Es ist $|f(x, \cdot)| \in \mathcal{L}^1(Y, \mu_Y)$ für μ_X-fast alle $x \in X$ und die Funktion $F : X \longrightarrow \mathbb{C}$ mit*

$$F(x) = \begin{cases} \int_Y |f(x,y)|\, d\mu(y) & : \text{falls } |f(x,\cdot)| \in \mathcal{L}^1(Y,\mu_Y) \\ 0 & : \text{sonst} \end{cases}$$

gehört zu $\mathcal{L}^1(X, \mu_X)$.

In diesem Fall gilt

$$\int_{X \times Y} f(x,y)\, d\mu(x,y) = \int_X \left(\int_Y f(x,y)\, d\mu_Y(y) \right) d\mu_X(x).$$

Bemerkung.

- Natürlich gilt entsprechendes, wenn die Rollen von X und Y vertauscht werden.
- Ist $f \geq 0$ messbar, so besagt der Satz, dass auf jeden Fall gilt

$$\int f(x,y) d\mu(x,y) = \int_X \int_Y f(x,y) d\mu_Y(x) d\mu_X(x),$$

wobei allerdings beide Seiten ∞ sein können.

Der Beweis würde recht viel Zeit kosten (ohne aber über die Aussage hinausgehenden wesentlichen Erkentnisgewinn zu liefern). Daher geben wir ihn hier nicht. Wir bemerken stattdessen, dass (i)$\Longrightarrow$(ii) als Satz von Fubini bekannt ist und (ii)$\Longrightarrow$(i) als Satz von Tonelli.

6. Die Transformationsformel für das Lebesguemaß

In diesem Abschnitt geht es um das Lebesguemaß im $\mathbb{R}^N$ und das höherdimensionale Analogon zur Substitutionsregel.

THEOREM. *Seien U und V offene Mengen im $\mathbb{R}^N$, deren Ränder Lebesguenullmengen sind. Sei $\phi : U \longrightarrow V$ stetig differenzierbar und es gebe abgeschlossene Nullmengen $N_1 \subseteq U$ und $N_2 \subseteq V$, so dass $\phi : U \setminus N_1 \longrightarrow V \setminus N_2$ bijektiv mit stetig differenzierbarer Inverser ist. Dann ist für jede integrierbares $f : V \longrightarrow \mathbb{C}$ auch $f \circ \phi : U \longrightarrow$ integrierbar und es gilt*

$$\int_U f \circ \phi(x) |\det D\phi(x)| d\lambda(x) = \int_V f(y)\, d\lambda(y).$$

BEMERKUNG. Im eindimensionalen Fall (und für stetiges f) handelt es sich im wesentlichen um die schon bekannte Substitutionsregel. Dem Vertauschen der Reihenfolge der Grenzen entspricht das Bilden des Betrages der Ableitung.

Die vielleicht wichtigste Anwendung der Transformationsformel liegt in der Benutzung von neuen (dem Problem besser angepassten) Koordinatensystemen. Wir studieren nun noch einige gängige Koordinatensysteme.

BEISPIEL (Polarkoordinaten in der Ebene). Sei

$$\Phi :\ [0,\infty) \times [0, 2\pi] \to \mathbb{R}^2,\ \Phi(r,\varphi) = \begin{pmatrix} r\cos(\varphi) \\ r\sin(\varphi) \end{pmatrix}.$$

Dann ist

$$\Phi :\ (0,\infty) \times (0, 2\pi) \to \mathbb{R}^2 \setminus \{(x,0) \mid\ x \geq 0\}$$

bijektiv und stetig differenzierbar mit

$$D\Phi(x,y) = \begin{pmatrix} \cos\varphi & -r\sin\varphi \\ \sin\varphi & r\cos\varphi \end{pmatrix}$$

Für die Funktionaldeterminante gilt

$$\det D\Phi(x,y) = \cos\varphi r \cos\varphi + r\sin\varphi \sin\varphi = r.$$

Damit ergibt sich also (mit dem Satz von Fubini/Tonelli)

$$\int_{\mathbb{R}^2} f(x,y) d\lambda(x,y) = \int_{[0,\infty)} r \left(\int_{[0,2\pi]} f(r\cos\phi, r\sin\phi) d\lambda(\phi) \right) d\lambda(r).$$

Ist f radialsymmetrisch d.h. $f(x,y) = \widetilde{f}(|(x,y)|)$, so folgt

$$\int_{\mathbb{R}^2} f(x,y) d\lambda(x,y) = 2\pi \int_{[0,\infty)} r\widetilde{f}(r) d\lambda(r).$$

Entsprechend folgt mit der abgeschlossenen Kugel B_R mit Radius R um den Ursprung also

$$\int_{B_R} f(x,y)d\lambda(x,y) = 2\pi \int_{[0,R]} r\widetilde{f}(r)d\lambda(r).$$

Damit erhält man insbesondere für die Fläche $F(R)$ der Kugel mit Radius R (d.h. $\widetilde{f} = 1_{[0,R]}$) also

$$F(R) = \int_{B_R} 1d\lambda(x,y) = 2\pi \int_{[0,R]} r \cdot 1d\lambda(r) = \pi R^2.$$

Beispiel (Polar-/Kugelkoordinaten in $\mathbb{R}^3$). Sei

$$\Phi: \ [0,\infty) \times [0,2\pi] \times [0,\pi] \to \mathbb{R}^3,\ \Phi(r,\varphi,\theta) = \begin{pmatrix} r\cos(\varphi)\sin(\theta) \\ r\sin(\varphi)\sin(\theta) \\ r\cos(\theta) \end{pmatrix}.$$

Dann ist

$$\Phi: \ (0,\infty) \times (0,2\pi) \times (o,\pi) \to \mathbb{R}^3 \setminus \{(x,0,z) \mid \ z \in \mathbb{R},\ x \geq 0\}$$

bijektiv und stetig differenzierbar mit der Jacobi-Matrix

$$D\Phi(r,\varphi,\theta) = \begin{pmatrix} \cos\varphi\sin\theta & -r\sin\varphi\sin\theta & r\cos\varphi\cos\theta \\ r\sin\varphi\sin\theta & r\cos\varphi\sin\theta & r\sin\varphi\cos\theta \\ \cos\theta & 0 & -r\sin\theta \end{pmatrix}$$

und der Funktionaldeterminante

$$\det D\Phi(r,\varphi,\theta) = -r^2\sin\theta$$

Damit ergibt sich also (mit dem Satz von Fubini/Tonelli)

$$\int_{\mathbb{R}^3} f(x,y,z)d\lambda(x,y,z)$$
$$= \int_{[0,\infty)} r^2 \left(\int_{[0,2\pi]} (\int_{[0,\pi)} \sin\theta f(r\cos\phi\sin\theta, r\sin\phi\sin\theta, r\cos\theta)d\lambda(\theta))d\lambda(\phi) \right) d\lambda(r).$$

Ist f radialsymmetrisch d.h. $f(x,y,z) = \widetilde{f}(|(x,y,z)|)$, so folgt

$$\int_{\mathbb{R}^3} f(x,y,z)d\lambda(x,y,z) = 4\pi \int_{[0,\infty)} r\widetilde{f}(r)d\lambda(r).$$

Entsprechend folgt mit der abgeschlossenen Kugel B_R im $\mathbb{R}^3$ mit Radius R um den Ursprung also

$$\int_{B_R} f(x,y,z)d\lambda(x,y,z) = 4\pi \int_{[0,R]} r^2\widetilde{f}(r)d\lambda(r).$$

Damit erhält man insbesondere für das Volumen $V(R)$ der Kugel mit Radius R (d.h. $\widetilde{f} = 1_{[0,R]}$) also

$$V(R) = \int_{B_R} 1d\lambda(x,y,z) = 4\pi \int_{[0,R]} r^2 \cdot 1d\lambda(r) = \frac{4}{3}\pi R^3.$$

Beispiel. Berechnen von $\int_{\mathbb{R}} e^{-t^2}\, d\lambda$.
Setze

$$I := \int_{\mathbb{R}} e^{-t^2} d\lambda; \text{ und } J := \int_{\mathbb{R}^2} e^{-|(x,y)|^2} \; d\lambda(x,y).$$

Dann gilt nach dem Satz von Fubini/Tonelli aber $I^2 = J$. Mittels Polarkoordinaten in der Ebene, des Satzes von Fubini/Tonelli und der Tatsache, dass für stetige Funktionen das Lebesgue Integral und das Riemann Integral übereinstimmen, können wir weiterhin J ausrechnen zu

$$\begin{aligned}
J &= \int_{(0,\infty)\times[0,2\pi]} re^{-r^2} d\lambda(\phi, r) \\
\text{(Fubini/Tonelli)} \quad &= 2\pi \int_{(0,\infty)} re^{-r^2} d\lambda(r) \\
\text{(Monotone Konv.)} \quad &= 2\pi \lim_{R\to\infty} \int_{(0,R)} re^{-r^2} d\lambda(r) \\
&= 2\pi \lim_{R\to\infty} \int_r^R re^{-r^2} dr \\
&= 2\pi \lim_{R\to\infty} (-\frac{1}{2} e^{-r^2}|_0^R) \\
&= \pi.
\end{aligned}$$

Damit folgt $I = \sqrt{\pi}$.