

Gewöhnliche Differentialgleichungen und Dynamische Systeme II

Prof. Roland Gunesch
Vorlesung

Universität Hamburg
Wintersemester 2008-2009

INHALTSVERZEICHNIS

Vorwort	1
Literatur	2
Einleitung: Was sind dynamische Systeme und Differentialgleichungen?	2
Grundbegriffe	3
Das Verhalten der Lösungen von eindimensionalen Differentialgleichungen	6
Kreisabbildungen	7
Rotationszahlen von Kreis-Homöomorphismen	9
Verhalten der Rotationszahl bei Konjugation	11
Rotationszahl als Limes	13
Ordnungserhaltung von Kreisabbildungen	14
Hyperbolizität	17
Spezifikation	17
Markov-Partitionen	18
Eigenschaften von Anosov-Diffeomorphismen	22
Ergodentheorie	26
Birkhoffs Ergodensatz (oder: Wie wir die Zukunft vorhersagen)	30
Maß-theoretisches Mischen	33
Billiards	35

VORWORT

Dieses Skript entsteht während der Vorlesung „Gewöhnliche Differentialgleichungen und Dynamische Systeme II“ im Wintersemester 2008-2009 an der Universität Hamburg. Dieser Veranstaltung vorausgegangen sind verschiedene Vorlesungen in verwandten Gebieten, insbesondere „Einführung in Dynamische Systeme“, „Dynamische Systeme“, „Gewöhnliche Differentialgleichungen“, „Geometrische Theorie von Differentialgleichungen“. Zu „Einführung in Dynamische Systeme“ gibt es auch ein Skript auf meiner Homepage, welches sich als Einstieg eignet.

Allerdings ist eine solche Vorbereitung nicht wirklich erforderlich für diese Vorlesung. Formale Voraussetzungen für diese Veranstaltung sind lediglich die Vorlesungen Lineare Algebra 1+2 und Analysis 1-3. Wer vorher eine der o.g. Veranstaltungen besucht hat, ist bestens vorbereitet.

Es gibt sehr viel mehr Material in jedem der Themenbereiche „Differentialgleichungen“ und „Dynamische Systeme“, als in ein oder zwei Semestern vermittelt werden kann. Daher wird hier nicht versucht, das gesamte enorme Wissen dieses Gebiets vollständig zu übermitteln. Dafür wird hier versucht, die Darstellung besonders einfach und klar zu halten und ausgewählte wichtige Themen sorgfältig darzustellen.

LITERATUR

Folgende Bücher sind empfehlenswert.

- Hasselblatt, B. & Katok, A.: *A First Course in Dynamics. With a panorama of recent developments.* Cambridge University Press
- Katok, A. & Hasselblatt, B.: *Introduction to the Modern Theory of Dynamical Systems.* Cambridge University Press
- Denker, M.: *Einführung in die Analysis dynamischer Systeme.* Springer
- Arrowsmith, D.K. & Place, C. M.: *An introduction to Dynamical Systems.* Cambridge University Press
- Pollicott, M. & Yuri, M.: *Dynamical systems and ergodic theory.* Cambridge University Press
- Brin, M. & Stuck, G.: *Introduction to Dynamical Systems.* Cambridge University Press
- Ott, E.: *Chaos in Dynamical Systems, 2nd edition.* Cambridge University Press

Im Laufe des Semesters kommen noch ein paar Einträge zu dieser Liste dazu. Dennoch bleibt eine Literaturliste in diesem Gebiet stets unvollständig.

EINLEITUNG: WAS SIND DYNAMISCHE SYSTEME UND DIFFERENTIALGLEICHUNGEN?

Dynamische Systeme und Differentialgleichungen sind die Lehre von allen Dingen, die sich mit der Zeit ändern. Das beinhaltet das Universum, das Leben und den ganzen Rest. Folgendes sind typische Beispiele, die untersucht werden:

- Himmelsmechanik (Ist das Sonnensystem stabil?),
- biologische Populationen,
- das Wetter,
- physikalische Pendel,

-
- Computersimulationen wie „game of life“ oder Computer selbst,
 - mathematische Iterationsverfahren, z.B. das Newton-Verfahren.

Allgemein werden hier insbesondere folgende zwei wichtige mathematische Objekte behandelt:

(1) Lösungen von Differentialgleichungen

$$\frac{dx}{dt} = f(x),$$

(2) Iteration von Abbildungen

$$f : X \rightarrow X,$$

also

$$x_{n+1} = f(x_n).$$

Hier werden einige typische Konzepte erklärt, z.B.

- Chaos,
- Ordnung,
- Vorhersagbarkeit,
- Stabilität und
- Instabilität,
- Attraktoren,
- Schmetterlingseffekt,
- Information und
- Entropie.

Diese Vorlesung erwartet die folgenden Vorkenntnisse:

- Analysis: Satz über implizite Funktionen, Differenzieren im \mathbb{R}^n , elementare Maßtheorie.
- Lineare Algebra: Konjugation und Äquivalenz von Matrizen, Jordan-Normalform

Wie verstehen wir gewöhnliche Differentialgleichungen?

GRUNDBEGRIFFE

Definition. Für eine Abbildung $f : X \rightarrow X$ auf einer Menge X schreiben wir $f^2 := f \circ f$, $f^3 := f \circ f \circ f$,

$$f^k := f \circ \dots \circ f \quad (k\text{-malige Verkettung von } f).$$

Denn wir werden die k -malige Verkettung von Abbildungen oft brauchen, die Multiplikation von Werten dagegen selten (und für letztere kann man ohnehin problemlos $f(x)^k$ schreiben, da dies kaum mit $f(x^k)$ zu verwechseln ist).

Wir nennen f^k auch die k -fache **Iteration** von f .

Definition. Das **Orbit** von $x \in X$ einer Abbildung f ist die Folge

$$(x, f(x), f^2(x), \dots) = (f^k(x))_{k \in \mathbb{N}_0}.$$

Dabei muss f nicht invertierbar sein; ist das aber der Fall, so ist f^k für alle $k \in \mathbb{Z}$ definiert und wir können das **Orbit von x unter der invertierbaren Abbildung f** definieren als

$$(f^k(x))_{k \in \mathbb{Z}}.$$

In diesem Fall heißt $(f^k(x))_{k \in \mathbb{N}_0}$ das **positive Semiorbit** von x unter f .

Definition. Ein **Fixpunkt** von f ist ein Punkt $x \in X$ mit

$$f(x) = x.$$

Ein Punkt x heißt **periodisch** mit **Periode k** , wenn gilt

$$f^k(x) = x.$$

Das ist offensichtlich genau dann der Fall, wenn x ein Fixpunkt von f^k ist. Es ist nicht nötig, dass k den kleinsten möglichen Wert hat; wenn doch, heißt k die **minimale Periode** von x .

Es gilt also:

Lemma. Wenn ein Punkt periodisch ist mit Periode $k \in \mathbb{N}$, dann auch mit $l \cdot k$ für alle $l \in \mathbb{N}$.

Definition. Ein **Fluss** φ auf einer Menge X ist eine Abbildung $X \times \mathbb{R} \rightarrow X$, $(x, t) \mapsto \varphi_t(x)$, so dass gilt:

- $\varphi_0 = \text{id}$,
- für alle $s, t \in \mathbb{R}$ gilt $\varphi_s \circ \varphi_t = \varphi_{s+t}$.

Üblicherweise wird gefordert, dass φ mindestens C^1 ist (in (x, t) , also beiden Variablen); in vielen Fällen ist φ glatt. Sinn macht die Definition auch, wenn φ nur C^0 ist.

Flüsse treten auf natürliche Weise auf, wenn wir Differentialgleichungen untersuchen:

Example. Wenn $\dot{x} = f(x)$ eine Differentialgleichung auf dem \mathbb{R}^n ist mit der Eigenschaft, dass jede Lösung $x = x(t)$ sich auf ganz \mathbb{R} fortsetzen lässt, also für alle Zeiten t definiert ist, dann können wir einen Fluss definieren durch

$$\varphi_t(x_0) := x(t),$$

wobei $x = x(t)$ diejenige Lösung von $\dot{x} = f(x)$ ist mit $x(0) = x_0$.

Definition. Für einen Fluss φ heißt φ_t die **Zeit- t -Abbildung** von φ .

Lemma. Wenn φ ein C^k -Fluss ist, dann ist die Zeit- t -Abbildung φ_t ein C^k -Diffeomorphismus für alle t .

Beweis. φ_t ist invertierbar mit Umkehrabbildung φ_{-t} , da $\varphi_{-t} \circ \varphi_t = \varphi_{-t+t} = \varphi_0 = \text{id}$. Mit φ ist auch φ_t und φ_{-t} eine C^k -Abbildung. \square

Die Begriffe von Fixpunkt und periodischem Orbit lassen sich sofort für Flüsse definieren:

Definition. Das **Orbit** von $x \in X$ eines Flusses φ ist die (überabzählbare) Sequenz

$$(\varphi_t(x))_{t \in \mathbb{R}}.$$

Statt Orbit sagt man auch die **Bahn** oder die **Trajektorie** von x .

Definition. Ein **Fixpunkt** von φ ist ein Punkt $x \in X$, so dass für alle $t \in \mathbb{R}$ gilt:

$$\varphi_t(x) = x.$$

Ein Punkt x heißt **periodisch** mit **Periode** τ , wenn gilt

$$\varphi_\tau(x) = x.$$

Wie vorhin ist es in dieser Definition nicht nötig, dass τ den kleinsten möglichen Wert hat; wenn doch, heißt τ die **minimale Periode** von x . Es gilt also wieder: Wenn ein Punkt periodisch ist mit Periode $\tau \in \mathbb{R}$, dann auch mit $l \cdot \tau$ für alle $l \in \mathbb{N}$.

Definition. Sei X metrischer (oder topologischer) Raum. Ein Fixpunkt x_0 einer Abbildung $f : X \rightarrow X$ oder eines Flusses $\varphi : X \times \mathbb{R} \rightarrow X$ heißt **stabil**, wenn für jede Umgebung U von x_0 eine offene Umgebung V von x_0 existiert, so dass für alle $x \in V$ gilt: Für alle $n \geq 0$ ist $f^n(x) \in U$, bzw. für alle $t \geq 0$ ist $\varphi_t(x) \in U$.

Example. Der Fluss der linearen Differentialgleichung $\dot{u} = Au$ im \mathbb{R}^n hat einen Fixpunkt $x_0 = 0$. Dieser ist stabil, wenn die Realteile aller Eigenwerte von A negativ oder 0 sind.

Definition. Sei X metrischer (oder topologischer) Raum. Ein Fixpunkt x_0 einer Abbildung $f : X \rightarrow X$ oder eines Flusses $\varphi : X \times \mathbb{R} \rightarrow X$ heißt **anziehend** bzw. **attrahierend**, wenn es eine offene Umgebung U von x_0 gibt, so dass für alle $x \in U$ gilt:

$$f^n(x) \rightarrow x_0 \quad \text{für } n \rightarrow +\infty \quad \text{bzw.}$$

$$\varphi_t(x) \rightarrow x_0 \quad \text{für } t \rightarrow +\infty.$$

Der Fixpunkt x_0 einer Abbildung, die außerdem invertierbar ist, heißt **abstoßend**, wenn er bei Zeitumkehr anziehend ist, d.h. wenn es eine offene Umgebung U von x_0 gibt, so dass für alle $x \in U$ gilt:

$$f^n(x) \rightarrow x_0 \quad \text{für } n \rightarrow -\infty \quad \text{bzw.}$$

$$\varphi_t(x) \rightarrow x_0 \quad \text{für } t \rightarrow -\infty.$$

Example. Der Fixpunkt $x_0 = 0$ des Flusses der linearen Differentialgleichung $\dot{u} = Au$ im \mathbb{R}^n ist anziehend, wenn alle Eigenwerte von A strikt negativen Realteil haben. Der Fixpunkt ist abstoßend, wenn die Realteile aller Eigenwerte positiv sind. Wenn Eigenwerte mit Realteilen verschiedener Vorzeichen existieren, ist der Fixpunkt weder anziehend noch abstoßend.

DAS VERHALTEN DER LÖSUNGEN VON EINDIMENSIONALEN
DIFFERENTIALGLEICHUNGEN

Mit dem folgenden Satz können wir das Verhalten von eindimensionalen Differentialgleichungen komplett verstehen:

Theorem. Sei $f : \mathbb{R} \rightarrow \mathbb{R}$ Lipschitz. Dann existiert ein Fluss φ für die Differentialgleichung $\dot{u} = f(u)$. Wenn $f(x) > 0$ ist, dann gilt

$$\lim_{t \rightarrow +\infty} \varphi_t(x) = \inf\{x_0 > x \mid f(x_0) = 0\} \quad \text{und} \\ \lim_{t \rightarrow -\infty} \varphi_t(x) = \sup\{x_0 < x \mid f(x_0) = 0\}.$$

Wenn $f(x) < 0$ ist, dann gilt

$$\lim_{t \rightarrow +\infty} \varphi_t(x) = \sup\{x_0 < x \mid f(x_0) = 0\} \quad \text{und} \\ \lim_{t \rightarrow -\infty} \varphi_t(x) = \inf\{x_0 > x \mid f(x_0) = 0\}.$$

D.h. die Lösung des Anfangswertproblems $\dot{u} = f(u)$, $u(0) = x$ konvergiert für $t \rightarrow \pm\infty$ gegen die nächste Ruhelage rechts von x oder links von x (abhängig vom Vorzeichen von $f(x)$ und dem von $\pm\infty$).

Beweis. Der Fluss existiert wegen der globalen Lipschitz-Bedingung. Wir betrachten den Fall $f(x) > 0$ und $t \rightarrow +\infty$; die anderen Fälle sind analog. Der Wert $\varphi_t(x)$ ist wachsend, solange $f(\varphi_t(x))$ positiv ist. Wir zeigen, dass dies für alle $t \in \mathbb{R}$ der Fall ist.

Wäre $f(\varphi_{t_0}(x)) = 0$ für irgendein $t_0 \in \mathbb{R}$, dann wäre $\varphi_{t_0}(x)$ ein Fixpunkt, und damit x ein Fixpunkt, was $f(x) > 0$ widerspräche und somit nicht sein kann. Wäre $f(\varphi_{t_0}(x)) < 0$ für irgendein $t_0 \in \mathbb{R}$, dann müsste wegen der Stetigkeit von f und des Flusses φ ein $t_1 \in (0, t_0)$ existieren mit $f(\varphi_{t_1}(x)) = 0$. Also können wir diesen Fall auch ausschließen.

Wir haben gezeigt, dass $f(\varphi_t(x)) > 0$ für alle $t \in \mathbb{R}$. Somit ist $f(\varphi_t(x))$ monoton wachsend in t und hat entweder einen endlichen Grenzwert $x_1 > x_0$ oder ist unbeschränkt. Wenn x_1 endlich ist, gilt: Für jedes $\varepsilon > 0$ ist für t groß genug die Zahl $\varphi_t(x)$ in der ε -Umgebung von x_1 und monoton wachsend. Daher ist notwendigerweise die Geschwindigkeit $f(\varphi_t(x)) = \frac{d}{dt}\varphi_t(x)$ beliebig klein. Deshalb gilt wegen der Stetigkeit von f , dass

$$\begin{aligned} f(x_1) &= f\left(\lim_{t \rightarrow \infty} \varphi_t(x)\right) \\ &= \lim_{t \rightarrow \infty} f(\varphi_t(x)) = 0. \end{aligned}$$

Also ist in diesem Fall x_1 ein Fixpunkt. Wie bereits gezeigt, kann $\varphi_t(x)$ keinen weiteren Fixpunkt überquert haben, also ist x_1 der erste Fixpunkt rechts von x . \square

Der eben bewiesene Satz ist deshalb beachtlich, weil im mehrdimensionalen Fall Aussagen über das Verhalten von Lösungen einer

Differentialgleichung sehr viel schwieriger zu machen sind. Es gibt dort auch außer Fixpunkten noch weitere Möglichkeiten für das Verhalten im Grenzwert.

KREISABBILDUNGEN

Im Folgenden sehen wir uns eine ganz spezielle Klasse von dynamischen Systemen an: Abbildungen auf dem Kreis. Diese sind einfach genug, so dass wir sie noch recht leicht analysieren können, haben aber andererseits schon viele interessante Eigenschaften, die zu untersuchen sich lohnt.

Zur Erinnerung: diskrete dynamische Systeme sind vom Typ

$$f : X \rightarrow X,$$

wobei X typischerweise ein metrischer (oder zumindest topologischer) Raum ist, oft mit „glatter Struktur“ (d.h. wir können differenzieren), und f ist typischerweise mindestens stetig, oft glatt und umkehrbar, d.h. Diffeomorphismus.

Für Kreisabbildungen untersuchen wir Abbildungen $f : X \rightarrow X$ auf

$$X = \mathbb{R}/\mathbb{Z},$$

d.h. $[0, 1)$ mit Addition modulo 1, dem sogenannten „Einheitskreis“ S^1 .

Remark. Genausogut könnten wir Abbildungen $g : Y \rightarrow Y$ studieren auf dem Raum $Y = \{z \in \mathbb{C} : |z| = 1\}$, der verwirrenderweise ebenfalls „Einheitskreis“ heißt, diesen Namen eigentlich eher verdient, und in anderen Büchern ebenfalls mit dem Symbol S^1 bezeichnet wird. Ob wir auf X oder Y arbeiten, macht keinen Unterschied, denn die Abbildung $h : X \rightarrow Y$, $h(x) = \exp(2\pi ix)$ konjugiert jedes solche $g : Y \rightarrow Y$ mit einem $f = h \circ g \circ h^{-1} : X \rightarrow X$. Wir verwenden stets X und nicht Y ; somit kann die Gruppenoperation additiv geschrieben werden (statt multiplikativ) und der Kreis hat Periode 1 (statt Periode 2π).

Sei also $f : S^1 \rightarrow S^1$ ein Homöomorphismus (d.h. umkehrbar mit f und f^{-1} stetig) und orientierungserhaltend.

Definition. Eine stetige Funktion $F : \mathbb{R} \rightarrow \mathbb{R}$ heißt ein **Lift** von f , wenn gilt:

$$f \circ \pi = \pi \circ F,$$

wobei $\pi : \mathbb{R} \rightarrow S^1$, $\pi(x) = [x]$ die Projektion von $x \in \mathbb{R}$ auf seine Äquivalenzklasse $[x]$ in S^1 ist. Das heißt, folgendes Diagramm kommutiert:

$$\begin{array}{ccc} \mathbb{R} & \xrightarrow{F} & \mathbb{R} \\ \pi \downarrow & & \downarrow \pi \\ S^1 & \xrightarrow{f} & S^1 \end{array}$$

Beispiele:

- (1) $f = \text{id}, F = \text{id}$,
- (2) $f = \text{id}, F = \text{id} + k, k \in \mathbb{Z}$,
- (3) $f = \text{id} + \alpha \pmod{1}, F = \text{id} + \alpha + k, \alpha \in \mathbb{R}, k \in \mathbb{Z}$.

Lemma. Für jeden Homöomorphismus f auf S^1 ist der Lift F von f eindeutig bestimmt bis auf Addition von konstanten Funktionen $k \in \mathbb{Z}$.

Beweis. Für konkretes $[x] \in S^1$ ist klar, dass $F(x)$ eindeutig bestimmt ist bis auf Addition von $k \in \mathbb{Z}$, denn dies trifft für x zu und wegen der Kommutativität von obigem Diagramm ist $[F(x)]$ gleich $f([x])$, also ist $F(x)$ eindeutig bestimmt bis auf Addition von $k \in \mathbb{Z}$.

Wir wählen nun $x \in \mathbb{R}$ fest. Wegen der Forderung, dass der Lift stetig sein muss, gibt es nur eine Fortsetzung von $F(x)$ zu einer auf ganz \mathbb{R} stetigen Funktion. \square

Remark. So ein Lift existiert immer.

Remark. Wegen Homöomorphie von f gilt dass $F(x+1) = F(x) + 1$ für alle $x \in \mathbb{R}$ gilt oder $F(x+1) = F(x) - 1$ für alle $x \in \mathbb{R}$ gilt. Da f orientierungserhaltend ist, scheidet der zweite Fall aus. Daraus folgt sofort:

$$F(x+k) = F(x) + k \quad \text{für alle } x \in \mathbb{R} \text{ und alle } k \in \mathbb{Z}.$$

Lemma. Wenn für $x, y \in \mathbb{R}$ gilt, dass

$$|x - y| \leq k \in \mathbb{N}_0,$$

dann gilt auch

$$|F(x) - F(y)| \leq k.$$

Beweis. Wegen Bijektivität von f gilt die Behauptung für $k = 1$: Der Graph von F kann auf jedem Intervall der Länge 1 nur um 1 wachsen, d.h. das Bild des Intervalls $(x, x+1)$ unter F ist enthalten im Intervall $(F(x), F(x)+1)$. Jetzt nutzen wir die Gleichung

$$F(x+m) = F(x) + m \quad (\text{für } x \in \mathbb{R}, m \in \mathbb{Z})$$

aus: Aus

$$|x - y| \leq k \in \mathbb{Z}$$

folgt

$$|x - m - y| \leq 1$$

für geeignetes $m \in \mathbb{Z}$, und zwar $m \in [0, k-1]$ für $x \geq y$ und $m \in [-(k-1), 0]$ für $x < y$. Weil $x' = x - m$ die Abschätzung $|x' - y| \leq 1$ erfüllt, gilt

$$|F(x-m) - F(y)| = |F(x') - F(y)| \leq 1$$

und deshalb

$$\begin{aligned} |F(x) - F(y)| &= |F(x) - F(x - m) + F(x - m) - F(y)| \\ &= |F(x') - F(y) + m| \\ &\leq 1 + |m| \\ &= k. \end{aligned}$$

□

Durch wiederholte Anwendung erhalten wir:

Corollary. Wenn für $x, y \in \mathbb{R}$ gilt, dass

$$|x - y| \leq k \in \mathbb{N}_0,$$

dann gilt auch für alle $n \in \mathbb{N}$, dass

$$|F^n(x) - F^n(y)| \leq k.$$

ROTATIONSZAHLEN VON KREIS-HOMÖOMORPHISMEN

Definition. Sei $f : S^1 \rightarrow S^1$ ein orientierungserhaltender Kreishomöomorphismus. Die **Rotationszahl** von f ist definiert durch

$$\rho(f) := \limsup_{n \rightarrow \infty} \frac{F^n(x) - x}{n} \pmod{1}$$

für einen (beliebigen) Lift F von f und für ein (beliebiges) $x \in S^1$.

Dieser Wert $\rho(f)$ hängt nicht von der Wahl von F ab, da sich verschiedene Lifts von f nur um Konstanten unterscheiden, und hängt auch nicht von x ab wegen vorigem Korollar.

Die Rotationszahl $\rho(f)$ misst die „durchschnittliche Verschiebung“ von x auf dem Kreis, wenn f darauf oft angewendet wird. Wobei diese „Zahl“ in Wirklichkeit ein Element von S^1 ist.

Aus der Definition folgt sofort:

Remark. $\rho(f^n) = n\rho(f)$ für alle $n \in \mathbb{N}$.

Wir werden jetzt sehen, dass f genau dann periodische Orbits hat, wenn die Rotationszahl Äquivalenzklasse von rationalen Zahlen ist.

Definition. Wir nennen $s \in S^1$ **rational**, wenn $s = [t]$ mit $t \in \mathbb{Q}$ ist, ansonsten **irrational**.

Theorem. Wenn ein orientierungserhaltender Kreishomöomorphismus f ein periodisches Orbit hat, dann ist die Rotationszahl $\rho(f)$ rational.

Genauer gilt: Wenn es $q \in \mathbb{Z}$ und $x \in S^1$ gibt mit $f^q(x) = x$, dann ist $\rho(f) \in \left(\frac{1}{q}\mathbb{Z}\right) / \mathbb{Z}$, das heißt

$$\rho(f) = \left[\frac{m}{q} \right] \quad \text{mit } m \in \mathbb{Z}.$$

Beweis. Setze für x dieses periodische Orbit ein: $f^q(x) = x$, also $F^q(x) = x + k$, $k \in \mathbb{Z}$. Es folgt also

$$F^{l \cdot q}(x) = x + l \cdot k \quad \text{für alle } l \in \mathbb{Z}.$$

Somit gilt

$$\frac{F^{lq}(x) - x}{lq} = \frac{lk}{lq} = \frac{k}{q}.$$

Für beliebiges $m \in \mathbb{N}$ ist $m = ql + r$ mit $r \in \{0, \dots, q-1\}$. Also gilt

$$\begin{aligned} \frac{F^m(x) - x}{m} &= \frac{F^{lq+r}(x) - x}{lq + r} \\ &= \frac{F^r(x + lk) - x}{lq + r} \\ &= \frac{F^r(x) + lk - x}{lq + r} \\ &\xrightarrow{l \rightarrow \infty} \frac{k}{q} \end{aligned}$$

wie behauptet. □

Die Umkehrung gilt auch:

Theorem. *Wenn ein orientierungserhaltender Kreishomöomorphismus f keine periodischen Orbits hat, dann ist seine Rotationszahl $\rho(f)$ irrational. D.h., wenn die Rotationszahl rational ist, dann hat f ein periodisches Orbit.*

Beweis. Wir nehmen also Rationalität der Rotationszahl an und folgern, dass es ein periodisches Orbit gibt. Wenn

$$\rho(f) = \left[\frac{p}{q} \right],$$

dann gilt

$$\rho(f^q) = q\rho(f) = [0].$$

Definiere

$$g := f^q.$$

Ein periodisches Orbit von f mit Periode q ist ein Fixpunkt von g . Wir wissen also, dass g Rotationszahl Null hat, und suchen einen Fixpunkt von g . Sei G ein Lift von g mit $G(0) \in [0, 1)$. Wegen der Bedingung $G(0) \in [0, 1)$ ist ein Fixpunkt von g auch ein Fixpunkt von G .

Annahme: g hat keinen Fixpunkt. Dann gilt entweder

$$G(x) - x > 0 \quad \text{für alle } x \in \mathbb{R}$$

oder

$$G(x) - x < 0 \quad \text{für alle } x \in \mathbb{R}.$$

Letztere Möglichkeit scheidet aus wegen der Bedingung $G(0) \in [0, 1)$.

Da $x \mapsto G(x) - x$ eine periodische stetige Funktion ist, also einer stetigen Funktion auf dem (kompakten) Kreis entspricht, wird das Minimum angenommen und ist positiv (wegen $G(x) - x > 0$ und Kompaktheit). Ebenso wird das Maximum angenommen und ist < 1 . Also gibt es ein $\varepsilon > 0$, so dass für alle $x \in R$ gilt:

$$G(x) - x \in (\varepsilon, 1 - \varepsilon).$$

Deswegen gilt

$$\begin{aligned} \limsup_{n \rightarrow \infty} \frac{G^n(x) - x}{n} &= \limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} G(G^i(x)) - G^i(x) \\ &= \limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} G(y_i) - y_i \quad (\text{mit } y_i := G^i(x)) \\ &\in (\varepsilon, 1 - \varepsilon) \\ &\neq 0 \pmod{1}. \end{aligned}$$

Dies ist ein Widerspruch dazu, dass g die Rotationszahl Null hat. \square

Um die Notation zu vereinfachen, unterscheiden wir hierbei nach Möglichkeit nicht mehr zwischen x und $[x]$ und schreiben für $[x]$ entweder „ $x \pmod{1}$ “ oder einfach x . Dabei müssen wir nur aufpassen, dass ein solches x nur bis auf Addition von ganzen Zahlen bestimmt ist; z.B. dürfen wir nicht aus $x_n > y_n + \frac{1}{2}$ versehentlich schließen, dass $\lim_n x_n \neq \lim_n y_n$ sei, denn es kann ja $\lim_n x_n = \lim_n y_n + 1$ sein, und natürlich ist $\lim_n y_n + 1 = \lim_n y_n \pmod{1}$.

VERHALTEN DER ROTATIONSZAHL BEI KONJUGATION

Wir wollen natürlich möglichst viele Abbildungen auf S^1 untersuchen. Welche konkreten Beispiele verstehen wir gut genug, um die Rotationszahl leicht berechnen zu können? Bis jetzt eigentlich recht wenige, denn die direkte Verwendung der Definition ist nur in wenigen Fällen leicht. Leicht ist es zum Beispiel bei der **Rotation** $R_\alpha : S^1 \rightarrow S^1$,

$$R_\alpha(x) = x + \alpha \pmod{1}$$

(bzw. $R_\alpha([x]) = [x + \alpha]$). Hier sehen wir sofort, dass $\rho(R_\alpha) = \alpha$ ist. Was wir brauchen, ist ein Mechanismus, um größere Klassen von Abbildungen auf einmal zu untersuchen. Hier ist einer: Wir untersuchen die Veränderung (oder deren Ausbleiben) der Rotationszahl bei Konjugation. Zur Erinnerung:

Definition. Für zwei diskrete dynamische Systeme $f : X \rightarrow X$ und $g : Z \rightarrow Z$ auf topologischen Räumen X, Z heißen f und g zueinander **konjugiert**, wenn gilt

$$g = h^{-1} \circ f \circ h$$

mit einem Homöomorphismus $h : X \rightarrow Z$.

Theorem. („Konjugation ändert die Rotationszahl nicht.“) Sei f wie bisher ein orientierungserhaltender Kreishomöomorphismus, h ebenfalls. Dann gilt:

$$\rho(h^{-1} \circ f \circ h) = \rho(f).$$

Mit anderen Worten, wenn $f : S^1 \rightarrow S^1$ und $g : S^1 \rightarrow S^1$ zueinander konjugiert sind, dann haben f und g dieselbe Rotationszahl.

Beweis. Sei F ein Lift von f und sei H ein Lift von h . Dann gilt $\pi \circ H = h \circ \pi$ und damit

$$\begin{aligned} \pi \circ H^{-1} &= h^{-1} \circ h \circ \pi \circ H^{-1} \\ &= h^{-1} \circ \pi \circ H \circ H^{-1} \\ &= h^{-1} \circ \pi, \end{aligned}$$

also ist H^{-1} ein Lift von h^{-1} .

Außerdem ist $H^{-1} \circ F \circ H$ ein Lift von $h^{-1} \circ f \circ h$, denn

$$\begin{aligned} \pi \circ H^{-1} \circ F \circ H &= h^{-1} \circ \pi \circ F \circ H \\ &= h^{-1} \circ f \circ \pi \circ H \\ &= h^{-1} \circ f \circ h \circ \pi. \end{aligned}$$

Wir müssen nun zeigen, dass der Term

$$|(H^{-1} \circ F \circ H)^n - F^n|$$

„nicht allzu groß“ ist, denn dies ist die Differenz der Terme, die in die Berechnung der Rotationszahl von $h^{-1} \circ f \circ h$ und der von f eingehen. Dazu wählen wir zuerst den Lift H von h so, dass $H(0) \in [0, 1)$ ist. Damit schätzen wir ab, dass für alle $x \in \mathbb{R}$ gilt, dass

$$|H(x) - x| \leq 2 \quad \text{und} \quad |y - H^{-1}(y)| \leq 2.$$

(und somit). Somit gilt auch

$$|F^n(H(x)) - F^n(x)| \leq 2.$$

Dadurch sehen wir, dass

$$\begin{aligned} |(H^{-1} \circ F \circ H)^n(x) - F^n(x)| &= |H^{-1}(F^n(H(x))) - F^n(x)| \\ &= |H^{-1}(F^n(H(x))) - F^n(H(x)) + F^n(H(x)) - F^n(x)| \\ &\leq 4. \end{aligned}$$

Folgerung:

$$\begin{aligned}
 \rho(h^{-1} \circ f \circ h) &= \limsup_{n \rightarrow \infty} \frac{(H^{-1} \circ F \circ H)^n(x) - x}{n} \\
 &= \limsup_{n \rightarrow \infty} \frac{F^n(x) - x}{n} \\
 &= \rho(f).
 \end{aligned}$$

Dieses Argument funktioniert mit beliebigem x , obwohl ein einziges x für die Rotationszahl schon genug wäre. \square

ROTATIONSZAHL ALS LIMES

Nachdem wir schon wissen, dass $\rho(f)$ weder von x noch von der Wahl des Lifts F von f abhängt, zeigen wir nun, dass der \limsup auch nicht von der Folge $n \rightarrow \infty$ abhängt, also der Limes existiert. Dies rechtfertigt auch die Bezeichnung „Rotations-Zahl“ (im Gegensatz zu „Rotations-Menge“ oder „Rotations-Intervall“).

Theorem. Die Rotationszahl von f ist gleich

$$\rho(f) = \lim_{n \rightarrow \infty} \frac{F^n(x) - x}{n} \pmod{1}$$

für einen beliebigen Lift F von f und einen beliebigen Punkt $x \in \mathbb{R}$. D.h., der Limes existiert immer.

Beweis. Für rationales $\rho(f)$ haben wir dies bereits bewiesen, denn wir haben gezeigt, dass in diesem Fall ein periodisches Orbit $[x]$ von f existiert. Also ist

$$F^n(x) = x + k \quad \text{mit } k \in \mathbb{Z}.$$

Genau wie im Beweis der Existenz eines periodischen Orbits von f gilt wieder

$$F^m(x) = F^{lq+r}(x) = F^r(x + lk) = F^r(x) + lk$$

und wieder gilt

$$\begin{aligned}
 \frac{F^m(x) - x}{m} &= \frac{F^{lq+r}(x) - x}{lq + r} \\
 &= \frac{F^r(x) + lk - x}{lq + r} \\
 &\xrightarrow{l \rightarrow \infty} \frac{k}{q}.
 \end{aligned}$$

Also existiert hier der Limes.

Bleibt noch der Fall, dass die Rotationszahl irrational ist, also keine periodischen Orbits existieren. Sei $k_n \in \mathbb{Z}$ so, dass

$$F^n(x) - x \in [k_n, k_n + 2]$$

für ein $x \in \mathbb{R}$ gilt. OBdA gilt dies auch für alle $x \in R$ gleichzeitig, denn $F^n - \text{id}$ ist eine periodische Funktion, die jedes Intervall der Länge 1 in ein Intervall von Länge höchstens 1 abbildet. Dann gilt:

$$\left| \frac{F^n(0) - 0}{n} - \frac{k_n}{n} \right| \leq \frac{2}{n}.$$

Also gilt für alle $m \in \mathbb{N}$, dass

$$\begin{aligned} F^{mn}(0) - 0 &= (F^n(F^{n(m-1)}(0)) - F^{n(m-1)}(0)) \\ &\quad + (F^{n(m-1)}(0) - F^{n(m-2)}(0)) \\ &\quad \vdots \\ &\quad + (F^{n \cdot 1}(0) - F^{n \cdot 0}(0)) \\ &\in [mk_n, m(k_n + 2)]. \end{aligned}$$

Also ist

$$\left| \frac{F^{mn}(0) - 0}{mn} - \frac{k_n}{n} \right| \leq \frac{2}{n}.$$

Somit gilt

$$\begin{aligned} \left| \frac{F^m(0) - 0}{m} - \frac{F^n(0) - 0}{n} \right| &\leq \left| \frac{F^m(0) - 0}{m} - \frac{k_m}{m} \right| \\ &\quad + \left| \frac{k_m}{m} - \frac{F^{mn}(0) - 0}{mn} \right| \\ &\quad + \left| \frac{F^{mn}(0) - 0}{mn} - \frac{k_n}{n} \right| \\ &\quad + \left| \frac{k_n}{n} - \frac{F^n(0) - 0}{n} \right| \\ &\leq 4 \left(\frac{1}{n} + \frac{1}{m} \right). \end{aligned}$$

Dies ist eine Cauchy-Folge, also konvergent. □

ORDNUNGSERHALTUNG VON KREISABBILDUNGEN

Sei f wie bisher ein orientierungserhaltender Kreishomöomorphismus. Wir wissen schon: Wenn die Rotationszahl $\rho(f)$ irrational ist, dann gibt es kein periodisches Orbit. Also ist jedes Orbit $(f^i(x))_{i \in \mathbb{Z}}$ von der Art, dass jedes $f^i(x)$ zwischen anderen Punkten des Orbits liegt.

Der Kreis ist (im Gegensatz zu \mathbb{R}) keine total geordnete Menge. Dennoch können wir von einer **Ordnung** von Punkten auf dem Kreis sprechen: 3 Punkte a, b, c auf dem Kreis können nämlich entweder in dieser Reihenfolge a, b, c oder in der Reihenfolge a, c, b liegen; im letzteren Fall hätten sie nicht dieselbe Ordnung auf dem Kreis.

Entsprechend gilt der Begriff der Ordnung auch für mehr als 3 Punkte, besonders für unendlich viele:

Definition. Wenn $f, g : S^1 \rightarrow S^1$ Kreisabbildungen sind, dann sagen wir, das Orbit von $x \in S^1$ unter f habe **dieselbe Ordnung** wie das Orbit von $y \in S^1$ unter g , wenn für alle $i, j, k \in \mathbb{Z}$ gilt: Die Ordnung der 3 Punkte

$$f^i(x), f^j(x), f^k(x)$$

(in dieser Reihenfolge) auf dem Kreis ist gleich der Ordnung von

$$g^i(y), g^j(y), g^k(y)$$

(in dieser Reihenfolge).

Theorem. Wenn ein orientierungserhaltender Kreishomöomorphismus f irrationale Rotationszahl $\rho(f)$ hat, dann ist die Ordnung von jedem Orbit von f gleich der Ordnung eines Orbits der Rotation $R_{\rho(f)}$.

Hierbei ist die Rotation R_α gegeben durch $R_\alpha(x) := x + \alpha \pmod{1}$. Für das im Satz erwähnte Orbit der Rotation R_α ist es egal, von welchem Startwert aus es beginnt, denn alle Orbits von R_α sind gleich bis auf Verschiebung.

Für den Beweis brauchen wir noch folgendes Lemma:

Lemma. Sei $\rho(f)$ irrational und F wie bisher ein Lift von f . Dann gilt:

(1) Wenn für irgendein $x \in \mathbb{R}$, $m_1, m_2, n_1, n_2 \in \mathbb{Z}$ gilt, dass

$$F^{n_1}(x) + m_1 < F^{n_2}(x) + m_2$$

ist, dann gilt auch für alle anderen $y \in \mathbb{R}$, dass

$$F^{n_1}(y) + m_1 < F^{n_2}(y) + m_2.$$

(2) Die Abbildung $H : \Omega \rightarrow \Lambda$ mit

$$\Omega := \{m + n\rho \mid m, n \in \mathbb{Z}\}, \quad \rho = \rho(f),$$

$$\Lambda := \{F^n(0) + m \mid m, n \in \mathbb{Z}\},$$

die definiert ist durch

$$H(m + n\rho) := F^n(0) + m,$$

erhält die Ordnung auf \mathbb{R} . (H ist wohldefiniert wegen der Irrationalität von ρ .)

Beweis. Zu (1): Wenn dies nicht gilt, dann gibt es $y \in \mathbb{R}$ mit $F^{n_1}(y) + m_1 \geq F^{n_2}(y) + m_2$, und wegen dem Zwischenwertsatz gibt es dann auch ein $y \in \mathbb{R}$ mit

$$F^{n_1}(y) + m_1 = F^{n_2}(y) + m_2.$$

Also ist $F^{n_1}(y) = F^{n_2}(y) + k$, $k \in \mathbb{Z}$. Damit hat $f^{n_1 - n_2}$ einen Fixpunkt, und damit wäre $\rho(f)$ rational. Widerspruch zur Annahme.

Zu (2): H ist bijektiv, weil aus $F^{n_1}(0) = F^{n_2}(0) + m_2 - m_1$ wieder folgt, dass 0 ein Fixpunkt von $f^{n_1-n_2}$ ist. Also ist H invertierbar. Es genügt zu zeigen: Die Umkehrfunktion H^{-1} ist monoton steigend.

Sei also $m_1, m_2, n_1, n_2 \in \mathbb{Z}$ gegeben mit

$$F^{n_1}(0) + m_1 < F^{n_2}(0) + m_2.$$

Zu zeigen ist nun, dass $m_1 + n_1\rho < m_2 + n_2\rho$ gilt.

Zunächst nehmen wir an, dass $n_1 > n_2$. Es gilt:

$$F^{n_1-n_2}(F^{n_2}(0)) - F^{n_2}(0) < m_2 - m_1.$$

Wegen (1) gilt diese Aussage auch, wenn wir $x := F^{n_2}(0)$ ersetzen durch $y = 0$, d.h.

$$F^{n_1-n_2}(0) - 0 < m_2 - m_1.$$

Hier können wir wieder (1) anwenden und 0 ersetzen durch $F^{n_1-n_2}(0)$. Dann erhalten wir

$$F^{2(n_1-n_2)}(0) - F^{n_1-n_2}(0) < m_2 - m_1.$$

Addition der letzten 2 Gleichungen gibt

$$F^{2(n_1-n_2)}(0) - 0 < 2(m_2 - m_1).$$

Per Induktion können wir genauso zeigen:

$$F^{k(n_1-n_2)}(0) - 0 < k(m_2 - m_1) \quad \forall k \in \mathbb{Z}.$$

Deshalb gilt

$$\begin{aligned} \rho(f) &= \lim_{n \rightarrow \infty} \frac{F^n(0) - 0}{n} \\ &= \lim_{k \rightarrow \infty} \frac{F^{k(n_1-n_2)}(0) - 0}{k(n_1-n_2)} \\ &\leq \frac{m_2 - m_1}{n_1 - n_2} \end{aligned}$$

und weil die linke Seite irrational ist, gilt strikte Ungleichung. Daraus folgt

$$\rho(n_1 - n_2) < m_2 - m_1,$$

also

$$m_1 + n_1\rho < m_2 + n_2\rho$$

wie gewünscht.

Wenn andererseits $n_1 < n_2$ gilt, dann folgt aus

$$F^{n_1}(0) + m_1 < F^{n_2}(0) + m_2,$$

dass

$$F^{n_2-n_1}(F^{n_1}(0)) - F^{n_1}(0) > m_1 - m_2$$

gilt. Per Induktion gilt damit

$$F^{k(n_2-n_1)}(0) - 0 > k(m_1 - m_2).$$

Dann folgt wie vorhin

$$\begin{aligned}\rho(f) &= \lim_{k \rightarrow \infty} \frac{F^{k(n_2 - n_1)}(0) - 0}{k(n_2 - n_1)} \\ &\geq \frac{m_1 - m_2}{n_2 - n_1}\end{aligned}$$

und damit wieder $\rho(n_1 - n_2) < m_2 - m_1$ und

$$m_1 + n_1\rho < m_2 + n_2\rho$$

wie behauptet. □

Damit können wir auch leicht den Satz beweisen:

Beweis. Die Ordnung der Orbits von f wird durch Λ beschrieben und die Ordnung der Orbits von $R_{\rho(f)}$ durch Ω . Nun folgt die Behauptung aus dem Lemma. □

HYPERBOLIZITÄT

SPEZIFIKATION

Wir wissen schon, dass wir Pseudo-Orbits nahe einer hyperbolischen Menge durch ein naheliegendes echtes Orbit „beschatten“ können. Allerdings setzte das bislang voraus, dass wir in jedem Schritt nur eine kleine Abweichung zulassen zwischen dem nächsten Punkt auf dem Pseudo-Orbit und dem Bild des aktuellen Punktes.

Jetzt sehen wir eine Methode, die es zulässt, dass die vorgegebenen Punkte beliebig weit von Bildern der vorigen Punkte entfernt sind, und dennoch eine Beschattungseigenschaft herauskommt: Wir werden sehen, dass wir mehrere endliche Orbitstuecke beliebig vorgeben können und dennoch ein echtes Orbit finden, dass beiden nahekommt. Und welches auch noch eine genau vorgeschriebene Zeit zwischen diesen Segmenten zubringt.

Definition. Für eine bijektive Abbildung $f : X \rightarrow X$ ist eine **Spezifikation** eine endliche Sammlung endlicher Teilmengen von \mathbb{Z} , d.h. $I_1 = \{a_1, \dots, b_1\}, \dots, I_N = \{a_N, \dots, b_N\}$, sowie eine Abbildung $P : \bigcup_{i=1}^N I_i \rightarrow X$, welche jedes I_i auf ein Orbitsegment abbildet, d.h. für $k, l \in I_i$ gilt $P(k) = f^{l-k}(P(l))$.

Die Spezifikation heißt **L -separiert**, wenn $a_{i+1} > b_i + L$ für alle $i = 1, \dots, N - 1$ gilt.

Theorem. Sei Λ eine lokal maximale hyperbolische Menge für einen topologisch mischenden Diffeomorphismus $f : U \rightarrow M$ auf einer Mannigfaltigkeit M . Dann gibt es für jedes $\varepsilon > 0$ ein $L = L(\varepsilon) < \infty$, so dass jede

L-separierte Spezifikation auf M von einem echten Orbit ε -beschattet wird, d.h. es gibt ein $x \in M$, so dass für alle $n \in \bigcup_{i=1}^N I_i$ gilt:

$$d(f^n(x), P(n)) < \varepsilon$$

und so dass für alle $k > \sum_{i=1}^N \text{Länge}(I_i) + NL$ ein periodisches Orbit mit Periode k existiert mit derselben Eigenschaft ($d(f^n(x), P(n)) < \varepsilon$ für alle $n \in \bigcup_{i=1}^N I_i$).

Remark. Das heißt, wir können zu mehreren vorgegebenen Orbitsegmenten ein Orbit finden, welches diese Segmente ε -genau approximiert, und welches auch noch exakt eine vorher festgelegte Zeit zwischen den Orbits zubringt. Dies ist ein wichtiges globales Ergebnis (während Beschattung dagegen lokal erfolgt).

Beweis. Sei $x_1 := P(b_1)$ der letzte Punkt der ersten Orbitsegments und $y_1 := P(a_2)$ der erste Punkt des zweiten Segments. Wegen topologischem Mischen schneiden sich $W^u(x)$ und $W^s(y)$ in einem Punkt z . Für k, l groß genug ist $f^k(z), f^{-l}(z)$ beliebig nahe an y_1, x_1 . Wähle $x := f^{-l-(b_1-a_1)}(z)$. Dann sind die ersten $b_1 - a_1$ Iterationen von x wegen Stetigkeit nahe an $P(a_1), \dots, P(b_1)$ und wegen dem Schnitt von $W^u(x)$ und $W^s(y)$ und der Wahl von k, l ist $f^{a_2}(x)$ auch nahe an $P(a_2)$. Somit auch auf dem zweiten Intervall I_2 .

Nun wiederholen wir die Prozedur: Sei $x_2 := P(b_2), y_2 := P(a_3)$. Finde $z \in W^u(x_2) \cap W^s(y_2)$, finde entsprechende k', l' , und sei $x := f^{-l'-(b_2-a_2)-l-k-(b_1-a_1)}$. Dieses x unterscheidet sich beliebig wenig vom vorigen. Nach endlich vielen Schritten ist das Verfahren abgeschlossen und das gefundene x hat ein allen Orbitsegmenten nahes Orbit. \square

MARKOV-PARTITIONEN

Die Idee hinter Markov-Partitionen ist es, ein dynamisches System auf einer Mannigfaltigkeit (also einem Raum mit überabzählbar vielen Punkten) zu reduzieren auf einen Shift auf endlich vielen Symbolen, d.h. den Raum in eine endliche Zahl Stücke zu zerlegen, die schon genügen, um die Dynamik zu verstehen. Eine **Markov-Partition** von Λ ist eine endliche „Partition“ von Λ durch Mengen R_1, \dots, R_N , die sich nur am Rand überlappen und die die Markov-Eigenschaft haben, dass die Abbildung f bestimmte Ränder von R_i wieder auf Ränder abbildet.

Zunächst etwas Wiederholung elementarer Topologie: Für eine Menge $A \subset M$ heißt eine Menge $B \subset M$ **offen relativ zu A** , wenn es eine in M offene Menge O gibt mit $B = A \cap O$. Notation: Das Innere einer Menge B relativ zu A bezeichnen wir mit

$$\text{Inn}_A(B),$$

den Rand mit

$$\partial_A B.$$

Erinnerung: Es gibt ein $\eta < \infty$, so dass gilt: Für alle $x, y \in \Lambda$ schneiden sich $W_\eta^u(x)$ und $W_\eta^s(y)$ höchstens in einem Punkt. Wenn auch noch $d(x, y) < \delta$ gilt, dann schneiden sie sich in genau einem Punkt, welcher dann mit

$$[x, y]$$

bezeichnet wird.

Definition. Sei $f : U \rightarrow M$ ein Diffeomorphismus mit lokal maximaler hyperbolischer Menge Λ . Eine Menge $R \subset \Lambda$ heißt ein **Rechteck**, wenn R Durchmesser $< \eta/10$ hat und für alle $x, y \in R$ gilt, dass

$$[x, y] \in R.$$

Wir benutzen die Notation

$$W_R^s(x) := R \cap W_\eta^s(x)$$

und analog für $W_R^u(x)$.

Eine **Markov-Partition** von Λ ist eine endliche „Partition“ von Λ durch Mengen R_1, \dots, R_N , welche:

- Abschluss Ihres Inneren bezüglich Λ sind, d.h. $R_i = \overline{\text{Inn}_\Lambda(R_i)}$,
- Rechtecke sind, d.h. abgeschlossen unter $[\cdot, \cdot]$,
- die Bedingung $\text{Inn}_\Lambda(R_i) \cap \text{Inn}_\Lambda(R_{j \neq i}) = \emptyset$ erfüllen,
- die Eigenschaft haben, dass wenn $x \in \text{Inn}(R_i)$ und $f(x) \in \text{Inn}(R_j)$, dann gilt

(1)

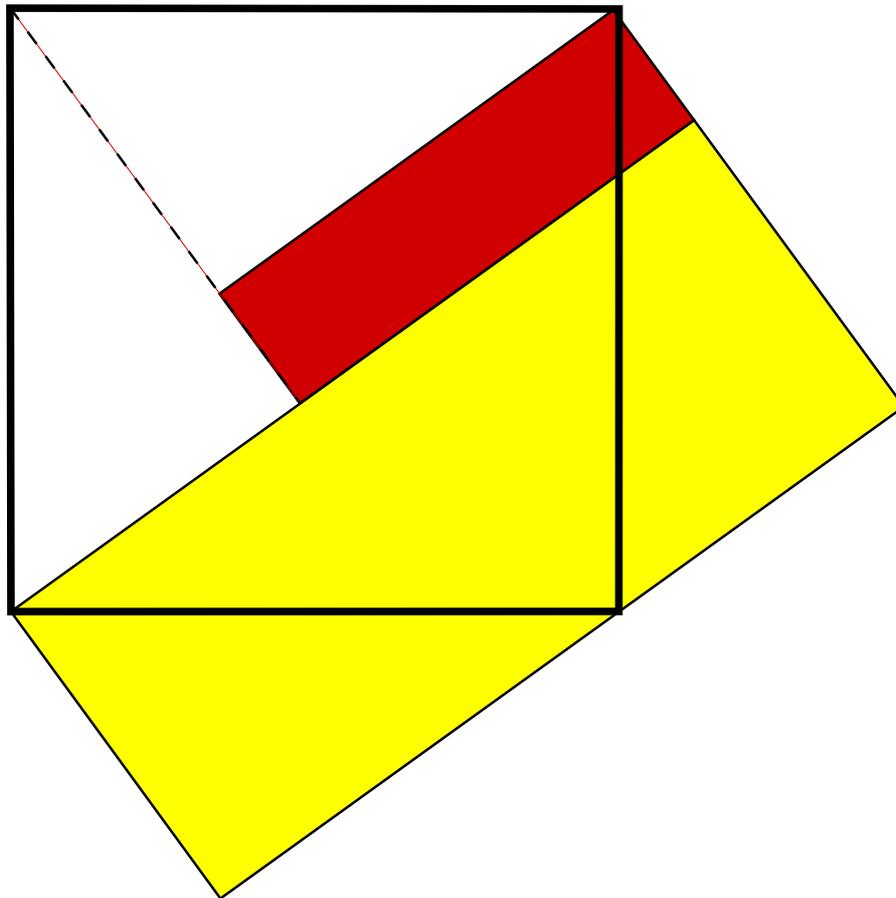
$$W_{R_j}^u(f(x)) \subset f(W_{R_i}^u(x)),$$

(2)

$$f(W_{R_i}^s(x)) \subset W_{R_j}^s(f(x)).$$

Remark. Die Bedingung $\text{Inn}_\Lambda(R_i) \cap \text{Inn}_\Lambda(R_{j \neq i}) = \emptyset$ ist schwächer als das, was bei der Definition einer *Partition* normalerweise gefordert wird, nämlich Disjunktheit der Elemente der Partition. Wir benutzen diese schwächere Forderung, damit wir für alle R_i abgeschlossene (somit kompakte) Mengen zulassen können.

Example. Sei f der Anosov-Automorphismus auf dem 2-Torus, gegeben durch Multiplikation mit $\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$. Wir wissen schon, dass f Eigenwerte $\frac{3 \pm \sqrt{5}}{2}$ hat und orthogonale Eigenvektoren. Im folgenden Diagramm ist eine Partition von T^2 in 2 Rechtecke eingezeichnet:



Remark. Für die Vorstellung ist es nützlich, sich Markov-Partitionen als euklidische Rechtecke wie in vorigem Beispiel vorzustellen. Dies ist allerdings nur ein vereinfachtes Bild. In Wirklichkeit ist es schon bei sehr einfachen Abbildungen, z.B. Automorphismen auf dem 3-Torus – d.h. dasselbe wie oben mit einer 3×3 -Matrix statt einer 2×2 -Matrix – so, dass die Elemente einer Markov-Partition keine glatten Quader mehr sind, sondern fraktale Mengen.

Man kann zeigen:

Auf einer kompakten lokal maximalen hyperbolischen Menge gibt es Markov-Partitionen von beliebig kleinem Durchmesser.

Damit sind wir nun in der Lage, zu verstehen, wozu Markov-Partitionen wirklich nützlich sind: Wie der folgende Satz sagt, lässt sich die Dynamik von f beschreiben durch einen Shift auf einem Symbolraum. Zunächst definieren wir den Symbolraum:

Definition. Für eine quadratische $N \times N$ Matrix A mit Einträgen 0 oder 1 ist der **bezüglich A zugelassene Folgenraum** gegeben durch

$$\Omega_A := \{\omega \in \Omega_N, \forall n \in \mathbb{Z} : \omega_n = i, \omega_{n+1} = j \text{ nur wenn } a_{ij} = 1\}.$$

Hierbei ist

$$\Omega_N = \{\omega = (\dots, \omega_{-1}, \omega_0, \omega_1, \dots) \mid \forall n \in \mathbb{Z} : \omega_n \in \{0, \dots, N-1\}\}.$$

Theorem. Wenn f eine lokal maximale hyperbolische Menge Λ hat und $\mathcal{R} = (R_1, \dots, R_N)$ eine Markov-Partition von genügend kleinem Durchmesser ist, dann ist die Abbildung

$$\varphi : \Omega_A \rightarrow \Lambda,$$

wobei A definiert ist durch

$$a_{ij} := \begin{cases} 1 & \text{für } f(R_i) \cap R_j \neq \emptyset \\ 0 & \text{sonst} \end{cases}$$

die definiert ist durch

$$\varphi(\omega) := \bigcap_{n \in \mathbb{Z}} f^{-n}(R_{\omega_n})$$

wohldefiniert, stetig, surjektiv und erfüllt

$$f \circ \varphi = \varphi \circ \sigma,$$

Außerdem gilt: φ ist injektiv auf $\varphi^{-1}(\Lambda \setminus \Lambda')$, wobei $\Lambda' = \bigcup_{n \in \mathbb{Z}} f^n(\partial_\Lambda \mathcal{R})$.

Beweis.

- (1) Wohldefiniertheit von φ : Jedes R_i ist eine abgeschlossene Teilmenge der kompakten Menge Λ . Wegen Definition von A ist für alle $n \in \mathbb{N}$ der Durchschnitt

$$f(R_{\omega_n}) \cap R_{\omega_{n+1}}$$

nichtleer. Also ist

$$\left(\bigcap_{|n| \leq k} f^{-n}(R_{\omega_n}) \right)_{k \in \mathbb{N}}$$

eine monoton fallende Folge von nichtleeren Mengen (d.h. jede Menge ist nichtleere Teilmenge der vorigen). Deswegen ist der Durchschnitt

$$\bigcap_{n \in \mathbb{Z}} f^{-n}(R_{\omega_n})$$

eine nichtleere geschlossene Menge. Wegen Expansivität von f kann diese Menge nicht mehr als einen Punkt enthalten, denn wenn $x, y \in \bigcap_{n \in \mathbb{Z}} f^{-n}(R_{\omega_n})$ und $\text{Durchmesser}(\mathcal{R}) < \varepsilon$, dann ist $d(f^n(x), f^n(y)) < \varepsilon$ für alle $n \in \mathbb{N}$, und Expansivität von f bedeutet, dass $d(f^n(x), f^n(y)) < \delta$ für alle $n \in \mathbb{N}$ nur möglich ist für $x = y$. Deswegen benötigen wir auch eine kleine Markov-Partition, um hier $\varepsilon < \delta$ wählen zu können.

-
- (2) Injektivität der Einschränkung von φ : Die Einschränkung von φ garantiert, dass $f^n(x)$ für kein $n \in \mathbb{N}$ auf dem Rand der Partition liegt. Da sich die Elemente der Partition nur am Rand schneiden können, gibt es zu jedem $n \in \mathbb{N}$ genau ein $i \in \{1, \dots, N\}$ mit $f^n(x) \in R_i$. Deswegen besteht $\varphi^{-1}(x)$ aus genau dem einen $\omega \in \Omega_A$, für welches für alle $n \in \mathbb{N}$ gilt, dass $\omega_n = i$ ist, wobei $f^n(x) \in R_i$.
- (3) Stetigkeit von φ : Für $\omega, \omega' \in \Omega_A$ ist $d(\omega, \omega')$ klein nur dann, wenn ω_n, ω'_n auf $\{n \in \mathbb{Z} \mid |n| < k\}$ übereinstimmen mit k gross. Wir haben vorhin schon gesehen, dass

$$\left(\bigcap_{|n| \leq k} f^{-n}(R_{\omega_n}) \right)_{k \in \mathbb{N}}$$

für $k \rightarrow \infty$ gegen einen Punkt konvergiert, also für k gross genug einen beliebig kleinen Durchmesser hat. Da $\varphi(\omega), \varphi(\omega')$ beide darin liegen, ist ihr Abstand auch beliebig klein für $d(\omega, \omega')$ klein genug.

- (4) φ erfüllt $f \circ \varphi = \varphi \circ \sigma$: Dies folgt direkt aus der Konstruktion:

$$\begin{aligned} f(\varphi(\omega)) &= f \left(\bigcap_{n \in \mathbb{Z}} f^{-n}(R_{\omega_n}) \right) \\ &= \bigcap_{n \in \mathbb{Z}} f^{-n+1}(R_{\omega_n}) \\ &= \bigcap_{n \in \mathbb{Z}} f^{-n}(R_{\omega_{n+1}}) \\ &= \bigcap_{n \in \mathbb{Z}} f^{-n}(R_{\sigma(\omega)_n}) \\ &= \varphi(\sigma(\omega)). \end{aligned}$$

□

EIGENSCHAFTEN VON ANOSOV-DIFFEOMORPHISMEN

Ein Diffeomorphismus $f : M \rightarrow M$ heißt **Anosov-Diffeomorphismus**, wenn ganz M eine hyperbolische Menge für f ist.

Da unsere Definition von „hyperbolische Menge“ unter anderem Kompaktheit fordert, setzen wir somit implizit voraus, dass M kompakt ist. Die Sprechweise „Sei $f : X \rightarrow X$ Anosov“ impliziert also: X ist kompakte Mannigfaltigkeit. Ein Anosov-Diffeomorphismus ist auch auf der ganzen Mannigfaltigkeit definiert, während bei allgemeinen hyperbolischen Mengen ausgereicht hat, dass $f : U \rightarrow M$ auf einer offenen Menge $U \subset M$ definiert war.

Theorem. Eigenschaften von Anosov-Diffeomorphismen:

Wenn $f : M \rightarrow M$ ein Anosov-Diffeomorphismus ist, existieren $\lambda \in (0, 1)$, $C_p < \infty$, $\varepsilon > 0$, $\delta > 0$ und für jedes $x \in M$ eine Spaltung von $T_x M$ in Unterräume $E^s(x) \oplus E^u(x)$, so dass:

- die Familien E^s und E^u sind f -invariant,
- für alle $v \in E^s$ und $n \in \mathbb{N}$ gilt $\|df^n v\| \leq \lambda^n \|v\|$,
- für alle $v \in E^u$ und $n \in \mathbb{N}$ gilt $\|df^{-n} v\| \leq \lambda^n \|v\|$,
- $W^s(x) = \{y \in M : d(f^n(x), f^n(y)) \rightarrow 0 \text{ für } n \rightarrow \infty\}$ erfüllt $d^s(f(x), f(y)) \leq \lambda d^s(x, y)$ für alle $y \in W^s(x)$,
- $W^u(x) = \{y \in M : d(f^{-n}(x), f^{-n}(y)) \rightarrow 0 \text{ für } n \rightarrow \infty\}$ erfüllt $d^u(f^{-1}(x), f^{-1}(y)) \leq \lambda d^u(x, y)$ für alle $y \in W^u(x)$,
- die Familien W^s und W^u sind invariant,
- $E^s(x)$ ist tangential an $W^s(x)$,
- $E^u(x)$ ist tangential an $W^u(x)$,
- für $d(x, y) < \delta$ besteht der Durchschnitt $W_\varepsilon^u(x) \cap W_\varepsilon^s(y)$ aus genau einem Punkt, genannt $[x, y]$,
- $[x, y]$ hängt stetig von x, y ab,
- $d^s([x, y], y) \leq C_p d(x, y)$ und $d^u(x, [x, y]) \leq C_p d(x, y)$.

Nun erforschen wir den Zusammenhang zwischen Dichtheit von periodischen Punkten, Dichtheit von (in-)stabilen Mannigfaltigkeiten und topologischem Mischungsverhalten bei Anosov-Diffeomorphismen.

Zur Vorbereitung etwas zur Dichtheit von periodischen Punkten:

Theorem. Sei $f : X \rightarrow X$ ein Anosov-Diffeomorphismus. Dann sind periodische Punkte dicht in der nichtwandernden Menge $NW(f)$.

Beweis. Für beliebiges $x \in NW(f)$ gilt: Für alle $\varepsilon > 0$ gibt es $N \in \mathbb{N}$ mit $f^N(B_\varepsilon(x)) \cap B_\varepsilon(x) \neq \emptyset$. Also gibt es ein y , das ε -nahe an x liegt, so dass $f^N(y)$ auch ε -nahe an x liegt. Somit ist das Orbit-Segment

$$y, f(y), f^2(y), \dots, f^N(y)$$

ein geschlossenes 2ε -Pseudoorbit. Nach dem Beschattungssatz gibt es in einer Umgebung von y einen periodischen Punkt. Dieser kann beliebig nahe an x gefunden werden, da ε beliebig klein gewählt werden kann. \square

Bei diesem Beweis haben wir die Anosov-Eigenschaft, dass die hyperbolische Menge der gesamte Raum ist, nur dazu verwendet, um x beliebig wählen zu können. Wenn wir statt eines Anosov-Diffeomorphismus eine beliebige Abbildung mit hyperbolischer Menge Λ nehmen, stimmt das Argument für $x \in \Lambda$ immer noch; allerdings ist das periodische Orbit nun nicht mehr notwendigerweise in Λ , sondern nur in einer kleinen Umgebung. Aus dieser Überlegung heraus formulieren wir folgenden Satz:

Theorem. Wenn Λ eine lokal maximale hyperbolische Menge von einem Diffeomorphismus $f : U \rightarrow M$ ist, dann sind periodische Punkte von $f|_\Lambda$ dicht in der nichtwandernden Menge $NW(f|_\Lambda)$.

Zur Erinnerung: Lokale Maximalität von $f : U \rightarrow M$ besagt, dass es eine offene Umgebung $V \subset U$ von Λ gibt, so dass Λ gleich dem Orbit vom Abschluss von V ist:

$$\Lambda = \bigcap_{i \in \mathbb{Z}} f^i(\overline{V}).$$

Äquivalent dazu ist, dass es eine offene Umgebung U' von Λ gibt, so dass für jede offene Umgebung $V \subset U'$ von Λ gilt, dass Λ gleich dem Orbit vom Abschluss von V ist, d.h. $\Lambda = \bigcap_{i \in \mathbb{Z}} f^i(\overline{V})$.

Diese Bedingung heißt deswegen „lokal maximal“, weil sie impliziert, dass man lokal, also in einer kleinen Umgebung U von Λ , keine Punkte zu Λ dazutun kann (also Λ vergrößern), so dass Λ immer noch invariant ist.

Beweis. Ähnlich wie vorhin: Für $x \in NW(f|_\Lambda)$ und alle $\varepsilon > 0$ gibt es $N \in \mathbb{N}$ mit $f^N((B_\varepsilon(x) \cap \Lambda)) \cap (B_\varepsilon(x) \cap \Lambda) \neq \emptyset$. Also gibt es ein y , das ε -nahe an x liegt, so dass $f^N(y)$ auch ε -nahe an x liegt. Wieder ist das Orbit-Segment $y, f(y), f^2(y), \dots, f^N(y)$ ein geschlossenes 2ε -Pseudoorbit in X . Nach dem Beschattungssatz gibt es in einer Umgebung von y in X einen periodischen Punkt p beliebig nahe an x .

Damit wissen wir schon, dass die periodischen Punkte von f dicht sind in Λ ; wir wissen aber noch nicht, ob auch die periodischen Punkte von $f|_\Lambda$ schon dicht sind in Λ . Es bleibt also noch zu zeigen, dass p auch in Λ liegt. Dazu benutzen wir die lokale Maximalität: p liegt schon beliebig nahe an Λ . Damit gilt das auch für die ersten n Iterierten von p (wegen Stetigkeit von f). Wegen Periodizität gilt das für das ganze Orbit von p . Also liegt das Orbit von p in einer Menge V in einer beliebig kleinen Umgebung von Λ . Somit liegt es in $\bigcap_{i \in \mathbb{Z}} \overline{V}$. Diese Menge ist aber gleich Λ . \square

Theorem. Sei $f : X \rightarrow X$ ein Anosov-Diffeomorphismus. Dann sind äquivalent:

- (1) $NW(f) = X$.
- (2) Für alle $x \in X$ ist $W^u(x)$ dicht in X .
- (3) Für alle $x \in X$ ist $W^s(x)$ dicht in X .
- (4) f ist topologisch transitiv.
- (5) f ist topologisch mischend.

Während des Beweises werden noch ein paar Lemmata auftauchen (und bewiesen).

Beweis. Zuerst zeigen wir (1) \Rightarrow (3):

Wähle in X eine $\varepsilon/4$ -dichte endliche Menge von periodischen Punkten

$$x_1, \dots, x_k.$$

Für

$$N := \prod_{i=1}^k \text{Periode}(x_i)$$

sind alle x_i periodisch mit (derselben) Periode N . Wir definieren $g := f^N$. Statt f können wir auch g untersuchen, denn die stabile Mannigfaltigkeit von f durch x ist genau die stabile Mannigfaltigkeit von g durch x ; genauso für die instabilen Mannigfaltigkeiten.

Lemma. *Es gibt ein $q \in \mathbb{N}$, so dass wenn $d(x_i, x_j) < \varepsilon/2$ und wenn für ein $y \in M$ gilt, dass $d(x_i, W^u(y)) < \varepsilon/2$, dann gilt für alle $n \in \mathbb{N}$, dass $d(x_j, g^{nq}(W^u(y))) < \varepsilon/2$.*

Beweis. Wähle z in $W^u(y) \cap W_\varepsilon^s(x_i)$. Dann gilt $d(g^T(z), x_i) < \varepsilon/2$ für ein $T \in \mathbb{N}$ (und auch für alle $T > T_1$, mit $T_1 \in \mathbb{N}$ unabhängig von z). Da gilt, dass $d(g^T(z), x_j) < \varepsilon$, existiert ein $w \in W^u(g^T(z)) \cap W_\varepsilon^s(x_j)$. Also ist $d(g^{T'}(w), x_j) < \varepsilon/2$ für ein $T' \in \mathbb{N}$ (und auch für alle $T' > T_2$ mit $T_2 \in \mathbb{N}$ unabhängig von w). Mit $q := T_1 + T_2$ folgt die Behauptung. \square

Damit können wir die Implikation (1) \Rightarrow (2) zeigen: Da die Menge $\{x_1, \dots, x_k\}$ endlich ist, können wir in endlich vielen ε -Sprüngen von jedem x_i zu jedem x_j gelangen. Bezeichnen wir mit K die maximal nötige Zahl. Dann ist wegen vorigem Lemma die Menge $g^{Kq}(W^u(y))$ eine ε -dichte Menge, denn die Abschätzung $d(x_j, g^{nq}(W^u(y))) < \varepsilon/2$ gilt auch für $j = i$, also $d(x_i, g^{nq}(W^u(y))) < \varepsilon/2$.

Wenn wir f mit f^{-1} vertauschen (was auch Anosov ist), vertauschen sich die Rollen von W^u , W^s . Also haben wir auch (1) \Rightarrow (3) bewiesen.

Jetzt noch eine Vorbereitung für den Beweis von (2) \Rightarrow (5):

Lemma. *Wenn für alle $y \in X$ gilt, dass $W^u(y)$ dicht ist in X , dann gibt es zu jedem $\varepsilon > 0$ ein $R(\varepsilon) < \infty$ so dass für alle $x \in X$ gilt: $W_{R(\varepsilon)}^u(x)$ ist ε -dicht in X .*

Beweis. Wir wissen schon, dass für jedes $x \in X$ die Menge $W^u(x)$ in X dicht ist, also ε -dicht für jedes $\varepsilon > 0$. Da

$$W^u(x) = \bigcup_{r \in \mathbb{N}} W_r^u(x),$$

muss also für jedes $\varepsilon > 0$ ein $r(\varepsilon, x) \in \mathbb{N}$ existieren, so dass $W_{r(\varepsilon, x)}^u(x)$ schon ε -dicht ist. Es bleibt noch zu zeigen, dass r unabhängig von x gewählt werden kann.

Seien x und ε schon gewählt. Die instabilen Mannigfaltigkeiten $(W^u(y))_{y \in X}$ bilden eine „Blätterung“ W^u , die stetig ist, in dem Sinn, dass festes r und für x' beliebig nahe bei x die Menge $W_r^u(x)$ auch beliebig nahe an der Menge $W_r^u(x')$ ist. D.h., es gibt ein $\delta > 0$, so dass für alle x' mit $d(x, x') < \delta$ die Menge $W_{r(\varepsilon/2, x)}^u(x')$ in X ε -dicht ist.

Wir finden nun zu jedem $x \in X$ so eine $\delta(x)$ -Umgebung. Alle zusammen bilden sie eine Überdeckung von X , also gibt es wegen Kompaktheit eine endliche Teilüberdeckung. Wir können also das Maximum der $r(\varepsilon/2, \tilde{x}_i)$ über die Mittelpunkte \tilde{x}_i dieser endlich vielen Bälle nehmen und es $R(\varepsilon)$ nennen. Dann ist $W_{R(\varepsilon)}^u(x)$ für alle x eine ε -dichte Menge. \square

Nun zeigen wir (2) \Rightarrow (5):

Zu zeigen ist: Für alle U, V offen in X (und nichtleer) gibt es \bar{n} , so dass für alle $n \geq \bar{n}$ gilt, dass $U \cap f^n(V) \neq \emptyset$. Da V offen ist, enthält es $W_\varepsilon^u(x)$ für ein $x \in X$ und ein $\varepsilon > 0$. Außerdem enthält U einen ε' -Ball (oBdA $\varepsilon' = \varepsilon$). Wir wissen schon, dass Abstände auf W^u sich unter Anwendung von f exponentiell vergrößern. Also gibt es $\bar{n} \in \mathbb{N}$, so dass $f^{\bar{n}}(W_\varepsilon^u(x)) \supset W_{R(\varepsilon)}^u(f^{\bar{n}}(x))$. Die Menge auf der rechten Seite ist ε -dicht, schneidet also U .

Genauso folgt (3) \Rightarrow (5), denn wenn alle $W^s(x)$ für f dicht sind, sind alle $W^u(x)$ für f^{-1} dicht, also gibt es für alle U, V offen und nichtleer ein $\bar{n} \in \mathbb{N}$, so dass für alle $n \geq \bar{n}$ gilt $U \cap (f^{-1})^n(V) \neq \emptyset$, also $V \cap f^n(U) \neq \emptyset$.

Die Implikation (5) \Rightarrow (4) folgt direkt aus der Definition von topologischem Mischen und topologischer Transitivität.

Die Folgerung (4) \Rightarrow (1) ist ebenfalls einfach, denn wenn f ein dichtes Orbit hat und x ein Punkt darauf, dann gilt für alle $y \in X$ und jede offene Umgebung U von y , dass es $n \in \mathbb{N}$, $m \in \mathbb{N}$, $m > n$ gibt mit $f^n(x) \in U$, $f^m(x) \in B_{d(y, f^n(x))/2}(y) \subset U$, und somit auch $f^{N:=m-n}(U) \cap U \neq \emptyset$. \square

ERGODENTHEORIE

Zunächst ein harmlos aussehender Satz, aus dem aber direkt die wichtigsten Ergodensätze folgen:

Theorem. Fundamentaler Ergodensatz: Sei (X, μ) ein Maßraum mit endlichem Gesamtmaß, $T : X \rightarrow X$ eine messbare Abbildung, welche T invariant läßt, d.h. für alle messbaren B gilt $\mu(T^{-1}(B)) = \mu(B)$, und sei $B \subset X$ eine (messbare) Teilmenge. Definiere

$$S_n(x) := \#\{i \in \{0, \dots, n-1\} \mid T^i(x) \in B\},$$

(also die Zahl der „Treffer“ in B unter dem Orbitsegment der Länge n mit Startpunkt x), und

$$A_n(x) := \frac{1}{n} S_n(x),$$

(also die relative Häufigkeit, die dieses Orbitsegment in B zubringt).

Dann gilt für μ -fast alle $x \in X$: Der Limes

$$\lim_{n \rightarrow \infty} A_n(x)$$

existiert.

Beweis. Definiere

$$\bar{A}(x) := \limsup_{n \rightarrow \infty} A_n(x),$$

$$\underline{A}(x) := \liminf_{n \rightarrow \infty} A_n(x).$$

Die Funktion \bar{A} ist T -invariant, d.h. $\bar{A}(T(x)) = \bar{A}(x)$, und das gilt natürlich auch für \underline{A} .

Wähle $\varepsilon > 0$ beliebig (klein) und definiere

$$\tau(x) := \min\{n \in \mathbb{N} \mid A_n(x) \geq \bar{A}(x) - \varepsilon\}.$$

τ hängt natürlich auch von ε ab, aber wir ändern ε bis ganz am Ende von diesem Beweis nicht.

Fall 1: τ ist essentiell beschränkt durch $M < \infty$, d.h. für μ -fast alle $x \in X$ gilt $\tau(x) \leq M$.

Wir betrachten für $n \in \mathbb{N}$ (groß) das Orbitsegment von Länge n ab einem beliebigen $x \in X$, d.h.

$$x, Tx, \dots, T^{n-1}x$$

und stellen fest, dass es $m_1 \leq M$ gibt, so dass $A_{m_1}(x) \geq \bar{A}(x) - \varepsilon$, denn wir können $m_1 = \tau(x)$ wählen. Dann betrachten wir das Orbitsegment von Länge M ab $T^{m_1}x$, d.h.

$$T^{m_1}x, Tx, \dots, T^{m_1+M-1}x$$

und stellen analog fest, dass es $m_2 \leq M$ gibt, so dass $A_{m_2}(T^{m_1}x) \geq \bar{A}(x) - \varepsilon$, denn wir können $m_2 = \tau(T^{m_1}x)$ wählen und es gilt (wie vorhin erwähnt) $\bar{A}(T^{m_1}x) = \bar{A}(x)$. Wir können mit diesem Verfahren fortfahren und das Orbitsegment von Länge n ab x , d.h.

$$x, Tx, \dots, T^{n-1}x$$

mit Teilsegmenten überdecken, auf denen $A_{m_i} \geq \bar{A}(x) - \varepsilon$ ist; das können wir mindestens solange tun, bis das verbleibende Stück

$$T^{m_1+\dots+m_j}x, \dots, T^n x$$

Länge $< M$ hat. Wir haben also vom Orbitsegment der Länge n ein Teil der Länge mindestens $n - M$ abgedeckt. Damit haben wir gezeigt, dass gilt:

$$S_n(x) \geq (n - M)(\bar{A}(x) - \varepsilon)$$

oder äquivalent $A_n(x) \geq (1 - \frac{M}{n})(\bar{A}(x) - \varepsilon)$.

Wenn wir das Integral von S_n bezüglich μ über ganz X berechnen, erhalten wir $n\mu(B)$, denn es gilt $S_1 = \chi_B$ (die Indikatorfunktion der Menge B) und somit

$$\int_X S_1(x) d\mu(x) = \mu(B),$$

$S_2 = \chi_B + \chi_{T^{-1}B}$ und wegen T -Invarianz von μ gilt somit $\int_X S_2(x) d\mu(x) = 2\mu(B)$, und analog $\int_X S_n(x) d\mu(x) = n\mu(B)$. Wir erhalten also die Gleichung

$$\mu(B) \geq (1 - \frac{M}{n}) \int_X \bar{A}(x) d\mu(x) - \varepsilon.$$

Hierbei haben wir übrigens verwendet, dass $\mu(X) < \infty$ ist. Durch Grenzübergang $n \rightarrow \infty$ folgt

$$\mu(B) \geq \int_X \bar{A}(x) d\mu(x) - \varepsilon.$$

Fall 2: τ ist nicht essentiell beschränkt. Dann nehmen wir die „schlechte Menge“ von Punkten x , auf denen $\tau(x) > M$ ist, noch zu B dazu, definieren also eine Menge

$$B' := B \cup \{x \in X \mid \tau(x) > M\}$$

und darauf

$$S'_n(x) := \#\{i \in \{0, \dots, n-1\} \mid T^i(x) \in B'\}.$$

Dann schließen wir: Es gilt analog zu vorhin

$$S'_n(x) \geq (n - M)(\bar{A}(x) - \varepsilon),$$

denn für $x \in X$ gilt: Entweder ist $\tau(x) \leq M$, und dann ist

$$A_{m_1}(x) \geq \bar{A}(x) - \varepsilon$$

für $m_1 = \tau(x) \leq M$. Oder es ist $\tau(x) > M$, und dann ist x schon in der „schlechten Menge“, also auch in B' , und deshalb ist $S_1(x) = 1$, also sicherlich

$$A_{m_1}(x) = 1 \geq \bar{A}(x) - \varepsilon$$

mit $m_1 = 1$. In beiden Fällen fahren wir dann mit $T^{m_1}x$ anstelle von x fort, wiederholen und erhalten schließlich die gewünschte Abschätzung $S'_n(x) \geq (n - M)(\bar{A}(x) - \varepsilon)$. Daraus schließen wir wieder

$$\mu(B') \geq \int_X \bar{A}(x) d\mu(x) - \varepsilon,$$

und daraus folgt

$$\mu(B) \geq \int_X \bar{A}(x) d\mu(x) - 2\varepsilon.$$

Letztere Gleichung gilt also immer, egal ob Fall 1 oder Fall 2 vorliegt.

Durch Vertauschen von B mit $X \setminus B$ erhalten wir

$$\mu(X \setminus B) \geq \int_X (1 - \underline{A}(x)) d\mu(x) - 2\varepsilon,$$

also

$$\mu(B) \leq \int_X \underline{A}(x) d\mu(x) + 2\varepsilon.$$

Da $\varepsilon > 0$ beliebig war, stellen wir durch Grenzwertübergang $\varepsilon \rightarrow 0$ jetzt fest, dass

$$\int_X \overline{A}(x) d\mu(x) \leq \mu(B) \leq \int_X \underline{A}(x) d\mu(x),$$

somit

$$\int_X (\underline{A}(x) - \overline{A}(x)) d\mu(x) \geq 0.$$

Da aber offensichtlich $\underline{A}(x) - \overline{A}(x) \leq 0$ für alle x gilt, muss gelten: $\underline{A}(x) = \overline{A}(x)$ für μ -fast alle $x \in X$. Also existiert der Limes

$$\lim_{n \rightarrow \infty} A_n(x)$$

für μ -fast alle $x \in X$, wie gefordert. □

Nun wenden wir uns Sätzen zu, die viel stärker klingen – und auch wirklich aussagekräftig sind – die aber eine direkte Folge aus dem fundamentalen Ergodensatz sind.

Definition. Das System (T, μ) heißt **ergodisch**, wenn für alle (messbaren) und T -invarianten Mengen $B \subset X$ gilt, dass B bezüglich μ Maß 0 hat oder das Komplement von B Maß 0 hat, also $\mu(B) = 0$ oder $\mu(X \setminus B) = 0$ ist. In diesem Fall benutzen wir auch die Sprechweise „ T ist ergodisch bezüglich μ “.

In der Literatur sieht man oft die Formulierung „ T ist ergodisch“ ohne Angabe eines Maßes, aber das ist etwas verwirrend und macht nur Sinn, wenn ein ganz bestimmtes Maß dazugedacht ist. Denn Ergodizität hängt von dem Maß ab; wenn T ergodisch ist bezüglich eines Maßes μ_1 , dann muss T nicht ergodisch sein bezüglich eines anderen Maßes μ_2 . Ebenso findet man häufig die Formulierung „ μ ist ergodisch“; offensichtlich macht das auch nur Sinn, wenn ein ganz bestimmtes T dazugedacht ist.

Lemma. *Folgende Aussagen sind äquivalent:*

- (1) (T, μ) ist ergodisch, d.h. aus $T^{-1}(B) = B$ folgt $\mu(B) = 0$ oder $\mu(X \setminus B) = 0$.
- (2) Wenn $f : X \rightarrow \mathbb{R}$ invariant ist unter T , d.h. wenn $f(T(x)) = f(x)$ für alle $x \in X$ gilt, dann ist f essentiell konstant.

(3) Der durch T gegebene lineare Operator

$$\hat{T} : L^1(\mu) \rightarrow L^1(\mu),$$

der definiert ist durch

$$(\hat{T}(f))(x) := f(T(x)),$$

also

$$\hat{T}(f) := f \circ T,$$

hat keine Eigenfunktionen außer Konstanten.

Die Äquivalenzen sind ganz leicht zu zeigen. Interessant ist hierbei zu bemerken, dass in (2) von „essentiell konstanten“ Funktionen und in (3) von „konstanten“ Funktionen geredet wird; dies liegt daran, dass eine essentiell konstante Funktion, die in L^1 liegt, in L^1 genau *gleich* einer konstanten Funktion ist, denn L^1 ist ja gerade die Menge von Äquivalenzklassen unter solchen Identifizierungen.

BIRKHOFFS ERGODENSATZ (ODER: WIE WIR DIE ZUKUNFT VORHERSAGEN)

Der fundamentale Ergodensatz besagt etwas über die relative Häufigkeit, mit welcher ein Orbit einer Abbildung T in eine bestimmte Menge trifft. In Anwendungen ist es häufig eher so, dass auf X (dem gesamten Raum) eine reelle Funktion f vorgegeben ist – nämlich eine **Messgröße**, welche in einem konkreten physikalischen Modell gemessen werden kann – und der Mittelwert dieser Funktion f längs des Orbits von T ist zu bestimmen. Dies können wir auch wirklich tun – und zwar in den wichtigen Fällen sogar dann, wenn wir fast nichts wissen über f , über T und über das Orbit. Kurz: Wir wissen praktisch nichts und schlussfolgern trotzdem praktisch alles!

Der folgende Ergodensatz von Birkhoff übersetzt den fundamentalen Ergodensatz in die Sprache von reellen Funktionen auf X :

Theorem. Birkhoff-Ergodensatz: Wenn (X, μ) ein Wahrscheinlichkeitsraum ist, $T : X \rightarrow X$ maßerhaltend bezüglich μ , und $f \in L^1(\mu)$ (d.h. $f : X \rightarrow \mathbb{R}$ integrierbar), dann existiert der Limes

$$A_f(x) := \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i(x))$$

für fast alle (bezüglich μ) Punkte $x \in X$, und der Limes existiert in $L^1(\mu)$.

Beweis. Den Fall $f = \chi_B$ kennen wir schon, denn da ist die Behauptung gerade die Aussage des fundamentalen Ergodensatzes. Der Term, der beim fundamentalen Ergodensatz S_n hieß, ist gerade unser Term $\sum_{i=0}^{n-1} f(T^i(x))$. Außerdem gilt die Behauptung sicherlich auch für eine endliche Linearkombination von solchen charakteristischen

Funktionen ist, also $f = \sum_{k=1}^N a_k \chi_{C_k}$ mit $a_k \in \mathbb{R}$ und C_k disjunkte messbare Mengen.

Wir wissen aus der Analysis, dass es zu jeder L^1 -Funktion f und für alle $\varepsilon > 0$ eine Funktion f_ε gibt, deren Werte eine diskrete Menge in \mathbb{R} sind und so dass $|f(x) - f_\varepsilon(x)| < \varepsilon$ für alle $x \in X$ gilt, also

$$f_\varepsilon = \sum_k a_k \chi_{C_k}.$$

Für den Fall, dass f essentiell beschränkt ist, gilt dann, dass f_ε nur endlich viele Werte annimmt, also $f_\varepsilon = \sum_{k=1}^N a_k \chi_{C_k}$ wie vorher, also existiert $A_{f_\varepsilon}(x)$ für fast alle x . Wegen der Abschätzung $|f(x) - f_\varepsilon(x)| < \varepsilon$ gilt dann

$$\overline{A}_f(x) := \limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i(x)) \leq \overline{A}_{f_\varepsilon}(x) + \varepsilon$$

und

$$\underline{A}_f(x) := \liminf_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i(x)) \geq \underline{A}_{f_\varepsilon}(x) - \varepsilon,$$

also existiert der Limes $A_f(x)$. Für den Fall, dass f nicht essentiell beschränkt ist, überlegen wir uns erst, dass wegen der Integrierbarkeit von f gilt, dass $\sum_k a_k \mu(C_k)$ endlich ist, denn

$$\sum_k |a_k| \mu(C_k) = \int_X |f_\varepsilon| d\mu < \varepsilon + \int_X |f| d\mu < \infty.$$

Deswegen gilt

$$\overline{A}_f(x) \leq \sum_k a_k A_{\chi_{C_k}}(x) + \varepsilon$$

und

$$\underline{A}_f(x) \geq \sum_k a_k A_{\chi_{C_k}}(x) - \varepsilon,$$

also existiert auch hier der Limes $A_f(x)$ für fast alle $x \in X$. Das zeigt die erste Behauptung.

Nun zeigen wir, dass A_f in $L^1(\mu)$ liegt. Der Einfachheit halber nehmen wir an, dass $f \geq 0$ ist; ansonsten ersetzen wir f durch $|f|$, und wenn $A_{|f|} \in L^1(\mu)$ gilt, dann gilt wegen $|A_f| \leq A_{|f|}$ auch $A_f \in L^1(\mu)$.

Es gilt

$$\begin{aligned} \int_X A_f d\mu &\leq \int_X \overline{A}_f d\mu \\ &\leq \sum_k a_k \int_X A_{\chi_{C_k}} d\mu + \varepsilon \\ &\leq \sum_k a_k \mu(C_k) + \varepsilon \\ &< \infty, \end{aligned}$$

also $A_f \in L^1(\mu)$ wie behauptet. \square

Übrigens benutzen wir hierbei die Notation $f \in L^1(\mu)$ und verstehen damit, dass f automatisch eine Funktion auf X ist; dies ist gerechtfertigt, denn wenn wir ein Maß μ haben, dann haben wir automatisch auch dessen Definitionsbereich (eine σ -Algebra) gegeben, da diese Information in μ enthalten ist, und natürlich ist die Menge X durch die σ -Algebra bestimmt, weil X das größte Element davon ist.

Bei diesem Beweis sehen wir: In Beweisen der Ergodentheorie benötigt es oft geraume Zeit, um ganz elementare (und nicht besonders interessant klingende) Dinge zu zeigen, z.B. dass das Zeitmittel A_f überhaupt definiert ist. Dagegen sind diejenigen Fakten, die uns brennend interessieren, z.B. dass das Zeitmittel fast überall von x unabhängig ist und gleich dem Raummittel ist, im Beweis ganz einfache Folgerungen aus den vorher langwierig gezeigten elementaren Dingen. Hier nun diese wichtige und einfache Folgerung:

Theorem. Wenn (T, μ) eine ergodische Transformation auf einem Wahrscheinlichkeitsraum (X, μ) ist (somit μ invariant unter $T : X \rightarrow X$), und $f \in L^1(\mu)$ ist, dann existiert für μ -fast alle $x \in X$ der Limes

$$A_f(x) := \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i(x)),$$

genannt das **Zeitmittel** von f (längs dem Orbit von x), und für μ -fast alle $x \in X$ ist das Zeitmittel von f längs dem Orbit von x gleich und hat den (konstanten) Wert

$$\int_X f d\mu,$$

genannt das **Raummittel** von f (über den gesamten Raum X).

Beweis. Die Funktion A_f ist T -invariant, also wegen der Ergodizität von (T, μ) essentiell konstant bezüglich μ . \square

Das ist nun ein reichlich erstaunliches Ergebnis! Es besagt Folgendes: Wir wollen den Wert einer Funktion (Messgröße) f längs Orbits einer Transformation T vorhersagen. Dabei wissen wir *praktisch nichts* über die Funktion f (denn L^1 ist eine Messgrößen eigentlich immer), wir wissen *sehr wenig* über die Transformation T (denn Ergodizität ist eine schwache Bedingung), und wir wissen *überhaupt nichts* über das (möglicherweise extrem komplizierte) Orbit von x , ja wir kennen *nicht einmal den Startwert* des Orbits. Und trotz alledem können wir *mit hundertprozentiger Wahrscheinlichkeit* und *ganz exakt* vorhersagen, wie der Wert von f im Mittel über das Orbit ist!

MASS-THEORETISCHES MISCHEN

Bislang kennen wir topologisches Mischen, topologische Transitivität (d.h. Dichtheit von einem Orbit) und Minimalität (d.h. Dichtheit von jedem Orbit). All dies sind *topologische* Eigenschaften, hängen also nur von der Struktur der offenen Mengen ab. Wenn wir (wie meist) auf einem metrischen Raum sind, dann hängen diese Eigenschaften von der Distanzfunktion ab.

Wenn wir nun ein System untersuchen wollen, das auf einem Maßraum definiert ist, dann stehen diese Begriffe nicht zur Verfügung, denn zwischen Maß und Topologie besteht erst einmal keine Verbindung; wir haben nur das Maß, um die Komplexität der Transformation zu bemessen.

Wir kennen schon Ergodizität (eine schwache Bedingung). Jetzt wenden wir uns dem Mischen zu, diesmal dem maß-theoretischen.

Was stellen wir uns also unter „Mischen“ vor? So etwas wie beim Anrühren von Farbe im Eimer: Wenn wir z.B. in einen 10-Liter Eimer 9 Liter gelbe und 1 Liter rote Farbe hineinschütten, dann nennen wir das ganze gut gemischt, wenn beliebige Teilmengen des Eimervolumens die Eigenschaft haben, dass darin 90% gelbe und 10% rote Farbteilchen sind. Dies führt zu folgender Definition:

Definition. Sei T eine Transformation eines Raums mit Maß μ und sei μ invariant unter T . Das System (T, μ) heißt (**maß-theoretisch**) **mischend**, wenn für alle messbaren Mengen A, B gilt:

$$\lim_{n \rightarrow \infty} \mu(A \cap T^n B) = \mu(A)\mu(B).$$

Der Zusatz „maß-theoretisch“ bei Mischen wird oft nicht ausgeschrieben. Wie bei Ergodizität ist auch dies nicht eine Eigenschaft von T oder μ allein, sondern vom 2-Tupel (T, μ) .

Mischen impliziert Ergodizität:

Lemma. Wenn das System (T, μ) *mischend* ist, dann ist (T, μ) *ergodisch*.

Beweis. Wenn (T, μ) *mischend* ist und $\mu(A) > 0$, $\mu(B) > 0$, dann gilt

$$\lim_{n \rightarrow \infty} \mu(A \cap T^n B) = \mu(A)\mu(B) > 0.$$

Nehmen wir an, (T, μ) sei nicht ergodisch. Dann gibt es solche disjunkte A, B , beide T -invariant, so dass $A \cup B = X$ gilt. Dann wäre

$$\mu(A \cap T^n B) = \mu(A \cap B) = \mu(\emptyset) = 0.$$

Widerspruch. □

Manchmal können wir Topologie und Maßtheorie aber doch in Verbindung bringen. Zum Beispiel dann, wenn das Maß auf allen offenen nichtleeren Mengen positiv ist. So etwas ist zum Beispiel der Fall, wenn das System auf einem euklidischen Raum (endliche

Teilmenge von \mathbb{R}^n oder n -Torus) definiert ist. Maß-theoretisches Mischen impliziert dann topologisches Mischen:

Lemma. *Wenn μ ein Maß auf einem topologischen Raum ist, für welches alle nichtleeren offenen Mengen messbar sind und positives Maß haben, und wenn (T, μ) maß-theoretisch mischend ist, dann ist (T, μ) auch topologisch mischend.*

Beweis. Wenn (T, μ) mischend ist und $\mu(A) > 0$, $\mu(B) > 0$, dann gilt wieder

$$\lim_{n \rightarrow \infty} \mu(A \cap T^n B) = \mu(A)\mu(B) > 0,$$

also gibt es ein $N \in \mathbb{N}$, so dass für $n > N$ gilt

$$\mu(A \cap T^n B) > 0,$$

also ist $A \cap T^n B$ nichtleer. \square

Als nächstes lernen wir ein Maß auf Symbolfolgen kennen. Dies ist ein Beispiel für ein mischendes System.

Definition. Auf dem Raum

$$\Omega_N = \{(\dots, \omega_{-1}, \omega_0, \omega_1, \dots) \mid \forall i \in \mathbb{N} : \omega_i \in \{0, \dots, N-1\}\}$$

der (zweiseitigen) Symbolfolgen im Raum mit N Symbolen ist das **Bernoulli-Maß** wie folgt definiert: Sei

$$p = (p_0, \dots, p_{N-1}) \in [0, 1]^N$$

ein Wahrscheinlichkeitsvektor, d.h. $\sum_{i=0}^{N-1} p_i = 1$. Für einen Zylinder

$$Z_{i_1, \dots, i_k}^{\alpha_1, \dots, \alpha_k} := \{\omega \in \Omega \mid \forall j \in \{1, \dots, k\} : \omega_{i_j} = \alpha_j\}$$

definieren wir

$$\mu_p(Z_{i_1, \dots, i_k}^{\alpha_1, \dots, \alpha_k}) := \prod_{j=1}^k p_{\alpha_j}.$$

Damit haben wir ein Maß auf der durch Zylinder erzeugten σ -Algebra. Dies ist ein Wahrscheinlichkeitsmaß, denn der ganze Raum Ω_N ist gleich dem Zylinder mit 0 Einträgen, hat also Maß $\prod_{j \in \emptyset} p_j = 1$.

Theorem. *Das Bernoulli-Maß ist mischend bezüglich dem Shift σ auf Ω_N .*

Beweis. Für zwei Zylinder

$$A = Z_{i_1, \dots, i_k}^{\alpha_1, \dots, \alpha_k}, \quad B = Z_{i'_1, \dots, i'_l}^{\beta_1, \dots, \beta_l}$$

gilt, dass es $N \in \mathbb{N}$ gibt, so dass für $n > N$ die Zylinder A und $\sigma^n B$ keine gemeinsamen Indices haben, also immer $i_j \neq i'_j$ ist. Das gilt nämlich für

$$N = \max \{i'_j \mid j' \in \{1, \dots, l\}\} - \min \{i_j \mid j \in \{1, \dots, k\}\}.$$

Also gilt für solche n , dass

$$A \cap \sigma^n B = Z_{i_1, \dots, i_k, i'_1-n, \dots, i'_l-n}^{\alpha_1, \dots, \alpha_k, \beta_1, \dots, \beta_l}$$

und daraus folgt sofort

$$\mu(A \cap \sigma^n B) = \mu(A)\mu(B).$$

Trivialerweise konvergiert also die linke Seite gegen $\mu(A)\mu(B)$. \square

BILLIARDS

Definition. Ein **Billiard** ist ein beschränktes Gebiet G im \mathbb{R}^2 mit glattem Rand ∂G (allgemeiner: mit stückweise glattem Rand).

Definition. Auf der Menge der Einheitstangentenvektoren an $G \cup \partial G$, bezeichnet mit $T_1(G \cup \partial G)$ definieren wir einen Fluss, den **Billiard-Fluss** φ , wie folgt: Im Inneren von G besteht der Fluss aus Geradeausgehen. Und eine Trajektorie, die am Rand anstößt, wird so reflektiert, dass eingehende und ausgehende Richtung die gleichen Winkel mit der Tangente haben (die Regel „Eingangswinkel gleich Ausgangswinkel“), wobei Eingangs- und Ausgangswinkel geeignet zu interpretieren sind. Formal haben wir folgende Definition

Für $(p, v) \in T_1 G$ (d.h. einen Vektor (p, v) mit Fußpunkt p im Inneren von G) definieren wir den **Billiard-Fluss** φ infinitesimal durch

$$\frac{d}{dt}\Big|_{t=0} \varphi_t((p, v)) := (v, 0),$$

d.h. der Fußpunkt p bewegt sich mit Geschwindigkeit v und der Vektor v ändert sich infinitesimal nicht. Für kleine t gilt dann, dass

$$\varphi_t((p, v)) = (p + tv, v).$$

Für $p \in \partial G$ und v derart, dass v ins Innere von G zeigt, definieren wir ebenfalls

$$\frac{d}{dt}\Big|_{t=0} \varphi_t((p, v)) := (v, 0).$$

Und für $p \in \partial G$ und v derart, dass v ins Äußere von G zeigt, definieren wir für kleine $t > 0$:

$$\varphi_t((p, v)) := (p + tw, w),$$

wobei w das Spiegelbild von v an der Tangente von ∂G ist (und somit in G hinein zeigt).

Für v tangential an den Rand ergeben die letzten beiden Formeln dasselbe, und deswegen ist unerheblich, welche wir wählen. Allerdings schließt die Analyse von Billiards typischerweise ohnehin aus, dass ein Orbit an einem Punkt tangential an den Rand ist, zumindest für den Fall, dass der Rand (überall) differenzierbar ist. Denn sonst läge das Orbit überall tangential an den Rand, würde also stets nur G periodisch umkreisen, was keine besonders interessante Dynamik ist.

Statt einem Fluss auf dem euklidischen Gebiet G können wir allgemeiner auch eine offene Teilmenge einer 2-dimensionalen Fläche

nehmen. Das gibt auch eine sehr interessante Dynamik: Für den Fluss im Inneren nehmen wir den *geodätischen Fluss*, den wir im folgenden noch kennenlernen (und welcher so funktioniert, dass Vektoren mit konstanter Geschwindigkeit wandern, und zwar auf „geradestmöglichen“ Kurven, sogenannten *Geodätischen*). Am Rand können wir Vektoren, die tangential an die Fläche sind und aus G heraus zeigen, wie bisher so reflektieren, dass sie immer noch tangential an die Fläche sind und in G hinein zeigen. Momentan beschränken wir uns aber noch auf den euklidischen Fall, der natürlich leichter zu untersuchen ist.

Da der Rand glatt (oder zumindest stückweise glatt ist), können wir ihn parameterisieren mit einer Funktion $a : \mathbb{R}/(L\mathbb{Z}) \rightarrow \partial G$, die definiert ist auf

$$\mathbb{R}/(L\mathbb{Z}) = [0, L) \pmod{L},$$

d.h. dem Kreis der Länge L , mit $L := \text{Länge}(\partial G)$.

Definition. Statt dem Billiard-Fluss untersuchen wir meist gleich die **Billiard-Abbildung**, welche Punkten auf dem Rand des Gebiets G sowie entsprechenden Richtungen (beschrieben durch den Winkel mit der Tangente), also einem Tupel (p, θ) den Punkt p' zuordnet, an welchem der Billiard-Fluss als nächstes den Rand trifft, sowie den entsprechenden Winkel θ' . Wenn wir den Rand mit dem Parameter s parametrisieren, können wir in die Billiard-Abbildung auch gleich den Parameterwert s statt dem Punkt p einsetzen. Wir erhalten dann eine Abbildung

$$f : ([0, L) \times (0, \pi)) \rightarrow ([0, L) \times (0, \pi)),$$

wobei das Intervall $[0, L)$ zu verstehen ist „ \pmod{L} “. Hierbei ist $\theta \in (0, \pi)$ der Winkel zwischen dem Orbit von φ am Rand mit der „positiv orientierten Tangente“ von ∂G .

f ist also auf einem Zylinder definiert. Dabei müssen wir uns merken, dass die Koordinate s für die zyklische und θ für die lineare Komponente steht, was anfangs etwas verwirrend scheint. (Würden wir für θ das ganze Intervall $[0, 2\pi)$ zulassen, dann wäre das natürlich auch eine zyklische Koordinate. Aber wir identifizieren θ mit $2\pi - \theta$, und das macht den Kreis zum Intervall.)

Example. Für $G =$ die Kreisscheibe von Radius ρ haben wir $L = 2\pi\rho$ und f ist der **lineare Twist**

$$f(s, \theta) = (s + \theta, \theta),$$

denn der Winkel θ ändert sich nicht. Für diese Abbildung sind alle Kreise

$$K_c := \{(s, \theta) \mid s \in [0, L), \theta = c\}$$

f -invariante Mengen.

Remark. Auch für jedes andere Billiard gilt, dass die Abbildung f eine **Twist**-Eigenschaft hat: Das Bild der senkrechten Linie $\{(0, \theta) \mid \theta \in (0, \pi)\}$ unter f ist eine Kurve, die $(0, 0)$ und $(0, \pi)$ verbindet und dabei einmal in θ -Richtung um den Zylinder herumgeht. Variationsansatz und Erzeugendenfunktion

Sei G nun ein strikt konvexes Gebiet mit überall differenzierbarem Rand.

Zuerst eine einfache Frage: Für welche Punkte p, p' gibt es ein Orbit der Billiard-Abbildung, in welchem p und p' hintereinander vorkommen?

Die Antwort ist natürlich einfach: Für alle $p \neq p'$. Das Orbit wird einfach gebildet aus der geraden Strecke zwischen p und p' (und kann vorwärts und rückwärts zu einem Orbit von unendlicher Länge fortgesetzt werden).

Als nächstes eine etwas schwierigere Frage: Wann liegen 3 Punkte p_{-1}, p_0, p_1 auf einem Orbit? Anders gefragt: Gegeben p_{-1} und p_1 , für welche p_0 gilt, dass ein Orbit durch p_{-1}, p_0, p_1 geht (hintereinander und in dieser Reihenfolge)?

Darauf gibt es zwei Antworten:

- (1) (Die offensichtliche Antwort:) p_0 muss so liegen, dass der Winkel der Linie von p_{-1} nach p_0 mit $T_{p_0} \partial G$ gleich dem Winkel von $T_{p_0} \partial G$ mit der Linie von p_0 nach p_1 ist.
- (2) (Die Variationsantwort:) p_0 muss so liegen, dass die Summe $d(p_{-1}, p_0) + d(p_0, p_1)$ extremal wird.

Die zweite Antwort ist äquivalent zur ersten. Dies lässt sich zwar auch durch geometrische Argumente zeigen, aber wir sehen gleich einen Beweis mittels sog. Erzeugendenfunktionen.

Definition. Die **Erzeugendenfunktion** des Billiards ist

$$H(s, s') = -d(a(s), a(s')).$$

Example. Für $G =$ die Kreisscheibe mit Radius 1 ergibt sich

$$H(s, s') = -2 \sin \left(\frac{1}{2}(s' - s) \right).$$

Lemma. Es gelten:

$$(1) \quad \frac{\partial}{\partial s'} H(s, s') = -\cos \theta',$$

$$(2) \quad \frac{\partial}{\partial s} H(s, s') = \cos \theta.$$

Beweis. Wir beweisen die erste Formel; die zweite kann analog bewiesen werden.

$$\begin{aligned}
\frac{\partial}{\partial s'} H(s, s') &= -\frac{d}{ds'} d(a(s), a(s')) \\
&= -\frac{d}{ds'} \sqrt{\langle a(s') - a(s), a(s') - a(s) \rangle} \\
&= -\frac{1}{2\sqrt{\langle a(s') - a(s), a(s') - a(s) \rangle}} \\
&\quad \cdot \frac{d}{ds'} \langle a(s') - a(s), a(s') - a(s) \rangle \\
&= -\frac{1}{2\|a(s') - a(s)\|} \cdot 2 \left\langle \frac{d}{ds'} a(s'), a(s') - a(s) \right\rangle \\
&= -\left\langle \frac{d}{ds'} a(s'), \frac{a(s') - a(s)}{\|a(s') - a(s)\|} \right\rangle \\
&= -\cos \theta',
\end{aligned}$$

da $\frac{d}{ds'} a(s')$ ein Vektor der Länge 1 ist. \square

Abgesehen von diesem algebraischen Beweis gibt es noch einen elementar-geometrischen.

Jetzt formulieren wir die Aussage von vorhin „Wenn ein Orbit durch p_{-1}, p_0, p_1 geht, muss p_0 so liegen, dass die Summe $d(p_{-1}, p_0) + d(p_0, p_1)$ extremal wird“ erneut:

Theorem. *Wenn ein Orbit der Billiard-Abbildung durch die 3 Punkte $p_{-1} = a(s_{-1}), p_0 = a(s_0), p_1 = a(s_1)$ geht, dann ist $s = s_0$ ein kritischer Punkt des Funktionals*

$$\mathcal{L}(s) := H(s_{-1}, s) + H(s, s_1),$$

d.h.

$$\frac{d}{ds} \Big|_{s=s_0} \mathcal{L}(s) = 0.$$

Beweis. Dies folgt sofort durch Zusammensetzen der beiden Formeln des Lemmas. Volumenerhaltung von Billiards \square

Im Folgenden betrachten wir ein Orbitstück, so dass der Punkt (s, θ) unter der Billiard-Abbildung f abgebildet wird auf (s', θ') . Das hat zur Auswirkung, dass s' jetzt eine von (s, θ) abhängige Variable ist.

Wenn wir die Koordinaten θ umschreiben mittels

$$r := -\cos \theta$$

und analog

$$r' := -\cos \theta',$$

dann vereinfachen sich die Gleichungen der Erzeugendenfunktion:

$$\frac{\partial}{\partial s'} H(s, s') = r',$$

$$\frac{\partial}{\partial s}H(s, s') = -r.$$

Wenn dann wie vorhin (s', θ') das Bild unter f von (s, θ) ist, dann können wir die erste Gleichung auch schreiben als

$$\frac{\partial}{\partial s'}H(s, s') = R = R(s, r).$$

In diesen Koordinaten können wir ein wichtiges Faktum über die Billiard-Abbildung zeigen:

Theorem. Die Billiard-Abbildung f ist volumenerhaltend in den Koordinaten (s, r) . Präziser formuliert: Die Abbildung

$$g : (s, r) \mapsto (S, R) = (S(s, r), R(s, r)),$$

die gegeben ist durch die Konjugation

$$g = h \circ f \circ h^{-1}$$

von f mit der Abbildung

$$h : \{(s, \theta) \in [0, L) \times (0, \pi)\} \rightarrow \{(s, r) \in [0, L) \times (-1, 1)\},$$

$$h((s, \theta)) := (s, -\cos \theta)$$

ist eine volumenerhaltende Abbildung (bezüglich des Standard-Volumens, d.h. Flächeninhalt, in \mathbb{R}^2). Die Abbildung ist außerdem auch orientierungserhaltend.

Beweis. Definiere

$$\tilde{H}(s, r) := H(s, S(s, r)).$$

Dann folgt:

$$\begin{aligned} \frac{\partial \tilde{H}(s, r)}{\partial s} &= -r + R(s, r) \frac{\partial S(s, r)}{\partial s} \\ \frac{\partial}{\partial r} \left(\frac{\partial \tilde{H}(s, r)}{\partial s} \right) &= -1 + R(s, r) \frac{\partial^2 S(s, r)}{\partial r \partial s} + \left(\frac{\partial R(s, r)}{\partial r} \right) \frac{\partial S(s, r)}{\partial s} \\ \frac{\partial \tilde{H}(s, r)}{\partial r} &= R(s, r) \frac{\partial S(s, r)}{\partial r} \\ \frac{\partial}{\partial s} \left(\frac{\partial \tilde{H}(s, r)}{\partial r} \right) &= \left(\frac{\partial R(s, r)}{\partial s} \right) \frac{\partial S(s, r)}{\partial r} + R(s, r) \frac{\partial^2 S(s, r)}{\partial s \partial r}. \end{aligned}$$

Gleichsetzen von

$$R(s, r) \frac{\partial^2 S(s, r)}{\partial s \partial r} = R(s, r) \frac{\partial^2 S(s, r)}{\partial r \partial s}$$

liefert

$$\left(\frac{\partial R(s, r)}{\partial r} \right) \frac{\partial S(s, r)}{\partial s} - \left(\frac{\partial R(s, r)}{\partial s} \right) \frac{\partial S(s, r)}{\partial r} = 1,$$

mit anderen Worten, die Jacobi-Determinante von g ist

$$\det \frac{\partial g(s, r)}{\partial (s, r)} = \det \frac{\partial (S(s, r), R(s, r))}{\partial (s, r)} = +1.$$

□

Remark. Die Abbildung h ist ein Diffeomorphismus auf seinem (offenen) Definitionsbereich, allerdings ist die Ableitung nach oben unbeschränkt für $\theta \rightarrow 0$ oder $\theta \rightarrow \pi$. Deswegen läßt sich h nicht auf $\theta = 0$ oder $\theta = \pi$ differenzierbar fortsetzen.

Außerdem sehen f und g nahe dieses Randes $\theta = 0$, $\theta = \pi$ verschieden aus: Die Abbildung g ist eine Twistabbildung, welche die Menge $\{(0, \theta) \mid \theta \in (0, \pi)\}$ in eine Kurve abbildet, welche bei $\theta = 0$ und $\theta = \pi$ eine senkrechte Tangente hat. Dagegen hat das Bild von $\{(0, \theta) \mid \theta \in (0, \pi)\}$ unter f endliche Steigung (welche gegeben ist durch die Krümmung von ∂G bei $s = 0$).