Ergodentheorie

Prof. Roland Gunesch

Universität Hamburg

Vorlesung im Wintersemester 2007/2008

Letzte Änderung an diesem Skript: 12.12.07

INHALTSVERZEICHNIS

1. Literatur

(1) Katok & Hasselblatt: Gründlich und sehr vollständig, behandelt auch die Ergodentheorie umfassend und präzise.
(2) Hasselblatt & Katok: Sehr gut zu lesen, generell empfehlenswert für Dynamische Systeme, allerdings nicht direkt ein Ergodentheoriebuch.
(3) Petersen: Empfehlenswert, da speziell für Ergodentheorie.
(4) Walters: Ein Klassiker zur Ergodentheorie, war früher Standardwerk; ist auf dem Stand von 1982, und seitdem hat sich einiges geändert.

Weitere Literatur wird an den entsprechenden Stellen noch bekannt gegeben.

2. Einleitung: Was ist Ergodentheorie?

Ergodentheorie ist ein Aspekt der Theorie von **Dynamischen Systemen**:

Gegeben sei eine Transformation $T : X \to X$, d.h. eine Abbildung von einer Menge X auf sich selbst. Dann können wir T mehrfach anwenden. Wir untersuchen im Folgenden das Verhalten von

$T^n(x) := T(T(\dots T(x))\dots)$ für $x \in X$. Dies ist ein **dynamisches System mit diskreter Zeit**. Außerdem gibt es noch **dynamische Systeme mit kontinuierlicher Zeit**: Das ist ein **Fluss** auf einer Menge X, d.h. eine Abbildung $\varphi : X \times \mathbb{R} \to X$,

$$(x,t) \mapsto \varphi(x,t) =: \varphi_t(x),$$

welche die Eigenschaften hat:

- $\varphi_0(x) = x$ für alle $x \in X$,
- $\varphi_s \circ \varphi_t = \varphi_{s+t}$.

Dazu kann der Raum X eine der folgenden Strukturen haben:

- Abstandsfunktion: Dann reden wir von **metrischen dynamischen Systemen**.
- Offene Mengen: Dann heißt es **topologische Dynamik**.
- Differenzierbare Struktur: Dazu gibt es **differenzierbare Dynamik**.
- Maß: Dann gibt es **Maß-Dynamik**, genauer gesagt **maßerhaltende dynamische Systeme**, und deren Studium heißt **Ergodentheorie.**

Das Studium von Ergodentheorie schließt das Betrachten von topologischen Strukturen (d.h. offenen Mengen) oder differenzierbaren Strukturen (d.h. Ableitungen) nicht aus. Im Gegenteil werden Fragen aus der Theorie dynamischer Systeme besonders interessant, wenn wir zwei oder mehr der obigen Strukturen haben.

Welche Fragen betrachten wir?

Beispielsweise interessieren wir uns für **Mittelwerte**

$$\frac{1}{n}\sum_{i=0}^{n-1} f(T^n(x))$$

wobei $T : X \to X$ wie oben ist und $f : X \to \mathbb{R}$ eine reelle Funktion ist. Für endliches $n \in \mathbb{N}$ existiert dieser Mittelwert natürlich, und uns interessiert, wann der Mittelwert

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

existiert. Hierbei ist es nützlich, ein **Maß** zu haben (d.h. eine Sigma-additive Funktion auf einer Sigma-Algebra auf X). Hier ein beispielhafter Satz (alle Definitionen folgen danach).

Birkhoffs Ergodensatz:

Theorem. *Sei $T : (X, \mu) \to (X, \mu)$ maßerhaltend. Seit $f : X \to \mathbb{R}$ in $L^1(\mu)$. Dann gilt für μ-fast alle $x \in X$ (d.h. alle $x \in X$ außer einer Menge von μ-Maß 0), dass*

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

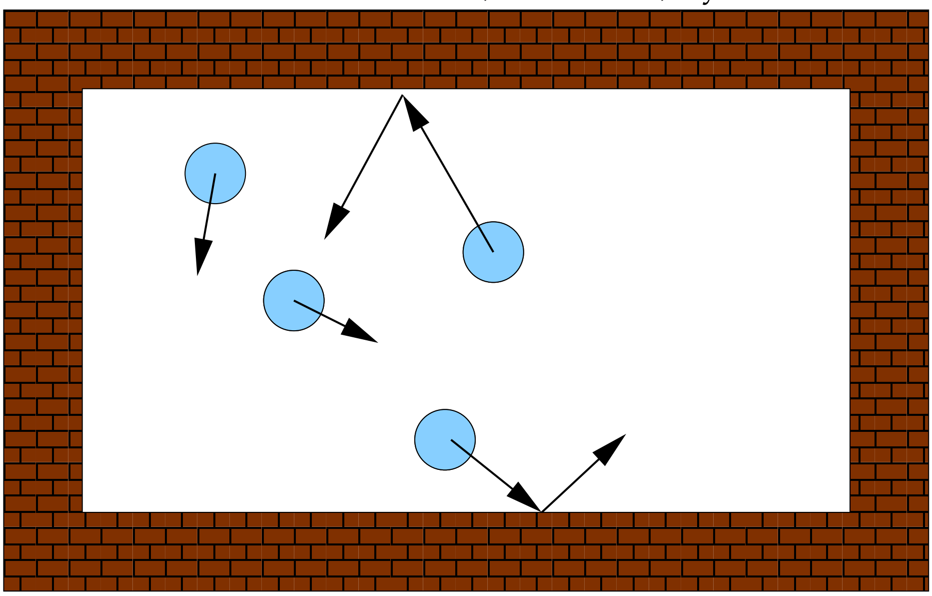
existiert. Ist T bezüglich μ auch ***ergodisch*** *(Erklärung später), dann ist der Grenzwert (fast überall) unabhängig von x.*

Solche Mittelwerte sind wichtig, weil es Vorgänge gibt, wo nur solche Mittelwerte beobachtbar sind.

Motivation für solche Fragen war unter das Studium der Thermodynamik in der theoretischen Physik.

2.1. **Thermodynamik, Gasdynamik.** Physiker haben sich für folgendes Problem interessiert: Gegeben ein Behälter mit Gasmolekülen darin, wie wird sich das Gas verhalten?

ABBILDUNG 1. Moleküle, Kollisionen, Dynamik



Wir könnten verrsuchen, dies mit Newtons Mechanik $F = m \cdot a$ zu verstehen: Gegeben alle Anfangspositionen und Geschwindigkeiten der Gaspartikel, können wir die Bewegung vorhersagen... bis zur nächsten Kollision. Dann müssen wir neue Bewegungsrichtungen ausrechnen.

Dummerweise geht die Dynamik ziemlich schnell, und Kollisionen treten häufig auf. Außerdem sind die oben erwähnten Anfangsdaten zu viele: in einem makroskopischen Behälter sind 10^{20} bis 10^{30} Moleküle. Dieses Verfahren wäre also nicht praktisch durchführbar. (Deswegen haben die Physiker auch ein paar Jahrhunderte nicht bemerkt, dass diese *deterministische Dynamik* das Verhalten kleiner Objekte wie Gasatome nicht richtig beschreibt. Seit Heisenberg und seiner Unschärferelation wissen wir, dass exakte „Anfangsdaten“ nicht existieren.)

Was wir aber mit unseren Gasmolekülen gut machen können, ist, makroskopische Größen zu betrachten: Z.B. *Druck* oder *Temperatur.* Die Temperatur ist in der Physik definiert als

$$\sum m_i ||\dot{x}_i||^2.$$

Als Druck (auf eine Fläche) verstehen wir im Wesentlichen, welchen Impuls pro Zeit Atome in eine bestimmte Richtung auf ein Stück Wand des Behälters übertragen. Da eine Messung bestimmte Zeit braucht (z.B. eine Millisekunde) und diese Größen sich derzeit sehr oft ändern messen wir in Wirklichkeit einen Mittelwert über sehr viele Werte. Deswegen macht es Sinn, solche asymptotischen Mittel zu untersuchen.

Das Wort **ergodisch** wurde vom Physiker Boltzmann kreiert. Es setzt sich zusammen aus den griechischen Worten für „Energie“ und „Weg“.

Die physikalische Theorie der Dynamik von Gasmolekülen ist übrigens im täglichen Leben nützlich. Zum einen gibt es das bekannde Gasgesetz

$$p \cdot V = \text{const.} \cdot T$$

(Druck ist proportional zur Temperatur, Volumen ist proportional zur Temperatur, Druck ist umgekehrt proportional zum Volumen). Außerdem klärt die Dynamik z.B. folgende Frage: Wenn wir unsere Kleidung mit *warmem* Wasser übergießen und uns in den *warmen* Wind stellen, wieso fühlt sich die nasse Kleidung dann alsbald *kalt* an, kälter als das ursprüngliche Wasser, die Umgebung und der Körper? Allein durch Wärmeleitung kann so etwas nicht passieren. Die Antwort ist: Die Flüssigkeitsatome tauschen Energie durch Stöße aus, werden schneller und langsamer, und am Rand des Wassers können nur die schnellen Atome sich lösen und davonfliegen. Zurück bleiben die langsamen, mit wenig kinetischer Energie, und Temperatur ist mittlere kinetische Energie.

2.2. **Informationstheorie.** Außerdem interessant ist **Informationstheorie**: Gegeben einen *Sender* von Information, einen *Kanal* zur Informationsübertragung (mit endlicher Geschwindigkeit und mit Rauschen oder dergleichen), und einen *Empfänger*, wieviel Information kann man über den Kanal übertragen, und wie? Z.B. bei Ihrem Internetzugang über DSL, über Modem oder dergleichen. Dies ist ein wichtiges Gebiet in der Technik. Es geht zurück auf den amerikanischen Ingenieur Claude Shannon, der in den 1940ern solche Fragen studiert hat.

Zum Beispiel wurde in der Computertechnik früher einmal mit *Modems* Daten übertragen, und in den 1980er Jahren wurde geglaubt, dass man nicht mehr als (sagen wir mal) 32 Kb/s übertragen könnte. Ein paar Jahre später gab es Modems, die schneller waren, und dann hieß es, OK, aber bei 56 Kb/s ist Schluss, mehr geht nun wirklich nicht über ein Telefonkabel. Und heute benutzen wir DSL, einen Faktor 10 (oder 100) schneller. Über genau dasselbe Telefonkabel.

Information wird u.A. mit Symbolräumen studiert: Sender und Empfänger benutzen eine endliche Menge von Symbolen (z.B 0 und 1), die nacheinander übertragen werden. Dazu betrachten wir die Menge der unendlichen Symbolfolgen

$$X = \{0,1\}^{\mathbb{N}} = \{(\dots, x_{-1}, x_0, x_1, \dots) \mid \forall i : x_i \in \{0,1\}\}$$

und die **Shift-Abbildung** $T : X \to X$,

$$(T(x))_n = x_{n+1},$$

die die Folge um eine Position nach links schiebt. Das sieht zunächst nach einer sehr simplen Abbildung aus, aber sie wird sehr interessant, wenn wir den Raum X mit einer Struktur versehen. Z.B. mit einer Metrik:

$$d(x,y) = \sum_{i\in\mathbb{Z}} 2^{-|i|}|x_i - y_i|.$$

Dann gilt z.B.: Die Menge an periodischen Punkten (d.h. $p \in X$ mit $\exists n \in \mathbb{N}$ so dass $T^n p = p$) ist dicht in X. Es gibt Punkte p, so dass $T^{\mathbb{N}} p$ dicht in X ist.

3. Definitionen

Im Folgenden seien alle Abbildungen messbar; dies schreiben wir nicht immer hin. Weiterhin seien alle Mengen, die in Definitionen usw. auftauchen, immer Element der passenden σ-Algebra.

Definition 3.1. Eine Abbildung $T : (X, \mu) \to (X, \mu)$ heißt **maßerhaltend** (bezüglich dem Maß μ), wenn für jedes (messbare) A gilt:

$$\mu(T^{-1}(A)) = \mu(A).$$

Wir sagen dann auch, dass das Maß μ unter T **invariant** ist.

In der Definition steht T^{-1} und nicht T, damit Abbildungen wie

$$T : [0, 1) \to [0, 1), \quad x \mapsto 2x \text{ (mod 1)}$$

maßerhaltend sind bezüglich dem Lebesgue-Maß. Denn für dieses T ist das Bild eines (kleinen) Intervalls ein Intervall doppelter Länge, hat also nicht dasselbe Maß. Das *Urbild* eines Intervalls dagegen besteht aus zwei Intervallen von genau halber Länge, hat also dasselbe Maß.

Für *invertierbare* Abbildungen T ist die Bedingung, dass T das Maß μ erhält, äquivalent dazu, dass für alle A gilt: $\mu(A) = \mu(T(A))$.

Definition 3.2. Sei $T : (X, \mu) \to (X, \mu)$ maßerhaltend. Dann heißt T **ergodisch** bezüglich μ, wenn für jede messbare Menge A gilt: Wenn A invariant ist unter T, d.h. $T^{-1}(A) = A$, dann gilt $\mu(A) = 0$ oder $\mu(A) = 1$. Wir sagen dann auch, das System (T, μ) ist ergodisch, oder, μ ist ergodisch bezüglich T.

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.

Wenn das Maß nicht endlich ist, muss in der Definition die Bedingung $\mu(A) = 1$ ersetzt werden durch die Bedingung $\mu(X \setminus A) = 0$. Auf solchen Räumen kann man auch Ergodentheorie machen, die **unendliche Ergodentheorie.** Wir erwähnen diese im Folgenden nicht, denn die dortigen Sätze sind entweder offensichtliche Verallgemeinerungen von denen, die wir hier behandeln (und dann können wir auch gleich bei der etwas einfacheren Notation bleiben) oder schwer zu beweisen.

Definition 3.3. T heißt **schwach mischend**, wenn für alle Mengen A, B gilt:

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \big|\mu(A \cap T^{-n}B) - \mu(A)\mu(B)\big| = 0.$$

Wo „schwach“ steht, gibt es meist auch ein „stark“, so auch in diesem Fall:

Definition 3.4. T heißt **stark mischend** bzw. **mischend**, wenn für alle Mengen A, B gilt:

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

Theorem 3.5.
- *Mischen impliziert schwaches Mischen.*
- *Schwaches Mischen impliziert Ergodizität.*

Beweis. Die erste Aussage ist klar, denn wenn eine Folge gegen 0 konvergiert, dann sind für jedes $\varepsilon > 0$ alle bis auf endlich viele Folgenglieder vom Betrag $< \varepsilon$, somit auch das Mittel über genügend viele Folgenglieder $< 2\varepsilon$. Daher konvergiert das Mittel gegen 0.

Für die zweite Aussage überlegen wir uns, dassaus der Definition von schwachem Mischen folgt, dass eine Teilfolge aus dem Mittel in der Definition gegen 0 konvergiert:

$$\lim_{k\to\infty} \mu(A \cap T^{-n_k}B) - \mu(A)\mu(B) = 0.$$

(Die Existenz einer solchen Teilfolge ist sogar äquivalent zu schwachem Mischen.) Aus der Existenz einer solchen Teilfolge folgt sofort Ergodizität: Wenn A eine T-invariante Menge ist und $B := X \setminus A$, dann ist $\mu(A \cap T^{-n_k}B) = \mu(A \cap B) = \emptyset$ und somit gilt

$$0 = \lim_{k\to\infty} \mu(A \cap T^{-n_k}B) - \mu(A)\mu(B) = \mu(A)\mu(B).$$

Also ist $\mu(A) = 0$ oder $\mu(X \setminus A) = 0$. □

4. Konstruktionsprinzipien für masserhaltende Abbildungen; einfache Sätze

4.1. **Produktabbildung.**

Für Abbildungen $T : X \to X$ und $S : Y \to Y$ ist die **Produktabbildung** $T \times S$ definiert durch

$$(T \times S)(x, y) = (T(x), S(y)).$$

Für ein Maß μ auf einem Raum X und ein Maß ν auf einem Raum Y gibt es das bekannte **Produktmaß** auf $X \times Y$.

Wenn μ vollständig ist (d.h. die zugehörige Sigma-Algebra enthält alle Teilmengen von allen Nullmengen) und ν auch, dann ist das Produkt der Sigma-Algebren zwar im Allgemeinen nicht vollständig. Aber wir können es natürlich einfach vervollständigen, und deswegen können wir auch gleich annehmen, dass $\mu \times \nu$ dann auch vollständig ist.

Lemma 4.1. *Wenn $T : (X, \mu) \to (X, \mu)$ und $S : (Y, \nu) \to (Y, \nu)$ maßerhaltend sind, dann ist auch die Produktabbildung*

$$T \times S : (X \times Y, \mu \times \nu) \to (X \times Y, \mu \times \nu)$$

maßerhaltend bezüglich dem ***Produktmaß****.*

4.2. **Fluss unter einem Graphen.** Sei $T : X \to X$ eine Abbildung (im folgenden Lemma maßerhaltend) und $f : X \to (0, \infty)$ eine Funktion (im folgenden Lemma messbar). Dann ist auf der Menge

$$\Big(X \times [0, \infty)|_{\text{unterhalb Graph}(f)}\Big) / \sim$$

(d.h. auf $((x, y) \in X \times [0, \infty) : y \leq f(x)) / \sim)$ mit der Äquivalenzrelation $(x, f(x)) \sim (T(x), 0)$ ein Fluss, nämlich der **Fluss unter dem Graphen** von f, definiert durch

$$\frac{d}{dt}(x, y) = (0, 1).$$

Der Fluss funktioniert also so, dass an jedem Punkt (x, y) die Flusslinien senkrecht nach oben gehen, bis der Graph von f erreicht ist, und dann springt die Flusslinie zu $(T(x), 0)$.

Lemma 4.2. *Wenn $T : (X, \mu) \to (X, \mu)$ maßerhaltend und $f : X \to (0, \infty)$ messbar ist, dann ist der Fluss unter dem Graphen von f maßerhaltend.*

4.3. **Induzierte Abbildung auf einer Untermenge.** Sei $T : X \to X$ maßerhaltend und $A \subset X$ eine Teilmenge mit $\mu(Y) > 0$, für die gelte, dass es für fast alle Punkte $a \in A$ ein $n = n(a) \in \mathbb{N}$, $n(a) > 0$, gebe mit $T^{n(a)}(a) \in A$. Dann ist die **induzierte Abbildung** definiert durch

$$T_A : A \to A, \quad a \mapsto T^{n(a)}(a).$$

Lemma 4.3. *Diese induzierte Abbildung ist wieder maßerhaltend (bzgl. demselben Maß μ).*

Die Existenz von $n(a)$ für fast alle $a \in A$ gilt übrigens immer für maßerhaltendes T; dieses Faktum heißt *Poincaré'scher Rekurrenzsatz.*

4.4. **Induzierte Abbildung auf einer Obermenge.** Sei $T : X \to X$ maßerhaltend und $X = Y_1 \supset Y_2 \supset \ldots$ eine verschachtelte Folge von Teilmengen. Wir bauen uns nun ein Objekt, das aus verschiedenen Schichten besteht: in der ersten Schicht steht Y_1, darüber in der zweiten Schicht Y_2, usw. Dazu definieren wir

$$\hat{X} = \bigcup_{i \in \mathbb{N}} X_i$$

wobei X_i eine Kopie von Y_i ist und alle X_i disjunkt sind. $\hat{X}$ heißt (je nach Autor bzw. Vorlesung) **skyscraper**, **Rokhlin tower**, **Buhné'sche mesepotamische Pyramide** oder **Kaltner'sche Torte**.

Auf jedem X_i gibt es ein Maß (nämlich die Einschränkung von dem auf X), und somit gibt es auch eins auf $\hat{X}$, das nicht notwendigerweise endlich sein muss.

Dazu gibt es die **induzierte Abbildung**

$$\hat{T} : \hat{X} \to \hat{X}$$

die definiert ist wie folgt: Das Bild vom Punkt x im Stockwerk i ist derselbe Punkt x im Stockwerk darüber, sofern er im Stockwerk darüber enthalten ist, und ansonsten $T(x)$ im untersten Stockwerk:

$$T(x, n) = \begin{cases} (x, n+1) & \text{falls } x \in X_{n+1}, \\ (T(x), 1) & \text{sonst.} \end{cases}$$

Lemma 4.4. *Wenn T maßerhaltend ist und die Folge Y_i wie oben, dann ist auch die induzierte Abbildung $\hat{T}$ maßerhaltend.*

Nun können wir ein wichtiges Problem der Ergodentheorie formulieren:

5. Das Isomorphismusproblem

Zwei maßerhaltende Systeme $T : (X, \mu) \to (X, \mu)$ und $S : (Y, \nu) \to (Y, \nu)$ heißen **isomorph** (und zwar **maß-theoretisch isomorph** – andere Strukturen als das Maß gibt es ja bisher nicht), wenn gilt: Es gibt einen Maß-Isomorphismus $\Psi : (X, \mu) \to (Y, \nu)$, der das Maß μ in das Maß ν überführt, d.h. für alle A gilt $\mu(\Psi^{-1}(A)) = \nu(A)$, und der dabei die Abbildung T in die Abbildung S überführt, d.h.

$$\Psi \circ T = S \circ \Psi.$$

Das folgende Diagramm kommutiert also:

$$\begin{array}{ccccc} & & T & & \\ & X & \to & X & \\ \Psi & \downarrow & & \downarrow & \Psi \\ & Y & \to & Y & \\ & & S & & \end{array}$$

Das **Isomorphismusproblem** lautet nun: Gegeben zwei Systeme T und S, sind sie isomorph? Gegeben T, welche S sind dazu isomorph?

Analoge Fragestellungen gibt es in der Mathematik häufig. Z.B. in der linearen Algebra: Wann sind zwei reelle Vektorräume (mit endlicher Dimension) linear isomorph? Antwort: Genau dann, wenn die Dimensionen gleich sind. Die Dimension ist in der linearen Algebra somit eine **Invariante** (d.h. eine Größe, die für isomorphe Vektorräume gleich ist) und sogar eine **vollständige Invariante** (d.h. wenn diese Größe für verschiedene Vektorräume gleich sind, dann folgt daraus, dass diese isomorph sind).

Analog in der Topologie: Wann sind die Räume $\mathbb{R}^n$ und $\mathbb{R}^m$ isomorph? Wobei „isomorph" hier bedeutet: homöomorph. Antwort: Genau dann, wenn $m = n$. Wieder ist die Dimension eine Invariante, und zwar wieder eine vollständige.

Analog in der Differentialgeometrie: Wann sind zwei Flächen (zweidimensional), die kompakt, randfrei und orientierbar sind, isomorph? Wobei „isomorph" hier bedeutet: diffeomorph (bijektiv und differenzierbar in beiden Richtungen). Antwort: Wenn sie dieselbe Zahl von Henkeln haben. Hier ist also die Zahl der Henkel eine (vollständige) Invariante.

So, und was ist die entsprechende Größe in der Ergodentheorie?

Invarianten gibt es in der Tat, insbesondere die **Entropie**, die wir noch kennenlernen werden. Diese ist keine vollständige Invariante, d.h. es gibt nicht-isomorphe Systeme mit gleicher Entropie. Besonders häufig ist das der Fall, wenn die Entropie 0 ist.

Und welche Größe ist nun eine *vollständige* Invariante für das (ergodentheoretische) Isomorphismusproblem? Niemand weiss es.

Es gibt in der Ergodentheorie zwei wesentliche Denkrichtungen: Die einen meinen, das Isomorphismusproblem sei das wichtigste (oder sogar das einzig wichtige) Thema in der Ergodentheorie. Diese Denkweise beschreibt die **theoretische Ergodentheorie**. Die anderen (z.B. wir in dieser Vorlesung) interessieren sich für Anwendungen der Ergodentheorie in anderen Bereichen der Mathematik. Diese Denkweise heißt **angewandte Ergodentheorie**.

6. Ergodizität: Äquivalente Bedinugungen

Hier betrachten wir ein paar äquivalente Charakterisierungen von Ergodizität:

Theorem 6.1. *Sei (X, μ) ein Wahrscheinlichkeitsraum (im folgenden W-Raum genannt). $T : (X, \mu) \to (X, \mu)$ sei maßerhaltend. Dann sind äquivalent:*

(1) *T ist ergodisch, d.h. die einzigen T-invarianten Mengen haben Maß 0 oder 1.*
(2) *Alle messbaren reellen Funktionen f, die T-invariant sind*

$$\text{(das heißt } f \circ T = f\text{),}$$

sind konstant.
(3) *Alle reellen Funktionen f in $L^p(\mu)$, die T-invariant sind*

$$\text{(das heißt } f \circ T = f\text{),}$$

sind konstant.

Diese Bedingungen sind äquivalent zu den Bedingungen, die wir erhalten, wenn wir die Invarianzbedingung

$$f \circ T = f$$

ersetzen durch die „Halbinvarianz"-Bedingung

$$f \circ T \geq f.$$

Alle diese Bedingungen sind äquivalent zu denselben Bedingungen mit den Worten „invariant" und „konstant" ersetzt durch „fast überall invariant" und „fast überall konstant".

Somit haben wir genaugenommen 10 äquivalente Bedingungen für Ergodizität angegeben. Wobei wir bei solchen Aussagen ohnehin nicht zwischen „Menge von Maß 0" und „leerer Menge" unterscheiden, da wir immer mod 0 rechnen, also sind die letzten 5 Bedingungen für uns ohnehin dasselbe wie die ersten 5.

Beweis. Wir zeigen gleichzeitig $(2) \Rightarrow (1)$ und $(3) \Rightarrow (1)$. Wenn E invariant ist, dann setze $f = \chi_E$. Da dies invariant ist, ist es konstant. Also konstant 0 oder 1, somit $E = \emptyset$ oder $E = X$ (mod 0).

Jetzt zeigen wir gleichzeitig $(1) \Rightarrow (2)$ und $(1) \Rightarrow (3)$. Sei f invariant und nicht konstant, dann gibt es $c \in \mathbb{R}$ mit

$$A = \{x : f(x) < c\}$$

und

$$B = \{x : f(x) \geq c\},$$

so dass $\mu(A) > 0, \mu(B) > 0$. Wegen der Invarianz von f sind auch die Mengen A und B invariant. Wegen Ergodizität haben sie also Maß 0 oder 1, im Widerspruch zu $\mu(A) > 0, \mu(B) > 0$.

Als nächstes überlegen wir uns, warum die bisherigen Argumente in diesem Beweis gültig bleiben, wenn die Invarianzbedingung $f \circ T = f$ durch

$$f \circ T \geq f$$

ersetzt wird:

Beim Schritt $(2) \Rightarrow (1)$ und $(3) \Rightarrow (1)$ ist $f = \chi_E$ invariant und somit insbesondere halbinvariant, also erst recht konstant.

Beim Schritt $(1) \Rightarrow (2)$ und $(1) \Rightarrow (3)$ gilt jetzt

$$x \in B \Rightarrow T(x) \in B$$

und

$$x \in A \Rightarrow T^{-1}(x) \in A.$$

Hier ist die Menge A immer noch enthalten in $T^{-1}(A)$, und B ist enthalten in $T(B)$. Da wir voraussetzen, dass T maßerhaltend ist, gilt $\mu(A) = \mu(T^{-1}A)$. Somit ist

$$\mu\left(T^{-1}(A) \setminus A\right) = 0,$$

also A in Wirklichkeit schon invariant unter T. Wegen Ergodizität hat A also Maß 0 oder 1, im Widerspruch zu $\mu(A) > 0$,$\mu(X \setminus A) > 0$. Wir haben also gezeigt: Jede reelle Funktion, die unter einer maßerhaltenden Abbildung *halbinvariant* ist, ist bis auf eine Nullmenge auch *invariant*. □

7. ERGODENSÄTZE

Hier betrachten wir die zwei wichtigsten Ergodensätze. Das sind zwei der wichtigsten Sätze aus der ganzen Ergodentheorie.

7.1. **Birkhoffs Ergodensatz (oder: Wie wir die Zukunft vorhersagen).** In Anwendungen ist es häufig 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!

Theorem 7.1. ***Birkhoffs Ergodensatz:*** *Sei (T, μ) ein maßerhaltendes System, μ ein W-Maß. Sei $f : X \to \mathbb{R}$ eine reelle Funktion und in $L^1(\mu)$, d.h. integrabel. Dann existiert für fast alle $x \in X$ die Funktion*

$$f^*(x) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i(x)).$$

Die Funktion f^* ist wieder in $L^1(\mu)$. Die Funktion f^* ist T-invariant, d.h.

$$f^*(x) = f^*(T(x))$$

für fast alle $x \in X$. Außerdem gilt:

$$\int_X f d\mu = \int_X f^* d\mu.$$

Der Beweis steht am Ende dieses Abschnitts über Ergodensätze. Zunächst sollten wir verstehen, warum dieser Satz so wichtig ist. Dazu zunächst eine wichtige Schlussfolgerung für den Fall, dass (T, μ) ergodisch ist. Diese Schlussfolgerung ist das wichtigste überhaupt an diesem Satz, und ist auch der Grund, warum dieser überhaupt *Ergoden*satz heißt. Wenn nämlich das System ergodisch ist, dann muss die invariante Funktion f^* konstant sein (bis auf eine μ-Nullmenge). Es gilt also

Corollary 7.2. *Sei T, μ, f wie vorhin, und außerdem sei (T, μ) ergodisch. Dann existiert für fast alle $x \in X$ das Mittel*

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

und hat für fast alle $x \in X$ denselben Wert, nämlich

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i(x)) = \int_X f d\mu.$$

Remark 7.3. 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 wisssen *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!

Beweis. Dass die (invariante) Funktion f^* konstant sein muss, folgt aus der Ergodizität. Diese Konstante hat die Eigenschaft $\int_X f d\mu = \int_X f^* d\mu$, und weil μ ein W-Maß ist, muss die Konstante gleich $\int_X f d\mu$ sein. □

Als Nächstes betrachten wir eine weitere Anwendung des Birkhoff'schen Ergodensatzes. Im Prinzip ist es meist schwer, über eine gegebene Zahl (z.B. $\pi = 3.14...$) Aussagen über die Häufigkeit der einzelnen Ziffern zu machen. Es sieht so aus, als wären diese alle gleich häufig, doch weiß niemand, wie das für eine konkrete Zahl zu beweisen ist. Für einige Zahlen sind die Ziffern sicher nicht gleich häufig, z.B. $\frac{14}{99} = 0.141414...$. Daher ist folgender Satz besonders interessant:

Theorem 7.4. ***Borels Satz über normale Zahlen:*** *Für fast alle $x \in [0,1]$ (bezüglich Lebesgue-Maß) gilt: Die Häufigkeit der Ziffern 0 und 1 in der Binärdarstellung von x ist gleich, d.h.*

$$\lim_{n\to\infty} \frac{1}{n} (\#\textit{Einsen in den ersten } n \textit{ Nachkommastellen}) = \frac{1}{2}.$$

Beweis. Betrachte die Funktion $f : [0,1) \to \mathbb{R}$, $f = \chi_{[1/2,1)}$. Dann gilt:

$$f(x) = \begin{cases} 0 & \text{erste Nachkommastelle ist } 0 \\ 1 & \text{erste Nachkommastelle ist } 1 \end{cases}.$$

Betrachte $T(x) = 2x \pmod 1$. Dann gilt:

$$f(T^i x) = \begin{cases} 0 & i\text{-te Nachkommastelle ist } 0 \\ 1 & i\text{-te Nachkommastelle ist } 1 \end{cases}.$$

Somit ist

$$\#\text{Einsen in den ersten } n \text{ Nachkommastellen} = \sum_{i=1}^{n} f(T^i(x)$$

und

$$\lim_{n\to\infty} \frac{1}{n} (\#\text{Einsen in den ersten } n \text{ Nachkommastellen})$$

$$= \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} f(T^i(x).$$

Letzteres ist gleich $\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i(x))$, d.h. dieselbe Summe mit verschobenen Indexgrenzen.

Da die Abbildung T ergodisch ist (Beweis später in Satz 7.5), gilt: für fast alle $x \in X$ ist

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i(x)) = \int_{[0,1)} f d\lambda = \frac{1}{2}.$$

□

Aufgabe: Gilt dies auch bezüglich anderer Basen, z.B. im Dezimalsystem?

Theorem 7.5. *Die Abbildung*

$$T : [0,1) \to [0,1), \quad x \mapsto 2x \quad \textit{(mod 1)}$$

ist mischend.

Beweis. Sei $\Omega = \{0,1\}^{\mathbb{N}}$,

$$\hat{T} : \Omega \to \Omega, \quad \hat{T}(x_0, x_1, x_2, \dots) = (x_1, x_2, \dots).$$

Dann sind T und $\hat{T}$ maß-theoretisch konjugiert mittels

$$x = 0.x_0x_1x_2 \dots \mapsto (x_0, x_1, x_2, \dots)$$

wobei x_i die i-te Binärziffer von x ist. Hierbei ist das Maß auf $[0,1)$ das Lebesguemaß, und das Maß ν auf Ω ist gegeben durch die Regel, dass ein Zylinder

$$\{(x_0, x_1, x_2, \dots) \mid x_{i_1} = a_1,\ x_{i_2} = a_2,\ \dots,\ x_{i_k} = a_k\}$$

mit k vorgeschriebenen Stellen das Maß 2^{-k} hat.

Wir zeigen sogar: $\hat{T}$ ist mischend. Daraus folgt Ergodizität von T.

Wir benutzen den Satz (Satz 9.9), dass es genügt, die Mischen-Eigenschaft nicht auf der ganzen σ-Algebra zu zeigen, sondern nur auf einer erzeugenden Semi-Algebra, d.h. einer Menge, für die der Durchschnitt zweier Elemente wieder in ihr liegt und für die das Komplement jedes Elements endliche Vereinigung von Elementen der Menge ist. Z.B. ist die Menge aller Zylinder eine Semi-Algebra, die die σ-Algebra für ν erzeugt.

Wenn Z ein Zylinder ist, der k vorgeschriebene Stellen unter den ersten $N_Z (= i_k)$ Stellen hat, und W ein Zylinder mit l vorgeschriebenen Stellen, dann ist

$$\nu(Z) = 2^{-k},$$
$$\nu(W) = 2^{-l},$$

und für $n > N_Z$ ist $W \cap \hat{T}^{-n}(Z)$ ein Zylinder mit $k + l$ vorgeschriebenen Stellen, also

$$\nu\left(W \cap \hat{T}^{-n}(Z)\right) = 2^{-(k+l)}.$$

Somit gilt natürlich

$$\lim_{n\to\infty} \nu\left(W \cap \hat{T}^{-n}(Z)\right) = \nu(W)\nu(Z).$$

□

7.2. **Der Ergodensatz von von Neumann.** (kein Tippfehler, da steht wirklich zweimal „von“!)

Etwa zeitgleich mit Birkhoff hat von Neumann einen Ergodensatz bewiesen (1932-33). Die Beweise von Birkhoff und von von Neumann waren unabhängig; es ist aber leicht, einen Satz aus dem anderen zu folgern. Das tun wir hier.

Corollary 7.6. ***von Neumanns Ergodensatz:*** *Sei (T, μ) ein maßerhaltendes System, μ ein W-Maß. Sei $p \in [0, \infty)$ und sei $f \in L^p(\mu)$. Dann existiert für fast alle $x \in X$ die Funktion*

$$f^*(x) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i(x)).$$

Die Funktion f^* ist in $L^p(\mu)$. Die Funktion f^* ist T-invariant, d.h. $f^*(x) = f^*(T(x))$ für fast alle $x \in X$, und es gilt $\int_X f d\mu = \int_X f^* d\mu$.

Weiterhin gilt: Das Mittel konvergiert in L^p, d.h.

$$\left\| \frac{1}{n} \sum_{i=0}^{n-1} f(T^i(x)) - f^*(x) \right\|_p \to 0.$$

Die letzte Behauptung (Konvergenz in L^p) ist die wichtigste Behauptung, und es ist die Stelle, an der sich dieser Ergodensatz von Birkhoffs unterscheidet.

Beweis. Zunächst einmal beweisen wir die Behauptungen für Funktionen $g \in L^\infty(\mu)$. Das ist zwar nicht ausreichend (denn L^∞ impliziert zwar L^p, jedenfalls bei W-Maß, aber nicht umgekehrt), aber wir verfahren so wie oft in der Analysis: Zuerst ersetzen wir alle Größen (hier die Funktion) durch solche, mit denen wir besser umgehen können, und dann zeigen wir, dass die entstandenen Fehlerterme harmlos sind.

Nach dem Birkhoff'schen Ergodensatz existiert für eine L^∞-Funktion g der Mittelwert g^* fast überall. Wenn g in L^∞ ist, also $|g| \le K$ bis auf einer Nullmenge, dann gilt dies auch für $g \circ T$, es gilt auch für

$$\frac{1}{n} \sum_{i=0}^{n-1} g \circ T^i,$$

und es gilt auch für

$$g^* = \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} g \circ T^i,$$

mit derselben Konstante K. Also ist die Differenz

$$\frac{1}{n}\sum_{i=0}^{n-1} g \circ T^i - g^*$$

eine beschränkte Funktionenfolge (mit Schranke $2K$), die wegen Birkhoffs Ergodensatz punktweise fast überall gegen 0 konvergiert, also

$$\frac{1}{n}\sum_{i=0}^{n-1} g(T^i(x)) - g^*(x) \to 0$$

und somit

$$\left|\frac{1}{n}\sum_{i=0}^{n-1} g(T^i(x)) - g^*(x)\right|^p \to 0.$$

Nach dem Satz über majorisierte Konvergenz gilt dann: Die punktweise (f.ü.) Konvergenz von

$$a_n := \left|\frac{1}{n}\sum_{i=0}^{n-1} g(T^i(x)) - g^*(x)\right|^p \to 0.$$

ist auch bezüglich der L^1-Norm, d.h.

$$\int_X a_n d\mu \to 0$$

(und wir dürfen Grenzübergang und Integration vertauschen). Also gilt

$$\left\|\frac{1}{n}\sum_{i=0}^{n-1} g(T^i(x)) - g^*(x)\right\|_p \to 0.$$

Insbesondere ist

$$S_n g = \frac{1}{n}\sum_{i=0}^{n-1} g \circ T^i$$

eine Cauchy-Folge in L^p, d.h. für alle $\varepsilon > 0$ gibt es $N(\varepsilon, g)$, so dass für alle $n > N(\varepsilon, g)$ und alle $k \in \mathbb{N}$ gilt

$$\|S_n g - S_{n+k} g\|_p < \varepsilon.$$

Sei jezt f die ursprünglich betrachtete L^p-Funktion. Es gilt wegen der Maßerhaltung, dass

$$\|f\|_p = \|f \circ T\|_p$$

und deshalb

$$\|S_n f\|_p \leq \frac{1}{n}\sum_{i=0}^{n-1} \left\|f \circ T^i\right\|_p = \|f\|_p\,.$$

Wir wollen zeigen, dass $S_n f$ eine Cauchyfolge in L^p ist. Für $\varepsilon > 0$ wählen wir ein $g \in L^\infty$ mit $\|f-g\|_p < \varepsilon$. Dann gilt

$$\|S_n f - S_n g\|_p = \|f-g\|_p\,,$$

und deshalb

$$\begin{aligned}&\|S_n f - S_{n+k} f\|_p\\ \leq&\|S_n f - S_n g\|_p + \|S_n g - S_{n+k} g\|_p + \|S_{n+k} g - S_{n+k} f\|_p\\ <&3\varepsilon.\end{aligned}$$

Also ist $S_n f$ Cauchy und somit konvergent in L^p gegen einen Grenzwert f^*.

f^* ist invariant, da gilt

$$\frac{n+1}{n} S_{n+1} f - S_n(f \circ T) = \frac{1}{n} f$$

und die linke Seite konvergiert gegen $f^* - f^* \circ T$, die rechte gegen 0. □

Diese Sätze sind ein wichtiger Teil der Ergodentheorie. Wie der Name vermuten läßt, sind sie auch wichtig für das Verständnis von ergodischen Abbildungen. Dazu noch eine (weitere) Charakterisierung von ergodischen Abbildungen:

Theorem 7.7. *Sei $T : (X,\mu) \to (X,\mu)$ eine maßerhaltende Abbildung. Dann sind äquivalent:*

(1) *T ist ergodisch.*
(2) *Für alle A, B aus der σ-Algebra von μ gilt:*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \mu\left(A \cap T^{-i}(B)\right) - \mu(A)\,\mu(B) = 0.$$

Beweis. $(1) \Rightarrow (2)$: Sei T ergodisch. Sei $f = \chi_B$, was eine L^1-Funktion ist. Nach Birkhoff gilt dann

$$\frac{1}{n} \sum_{i=0}^{n-1} \chi_B(T^i x) \to \int_X \chi_B d\mu = \int_X \chi_B(y) d\mu(y) = \mu(B).$$

Multiplikation mit χ_A gibt

$$\frac{1}{n} \sum_{i=0}^{n-1} \chi_B(T^i x)\chi_A(x) \to \int_X \chi_B(y)\chi_A(x) d\mu(y) = \mu(B)\chi_A(x).$$

Der Term

$$\frac{1}{n}\sum_{i=0}^{n-1}\chi_B(T^ix)\chi_A(x) = \chi_A(x)\frac{1}{n}\sum_{i=0}^{n-1}\chi_B(T^ix)$$

konvergiert nach Birkhoff punktweise gegen eine L^1-Funktion. Die können wir integrieren und (wieder nach Birkhoff) Integral und Limes vertauschen. Also:

$$\frac{1}{n}\sum_{i=0}^{n-1}\mu(A\cap T^{-i}B) \to \mu(B)\mu(A).$$

$(2) \Rightarrow (1)$: Sei umgekehrt die Gleichung in (2) erfüllt. Sei E eine invariante Menge. Setze $A = E = B$ in (2). Dann ist die linke Seite gleich

$$\frac{1}{n}\sum_{i=0}^{n-1}\mu(E\cap T^{-i}E) = \frac{1}{n}\sum_{i=0}^{n-1}\mu(E) = \mu(E)$$

und die rechte Seite gleich $\mu(E)^2$. Die Aussage $\mu(E) \to \mu(E)^2$ impliziert $\mu(E) \in \{0,1\}$. □

8. Beweis des Birkhoff'schen Ergodensatzes

8.1. **Der fundamentale Ergodensatz.** Zunächst ein harmlos aussehender Satz, aus dem aber direkt die wichtigsten Ergodensätze folgen. Dieser **fundamentale Ergodensatz** besagt etwas über die relative Häufigkeit, mit welcher ein Orbit einer Abbildung T in eine bestimmte Menge trifft.

Theorem 8.1. ***Fundamentaler Ergodensatz:*** *Sei (X,μ) ein Maßraum mit endlichem Gesamtmaß, $T : X \to X$ 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*

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

(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

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

existiert.

Beweis. Definiere

$$\overline{A}(x) := \limsup_{n\to\infty} A_n(x),$$
$$\underline{A}(x) := \liminf_{n\to\infty} A_n(x).$$

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

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

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

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.

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

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

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

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

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

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

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

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:

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

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

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

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

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

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

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

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

und darauf

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

Dann schließen wir: Es gilt analog zu vorhin

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

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

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

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

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

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 $S'_n(x) \geq (n - M)(\overline{A}(x) - \varepsilon)$. Daraus schließen wir wieder

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

und daraus folgt

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

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 $\varepsilon \to 0$ 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

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

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

8.2. **Folgerung von Birkhoffs Ergodensatz aus dem fundamentalen Ergodensatz.** Der fundamentale Ergodensatz besagt etwas über die relative Häufigkeit, mit welcher ein Orbit einer Abbildung T in eine bestimmte Menge trifft. Der Ergodensatz von Birkhoff übersetzt den fundamentalen Ergodensatz in die Sprache von reellen Funktionen auf X. Hier in passender Notation:

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

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

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

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 vorhin, also existiert $A_{f_\varepsilon}(x)$ für fast alle x. Wegen der Abschätzung $|f(x) - f_\varepsilon(x)| < \varepsilon$ gilt dann

$$\overline{A_f}(x) := \limsup_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i(x)) \leq \overline{A_f}(x) + \varepsilon$$

und

$$\underline{A_f}(x) := \liminf_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i(x)) \geq \overline{A_f}(x) - \varepsilon,$$

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.

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:

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

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

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).

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

Als nächstes untersuchen wir *schwach mischende* und *mischende* Abbildungen.

9. Schwaches Mischen: Äquivalente Bedinugungen

Zur Erinnerung: T heißt **schwach mischend**, wenn für alle Mengen A, B gilt:

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \left|\mu(A \cap T^{-i}B) - \mu(A)\mu(B)\right| = 0.$$

Theorem 9.1. ***(Alternative Charakterisierungen von schwachem Mischen.)*** *Sei $T : (X, \mathcal{A}, \mu) \to (X, \mathcal{A}, \mu)$ eine maßerhaltende Abbildung. Die folgenden Aussagen sind äquivalent:*

(1) *T erfüllt: für alle Mengen $A, B \in \mathcal{A}$ gilt:*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \left|\mu(A \cap T^{-i}B) - \mu(A)\mu(B)\right| = 0.$$

(2) *Für alle $A, B \in \mathcal{A}$ gibt es eine Menge $J \subset \mathbb{N}$ der Dichte 0 (Definition folgt), so dass gilt*

$$\lim_{\substack{n\to\infty \\ n\notin J}} \mu(A \cap T^{-i}B) = \mu(A)\mu(B).$$

(3) *T erfüllt: für alle Mengen $A, B \in \mathcal{A}$ gilt:*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \left|\mu(A \cap T^{-i}B) - \mu(A)\mu(B)\right|^2 = 0.$$

Hierbei kann der Exponent statt 2 auch eine beliebige Zahl $p \in (0, \infty)$ sein.

Definition 9.2. Für eine Menge $I \subset \mathbb{N}$ definiere

$$\alpha(I, n) := \#(I \cap [0, n])$$

und die Dichte ρ von I bis n als

$$\rho(I, n) = \frac{1}{n}\alpha(I, n).$$

Eine Menge $J \subset \mathbb{N}$ hat **Dichte 0**, wenn gilt

$$\lim_{n\to\infty} \rho(J, n) = 0.$$

Mittels $a_i := \mu(A \cap T^{-i}B) - \mu(A)\mu(B)$ (was immer im Intervall $[-1, 1]$ liegt) reduziert sich der Beweis auf folgenden Satz:

Theorem 9.3. *Sei $(a_i)_{i\in\mathbb{N}}$ eine beschränkte Folge von reellen Zahlen. Dann sind äquivalent:*

(1)

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} |a_i| = 0.$$

(2) *Es gibt eine Menge $J \subset \mathbb{N}$ der Dichte 0 (Definition folgt), so dass gilt*

$$\lim_{\substack{n\to\infty \\ n\notin J}} a_n = 0.$$

(3)

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} |a_i|^2 = 0.$$

Hierbei kann der Exponent statt 2 auch eine beliebige Zahl $p \in (0, \infty)$ *sein.*

Beweis. $(1) \Rightarrow (2)$: Sei $J_k := \{n \mid a_n > 1/k\}$. Dann ist jedes J_k eine Menge der Dichte 0, denn

$$0 \overset{n\to\infty}{\longleftarrow} \frac{1}{n} \sum_{i=0}^{n-1} |a_i| \geq \frac{1}{n} \sum_{i \in J_k \cap [0,n)} \frac{1}{k} = \frac{1}{k} \rho(J_k, n-1)$$

und für k fest, $n \to \infty$ muss somit gelten $\rho(J_k, n) \to 0$.

Finde eine Folge $0 = l_0 < l_1 < l_2 < \dots$ in $\mathbb{N}$, so dass für alle $n \geq l_k$ gilt, dass $\rho(J_k, n) < \frac{1}{k}$.

Setze

$$J := \bigcup_{k \in \mathbb{N}} \left(J_k \cap [l_k, l_{k+1})\right).$$

Wir werden zeigen, dass J Dichte 0 hat.

Zuerst stellen wir fest, dass die J_k geschachtelt sind:

$$J_1 \subset J_2 \subset J_3 \subset \dots .$$

Deswegen dürfen wir in der folgenden Teilmengenrelation und Ungleichung das k vergrößern.

Für $n \in [l_k, l_{k+1})$ gilt $J \cap [0, n] \subset J_{k+1} \cap [0, n]$, und daher gilt [1]

$$\rho(J, n) \leq \rho(J_{k+1}, n) < \frac{1}{k+1}.$$

Da k beliebig ist, gilt somit

$$\lim_{n\to\infty} \rho(J, n) = 0.$$

Nun zeigen wir, dass a_n auf $\mathbb{N} \setminus J$ gegen 0 konvergiert: Für $n \in [l_k, l_{k+1})$ ist entweder $n \in J$ oder es gilt $a_n < \frac{1}{k}$. Somit gilt für alle $n \in [l_k, \infty)$,dass $a_n < \frac{1}{k}$ ist. Somit

$$\lim a_n = 0.$$

$(2) \Rightarrow (1)$: Es gilt allgemein: Wenn a_n gegen 0 konvergiert, dann auch die Mittelwerte

$$\frac{1}{n} \sum_{i=0}^{n-1} a_i.$$

[1] Schauer'sche Vereinfachung

Denn alle Werte mit genügend großem n ($n > n_0$) sind $< \varepsilon$, und im Mittelwert haben n_0 Werte beliebig wenig Einfluss für $n \to \infty$ und die verbleibenden sind $< \varepsilon$, somit der Mittelwert irgendwann $< 2\varepsilon$. Also konvergiert auch so ein Mittelwert gegen 0.

Jetzt ist

$$\begin{aligned}\frac{1}{n}\sum_{i=0}^{n-1} a_i = &\frac{1}{n}\sum_{i\in[0,n)\cap J} a_i + \frac{1}{n}\sum_{i\in[0,n)\setminus J} a_i\\ \leq &\frac{1}{n}\sum_{i\in[0,n)\cap J} K + \frac{1}{n}\sum_{i\in[0,n)\setminus J} a_i\\ = &\rho(J, n-1)\cdot K + \frac{1}{n}\sum_{i\in[0,n)\setminus J} a_i\\ \overset{n\to\infty}{\longrightarrow} &0,\end{aligned}$$

da beide Summanden gegen 0 konvergieren.

$(2) \Leftrightarrow (3)$: Die Bedingung $a_n \to 0$ in Aussage (2) ändert sich nicht, wenn wir a_n durch $b_n = |a_n|^p$ ersetzen. Einsetzen von b_n statt a_n in Aussage (1) liefert Aussage (3) . □

Die Menge $J = J(A, B)$ in obigem Beweis kann im Allgemeinen von A, B abhängen. Folgender Satz sagt, dass wir in wichtigen Fällen auch ein J finden können, das für alle A, B funktioniert, also unabhängig von diesen ist.

Definition 9.4. Eine σ-Algebra $\mathcal{A}$ hat eine **abzählbare Basis** $(B_i)_{i\in\mathbb{N}}$ (mit $B_i \in \mathcal{A}$), wenn für alle $A \in \mathcal{A}$ und alle $\varepsilon > 0$ ein $i \in \mathbb{N}$ existiert mit $\mu\left(A \triangle B_i\right) < \varepsilon$.

Theorem 9.5. *Sei $T : (X, \mathcal{A}, \mu) \to (X, \mathcal{A}, \mu)$ eine maßerhaltende Abbildung, und die σ-Algebra habe zusätzlich eine abzählbare Basis. Dann sind folgende Bedingungen äquivalent (und äquivalent zu schwachem Mischen):*

(1) *Für alle $A, B \in \mathcal{A}$ gibt es eine Menge $J = J(A, B) \subset \mathbb{N}$ der Dichte 0, so dass gilt*

$$\lim_{\substack{n\to\infty\\ n\notin J}} \mu(A \cap T^{-n}B) = \mu(A)\mu(B).$$

(2) *Es gibt eine Menge $J \subset \mathbb{N}$ der Dichte 0, so dass für alle $A, B \in \mathcal{A}$ gilt*

$$\lim_{\substack{n\to\infty\\ n\notin J}} \mu(A \cap T^{-n}B) = \mu(A)\mu(B).$$

Beweis. $(2) \Rightarrow (1)$ ist klar. $(1) \Rightarrow (2)$: Für je zwei Basiselemente $A = B_k$, $B = B_l$ gibt es ein $J = J_{kl}$ der Dichte 0 mit

$$\lim_{\substack{n\to\infty \\ n\notin J_{kl}}} \mu(B_k \cap T^{-n}B_l) = \mu(B_k)\mu(B_l).$$

Für festes k, l ist $\mu(B_k \cap T^{-n}B_l) - \mu(A_k)\mu(B_l) \in [-1, 1]$ als Funktion von n eine beschränkte Folge. Wegen dem schwachen Mischen gilt

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} \left|\mu(B_k \cap T^{-n}B_l) - \mu(A_k)\mu(B_l)\right| = 0.$$

Sei

$$a_n = \sum_{k,l\in\mathbb{N}} \frac{1}{2^{k+l}} \left|\mu(B_k \cap T^{-n}B_l) - \mu(A_k)\mu(B_l)\right|.$$

Dann konvergiert

$$\frac{1}{n}\sum_{i=0}^{n-1} a_n = \frac{1}{n}\sum_{i=0}^{n-1} |a_n|$$

gegen 0, denn für jedes $\varepsilon > 0$ ist für genügend großes N_0 die Summe

$$\sum_{k+l\in\mathbb{N},\ k+l\geq N_0} \frac{1}{2^{k+l}}$$

kleiner als ε, und die ersten N_0 (somit endlich vielen) Folgen $a_0, \ldots, a_{N_0-1}$konvergieren gleichmäßig gegen 0, d.h. sind ab einem $n_0 \in \mathbb{N}$ auch $< \varepsilon$.

Wir wissen schon, dass die Bedingung

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} |a_n| = 0$$

äquivalent dazu ist, dass es eine Menge $J \subset \mathbb{N}$ der Dichte 0 gibt, so dass gilt

$$\lim_{\substack{n\to\infty \\ n\notin J}} a_n = 0.$$

Deswegen muss für jedes k, l gelten:

$$\lim_{\substack{n\to\infty \\ n\notin J}} \mu(B_k \cap T^{-n}B_l) - \mu(A_k)\mu(B_l) = 0.$$

D.h. wir haben ein J gefunden, das für alle Basiselemente funktioniert.

Um die Behauptung für beliebige $A, B \in \mathcal{A}$ zu zeigen, finden wir Basiselemente B_k, B_l mit $\mu(A \triangle B_k) < \varepsilon$, $\mu(B \triangle B_l) < \varepsilon$ und stellen fest, dass alle Behauptungen in diesem Beweis bis auf einen Fehlerterm $< 4\varepsilon$ gelten. $\square$

In der Praxis ist es lästig, die Eigenschaften von Ergodizität, schwachem Mischen und starkem Mischen für *alle* A, B der σ-Algebra $\mathcal{A}$ überprüfen zu müssen. Besser ist es, dies nur für eine bestimmte, leicht handhabbare Untermenge tun zu müssen. Dies leistet folgender Satz. Zuvor noch ein paar Definitionen.

Definition 9.6. Eine Menge L von Teilmengen von X heißt **Algebra**, wenn gilt:

- $\emptyset \in L$,
- $A, B \in L$ impliziert $A \cap B \in L$,
- $A \in L$ impliziert $X \setminus A \in L$.

Das ist natürlich weniger als eine σ-Algebra. Es geht noch besser:

Definition 9.7. Eine L von Teilmengen von X heißt **Semi-Algebra**, wenn gilt:

- $\emptyset \in L$,
- $A, B \in L$ impliziert $A \cap B \in L$,
- $A \in L$ impliziert, dass $X \setminus A$ endliche Vereinigung von Elementen in L ist.

Beispielsweise ist die Menge der Zylinder eine Semi-Algebra in der σ-Algebra von allen Mengen in $\{0,1\}^{\mathbb{N}}$.

Definition 9.8. Die Semi-Algebra L **erzeugt** die σ-Algebra $\mathcal{A}$ **bezüglich dem Maß** μ, wenn es für jedes $A \in \mathcal{A}$ und jedes $\varepsilon > 0$ ein $B \in L$ gibt mit

$$\mu\,(A \triangle B) < \varepsilon.$$

Beispiel wie vorhin: Die Menge der Zylinder erzeugt alle Mengen in $\{0,1\}^{\mathbb{N}}$ bezüglich Bernoulli-Maß.

Theorem 9.9. *Sei $T : (X, \mathcal{A}, \mu) \to (X, \mathcal{A}, \mu)$ eine maßerhaltende Abbildung. Sei L eine Semi-Algebra, welche $\mathcal{A}$ erzeugt. Dann genügt es, die Eigenschaften für*

- *Ergodizität:*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \big(\mu(A \cap T^{-i}B) - \mu(A)\mu(B)\big) = 0,$$

- *schwachem Mischen:*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \big|\mu(A \cap T^{-i}B) - \mu(A)\mu(B)\big| = 0$$

- *und starkem Mischen:*
$$\lim_{n\to\infty} \mu(A \cap T^{-n}B) - \mu(A)\mu(B) = 0$$

für alle A, B in L zu überprüfen.

Der Beweis geht durch einfache Approximation.

10. Wie „typisch" ist schwaches Mischen und starkes Mischen?

In diesem Abschnitt werden wir verstehen, warum eine maßerhaltende Abbildung „typischerweise" (Erklärung folgt) schwach mischend und gleichzeitig nicht stark mischend ist.

Definition 10.1. Ein topologischer Raum heißt **Baire-Raum**, wenn eine der folgenden äquivalenten Eigenschaften zutrifft:

- Jeder abzählbare Durchschnitt von offenen dichten Mengen ist dicht.
- Jede abzählbare Vereinigung von nirgends dichten Mengen (das sind solche, deren Abschluss ein leeres Inneres hat) ist nirgends dicht.

In einem Baire-Raum heißt eine Menge **dünn** oder **mager** oder **erste Baire-Kategorie**, wenn sie abzählbare Vereinigung von nirgends dichten Mengen ist. Ansonsten heißt sie **dick** oder **fett** oder **zweite Baire-Kategorie**.

Remark 10.2. Die Baire'schen Kategorien haben in Wirklichkeit nichts mit Kategorientheorie zu tun.

Definition 10.3. Sei $(X, \mathcal{A}, \mu)$ ein W-Raum. Die **schwache Topologie** auf der Menge $\mathcal{M}$ aller maßerhaltenden Abbildugen $T : (X, \mathcal{A}, \mu) \to (X, \mathcal{A}, \mu)$ ist definiert wie folgt: Eine Folge $(T_i)_{i\in\mathbb{N}}$ von Abbildungen $T_i \in \mathcal{M}$ ist konvergent und hat einen Limes $T \in \mathcal{M}$, also
$$\lim_{i\to\infty} T_i = T,$$
genau dann, wenn für alle $E \in \mathcal{A}$ gilt, dass
$$\lim_{i\to\infty} \mu\left(T_i\left(E\right) \triangle T\left(E\right)\right) = 0.$$

Theorem 10.4. *Sei $(X, \mathcal{A}, \mu)$ ein W-Raum. In der Menge aller maßerhaltenden Abbildungen auf $(X, \mathcal{A}, \mu)$ ist die Menge der schwach mischenden Abbildungen eine dicke Menge (zweite Baire-Kategorie), und die Menge der stark mischenden Abbiludngen ist eine dünne Menge (erste Baire-Kategorie).*

10.1. **Eine Abbildung, die ergodisch und nicht schwach mischend.** Sei $(X,\mu) = ([0,1),\lambda)$ das Einheitsintervall mit Lebesgue-Maß und $R_\alpha : ([0,1),\lambda) \to ([0,1),\lambda)$ die Kreisrotation

$$R_\alpha(x) = x + \alpha \pmod 1.$$

(Die Bezeichnungen „Kreis", „Rotation" erschließen sich besser in der multiplikativen Darstellung.) R_α ist für $\alpha \notin \mathbb{Q}$ ergodisch. Wir zeigen jetzt, dass R_α nicht schwach mischend ist. Für $\alpha \in \mathbb{Q}$ ist das ohnehin klar, also betrachten wir $\alpha \notin \mathbb{Q}$. Sei $A = [0,\varepsilon)$ und $B = \left[\frac{1}{2}, \frac{1}{2} + \varepsilon\right)$. Dann ist für jedes $n \in \mathbb{N}$ entweder $A \cap T^{-n}B = \emptyset$ oder $A \cap T^{-n}A = \emptyset$. Also gilt mindestens eine der folgenden beiden Zeilen:

$$\frac{1}{n}\sum_{i=0}^{n-1} \left|\mu\left(A \cap T^{-n}B\right) - \mu(A)\mu(B)\right| \geq \frac{1}{2}\mu(A)\mu(B),$$

$$\frac{1}{n}\sum_{i=0}^{n-1} \left|\mu\left(A \cap T^{-n}A\right) - \mu(A)\mu(A)\right| \geq \frac{1}{2}\mu(A)\mu(A).$$

Also kann in keiner dieser beiden Zeilen die linke Seite für $n \to \infty$ gegen 0 konvergieren, was für schwaches Mischen erforderlich wäre.

10.2. **Konstruktion von schwach mischenden Abbildungen.** Wir kennen schon Abbildungen, die ergodisch sind und nicht schwach mischend. Es gibt Konstruktionen für Abbildungen, die schwach mischend sind und nicht stark mischend (z.B. von Katok und Stepin), und es gibt solche sogar auf jeder Mannigfaltigkeit mit Kreisoperation (Gunesch und Katok). Diese Konstruktionen sind aber nicht allzu explizit, denn sie haben viele frei wählbare Parameter, und daher braucht es viel „Platz zum Schreiben", um Abbildungen mit gewünschten Eigenschaften zu erzeugen. Der Vorteil dabei ist allerdings ebenfalls, dass die Konstruktionen viele frei wählbare Parameter hat und somit viele verschiedene gewünschte Eigenschaften eingebaut werden können.

11. L^2-Charakterisierung von Ergodizität, schwachem und starkem Mischen

Zur Erinnerung: Auf $L^2(\mu)$ (als Raum von reellen Funktionen) ist ein reelles Skalarprodukt definiert durch

$$\langle f, g\rangle := \int_X f \cdot g d\mu = \int_X f(x)g(x)d\mu(x).$$

Dadurch erhalten wir auch eine Norm $\|\| = \|\|_2 = \|\|_{L^2(\mu)}$. Für komplexe Funktionen würden wir $\int_X f \cdot \bar{g} d\mu$ nehmen. Im Folgenden beschreiben wir nur den reellen Fall; der komplexe ergibt sich durch offensichtliche Modifkation.

Theorem 11.1. *Sei $T : (X, \mu) \to (X, \mu)$ eine maßerhaltende Abbildung. Dann gelten folgende Äquivalenzen.*

(1) *Es sind äquivalent:*
 (a) *(T, μ) ist ergodisch.*
 (b) *Für alle $f, g \in L^2(\mu)$ gilt:*
$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \left\langle U_T^i f, g \right\rangle = \langle f, 1\rangle \cdot \langle 1, g\rangle .$$
 (c) *Für alle $f \in L^2(\mu)$ gilt:*
$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \left\langle U_T^i f, f \right\rangle = \langle f, 1\rangle^2 .$$
(2) *Es sind äquivalent:*
 (a) *(T, μ) ist schwach mischend.*
 (b) *Für alle $f, g \in L^2(\mu)$ gilt:*
$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \left| \left\langle U_T^i f, g \right\rangle - \langle f, 1\rangle \cdot \langle 1, g\rangle \right| = 0.$$
 (c) *Für alle $f \in L^2(\mu)$ gilt:*
$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \left| \left\langle U_T^i f, f \right\rangle - \langle f, 1\rangle^2 \right| .$$
 (d) *Für alle $f, g \in L^2(\mu)$ gilt:*
$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \left| \left\langle U_T^i f, g \right\rangle - \langle f, 1\rangle \cdot \langle 1, g\rangle \right|^2 = 0.$$
 Ebenso mit dem Exponenten 2 ersetzt durch beliebiges $p \in (0, \infty)$.
(3) *Es sind äquivalent:*
 (a) *(T, μ) ist mischend.*
 (b) *Für alle $f, g \in L^2(\mu)$ gilt:*
$$\lim_{n\to\infty} \langle U_T^n f, g\rangle = \langle f, 1\rangle \cdot \langle 1, g\rangle .$$
 (c) *Für alle $f \in L^2(\mu)$ gilt:*
$$\lim_{n\to\infty} \langle U_T^n f, f\rangle = \langle f, 1\rangle^2 .$$

Beweis. Wir beweisen (3). Die anderen Aussagen folgen durch Hinzufügen von Summenzeichen und Mittelwerten; bei (2)(d) wird die schon bekannte Beschreibung von schwach mischenden Abbildungen in Satz 9.1 benutzt.

(b)⇒(a): Einsetzen von $f = \chi_A$, $g = \chi_B$ liefert

$$\lim_{n\to\infty} \mu\left(T^{-n}(A) \cap B\right) = \mu(A)\,\mu(B).$$

(a)⇒(c): Die Aussage (a), also

$$\lim_{n\to\infty} \mu\left(T^{-n}(A) \cap B\right) = \mu(A)\,\mu(B),$$

zeigt die Gleichung

$$\lim_{n\to\infty} \langle U_T^n f, g\rangle = \langle f, 1\rangle \cdot \langle 1, g\rangle \qquad (*)$$

für charakteristische Funktionen f, g.

Durch endliche Summenbildung und Ausnutzung der Bilinearität von $\langle , \rangle$ folgt, dass die Gleichung (*) für *einfache* Funktionen g gilt. (Zur Erinnerung: Eine Funktion heißt *einfach*, wenn sie nur endlich viele Werte annimmt.) Summenbildung angewendet auf f zeigt, dass Gleichung (*) immer noch gilt, wenn f auch eine einfache Funktion ist. Insbesondere haben wir

$$\lim_{n\to\infty} \langle U_T^n f, f\rangle = \langle f, 1\rangle^2$$

für einfache Funktionen f gezeigt.

Sei nun $f \in L^2(\mu)$ beliebig. Wähle eine einfache Funktion h mit $\|f - h\| < \varepsilon$. Dann gilt

$$\begin{aligned} |\langle U_T^n f, f\rangle - \langle f, 1\rangle \cdot \langle 1, f\rangle| \leq\ & |\langle U_T^n f, f\rangle - \langle U_T^n f, h\rangle| \\ & + |\langle U_T^n f, h\rangle - \langle U_T^n h, h\rangle| \\ & + |\langle U_T^n h, h\rangle - \langle h, 1\rangle \cdot \langle 1, h\rangle| \\ & + |\langle h, 1\rangle \cdot \langle 1, h\rangle - \langle f, 1\rangle \cdot \langle 1, h\rangle| \\ & + |\langle f, 1\rangle \cdot \langle 1, h\rangle - \langle f, 1\rangle \cdot \langle 1, f\rangle|. \end{aligned}$$

Der mittlere Summand ist beliebig klein, da $\lim_{n\to\infty} \langle U_T^n h, h\rangle = \langle h, 1\rangle^2$ für einfache Funktionen h schon gilt. Alle anderen Summanden sind $\leq \|f - h\| \cdot$ const., wobei const$\leq 2\|f\|$ ist. Da $\|f - h\| < \varepsilon$ ist, folgt die Behauptung.

(c)⇒(b): Zuerst überlegen wir uns, dass es einen eindeutigen linearen Unterraum

$$\textbf{Inv}(f) \subset L^2(\mu)$$

gibt, der f enthält, alle konstanten Funktionen enthält, der abgeschlossen ist in L^2 un der die Eigenschaft hat, dass

$$U_T\left(\mathbf{Inv}\left(f\right)\right) \subset \mathbf{Inv}(f)$$

ist: Sei R die konvexen Hülle aller $U_T^n f$ und aller Konstanten. Dann ist der Abschluss $\bar{R}$ von R ein wohldefinierter Unterraum. Es gilt $U_T R \subset R$, denn für $r \in R$ gilt $r = \lim_{k\in\mathbb{N}} r_k$ mit

$$r_k = \frac{1}{k+1}\left(a_0 + \sum_{i=1}^{k} a_i \cdot U_T^{b_i} f\right),$$

wobei a_i reelle Konstanten und $b_i \in \mathbb{N}$, und es folgt

$$U_T r = \lim_{k\in\mathbb{N}} U_T r_k = \frac{1}{k+1}\left(a_0 + \sum_{i=1}^{k} a_i \cdot U_T^{b_i+1} f\right),$$

also $U_T r \in R$. Weiterhin gilt $U_T \bar{R} \subset \bar{R}$, denn für $r \in \bar{R}$ gilt $r = \lim_n r_n$ mit $r_n \in R$ und deswegen $U_T r = \lim_n U_T r_n$. Deswegen hat der Raum Raum $\mathbf{Inv}(f) := \bar{R}$ die gewünschte Invarianzeigenschaft.

Weiterhin definieren wir den Raum $\mathbf{Misch}(f)$ der Funktionen g, die mit f korrekt mischen:

$$\mathbf{Misch}(f) := \left\{g \in L^2\left(\mu\right) : \lim_{n\to\infty} \langle U_T f, g\rangle - \langle f, 1\rangle \cdot \langle 1, g\rangle = 0\right\}.$$

Wir wollen zeigen, dass $\mathbf{Misch}(f)$ gleich dem ganzen Raum $L^2\left(\mu\right)$ ist.

Zunächst zeigen wir $\mathrm{Inv}(f) \subset \mathbf{Misch}(f)$: Es gilt $U_T^n f \in \mathbf{Misch}(f)$, und sicherlich liegen die Konstanten in $\mathbf{Misch}(f)$. Da die Gleichung in der Definition von $\mathbf{Misch}(f)$ linear ist, liegt auch jede endliche konvexe Kombination von $\{U_T^n f\}_{n\in\mathbb{N}} \cup \{\text{Konstanten}\}$ in $\mathbf{Misch}(f)$. Da die Gleichung in der Definition von $\mathbf{Misch}(f)$ stetig in g ist, liegt auch jede unendliche konvexe Kombination von $\{U_T^n f\}_{n\in\mathbb{N}} \cup \{\text{Konstanten}\}$ in $\mathbf{Misch}(f)$, also die Menge R. Aus demselben Stetigkeitsgrund liegt auch $\bar{R}$ in $\mathbf{Misch}(f)$.

Weiterhin zeigen wir, dass $\left(\mathbf{Inv}(f)\right)^{\perp}$ (das orthogonale Komplement im Hilbertraum $L^2\left(\mu\right)$) auch in $\mathbf{Misch}(f)$ liegt: Wenn $g \in \left(\mathbf{Inv}(f)\right)^{\perp}$, dann gilt $\langle 1, g\rangle = 0$ und

$$\langle U_T^n f, g\rangle = 0$$

für alle $n \in \mathbb{N}$. Also ist

$$\langle U_T f, g\rangle - \langle f, 1\rangle \cdot \langle 1, g\rangle = 0.$$

Somit ist die Behauptung $\mathbf{Misch}(f) = L^2\left(\mu\right)$ gezeigt. $\square$

Nun interessieren wir uns für Ergodizität, schwaches und starkes Mischen von Produktabbildungen. Zur Erinnerung: Das Produktmaß $\mu \times v$ ist definiert durch $(\mu \times v)(A \times B) = \mu(A)\, v(B)$ auf Produktmengen $A \times B$, und auf der Produkt-σ-Algebra, die durch $\mathcal{A} \times \mathcal{B}$ erzeugt wird, ist das Produktmaß definiert durch Approximation einer beliebigen Menge durch abzählbare Vereinigung von Produktmengen.

Remark 11.2. Das heißt (und dies ist sehr interessant und praktisch), dass wir Eigenschaften wie Ergodizität, schwaches und starkes Mischen beweisen können, indem wir sie nur für Mengen beweisen, welche Produktmengen sind.

Theorem 11.3. *Es sind äquivalent:*

(1) (T, μ) *ist stark mischend.*
(2) $(T \times T, \mu \times \mu)$ *ist stark mischend.*

Beweis. Wie in der Bemerkung eben gesagt, reicht es, Mengen der Produktform $A \times A'$, $B \times B'$ zu prüfen.

(1)⇒(2): Es gilt

$$\begin{aligned}
&(\mu \times \mu)\left((A \times A') \cap (T \times T)^{-n}(B \times B')\right)\\
&\quad= \mu\left(A \cap T^{-1}(B)\right) \mu\left(A' \cap T^{-n}(B')\right)\\
&\quad\to \mu(A)\, \mu(B)\, \mu(A')\, \mu(B')\\
&\quad= (\mu \times \mu)(A \times A') \cdot (\mu \times \mu)(B \times B') .
\end{aligned}$$

(2)⇒(1):

$$\begin{aligned}
&\mu(A \times T^{-n}(B))\\
&\quad= (\mu \times \mu)\left((A \times X) \cap (T \times T)^{-n}(B \times X)\right)\\
&\quad\to (\mu \times \mu)(A \times X) \cdot (\mu \times \mu)(B \times X)\\
&\quad= \mu(A)\, \mu(B).
\end{aligned}$$

□

Theorem 11.4. *Die folgenden Aussagen sind äquivalent:*

(1) (T, μ) *ist schwach mischend.*
(2) $(T \times T, \mu \times \mu)$ *ist schwach mischend.*
(3) $(T \times T, \mu \times \mu)$ *ist ergodisch.*

Theorem 11.5. *Weiterhin gilt: Wenn* $(T \times T, \mu \times \mu)$ *ergodisch ist, dann ist auch* (T, μ) *ergodisch.*

Remark 11.6. Die Umkehrung des letzten Satzes gilt nicht: Daraus, dass (T, μ) ergodisch ist, folgt im Allgemeinen nicht, dass $(T \times T, \mu \times \mu)$ ergodisch ist.

Wir beweisen die beiden vorigen Sätze auf einmal.

Beweis. $(2) \Rightarrow (3)$ ist klar.

$(1) \Rightarrow (2)$: Für alle Mengen $A, B \in \mathcal{A}$ gibt es $J(A, B) \subset \mathbb{N}$ der Dichte 0 mit

$$\lim_{\substack{n\to\infty \\ n \notin J(A,B)}} \mu\left(A \cap T^{-n}(B)\right) = \mu(A)\,\mu(B)$$

und analog $J(A', B')$ für A', B'. Sei

$$\bar{J} = J(A \times A', B \times B') := J(A, B) \cup J(A', B').$$

Dann hat $\bar{J}$ immer noch Dichte 0, und es gilt

$$\begin{aligned}
&\lim_{\substack{n\to\infty \\ n \notin \bar{J}}} (\mu \times \mu)\left((A \times A') \cap T^{-n}(B \times B')\right) \\
&\quad = \lim_{\substack{n\to\infty \\ n \notin \bar{J}}} \mu(A \cap T^{-n}(B))\mu\left(A' \cap T^{-n}(B')\right) \\
&\quad = \mu(A)\mu(B)\mu(A')\,\mu(B') \\
&\quad = (\mu \times \mu)(A \times A') \cdot (\mu \times \mu)(B \times B').
\end{aligned}$$

$(3) \Rightarrow (1)$: Ergodizität von $(T \times T, \mu \times \mu)$ heißt, dass die Mittelwerte der Abweichungen (ohne Absolutbeträge) gegen 0 konvergieren, d.h.

$$\begin{aligned}
&\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} (\mu \times \mu)\left((A \times A') \cap (T \times T)^{-n}(B \times B')\right) \\
&\quad = (\mu \times \mu)(A \times A') \cdot (\mu \times \mu)(B \times B').
\end{aligned}$$

Einsetzen von $A' = X = B'$ liefert

$$\begin{aligned}
&\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \mu(A \cap T^{-n}(B)) \\
&= \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} (\mu \times \mu)\left((A \times X) \cap (T \times T)^{-n}(B \times X)\right) \\
&= (\mu \times \mu)(A \times X) \cdot (\mu \times \mu)(B \times X) \\
&= \mu(A)\,\mu(B).
\end{aligned}$$

Womit nebenbei Satz 11.5 bewiesen ist.

Einsetzen von $A = A'$, $B = B'$ liefert

$$\begin{aligned}
&\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}\mu(A\cap T^{-n}(B))^2\\
&=\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}(\mu\times\mu)\left((A\times A)\cap(T\times T)^{-n}(B\times B)\right)\\
&=(\mu\times\mu)(A\times A)\cdot(\mu\times\mu)(B\times B)\\
&=\mu(A)^2\,\mu(B)^2\,.
\end{aligned}$$

Nun beweisen wir schwaches Mischen:

$$\begin{aligned}
&\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}(\mu(A\cap T^{-n}(B))-\mu(A)\mu(B))^2\\
&=\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}((\mu(A\cap T^{-n}(B)))^2\\
&\qquad-2(\mu(A\cap T^{-n}(B))\mu(A)\mu(B)\\
&\qquad+\mu(A)^2\mu(B)^2)\\
&=\mu(A)^2\mu(B)^2-2\mu(A)^2\mu(B)^2+\mu(A)^2\mu(B)^2\\
&=0.
\end{aligned}$$

$\square$