Transfer-Operatoren für dynamische Systeme
Vorlesung im WS 2008/09

von Gerhard Keller

Inhaltsverzeichnis

1 Markov-Operatoren und Frobenius-Perron Operatoren

Dieser Abschnitt stützt sich auf Kapitel 3 in [LM85].

1.1 Markov-Operatoren auf L^1-Räumen

Definition 1.1. *Sei* $(X, \mathcal{A}, \mu)$ *ein Maßraum.* $P\colon L^1_\mu \to L^1_\mu$ *ist ein* Markov-Operator*, falls für jedes* $0 \leqslant f \in L^1_\mu$ *gilt:* $Pf \geqslant 0$ *und* $\|Pf\| = \|f\|$*. Hier ist* $\|f\| := \mu(|f|) := \int |f| d\mu$*.*

Notierung 1.2. *Von jetzt an sei* P *in diesem Kapitel immer ein Markov-Operator,* L^1 *bezeichne* L^1_μ*, und* $D_\mu := \{f \in L^1_\mu\colon f \geqslant 0, \|f\| = 1\}$ *sei die Menge der Wahrscheinlichkeitsdichten bzgl.* μ*.*

Proposition 1.3. *Für $f, g \in L^1$ gilt:*

a) $f \leqslant g \Rightarrow Pf \leqslant Pg$

b) $(Pf)^+ \leqslant Pf^+$, $(Pf)^- \leqslant Pf^-$

c) $|Pf| \leqslant P|f|$

d) $\|Pf\| \leqslant \|f\|$

Bemerkung 1.4. Für die n-te Iterierte von P folgt daraus:

$$\|P^n f\| = \|P(P^{n-1}f)\| \leqslant \|P^{n-1}f\| \leqslant \ldots \leqslant \|f\|.$$

Sind f und g zwei Wahrscheinlichkeitsdichten, so folgt sofort:

$$\|P^n f - P^n g\| = \|P^n(f-g)\| \leqslant \|f-g\|.$$

Notierung 1.5. *Der* Träger *von f,* supp $f := \{x\colon f(x) \neq 0\}$*, ist bis auf μ-Nullmengen eindeutig bestimmt.*

Proposition 1.6. *Sei $f \in L^1$. Äquivalent sind:*

i. $\|Pf\| = \|f\|$

ii. $(Pf)^+ = Pf^+$

iii. $(Pf)^- = Pf^-$

iv. Pf^+ *und* Pf^- *haben disjunkte Träger.*

Definition 1.7. $f \in L^1$ *ist ein* Fixpunkt *von P, falls $Pf = f$. Ist f eine Wahrscheinlichkeitsdichte, so spricht man auch von einer* invarianten Dichte.

Proposition 1.8. *Ist $Pf = f$, so sind auch $Pf^+ = f^+$ und $Pf^- = f^-$.*

1.2 Frobenius-Perron Operatoren

In diesem Abschnitt konzentrieren wir uns auf eine wichtige Beispielklasse von Markov-Operatoren, nämlich auf die sogenannten Frobenius[1.1]-Perron[1.2] Operatoren. Dazu sei $S\colon X \to X$ eine messbare (bzgl. $\mathcal{A}$), nicht singuläre (bzgl. μ) Abbildung. *Nicht singulär* bedeutet, dass $\mu(S^{-1}A) = 0$ für alle $A \in \mathcal{A}$ mit $\mu(A) = 0$.

Ist $0 \leqslant f \in L^1_\mu$, so ist $(f\mu) \circ S^{-1} \ll \mu$, denn aus $\mu(A) = 0$ folgt $(f\mu)(S^{-1}A) = \int_{S^{-1}A} f\, d\mu = 0$. Also existiert (nach dem Satz von Radon-Nikodym) der *Frobenius-Perron Operator* von S

$$Pf := \frac{d((f\mu) \circ S^{-1})}{d\mu}.$$

Für beliebige $f \in L^1$ setzt man $Pf := Pf^+ - Pf^-$.

Proposition 1.9. *Sei P der F-P Operator zu S.*

a) $f \in L^1 \Rightarrow Pf \in L^1$.

b) $P\colon L^1 \to L^1$ *ist linear.*

c) *P ist positive, d.h.* $f \geqslant 0 \Rightarrow Pf \geqslant 0$.

d) *Für $f \in L^1$ und $g \in L^\infty$ gilt: $\int Pf \cdot g\, d\mu = \int (f \circ S) \cdot g\, d\mu$. Insbesondere ist für $A \in \mathcal{A}$ (setze $g = 1_A$): $\int_A Pf\, d\mu = \int_{S^{-1}A} f\, d\mu$. Durch die Gültigkeit jeder dieser beiden Gleichungen wird Pf als Element von L^1_μ eindeutig bestimmt.*

e) *P^n ist der F-P Operator zu S^n, der n-fachen Komposition von S mit sich selbst.*

1.1. Ferdinand Georg Frobenius, 1849-1917, lehrte Mathematik in Berlin und Zürich.

1.2. Oskar Perron, 1880-1975, lehrte Mathematik in Heidelberg und München.

In vielen Fällen ist eine explizitere Darstellung von P möglich. Dazu nehmen wir jetzt an, dass $X \subseteq \mathbb{R}^d$, dass $X_i \subseteq X$ $(i \in I)$ offen und paarweise disjunkt sind mit endlicher oder abzählbarer Indexmenge I, und dass $\mu(X \setminus \bigcup_{i\in I} X_i) = 0$, wo μ das Lebesgue-Maß auf X ist. $S|_{X_i}: X_i \to S(X_i)$ sei differenzierbar und invertierbar für jedes $i \in I$, und es sei $JS(x) := \det(DS(x)) \neq 0$ für jedes $i \in I$ und jedes $x \in X_i$. In dieser Situation gilt:

Proposition 1.10. *Für jedes $f \in L^1_\mu$ und μ-fast jedes $x \in X$ ist*

$$Pf(x) = \sum_{i\in I} 1_{S(X_i)}(x) \cdot \left(\frac{f}{|JS|}\right)(S^{-1}_{X_i} x) \tag{1.1}$$

Bemerkung 1.11. Ist $h \in D_\mu$ ein Fixpunkt von P, so ist $h\mu$ ein S-invariantes Wahrscheinlichkeitsmaß, denn für $A \in \mathcal{A}$ ist $\int_{S^{-1}A} h\, d\mu = \int_A Ph\, d\mu = \int_A h\, d\mu$.

.

Beispiel 1.12. [Die Gauss-Abbildung, auch Kettenbruchtransfsormation] Sei $X = [0,1]$, μ das Lebesgue-Maß auf $[0,1]$, $S: [0,1] \to [0,1]$, $S(x) = \frac{1}{x} \bmod 1$, also $X_i = (\frac{1}{i+1}, \frac{1}{i}]$ für $i \in I := \mathbb{N}$. Es ist

$$Pf(x) = \sum_{i=1}^{\infty} f(\frac{1}{x+i}) \cdot \frac{1}{(x+i)^2}$$

Die Wahrscheinlichkeitsdichte $h(x) = \frac{1}{\log 2} \cdot \frac{1}{x+1}$ ist invariant unter P.

Anmerkung 1.13. Die Gauss-Abbildung ist eng mit Kettenbruchentwicklungen verknüpft. Hier ist die Grundidee: Seien $a_1, a_2, \ldots$ natürliche Zahlen. Bezeichne

$$[a_1, \ldots, a_n] = \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{\ddots + \cfrac{1}{a_n}}}}$$

Man kann zeigen, dass $[a_1, a_2, \ldots] := \lim_{n\to\infty} [a_1, \ldots, a_n]$ immer existiert, dass jedes $x \in (0, 1]$ genau eine solche Kettenbruchentwicklung $x = [a_1, a_2, \ldots]$ hat, und dass gilt:

- $x \in (0,1] \cap \mathbb{Q}$ genau dann, wenn $\exists n \in \mathbb{N}\, \exists a_1, \ldots, a_n \in \mathbb{N}$ so dass $x = [a_1, \ldots, a_n]$.
- $x = [a_1, a_2, \ldots]$ ist eine quadratische Irrationalzahl in $(0, 1]$ genau dann, wenn die Koeffizeintenfolge schließlich periodisch ist.

Dazu ein Beispiel: $x = [1, 1, 1, \ldots]$ erfüllt offensichtlich die Gleichung $x = \frac{1}{1+x}$. Also ist $x = \frac{\sqrt{5}-1}{2}$. Allgemeiner gilt immer $x = \frac{1}{S(x) + a_1}$, wobei a_1 der ganzzahlige Anteil von $\frac{1}{x}$ ist, also $a_1 = \psi(x) := \frac{1}{x} - S(x)$. Definiert man allgemein $a_n = \psi(S^{n-1}x)$, so sieht man leicht, dass $x = [a_1, a_2, \ldots]$. Die Koeffizientenfolge wird also durch den Orbit von x unter der Gauss-Abbildung bestimmt. In der Tat gilt für jedes $k \in \mathbb{N}$ und Lebesgue-fast jedes $x \in (0,1]$:

$$\lim_{n\to\infty} \frac{1}{n} \sharp\{j \in \{1, \ldots, n\}: a_j = k\} = \frac{1}{\log 2} \int_{\frac{1}{k+1}}^{\frac{1}{k}} \frac{1}{x+1}\, dx = \frac{\log(1 + \frac{1}{k(k+2)})}{\log 2}$$

Das folgt aus dem Birkhoffschen Ergodensatz angewandt auf die $h\mu$-invariante Abbildung S, sobald man gezeigt hat, dass S für $h\mu$ ergodisch ist.

1.3 Koopmann-Operatoren

Seien $(X, \mathcal{A}, \mu)$ und $S: X \to X$ wie bisher. Der *Koopmann-Operator* $U: L^\infty_\mu \to L^\infty_\mu$ zu S ist definiert durch $Uf(x) = f(S(x))$. Er hat folgende Eigenschaften:

a) U ist linear

b) $\|Uf\|_\infty \leqslant \|f\|_\infty$ für alle $f \in L^\infty_\mu$

c) $\langle Pf, g\rangle = \langle f, Ug\rangle$ für alle $f \in L^1_\mu$, $g \in L^\infty_\mu$, wo $\langle f, g\rangle = \int fg\, d\mu$

Definition 1.14. *Seien f_n, $f \in L^1$. (f_n) konvergiert schwach gegen f (in Zeichen: $f_n \xrightarrow{w} f$), falls $\lim_{n\to\infty} \langle f_n, g\rangle = \langle f, g\rangle$ für alle $g \in L^\infty$.*

Proposition 1.15. *Aus $f_n \xrightarrow{w} f$ folgt $Pf_n \xrightarrow{w} Pf$.*

Proposition 1.16. *Für $f \in L^1_\mu$ und $g \in L^\infty_\mu$ ist $P(f \cdot Ug) = g \cdot Pf$.*

Beweis. Für jedes $A \in \mathcal{A}$ ist

$$\int_A P(f \cdot Ug)\, d\mu = \int_{S^{-1}A} f \cdot Ug d\mu = \int_X (1_A\, g) \circ S \cdot f d\mu = \int_A g \cdot Pf\, d\mu$$

□

2 Ergodizität

In diesem Abschnitt präsentiere ich einige Auszüge aus Kapitel 4 in [LM85]. Zum vertieften Verständnis und für weitere Beispiele sollten Sie die Abschnitte 4.1 - 4.4 dieses Kapitels lesen. Im ganzen Abschnitt sei wieder $(X, \mathcal{A}, \mu)$ ein Maßraum und $S: X \to X$ messbar.

2.1 Invariante Maße, maßerhaltende Transformationen

Definition 2.1. *Die Abbildung S heißt* maßerhaltend, *falls $\mu(S^{-1}(A)) = \mu(A)$ für alle $A \in \mathcal{A}$. Das Maß μ heißt dann* invariant *unter S. Kurz: $\mu \circ S^{-1} = \mu$.*

Proposition 2.2. *Sei S nicht-singulär, P der FP-Operator zu S und $h \in D_\mu$. Dann ist $h\,\mu$ invariant unter S genau dann, wenn $Ph = h$.*

In Beispiel 1.12 hatten wir bereits die Gauss-Transformation mit ihrer invarianten Dichte $\frac{1}{\log 2} \cdot \frac{1}{x+1}$ kennen gelernt. Hier sind ein paar weitere Beispiele:

Beispiel 2.3. [r − adische Transformation] S: $[0, 1] \to [0, 1]$, $S(x) = rx \bmod 1$ ($r \geqslant 2$ ganzzahlig). Ist μ das Lebesgue-Maß, so ist $P1 = 1$, d.h. das Lebesgue-Maß ist invariant. (Aus Gleichung (1.1) folgt sofort, dass $P1 = 1$.)

Beispiel 2.4. [Quadratische Abbildungen] S: $[0, 1] \to [0, 1]$, $S(x) = ax\,(1 - x)$ mit $0 < a \leqslant 4$, und μ wieder das Lebesgue-Maß auf [0, 1]. Im Fall $a = 4$ (*volle quadratische Abbildung*, da S dann surjektiv) ist $h(x) = \frac{1}{\pi\sqrt{x\,(1-x)}}$ eine invariante Wahrscheinlickeitsdichte. Das rechnet man wieder unter Benutzung von Gleichung (1.1) leicht nach. Für andere Parameter a sieht die Situation ganz anders aus.

Bemerkung 2.5. Die Für die r-adische Abbildung ist $|S'| = r > 1$, und für die Gauss-Abbildung ist zwar nur $|S'(x)| = \frac{1}{x^2} \geqslant 1$, aber schon für die zweite Iterierte rechnet man leicht nach, dass $|(S^2)'| \geqslant 2 > 1$. Deshalb nennt man beide Abbildungen *stückweise expandierend*. Für quadratische Abbildungen ist das nicht der Fall, denn $S'(\frac{1}{2}) = 0$, und aus der Kettenregel folgt für jedes n, dass $(S^n)'(x) = 0$ sobald $S^j(x) = \frac{1}{2}$ für ein $j \in \{0, \ldots, n-1\}$.

Beispiel 2.6. [Arnold's cat map] Jetzt sei $X = \mathbb{T}^2 := \mathbb{R}^2/\mathbb{Z}^2$ der zweidimensionale Torus und S: $X \to X$ der folgende hyperbolische Torus-Automorphismus: $S(x, y) = (x + y, x + 2\,y) \bmod 1$. Die Abbildung S wird durch die Matrix $A = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$ beschrieben. Da die durch A gegebene lineare Abbildung des $\mathbb{R}^2$ das Lebesgue-Maß erhält ($\det(A) = 1$) und $\mathbb{Z}^2$ bijektiv auf sich abbildet (auch $A^{-1} = \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}$ ist ganzzahlig), ist $S: X \to X$ bijektiv und erhält das Lebesgue-Maß auf X. Wir bemerken noch, dass A die Eigenwerte $\lambda^+ = \frac{3}{2} + \frac{\sqrt{5}}{2} > 1$ und $0 < \lambda^- = \frac{3}{2} - \frac{\sqrt{5}}{2} < 1$ hat, so dass S je eine, durch die zugehörigen Eigenvektoren v^+ und v^- von A gegebene, expandierende und kontrahierende Richtung hat. Damit gilt $|DS(v^+)| = \lambda^+ > 1$ und $0 < |DS(v^-)| = \lambda^- < 1$. Man sagt, S ist *gleichmäßig hyperbolisch*.

Diese und einige weitere Beispiele findet man in Abschnitt 4.1 von [LM85].

Im weiteren werden wir uns hauptsächlich auf (stückweise) expandierende Abbildungen beschränken, da ihre Frobenius-Perron Operatoren am besten zugänglich sind. Aber mit mehr Aufwand können auch viele quadratische Abbildungen und viele hyperbolische Abbildungen mit den gleichen Ideen behandelt werden.

2.2 Ergodische Transformationen

Definition 2.7. *S heißt* ergodisch*, falls für jedes $A \in \mathcal{A}$ gilt: $S^{-1}(A) = A \bmod \mu$ impliziert $\mu(A) = 0$ oder $= 1$.*

Proposition 2.8. *Sei S eine nicht-singuläre Transformation von $(X, \mathcal{A}, \mu)$ mit FP-Operator P.*

a) Ist S ergodisch, so hat P höchstens einen Fixpunkt $h = Ph \in D_\mu$.

b) Hat P genau einen Fixpunkt $h \in D_\mu$ und ist $h(x) > 0$ μ-f.s., so ist S ergodisch.

c) Hat P genau einen Fixpunkt $h \in D_\mu$, so ist $h\mu$ das einzige zu μ absolut stetige invariante W'maß von S. Es ist ergodisch.

Beweis. Hier ist eine Skizze für a) und b). Details findet man in Theorem 4.2.2. aus [LM85]. a) Habe P verschiedene Fixpunkte $h_1, h_2 \in D_\mu$. Dann ist $g = h_1 - h_2$ ein Fixpunkt von P und damit auch g^+ und g^-. Da $\int g\,d\mu = 0$, ist $\int g^+\,d\mu = \int g^-\,d\mu > 0$ und g^+ und g^- haben disjunkte Träger. Aus diesen Trägern konstruiert man zwei disjunkte invariante Mengen positiven Maßes – ein Widerspruch zur Ergodizität von S.
b) Wäre S nicht ergodisch, so gäbe es ein $A = S^{-1}A$ mit $0 < \mu(A) < 1$. Sei $g = (\int_A h\,d\mu)^{-1} \cdot 1_A \cdot h$. Dann ist $Pg = \mu(A)^{-1} P(1_A \circ S \cdot h) = \mu(A)^{-1} 1_A Ph = g$ ein weiterer Fixpunkt in D_μ.
c) Da jedes zu μ absolut stetige S-invariante W'maß einen Fixpunkt aus D_μ von P als Dichte hat, ist $h\mu$ das einzige solche Maß. Ist $S^{-1}A = A$, und wäre $0 < \int_A h\,d\mu < 1$, so wäre g wie in b) mit dem gleichen Argument ein weiterer Fixpunkt von P. □

Bemerkung 2.9. Sei S ergodisch und $Ph = h \in D_\mu$. Dann folgt aus em Birkhoffschen Ergodensatz für alle $f \in L^1_{h\mu}$ (und damit insbesondere für alle $f \in L^\infty_\mu$):

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(S^k(x)) = \int_X f h\,d\mu \quad \text{für } h\mu\text{-fast alle } x \in X.$$

(Es gilt nicht automatisch für μ-fast alle $x \in X$!)

3 Die Existenz invarianter Dichten

3.1 Die Grundidee

Details zu diesem Abschnitt findet man in den Abschnitten 5.1 und 5.2 von [BG97]. Sei wieder ganz allgemein P ein Markov-Operator auf L^1_μ. für $f \in L^1_\mu$ und $n \in \mathbb{N}$ bezeichne

$$A_n f := \frac{1}{n} \sum_{k=0}^{n-1} P^k f.$$

Beachte, dass $\|A_n f\| \leqslant \|f\|$ für alle n, also $\|A_n f\| \leqslant 1$ falls $f \in D_\mu$. Wir benötigen folgende Begriffsbildung aus der Funktionalanalysis:

Definition 3.1. *Eine Folge $(f_n)_{n\in\mathbb{N}}$ ist* schwach präkompakt*, falls jede Teilfolge eine schwach konvergente Teilfolge hat. (Für schwache Konvergenz siehe Definition 1.14.)*

Proposition 3.2. *Ist P: $L^1_\mu \to L^1_\mu$ ein Markov-Operator, $f \in D_\mu$, und ist die Folge $(A_n f)_{n\in\mathbb{N}}$ schwach präkompakt, so hat P mindestens einen Fixpunkt in D_μ.*

Beweis. Sei h schwacher Limes der Teilfolge $(A_{n_k} f)$. Da P stetig ist ($\|P\phi\| \leqslant \|\phi\|$ für alle $\phi \in L^1_\mu$!), gilt für jedes $g \in L^\infty_\mu$ nach Definition 1.14 und Proposition 1.15 zur schwachen Konverge

$$\langle Ph, g\rangle = \lim_{k\to\infty} \left\langle \frac{1}{n_k} P A_{n_k} f, g \right\rangle = \lim_{k\to\infty} \frac{1}{n_k} \langle A_{n_k} f - f + P^{n_k} f, g\rangle \, d\mu = \langle h, g\rangle ,$$

also $Ph = h$. □

Anmerkung 3.3. Eine Folge $(f_n)_{n\in\mathbb{N}}$ ist schwach präkompakt genau dann, wenn sie gleichgradig integrierbar ist. (Das steht z.B. in Theorem IV.8.9 bei Dunford & Schwatz [DS57]

Beispiel 3.4. [Zur Gauss-Abbildung] Sei $x = [a_1, a_2, \ldots]$, $[a_1, \ldots, a_n] = \frac{p_n}{q_n}$. Ohne Beweis halten wir fest: $|x - \frac{p_n}{q_n}| < \frac{1}{q_n\, q_{n+1}}$ und $q_n > (a_n + 1)\, q_{n-2}$, siehe das Skript [Sau05]. Iterativ folgt:

$$q_{n+1}\, q_n > C_x \prod_{k=1}^{n+1} (a_k + 1) .$$

Daher ist, wenn wir auch die Ergodizität der Kettenbruchtransformation S „glauben",

$$\liminf_{n\to\infty} \frac{1}{n} \log(q_n q_{n+1}) \geqslant \liminf_{n\to\infty} \frac{1}{n+1} \sum_{k=1}^{n+1} \log(1 + \psi(S^k x)) = \frac{1}{\log 2} \int_0^1 \frac{\log(1 + \psi(x))}{1 + x} \, dx$$

mit $\psi(x) = \frac{1}{x} - S(x)$.

3.2 Die Beobachtung von Lasota und Yorke

Eine Variante des Hauptergebnisses dieses Abschnitts, Satz 3.9, findet man in Abschnitt 5.2 in [BG97].

Während schwache Präkompaktheit von Folgen $(A_n f)$ nur relative schwache Ergebnisse liefert (z.B. die Existenz von Fixpunkten von P in D_μ, aber keine interessanten Aussagen über die Menge aller Fixpunkte wie z.B. die Eindeutigkeit eines Fixpunktes), liefert starke Präkompaktheit (das ist Präkompaktheit in der von der Norm erzeugten Metrik) viel stärkere Aussagen. Eine Art von Durchbruch in diese Richtung war die Arbeit von Lasota und Yorke aus dem Jahr 1973 [LY73]. In ihr wurden zum ersten Mal *Funktionen von beschränkter Variation* zur Untersuchung ergodentheoretischer Eigenschaften dynamischer Systeme eingesetzt. Wir stellen hier die wichtigsten Eigenschaftgen dieser Funktionen zusammen. Eine Reihe von Details findet man in Abschnitt 2.3 von [BG97].

Definition 3.5.

a) *Für eine Funktion $f: \mathbb{R} \to \mathbb{R}$ ist die* Variation *von f definiert als*

$$V(f) := \sup \left\{ \sum_{i=1}^{n} |f(a_i) - f(a_{i-1})| \colon a_0 < a_1 < \ldots < a_n \right\}$$

b) *Für ein Intervall $X \subset \mathbb{R}$ und $f: X \to \mathbb{R}$ definieren wir*

$$V_X(f) := V(f \cdot 1_X)$$

c) *Sei X ein kompaktes Intervall, m das Lebesgue-Maß auf X. Für $f \in L^1_m(X)$ definieren wir*

$$V_X(f) := \inf \left\{ V_X(\tilde{f}) \colon \tilde{f} \textit{ Repräsentant von } f \right\}$$

d) $\mathrm{BV}(X) := \{f \in L^1_m(X) \colon V_X(f) < \infty\}$.

Bemerkung 3.6.

a) Für alle $f\colon X\to\mathbb{R}$ gilt $\|f\|_1\leqslant m(X)\,\|f\|_\infty\leqslant\frac{1}{2}m(X)\,V_X(f)$ (auch wenn einer oder mehrerer dieser Werte $=\infty$ ist).

b) Ist $f\colon X\to\mathbb{R}$ differenzierbar, $X=[a,b]$, so ist $V_X(f)=|f(a)|+|f(b)|+\int_a^b|f'|\,dm.$ und $f'\in L^1_m(X)$,

Die für uns wichtigste Eigenschaft der Variation ist

Proposition 3.7. [Satz von Helly] *Ist $(f_n)_{n\in\mathbb{N}}$ eine Folge in $\mathrm{BV}(X)$ mit $\sup_n V_X(f_n)<\infty$, so ist $(f_n)_{n\in\mathbb{N}}$ stark präkompakt in $L^1_m(X)$. Für jeden Hüfungspunkt f der Folge gilt:*

$$V_X(f)\leqslant\limsup_{n\to\infty}V_X(f_n)$$

(Einen Beweis findet man in Kapitel VIII,§4 von [Nat75]. Dort wird tatsächlich eine etwas stärkere Aussage bewiesen.)

Lasota und Yorke untersuchten nun Frobenius-Perron Operatoren zu stückweise expandierenden Abbildungen eines Intervalls.

Definition 3.8. *Sei X ein kompaktes Intervall. Eine Abbildung $T\colon X\to X$ heißt* stückweise expandierend, *falls es eine endliche oder abzählbare Zerlegung $(X_i)_{i\in I}$ von X (modulo Lebesgue-Nullemngen) in offene Intervalle X_i gibt, so dass*

i. $T|_{X_i}$ stetig differenzierbar und monoton $(i\in I)$

ii. Es gibt ein $\alpha>1$ so dass $|T'(x)|\geqslant\alpha$ für alle $x\in\bigcup_{i\in I}X_i$.

iii. $V_I(g)<\infty$, wo $g\colon X\to\mathbb{R}$ folgendermaßen definiert ist:

$$g(x)=\begin{cases}1/|T'(x)| & \text{falls } x\in\bigcup_{i\in I}X_i\\ 0 & \text{sonst}\end{cases}$$

Bedingung *iii.* dieser Definition ist z.B. erfüllt, wenn es nur endlich viele X_i gibt, Bedingungen *i.* und *ii.* erfült sind und wenn $\sup\{|T''(x)|\colon x\in\bigcup_{i\in I}X_i\}<\infty$. (Das folgt aus Bemerkung 3.6.)

Satz 3.9. [Lasota und Yorke, Version von Rychlik [Ryc83]] *Sei $T\colon X\to X$ stückweise expandierend, P der Frobenius-Perron Operator zu T. Dann gibt es Konstanten $r\in(0,1)$ und $C>0$, so dass für alle $f\in\mathrm{BV}(X)$ und alle $n\in\mathbb{N}$ gilt:*

$$V_X(P^nf)\leqslant C(r^nV_X(f)+\|f\|_1)\,. \tag{3.1}$$

Insbesondere ist $\sup_n V_X(P^nf)<\infty$.

Bemerkung 3.10. Da jede präkompakte Folge auch schwach präkompakt ist, folgt aus diesem Satz zusammen mit Propositionen 3.7 und 3.2, dass P einen Fixpunkt in D_m hat.

Wir werden den Satz nur unter der etwas stärkeren Voraussetzung beweisen, dass g auf jedem Intervall X_i *Lipschitz-stetig* ist und die Lipschitzkonstanten der $g|_{X_i}$ beschränkt sind. Genauer nehmen wir an:

$$\widetilde{\mathrm{Lip}}(g):=\sup_{i\in I}\mathrm{Lip}(g|_{X_i})<\infty,\ \ \mathrm{Lip}(g|_{X_i}):=\sup_{x,y\in X_i,x\neq y}\frac{|g(y)-g(x)|}{|y-x|} \tag{3.2}$$

Der Beweis des Satzes stützt sich auf eine Reihe von Lemmas.

Lemma 3.11. *Ist $J\subset\mathbb{R}$ ein Intervall und sind $u,v\in\mathrm{BV}(J)$, so gilt*

i. $V_J(u\,v)\leqslant V_J(u)\,\|v\|_\infty+\|u\|_\infty V_J(v)$

ii. $V_J(u\,v)\leqslant V_J(u)\,\|v\|_\infty+\int_J|u|\,dm\cdot\mathrm{Lip}(v|_J)$.

Beweis. (Ich gebe den Beweis hier an, da der Zusatzterm, der in der Vorlesung in Abschätzung *ii.* aufgetaucht ist, doch nicht nötig ist.) Sei $\delta > 0$, $x_0 < x_1 < \ldots < x_n$ und o.B.d.A $|x_i - x_{i-1}| < \delta$ für alle $i \in I$ sowie $x_0, x_n \notin J$ aber $x_1, \ldots, x_{n-1} \in J$. Dann ist

$$\begin{aligned}
&\sum_{i=1}^{n} |(uv1_J)(x_i) - (uv1_J)(x_{i-1})| \\
\leqslant\ & |(uv)(x_1)| + |(uv)(x_{n-1})| + \sum_{i=2}^{n-1} |u(x_i) - u(x_{i-1})|\, |v(x_i)| + \sum_{i=2}^{n-1} |u(x_i)|\, |v(x_i) - v(x_{i-1})| \\
\leqslant\ & \|v|_J\|_\infty \sum_{i=1}^{n} |(u1_J)(x_i) - (u1_J)(x_{i-1})| + \sum_{i=2}^{n-1} |u(x_i)|\, |v(x_i) - v(x_{i-1})| \\
\leqslant\ & V_J(u)\, \|v\|_\infty + \sum_{i=2}^{n-1} |u(x_i)|\, |v(x_i) - v(x_{i-1})|
\end{aligned}$$

Daraus folgt sofort Aussage *i.* Um Aussage *ii.* zu zeigen, fahren wir so fort:

$$\begin{aligned}
&\sum_{i=2}^{n-1} |u(x_i)|\, |v(x_i) - v(x_{i-1})| \\
\leqslant\ & \mathrm{Lip}(v|_J) \left(\sum_{i=2}^{n-1} \min_{(x_{i-1}, x_i)} |u| \cdot |x_i - x_{i-1}| + \sum_{i=2}^{n-1} \mathrm{osc}_{(x_{i-1}, x_i)}(u) \cdot \delta \right) \\
\leqslant\ & \mathrm{Lip}(v|_J) \left(\int_X |u|\, dm + \delta V(u|_X) \right)
\end{aligned}$$

Aussage *ii.* folgt nun, da $\delta > 0$ beliebig gewählt werden konnte. □

Lemma 3.12. *Unter den Voraussetzungen von Satz 3.9 und Annahme (3.2) gilt: Für jedes $\delta > 0$ gibt es ein $B_\delta > 0$, so dass*

$$V_X(Pf) \leqslant \left(\frac{2}{\alpha} + \delta \right) V_X(f) + B_\delta\, \|f\|_1$$

wobei $\frac{1}{\alpha} = \|g\|_\infty < 1$. Ist I endlich, so kann $\delta = 0$ und $B_0 = \mathrm{Lip}(g) + \frac{2}{\alpha} \max_{i \in I} m(X_i)^{-1}$ gewählt werden.

Beweis. Der Beweis dieses Lemmas kann nun wie in der Vorlesung erfolgen. Insbesondere bleibt der Faktor $\frac{2}{\alpha}$ unverändert. □

Ist $\alpha > 2$, so folgt aus diesem Lemma bereits Satz 3.9 mit $r = \frac{2}{\alpha}$ und $C = \frac{B}{1 - 2/\alpha}$. (Beachte, dass $\|P^k f\|_1 \leqslant \|f\|_1$.) Andernfalls, d.h. wenn nur $\alpha > 1$ gilt, gibt es ein $k > 0$ derart, dass $\alpha^k > 2$, und man kann leicht zeigen, dass T^k stückweise expandierend ist mit $|(T^k)'(x)| \geqslant \alpha^k > 2$ an allen Stellen, an denen diese Ableitung wohldefiniert ist. Dann kann man Lemma 3.12 auf P^k anstatt P anwenden und erhält mit einer leichten Zusatzüberlegung ebenfalls die Aussage des Satzes, insbesondere also $\sup_n V_X(P^n f) < \infty$ und damit die Existenz eines Fixpunktes von P in D_m. Für die Details siehe Abschnitt 5.2 in [BG97].

Bevor wir in Kapitel ? zeigen werden, wie aus der Ungleichung (3.1), die oft als *Lasota-Yorke-Ungleichung* bezeichnet wird, viel stärkere Eigenschaften als nur die Existenz invarianter Dichten folgt, wollen wir im nächsten Kapitel eine andere Klasse von Markov-Operatoren kennenlernen.

4 Markov-Operatoren für Markov-Prozesse

In diesem Abschnitt skizzieren wir eine andere Art von Markov-Operatoren, mit denen viele Markov-Prozesse in diskreter Zeit beschrieben werden können.

Sei (M, d) ein kompakter metrischer Raum mit Borel-σ-Algebra $\mathcal{M}$. Wir betrachten eine M-wertige Markov-Kette X_0, X_1, X_2, ... mit beliebig gegebenem X_0 (oft ist $X_0 = x_0$ konstant) und Übergangskern K, der die Prozess-Verteilung beschreibt. Genauer: Sei $\mathcal{F}_n$ die von X_0, ..., X_n erzeugte σ-Algebra. Dann gelte für alle $n = 1, 2, 3, \ldots$

$$P(X_n \in A | \mathcal{F}_{n-1}) = P(X_n \in A | X_{n-1}) = K(X_{n-1}, A) \text{ für alle } A \in \mathcal{M}. \tag{4.1}$$

Wir nehmen an, dass K die *Feller*-Eigenschaft hat, d.h. dass der folgende Operator Q tatsächlich stetige auf stetige Funktionen abbildet:

$$Q: C(M) \to C(M),\ Qf(x) := \int_M f(y) K(x, dy)$$

Lemma 4.1. *Der Operator Q hat folgende Eigenschaften (für $f \in C(M)$):*

a) *Q ist linear.*

b) *Q ist positiv, d.h. $Qf \geqslant 0$ falls $f \geqslant 0$.*

c) *$Q1 = 1$*

d) *$\|Qf\|_\infty \leqslant \|f\|_\infty$*

e) *$\mathbb{E}[f(X_n) | \mathcal{F}_{n-1}] = (Qf)(X_{n-1})$*

Beweis. a) - d) sind offensichtlich. Für Indikatorfunktionen $f = 1_A$ folgt e) sofort aus (4.1) und (4.1), für allgemeine f dann durch Linearität und monotone Stetigkeit des Integrals. □

Bemerkung 4.2. Da $Q1 = 1$ ist die Existenz von Eigenfunktionen kein Problem. Interessant ist die Frage nach Fixpunkten für den dualen Operator Q^*, d.h. die Frage nach der Existenz (und evtl. Eindeutigkeit) eines invarianten Wahrscheinlichkeitsmaßes ν für Q, d.h. eines Borel-Wahrscheinlichkeitsmaßes ν auf M, für das $\int_M Qf d\nu = \int_M f d\nu$ für alle $f \in C(M)$ mit kompaktem Träger. Äquivalent dazu ist: Hat X_0 Verteilung ν, so haben auch alle X_n Verteilung ν. Genauere Kenntnisse über Q ergeben daher genauere Informationen über die stochastischen Eigenschaften der Markov-Kette $(X_n)_{n \geqslant 0}$.

Nun betrachten wir spezielle Markov-Prozesse, die oft als *iterierte Funktionensysteme* bezeichnet werden. Dazu sei eine endliche oder abzählbare Familie $(\tau_i: i \in I)$ von Kontraktionen τ_i: $M \to M$ gegeben. Die Kontraktionsfaktoren sind gerade die Lipschitzkonstanten der Abbildungen,

$$\mathrm{Lip}(\tau_i) := \sup_{x, y \in M, x \neq y} \frac{d(\tau_i(y), \tau_i(x))}{d(y, x)}$$

Dazu seien Lipschitz-stetige Funktionen $p_i: M \to [0, 1]$ gegeben, die $\sum_{i \in I} p_i(x) = 1$ für alle $x \in M$ und $\sum_{i \in I} \mathrm{Lip}(p_i) < \infty$ erfüllen. Der Markov-Prozess kann nun anschaulich so beschrieben werden: Vom Zustand X_n geht man mit Wahrscheinlichkeit $p_i(X_n)$ in den Zustand $X_{n+1} = \tau_i(X_n)$ über $(i \in I)$. Wir betrachten also von nun an den Kern

$$K(x, .) = \sum_{i \in I} p_i(x) \delta_{\tau_i(x)} \tag{4.2}$$

dessen Markov-Operator gegeben ist durch die Gleichung

$$(Qf)(x) = \sum_{i \in I} p_i(x)\, f(\tau_i(x))\,. \tag{4.3}$$

Beispiel 4.3. [Cantor-Menge] Das wohl bekannteste Beispiel für diese Situation ist folgendes:

$M = [0, 1]$, $I = \{1, 2\}$, $\tau_1(x) = \frac{x}{3}$, $\tau_2(x) = \frac{x+2}{3}$, $p_1 = p_2 \equiv \frac{1}{2}$. In diesem Fall hat der Operator Q genau ein invariantes W'maß ν – es ist die „Gleichverteilung“ auf der Standard-Cantormenge C, die Hausdorff-Dimension $\frac{\log 2}{\log 3}$ hat. (X_n) ist ein Markov-Prozess, der fast sicher gegen C konvergiert in dem Sinn, dass $\lim_{n \to \infty} \mathrm{dist}(X_n, C) = 0$ fast sicher.

Beispiel 4.4. [Sierpinski-Dreieck, Barnsley-Farn] Eine zweidimensionale Variante des vorherigen Beispiels ist das *Sierpinski-Dreieck*. Hier ist M ein gelichseitiges Dreieck mit Eckpunkten A_i, $i \in I = \{1, 2, 3\}$, die τ_i sind Kontraktionen mit Kontraktionsfaktor $\frac{1}{2}$ und Kontraktionszentrum A_i $(i \in I)$ und $p_1 = p_2 = p_3 \equiv \frac{1}{3}$. Siehe z.B. http://de.wikipedia.org/wiki/Sierpinski-Dreieck . Das Maß ν ist hier die Gleichverteilung auf einer Cantor-artigen Teilmenge S des Dreiecks, die Hausdorff-Dimension $\frac{\log 3}{\log 2}$ hat, und (X_n) konvergiert fast sicher gegen S. Ein anderes 2-dimensionales Beispiel ist *Barnsley's Farn*, siehe z.B. http://mathworld.wolfram.com/BarnsleysFern.html .

Beispiel 4.5. [Ruellescher Transfer-Operator] Sei I endlich oder abzählbar und $M = I^{\mathbb{N}_0}$ mit der Metrik $d(x, y) = 2^{-N(x,y)}$, $N(x, y) := \inf\{k \in \mathbb{N}_0: x_k \neq y_k\}$. Die $\tau_i: M \to M$ seien definiert als $\tau_i(x_0, x_1, x_2, \ldots) = (i, x_0, x_1, \ldots) =: (i, x)$. Dann ist $(Q f)(x) = \sum_{i \in I} p_i(x) f(i, x)$.

Sei nun $L(M)$ der Raum aller Lipschitz-stetigen Abbildungen von M nach $\mathbb{R}$ versehen mit der Norm $\|f\|_{\text{Lip}} := \text{Lip}(f) + \|f\|_\infty$.

Satz 4.6. [Doeblin-Fortet, Ruelle, Keane, ...] *Sei der Markov-Operator Q mit dem Kern aus (4.2) gebildet, r:=$\sup_{i \in I} \text{Lip}(\tau_i)$ und $C := \sum_{i \in I} \text{Lip}(p_i)$. Dann ist $Q(L(M)) \subseteq L(M)$, und für alle $f \in L(M)$ gilt:*

$$\|Q f\|_{\text{Lip}} \leqslant r \, \|f\|_{\text{Lip}} + C \, \|f\|_\infty \tag{4.4}$$

(Diese Ungleichung wird von einigen Autoren als Doeblin[4.1]-Fortet *Ungleichung bezeichnet.*

Ist $r < 1$, so ist man in einer Situation analog derjenigen aus dem Satz 3.9 von Lasota und Yorke, denn der Satz von Arzela und Ascoli garantiert, dass jede Folge von $f_n \in L(M)$ mit $\sup_n \|f\|_{\text{Lip}} < \infty$ relativ kompakt in $(C^0(M), \|.\|_\infty)$ ist und dass für jeden Häufungspunkt f der Folge gilt: $\|f\|_{\text{Lip}} \leqslant \sup_n \|f_n\|_{\text{Lip}}$.

5 Quasikompakte (Markov-)Operatoren

In diesem Kapitel arbeiten wir mit Banach-Räumen über $\mathbb{C}$. Da unsere üblichen Räume wie L^1_m, BV oder Lip zunächst Räume reeller Funktionen sind, komplexifizieren wir sie und die betrachteten Operatoren in der üblichen Weise, ohne das jeweils explizit zu erwähnen. Am Beispiel $\text{BV}(X)$ konkretisiert betrachten wir also den Raum $\text{BV}^{\mathbb{C}} := \{u + i v: u, v \in \text{BV}(X)\}$ mit Norm $\|u + i v\| = (V_X(u)^2 + V_X(v)^2)^{1/2}$. Mit den üblichen Rechenregeln für komplexe Zahlen ist das tatsächlich eine Norm (insbesondere gilt $\|\lambda(u + i v)\| = |\lambda| \, \|u + i v\|$ für alle $\lambda \in \mathbb{C}$). Der F-P Operator P wird kanonisch zu einem linearen Operator auf dem Banachraum $(\text{BV}^{\mathbb{C}} \|.\|)$ fortgesetzt, $P(u + i v) = P u + i P v$.

5.1 Ein paar Grundlagen über Banach-Räume

Proposition 5.1. [Theorem IV.3.5 in [DS57]]
$\mathcal{B}$ ist endlich-dimensional genau dann, wenn die Einheitskugel von $\mathcal{B}$ kompakt ist.

Proposition 5.2. [Lemma VII.4.3 in [DS57]]
Seien $U \subset V$ abgeschlossene lineare Teilräume von $\mathcal{B}$, U ein echter Unterraum von V. Dann gibt es zu jdem $\varepsilon > 0$ ein $v \in V$ mit $\|v\| = 1$ und $\inf\{\|v - u\|: u \in U\} > 1 - \varepsilon$.

Bemerkung 5.3. Man kann Proposition 5.1 auch leicht direkt aus Proposition 5.2 folgern, indem man eine Folge $U_0 \subset U_1 \subset U_2 \subset \ldots$ und Vektoren $u_n \in U_n$ mit $\inf\{\|u - u_{n+1}\|: u \in U_n\} > \frac{1}{2}$ erzeugt.

4.1. Wolfgang Doeblin, 1915-1940, Sohn des Schriftstellers Alfred Döblin, starb 1940 als französischer Soldat, hat in nur wenigen Jahren fundamentale Beiträge zur Wahrscheinlichkeitstheorie geleistet. Siehe auch http://de.wikipedia.org/wiki/Wolfgang_Döblin.

Sei nun $Q: \mathcal{B} \to \mathcal{B}$ ein beschränkter linearer Operator.

Definition 5.4. [Spektrum usw.]

a) *Die* Resolventenmenge $R(Q)$ *ist definiert als die Menge aller* $z \in \mathbb{C}$, *für die* $(zI - Q)^{-1}$ *als beschränkter linearer Operator von* $\mathcal{B}$ *nach* $\mathcal{B}$ *existiert.*

b) *Das* Spektrum *von* Q *ist* $\sigma(Q) := \mathbb{C} \setminus R(Q)$.

c) *Der* Spektralradius *von* Q *ist* $r(Q) := \sup\{|z|: z \in \sigma(Q)\}$.

Proposition 5.5. [Lemma VII.3.4 in [DS57]]

$$r(Q) = \lim_{n \to \infty} \|Q^n\|^{1/n} = \inf_{n \in \mathbb{N}} \|Q^n\|^{1/n}, \text{ insbesondere ist } r(Q^k) = (r(Q))^k \; \forall k \in \mathbb{N}.$$

Der Beweis beruht auf folgendem Lemma:

Lemma 5.6. [Lemma VII.3.2 in [DS57]] $R(Q) \subset \mathbb{C}$ *ist offen und* $(zI - Q)^{-1}: R(Q) \to \mathrm{BL}(\mathcal{B}, \mathcal{B})$ *ist analytisch.*

5.2 Quasikompaktheit

In diesem Abschnitt stützen wir uns ganz auf Kapitel XIV in [HH01]. Sei weiterhin $Q: \mathcal{B} \to \mathcal{B}$ ein beschränkter linearer Operator.

Definition 5.7. [Essentieller Spektralradius]

a) *Der* essentielle Spektralradius $r_e(Q)$ *von* Q *ist das Infimum aller* $\rho > 0$ *für die es eine Zerlegung*

$$\mathcal{B} = F_\rho \oplus H_\rho$$

in abgeschlossene Q*-invariante Teilräume gibt, so dass gilt:*

i. $\dim(F_\rho) < \infty$,

ii. $Q|_{F_\rho}$ *hat nur Eigenwerte vom Betrag* $\geqslant \rho$,

iii. $r(Q|_{H_\rho}) < \rho$.

Insbesondere ist $r_e(Q) \leqslant r(Q)$. *[Wähle* $F_\rho = \{0\}$ *und* $H_\rho = \mathcal{B}$ *für* $\rho > r(Q)$.*]*

b) *Ist* $r_e(Q) < r(Q)$, *so heißt* Q quasikompakt.

Proposition 5.8. *In der Situation der obigen Definition sei* $\{\lambda_1, \ldots, \lambda_t\}$ *die Menge der Eigenwerte von* $Q|_{F_\rho}$, *d.h. der Eigenwerte von* Q *mit Betrag* $\geqslant \rho$. *Dann gilt:*

i. $F_\rho = \bigoplus_{k=1}^t \left(\bigcup_{j=1}^\infty \ker(\lambda_k I - Q)^j \right)$

ii. $H_\rho = \left\{ f \in \mathcal{B}: \limsup_{n \to \infty} \|Q^n f\|^{1/n} < \rho \right\}$

iii. *Für alle* $k \in \mathbb{N}$ *ist* $r_e(Q^k) = (r_e(Q))^k$.

Satz 5.9. [Kriterium für Quasikompaktheit] *Sei* $|.|$ *eine weitere Norm auf* $\mathcal{B}$, $|f| \leqslant \|f\|$ *für alle* $f \in \mathcal{B}$. *(Es kann auch nur eine Semi-Norm sein.)* $Q: \mathcal{B} \to \mathcal{B}$ *habe folgende Eigenschaften:*

i. $Q\{f \in \mathcal{B}: \|f\| \leqslant 1\}$ *ist total beschränkt (d.h. relativ kompakt) in* $(\mathcal{B}, |.|)$,

ii. Q *ist auch bzgl. der* $|.|$*–Norm ein beschränkter Operator,*

iii. *Es gibt ein* $k \in \mathbb{N}$ *und reelle Konstanten* $r, R > 0$, *so dass* $r < r(Q)$ *und dass für alle* $f \in \mathcal{B}$

$$\|Q^k f\| \leqslant r^k \|f\| + R|f|.$$

Dann ist Q *quasikompakt und* $r_e(Q) \leqslant r$.

Dieser Satz geht wohl auf einen (etwas schwächeren) Satz von Ionescu Tulcea und Marinescu [TM50] aus dem Jahr 1950 zurück.

5.3 Frobenius-Perron Operatoren, die eine Lasota-Yorke-Ungleichung erfüllen

Sei P: $L_m^1 \to L_m^1$ ein F-P Operator wie in Satz 3.9, m das normierte Lebesgue-Maß auf dem Intervall X. Dann hat P eine invariante Wahrscheinlichkeitsdichte h, so dass $\lambda_1 = 1$ ein Eigenwert von P ist. Aus Satz 5.9 folgt, dass λ_1 endliche Vielfachheit hat und dass es außer $\lambda_1 = 1$ höchstens endlich viele weitere Eigenwerte vom Betrag 1 gibt. Das folgende Lemma zeigt, dass diese Eigenwerte alle Einheitswurzeln sind.

Lemma 5.10. *In der obigen Situation gilt:*

i. *Ist λ ein Eigenwert von P mit Betrag 1 (kurz: peripherer Eigenwert), so ist auch λ^j für jedes $j \in \mathbb{N}$ ein peripherer Eigenwert.*

ii. *Jeder periphere Eigenwert λ ist eine Einheitswurzel.*

iii. *Es gibt ein $k \in \mathbb{N}$ so dass $\lambda^k = 1$ für jeden peripheren Eingenwert von P.*

iv. *Ist λ ein peripherer Eigenwert und $j \in \mathbb{N}$, so ist $\ker\big((\lambda I - P)^j\big) = \ker(\lambda I - P)$.*

v. *Es gibt ein $k \in \mathbb{N}$ und ein $\rho \in (r_e(Q), r(Q))$ so dass $\mathcal{B} = \ker(I - P^k) \oplus H_\rho$.*

Beweis. *i.* Sei $P\varphi = \lambda\varphi$ mit $|\lambda| = 1$ und $\|\varphi\|_1 = 1$. Da

$$|\varphi(x)| = |P\varphi(x)| \leqslant (P|\varphi|)(x) \quad \text{und} \quad \int_X |\varphi|\, dm = \int_X P|\varphi|\, dm\,,$$

ist $|\varphi| = P|\varphi|$ m-f.s., d.h. $|\varphi|$ ist eine invariante Wahrscheinlichkeitsdichte auf X. Setze $\psi := \frac{\varphi}{|\varphi|}$. (Ist $\varphi(x) = 0$, so setzt man auch $\psi(x) = 0$.) Dann ist

$$\int_X \frac{\psi}{\lambda \cdot \psi \circ T} |\varphi|\, dm = \frac{1}{\lambda} \int_X P\bigg(\frac{\psi}{\psi \circ T} |\varphi|\bigg) dm = \frac{1}{\lambda} \int_X \frac{P\varphi}{\psi}\, dm = \int_X \frac{\varphi}{\psi}\, dm = \int_X |\varphi|\, dm \tag{5.1}$$

(Dabei wurde für die zweite Identität Proposition 1.16 verwandt.) Für alle $x \in X$ mit $\varphi(x) \neq 0$ ist $|\psi(x)| = 1$ und daher auch $\left|\frac{\psi(x)}{\lambda \cdot \psi(Tx)}\right| = 1$. Zusammen mit Gleichung (5.1) folgt daraus $\psi(x) = \lambda \cdot \psi(Tx)$ für alle x mit $\varphi(x) \neq 0$. Daher ist für jedes $j \in \mathbb{N}$:

$$P(\psi^j |\varphi|) = P(\psi^{j-1}\, \psi\, |\varphi|) = \lambda^{j-1} P(\psi^{j-1} \circ T \cdot \psi\, |\varphi|) = \lambda^{j-1}\, \psi^{j-1} P(\varphi) = \lambda^j\, \psi^{j-1}\, \varphi = \lambda^j\, \psi^j\, |\varphi|.$$

Also ist $\psi^j |\varphi|$ eine Eigenfunktion zum Eigenwert λ^j.
ii. und *iii.* folgen aus *i.*, da P nur endlich viele periphere Eigenwert hat.
iv. beweist man durch vollständige Induktion: Für $j = 1$ ist die Aussage trivial. Angenommen, sie gilt für $j - 1$. Sei dann $(\lambda I - P)^j(f) = 0$. Setze $g := (\lambda I - P)(f)$. Dann ist $g \in \ker\big((\lambda I - P)^{j-1}\big) = \ker(\lambda I - P)$, also $Pg = \lambda g$. Es folgt

$$P^n f = P^{n-1}(\lambda f - g) = \lambda P^{n-1} f - \lambda^{n-1} g = \ldots = \lambda^n f - n\, \lambda^{n-1}\, g$$

und daher $V_X(g) = V_X(\lambda^{n-1} g) \leqslant \frac{1}{n}(V_X(\lambda^n f) + V_X(P^n f)) \leqslant \frac{1}{n}((1 + C r^n)\, V_X(f) + C\, \|f\|_1) \to 0$ mit $n \to \infty$. Also ist $g = 0$, d.h. $f \in \ker(\lambda I - P)$.
v. folgt aus *iii.* und *iv.* □

Bemerkung 5.11.

a) Ist $\lambda_1 = 1$ ein einfacher Eigenwert und gibt es keine weiteren Eigenwerte vom Betrag 1, so gibt es ein $\rho \in (0, 1)$ derart, dass $\mathrm{BV}(X) = \mathbb{C}\, h \oplus H_\rho$. Dabei ist $h \geqslant 0$ die eindeutige invariante Wahrscheinlichkeitsdichte und $H_\rho = \{f \in \mathrm{BV}(X)\colon m(f) = 0\}$. In der Tat gilt für $f \in H_\rho$: $\quad m(f) = m(P^n f) = \mathcal{O}(\rho^n) \to 0$ mit $n \to \infty$, also $m(f) = 0$. Umgekehrt sei $f = \alpha h + v \in \mathrm{BV}(X)$ mit $v \in H_\rho$ und $m(f) = 0$. Dann ist $m(\alpha h + P^n v) = m(P^n f) = m(f) = 0$ für alle n, also $\alpha = \alpha m(h) = -\, m(P^n v) = \mathcal{O}(\rho^n) \to 0$, also $\alpha = 0$ und damit $f = v \in H_\rho$.

b) Wenn wir in konkreten Beispielen zeigen wollen, dass $\lambda_1 = 1$ tatsächlich ein einfacher Eigenwert ist und dass keine weiteren Eigenwerte vom Betrag 1 existieren, so reicht es wegen Lemma 5.10*iii* zu zeigen, dass $\lambda_1 = 1$ für jedes $k \in \mathbb{N}$ ein einfacher Eigenwert von P^k ist. Denn ist $Pf = \lambda f$ mit $|\lambda| = 1$, so folgt für das k aus Lemma 5.10*iii*, dass $P^k f = f$, und aus der Voraussetzung folgt, dass $f = h$ die eindeutige invariante Wahrscheinlichkeitsdichte von P ist.

c) Wegen Lemma 5.10*v* gibt es ein $\rho \in (0,1)$, so dass $\mathrm{BV}(X) = \ker(I - P^k) \oplus H_\rho$.

Bemerkung 5.12. Für ρ wie in Bemerkung 5.11 gilt zum Beispiel folgendes: Sind $u \in \mathrm{BV}(X)$, $v \in L^1_m(X)$, so fällt die Kovarianz zwischen den Zufallsvariablen u und $v \circ T^n$ auf (X, m) exponentiell in n, denn mit der Notation $u_0 := u - \mathbb{E}_m[u]$ ist $u_0 \in H_\rho$ und daher

$$\mathrm{Cov}(u, v \circ T^n) = \int_X u_0 \cdot v \circ T^n \, dm = \int_X P^n u_0 \cdot v \, dm = \mathcal{O}(\rho^n \, V_X(u_0) \|v\|_1) \tag{5.2}$$

Mit dieser Eigenschaft kann man z.B. leicht nachrechnen, dass die asymptotische Varianz folgender, wie im Zentralen Grenzwertsatz normierter, Partialsummen existiert:

$$\sigma^2_{\mathrm{asymp}}(u) := \lim_{n \to \infty} \mathrm{Varianz}\left(\frac{1}{\sqrt{n}} \sum_{k=0}^{n-1} u \circ T^k \right) = \mathrm{Varianz}(u) + 2 \sum_{k=1}^{\infty} \mathrm{Cov}(u, u \circ T^k) < \infty \, ,$$

wobei die $u \circ T^k$ als Zufallsvariablen auf dem Wahrscheinlichkeitsraum $(X, \mathcal{B}, m)$ aufzufassen sind ($\mathcal{B}$ ist die Borel-σ-Algebra auf X).

Man kann zeigen, dass $\sigma^2_{\mathrm{asymp}}(u) = 0$ genau dann, wenn es ein $\varphi \in \mathrm{BV}(X)$ gibt, für das $u = \varphi \circ T - \varphi$.

Lemma 5.13. *Ist* $\mathrm{BV}(X) = \ker(I - P^k) \oplus H_\rho$*, so gibt es eine Basis* $h_1, \ldots, h_d$ *von* $\ker(I - P^k)$ *derart, dass alle* $h_i \geqslant 0$ *und dass* $h_i \wedge h_j = 0$ *für* $i \neq j$*.*

Beweis. O.B.d.A. $k = 1$. Sei $N := \ker(I - P)$, $h_1, \ldots, h_d$ eine Basis von N. Aus Proposition 1.8 folgt, dass mit h_i auch h_i^+ und h_i^- zu N gehören. Da $h_i = h_i^+ - h_i^-$, können wir o.B.d.A. annahmen, dass alle $h_i \geqslant 0$. Sei $f_{ij} := h_i \wedge h_j$. Da $P \geqslant 0$, ist $Pf_{ij} \leqslant Ph_i = h_i$ und analog $Pf_{ij} \leqslant h_j$, also $Pf_{ij} \leqslant h_i \wedge h_j = f_{ij}$. Wegen $\int Pf_{ij} \, dm = \int f_{ij} \, dm$ folgt daraus $Pf_{ij} = f_{ij}$, also $f_{ij} \in N$. Also ist entweder $f_{ij} = 0$ oder man kann in der Basis $\{h_1, \ldots, h_d\}$ eines der h_i durch f_{ij} ersetzen. Mit der neuen Basis kann man auf gleiche Weise fortfahren, so lange es in ihr Funktionen mit nicht disjunkten Trägern gibt. Der Prozess bricht nach endlich vielen Schritten ab, da alle dabei vorkommenden Basiselemente von der Form $h_{i_1} \wedge h_{i_2} \wedge \ldots \wedge h_{i_k}$ mit Elementen h_{i_j} aus der anfänglichen Basis sind. □

Beispiel 5.14. [Gauss-Abbildung] Um für die Gauss-Abbildung den exponentiellen Korrelationsabfall nachzuweisen, bleibt also zu zeigen, dass $\lambda_1 = 1$ für jedes $k \in \mathbb{N}$ ein einfacher Eigenwert von ist. Wegen Lemma 5.13 reicht es zu zeigen, dass es keine $h_1, h_2 \in \ker(I - P^k)$ mit disjunkten Trägern gibt. Angenommen es gibt solche Dichten. Da $h_1, h_2 \in \mathrm{BV}(X)$, existieren bei beiden Dichten überall die rechts- und die linksseitigen Grenzwerte. Insbesondere gibt es ein Intervall $J \subset X$, so dass $h_1|_J > 0$. Für ausreichend großes $n \in \mathbb{N}$ enthält J ein Intervall Z, das unter einem einzigen monotonen Zweig von T^n auf ganz $X = [0, 1]$ abgebildet wird. Aus der expliziten Darstellung von $P^n = P_{T^n}$ in Gleichung (1.1) folgt, dass $h_1 = P^n h_1 > 0$ auf ganz X, im Widerspruch dazu, dass h_2 disjunkten Träger zu h_1 hat.

5.4 Markov-Operatoren, die eine Doeblin-Fortet Ungleichung erfüllen

Wir beschränken uns hier auf Markov-Operatoren $Qf = \sum_{i \in I} p_i \cdot f \circ \tau_i$ vom Typ „iterierte Funktionensysteme“ wie in (4.3) und nehmen an, dass die Voraussetzungen der Sätze 4.6 und 5.9 erfüllt sind, insbesondere also, dass $r := \sup_{i \in I} \mathrm{Lip}(\tau_i) < 1$. Dann ist $Q\colon L(M) \to L(M)$ quasikompakt auf dem Raum $L(M)$ aller Lipschitz-stetigen Funktionen vom kompakten metrischen Raum (M, d) nach $\mathbb{C}$. Wir wissen, dass $r(Q) = 1$ und dass $Q1 = 1$, so dass 1 ein Eigenwert und $r_e(Q) < 1$ ist.

Für $\boldsymbol{i} = (i_1, i_2, ...) \in I^{\mathbb{N}}$ sei $\tau_{\boldsymbol{i}}^n := \tau_{i_n} \circ ... \circ \tau_{i_1}$. Wenn sinnvoll benutzen wir die gleiche Notation auch für endliche Indexfolgen. Für $x \in M$ sei

$$S(x) := \bigcap_{m>0} \overline{\{\tau_{\boldsymbol{i}}^n(x) \colon \boldsymbol{i} \in I^{\mathbb{N}} \text{ und } n \geqslant m\}}$$

Da $d(\tau_{\boldsymbol{i}}^n(x), \tau_{\boldsymbol{i}}^n(y)) \leqslant r^n\, d(x, y) \to 0$ für alle $x, y \in M$, ist $S := S(x) = S(y)$ für alle $x, y \in M$.

Proposition 5.15. *Ist $p_i(x) > 0$ für alle $i \in I$ und $x \in M$, so ist 1 der einzige periphere Eigenwert und 1 ist einfach.*

Beweis. Sei $Qf = \lambda f$, $0 \neq f \in L(M)$, $|\lambda| = 1$. Da f stetig ist, gibt es ein $x \in M$ mit $|f(x)| = \|f\|_\infty$. Ersetzt man f durch ein geeignetes komplexes Vielfaches, kann man annehmen, dass $f(x) = \|f\|_\infty = 1$. Also ist für alle $n \in \mathbb{N}$

$$1 = \lambda^{-n}(Q^n f)(x) = \lambda^{-n} \sum_{\boldsymbol{i} \in I^n} p_{\boldsymbol{i}}^n(x) \cdot f(\tau_{\boldsymbol{i}}^n(x)) \tag{5.3}$$

mit $p_{\boldsymbol{i}}^n(x) = p_{i_n}(\tau_{\boldsymbol{i}}^{n-1}(x)) \cdot ... \cdot p_{i_1}(x)$. Da $|\lambda| = 1$, $|f| \leqslant 1$ und alle $p_{\boldsymbol{i}}^n(x) > 0$ mit $\sum_{\boldsymbol{i} \in I^n} p_{\boldsymbol{i}}^n(x) = 1$, folgt

$$f(\tau_{\boldsymbol{i}}^n(x)) = \lambda^n \text{ für alle } \boldsymbol{i} \in I^n \text{ und alle } x \in M. \tag{5.4}$$

Ist speziell $\boldsymbol{i} \in I^{\mathbb{N}}$ eine konstante Indexfolge, so konvergiert $\tau_{\boldsymbol{i}}^n(x)$ für $n \to \infty$ gegen den eindeutigen Fixpunkt $x_{\boldsymbol{i}}$ von $\tau_{\boldsymbol{i}}$, und aus der Stetigkeit von f folgt, dass $\lim_{n\to\infty} \lambda^n = f(x_{\boldsymbol{i}})$. Also ist $\lambda = 1$, und aus (5.4) folgt, dass $f|_S = 1$. Da

$$f(y) = (Q^n f)(y) = \sum_{\boldsymbol{i} \in I^n} p_{\boldsymbol{i}}^n(y) \cdot f(\tau_{\boldsymbol{i}}^n(y)) \text{ für alle } n \in \mathbb{N}, \boldsymbol{i} \in I^n \text{ und } y \in M,$$

folgt aus der Stetigkeit von f nun ebenfalls, dass $f = 1$ auf ganz M. Damit ist gezeigt, dass 1 der einzige periphere Eigenwert von Q ist und dass 1 ein halbeinfacher Eigenwert ist. Der Beweis, dass 1 tatsächlich ein einfacher Eigenwert ist, erfolgt ähnlich wie in Lemma 5.10*iv*. □

Proposition 5.16. *Unter den obigen Voraussetzungen gibt es ein Borel-Wahrscheinlichkeitsmaß ν auf M und ein $\rho \in (r, 1)$ derart, dass $L(M) = \mathbb{C} \cdot 1 \oplus \ker(\nu)$, $\nu Q = \nu$ und $r(Q|_{\ker(\nu)}) < \rho < 1$. (Dabei ist $\ker(\nu) = \{f \in L(M) \colon \nu(f) = 0\}$.) Insbesondere gilt für jedes $f \in \ker(\nu)$:*

$$\|Q^n f\|_{\mathrm{Lip}} \leqslant C \cdot \rho^n \cdot \|f\|_{\mathrm{Lip}} \tag{5.5}$$

mit einer von n und f unabhängigen Konstante C. Das Maß ν ist durch $\nu(f) \cdot 1 = \lim_{n\to\infty} Q^n f$ eindeutig bestimmt.

Beweis. Aus Proposition 5.15 folgt die Existenz eines $\rho \in (0, 1)$ für das $L(M) = F_\rho \oplus H_\rho$ wie in Definition 5.7 mit $F_\rho = \mathbb{C} \cdot 1$ und $r(Q|_{H_\rho}) \leqslant \rho < 1$. Wir definieren ν zunächst als stetiges lineares Funktional von $L(M)$ nach $\mathbb{C}$: Jedes $f \in L(M)$ lässt sich schreiben als $f = \alpha_f \cdot 1 + (f - \alpha_f \cdot 1)$ mit $(f - \alpha_f \cdot 1) \in H_\rho$, so dass $\nu(f) \cdot 1 := \alpha_f \cdot 1 = \lim_{n\to\infty} Q^n f$ wohldefiniert ist. Da Q nach Voraussetzung die Doeblin-Fortet Ungleichung (4.4) erfüllt, folgt

$$|\nu(f)| = \lim_{n\to\infty} \|Q^n f\|_{\mathrm{Lip}} \leqslant \frac{C}{1-r} \|f\|_\infty,$$

und da $L(M)$ dicht in $(C(M), \|.\|_\infty)$ liegt, kann ν eindeutig zu einem stetigen linearen Funktional auf $(C(M), \|.\|_\infty)$ fortgesetzt werden. Für $f \geqslant 0$ ist auch $\nu(f) = \lim_{n\to\infty} Q^n f \geqslant 0$. Daher folgt aus dem Riesz'schen Darstellungssatz (z.B. Theorem 2.14,[Rud70]), dass ν durch ein endliches Borel-Maß auf (M, d) gegeben ist. Da $\nu(1) \cdot 1 = \lim_{n\to\infty} Q^n 1 = 1$, ist ν sogar ein Wahrscheinlichkeitsmaß.

Aus Proposition 5.8*ii* und (5.5) folgt nun $\ker(\nu) \subseteq H_\rho$, und da $\mathrm{codim}(\ker(\nu)) = 1$, muss $\ker(\nu) = H_\rho$ sein. □

Proposition 5.17. *Unter den obigen Voraussetzungen ist ν das einzige invariante W'Maß für die durch (4.2) und (4.3) definierte Markovkette $(X_n)_{n\geqslant 0}$. Für jedes $f\in L(M)$ ist*

$$\left|\mathbb{E}[f(X_n)|\mathcal{F}_{n-k}]-\nu(f)\right|\leqslant C\cdot\rho^k\cdot|<f\|_{\mathrm{Lip}}$$

Beweis. Nach Lemma 4.1e) ist

$$\mathbb{E}[f(X_n)|\mathcal{F}_{n-k}]=\mathbb{E}[f(X_n)|X_{n-k}]=(Q^kf)(X_{n-k}).$$

Sei nun ν die Verteilung von X_{n-1}. Dann ist $\mathbb{E}[f(X_n)]=\mathbb{E}[(Qf)(X_{n-1})]=\nu(Qf)=\nu(f)$, d.h. auch X_n hat Verteilung ν und ν ist also ein invariantes W'Maß für die Markov-Kette. Ist μ ein weiteres invariantes W'Maß für die X_n, so ist

$$\mu(f)=\mathbb{E}[f(X_n)]=\mathbb{E}[\mathbb{E}[f(X_n)|\mathcal{F}_{n-k}]]=\mathbb{E}[(Q^kf)(X_{n-k})]=\mu(Q^kf)\overset{k\to\infty}{\to}\mu(\nu(f)\cdot 1)=\nu(f),$$

also $\mu=\nu$. Schließlich ist $|\mathbb{E}[f(X_n)|\mathcal{F}_{n-k}]-\nu(f)|=|(Q^k(f-\nu(f)\cdot 1))(X_{n-k})|\leqslant C\cdot\rho^k\cdot\|f\|_{\mathrm{Lip}}$. □

Bemerkung 5.18. Ist $M=I^{\mathbb{N}_0}$ wie in Beispiel 4.5 und ist $A\subseteq M$ eine Zylindermenge, so ist insbesondere $|P(X_n\in A|\mathcal{F}_{n-k})-\nu(A)|\leqslant C_A\cdot\rho^k$ mit einer von A abhängigen Konstanten, denn Indikatorfunktionen von Zylindermengen sind Lipschitz-stetig bzgl. der in Beispiel 4.5 eingeführten Metrik $d(x,y)=2^{-N(x,y)}$ auf dem Folgenraum M. In der Tat ist die Lipschitz-Konstante eines Zylinders $[a_0,\ldots,a_N]$ gleich 2^N.

6 Klassische Störungstheorie für quasikompakte Operatoren

6.1 Ein allgemeiner Störungssatz

Sei Q ein quasikompakter Operator auf einem Banachraum $\mathcal{B}$. Wir nehmen an, dass das periphere Spektrum von Q nur aus dem einfachen Eigenwert $r(Q)$ besteht, so wie es bei vielen Markov-Operatoren oder Frobenius-Perron-Operatoren mit $r(Q)=1$ der Fall ist.[6.1] Ist nun Q eingebettet in eine Schar von Operatoren $Q_t\colon\mathcal{B}\to\mathcal{B}$ $(-\delta<t<\delta)$ mit $Q_0=Q$ und ist die Abbildung $t\mapsto Q_t$ m-fach stetig differenzierbar als Abbildung von $(-\delta,\delta)$ nach $L(\mathcal{B},\mathcal{B})$ (den Banachraum der beschränkten linearen Operatoren auf $\mathcal{B}$), so haben auch die Q_t einen isolierten einfachen peripheren Eigenwert λ_t und es ist $t\mapsto\lambda_t$ m-fach stetig differenzierbar mit $\lambda_0=r(Q)$. Genauer:

Satz 6.1. [Theorem III.8 und Korollar III.11 in [HH01]]
Sei die Familie $(Q_t)_{|t|<\delta}$ von Operatoren auf $\mathcal{B}$ wie eben beschrieben. Dann gibt es ein $\rho\in(0,r(Q))$, ein $\delta_0\in(0,\delta)$ und m-fach stetig differenzierbare Abbildungen

- $\lambda\colon(-\delta_0,\delta_0)\to\mathbb{R}$,
- $\varphi\colon(-\delta_0,\delta_0)\to\mathcal{B}$,
- $\nu\colon(-\delta_0,\delta_0)\to\mathcal{B}^*$ *und*
- $N\colon(-\delta_0,\delta_0)\to L(\mathcal{B},\mathcal{B})$

derart, dass für alle $t\in(-\delta_0,\delta_0)$ gilt:

a) $Q_t\varphi_t=\lambda_t\varphi_t$

b) $\nu_tQ_t=\lambda_t\nu_t$

c) $\langle\nu_t,\varphi_t\rangle=1$

d) $Q_t^n=\lambda_t^n\cdot\varphi_t\otimes\nu_t+N_t^n$ *für alle $n\in\mathbb{N}$ und*

6.1. Ein Resultat, das auch im Fall mehrerer peripherer Eigenwerte anwendbar ist, findet man in [HH01], Theorem III.8.

e) $r(N_t) < \rho$, *genauer sogar:* $\|N_t^n\| \leqslant C\,\rho^n$ *für alle* n *mit einer von* t *unabhängigen Konstanten* C.

Außerdem gilt mit $Q_0' := \frac{dQ}{dt}_{|t=0}$ *und* $Q_0'' := \frac{d^2Q}{dt^2}_{|t=0}$*:*

$$\frac{d\lambda}{dt}_{|t=0} = \left\langle \nu_0, Q_0'(\varphi_0) \right\rangle \text{ und} \tag{6.1}$$

$$\frac{d^2\lambda}{dt^2}_{|t=0} = \left\langle \nu_0, Q_0''(\varphi_0) \right\rangle + 2\left\langle \nu_0, Q_0'(\lambda_0 - N_0)^{-1}(1 - \varphi_0 \otimes \nu_0) Q_0'\varphi_0 \right\rangle \tag{6.2}$$

Beweis. Wir leiten nur die Formeln für die Ableitungen her: Durch zweimaliges differenzieren von $Q_t\varphi_t = \lambda_t\varphi_t$ folgt:

$$Q_t'\varphi_t + Q_t\varphi_t' = \lambda_t'\varphi_t + \lambda_t\varphi_t' \text{ und} \tag{6.3}$$

$$Q_t''\varphi_t + 2Q_t'\varphi_t' + Q_t\varphi_t'' = \lambda_t''\varphi_t + 2\lambda_t'\varphi_t' + \lambda_t\varphi_t''. \tag{6.4}$$

Ausgewertet bei $t = 0$ ergibt das unter ν_0 (der Index 0 wird von nun an unterdrückt):

$$\langle \nu, Q'\varphi \rangle + \lambda\langle \nu, \varphi' \rangle = \lambda' + \lambda\langle \nu, \varphi' \rangle \text{ und} \tag{6.5}$$

$$\langle \nu, Q''\varphi \rangle + 2\langle \nu, Q'\varphi' \rangle + \lambda\langle \nu, \varphi'' \rangle = \lambda'' + 2\lambda'\langle \nu, \varphi' \rangle + \lambda\langle \nu, \varphi'' \rangle \tag{6.6}$$

Aus (6.5) folgt sofort Gleichung (6.1). Aus (6.6) folgt

$$\lambda'' = \langle \nu, Q''\varphi \rangle + 2\langle \nu, (Q' - \lambda')\varphi' \rangle$$

und daraus folgt (6.2), denn aus (6.3) folgt zunächst, dass

$$\begin{aligned}(1 - \varphi \otimes \nu)(Q' - \lambda')\varphi &= (1 - \varphi \otimes \nu)(\lambda - Q)\varphi' = (\lambda - Q)(1 - \varphi \otimes \nu)\varphi' \\ &= (\lambda - N)(1 - \varphi \otimes \nu)\varphi'\end{aligned}$$

und daher (unter Beachtung von (6.5))

$$\begin{aligned}\langle \nu, (Q' - \lambda')\varphi' \rangle &= \langle \nu Q', (1 - \varphi \otimes \nu)\varphi' \rangle \\ &= \left\langle \nu, Q'(\lambda - N)^{-1}(1 - \varphi \otimes \nu)(Q' - \lambda')\varphi \right\rangle \\ &= \left\langle \nu, Q'(\lambda - N)^{-1}(1 - \varphi \otimes \nu)Q'\varphi \right\rangle,\end{aligned}$$

also Gleichung (6.2). □

6.2 Zentraler Grenzwertsatz durch Störungstheorie

Sei nun $(X_n)_{n \in \bar{N}}$ eine Markovkette mit Werten im kompakten metrischen Raum (M, d) wie in Kapitel 4. Wir nehmen an, dass ihr Markov-Operator Q: $L(M) \to L(M)$ quasikompakt ist und dass $r(Q) = 1$ ein einfacher Eigenwert ist und der Rest des Spektrums in einer Kreisscheibe vom Radius < 1 liegt. Dann gilt Satz 6.1. Sei nun f: $M \to \mathbb{R}$ Lipschitz-stetig mit $\mathbb{E}[f(X_0)] = 0$ unter der invarianten Verteilung des Prozesses. Das heißt mit der Darstellung $Q = \varphi_0 \otimes \nu_0 + N_0$ wie im vorherigen Abschnitt, dass $\langle \nu_0, f \rangle = 0$. (Beachte, dass hier $\varphi_0 = 1$.) Sei $S_n := \sum_{k=0}^{n-1} f(X_k)$. Dann ist $\mathbb{E}[S_n] = 0$ für alle n, und wir wollen zeigen, dass dieser Partialsummenprozess einem zentralen Grenzwertsatz genügt:

Satz 6.2. *Unter den obigen Voraussetzungen gilt:* $\mathcal{L}(n^{-1/2}\mathcal{S}_n) \Rightarrow \mathcal{N}(0, \sigma^2)$ *mit*

$$\begin{aligned}\sigma^2 &= -\frac{d^2\,\lambda_t}{dt^2}_{|t=0} \\ &= \mathbb{E}[(f(X_0))^2] + 2\sum_{k=1}^{\infty} \mathbb{E}[f(X_0)f(X_k)]\end{aligned}$$

Bemerkung: *Eigentlich sollte es heißen:* $\sigma^2 = -\frac{d^2 \log(\lambda_t)}{dt^2}_{|t=0}$, *aber da* $\lambda_0 = 1$ *und, wie wir unten sehen werden,* $\frac{d\lambda_t}{dt}_{|t=0} = 0$ *ist, ist das gleich* $-\frac{d^2\lambda_t}{dt^2}_{|t=0}$.

Beweis. Für $t\in\mathbb{R}$ und $x\in M$ sei $g_t(x)=\exp(it\cdot f(x))$. Dann ist $t\mapsto\mathbb{E}[\prod_{k=0}^{n-1} g_{tn^{-1/2}}(X_k)]$ die charakteristische Funktion von $n^{-1/2}S_n$, und wir müssen zeigen, dass sie für jedes $t\in\mathbb{R}$ mit $n\to\infty$ gegen $\exp(-\frac{\sigma^2t^2}{2})$ konvergiert. Wir definieren die Familie von Operatoren $Q_t\colon L(M)\to L(M)$ durch $Q_t\varphi:=Q(g_t\,\varphi)$. Da $Q_0=Q$,

$$\begin{aligned}\tfrac{1}{\delta}(Q_{t+\delta}\varphi-Q_t\varphi)-iQ_t(f\varphi) &= \tfrac{1}{\delta}Q((g_{t+\delta}-g_t)\varphi)-iQ_t(f\varphi)=Q_t\Big(\tfrac{1}{\delta}(g_\delta-1)\varphi\Big)-iQ_t(f\varphi)\\ &= Q_t\Big(\Big(\tfrac{1}{\delta}(e^{i\delta f}-1)-if\Big)\,\varphi\Big)\end{aligned}$$

und da $\mathrm{Lip}(\frac{1}{\delta}(e^{i\delta f(x)}-1)-if(x))\leqslant\mathrm{Lip}(f)\cdot\mathcal{O}(\delta\,\|f\|_\infty)$, folgt aus der Beschränktheit von Q_t als Operator auf $L(M)$, dass $\|\frac{1}{\delta}(Q_{t+\delta}\varphi-Q_t\varphi)-iQ_t(f\varphi)\|_{\mathrm{Lip}}\leqslant\|f\|_{\mathrm{Lip}}^2\cdot\|\varphi\|_{\mathrm{Lip}}\cdot\mathcal{O}(\delta)$, so dass die Familie $(Q_t)_{t\in\mathbb{R}}$ differenzierbar ist und $\frac{d}{dt}(Q_t\varphi)=i\,Q_t(f\,\varphi)$. Daraus folgt durch nochmaliges Differenzieren sofort, dass $\frac{d^2}{dt^2}(Q_t\varphi)=-\,Q_t(f^2\varphi)$.

Da $\varphi_0\equiv1$ ist, folgt aus (6.1) und (6.2):

$$\begin{aligned}\frac{d\lambda_t}{dt}_{|t=0} &= i\,\langle\nu_0,Qf\rangle=i\,\langle\nu_0,f\rangle=0\quad\text{und}\\ \frac{d^2\lambda_t}{dt^2}_{|t=0} &= -\big\langle\nu_0,Qf^2\big\rangle-2\,\big\langle\nu_0,Q(f\cdot(1-N_0)^{-1}(1-1\otimes\nu_0)Qf)\big\rangle\\ &= -\big\langle\nu_0,f^2\big\rangle-2\,\big\langle\nu_0,f\cdot(1-N_0)^{-1}N_0\,f\big\rangle\\ &= -\left(\mathbb{E}[(f(X_0))^2]+2\sum_{k=1}^{\infty}\mathbb{E}[f(X_0)(Q^kf)(X_0)]\right)\\ &= -\sigma^2\end{aligned}$$

da aus $\langle\nu_0,f\rangle=0$ folgt dass $Q^kf=N_0^kf$ und da $(Q^kf)(X_0)=\mathbb{E}[f(X_k)|\mathcal{F}_0]$. Also folgt (zumindest formal, wir übergehen einige detaillierte Konvergenzbetrachtungen): Mit $t_n:=tn^{-1/2}$ ist

$$\begin{aligned}\mathbb{E}\left[\prod_{k=0}^{n-1}g_{tn^{-1/2}}(X_k)\right] &= \mathbb{E}\left[\prod_{k=0}^{n-2}g_{t_n}(X_k)\cdot\mathbb{E}[g_{t_n}(X_{n-1})|\mathcal{F}_{n-2}]\right]\\ &= \mathbb{E}\left[\prod_{k=0}^{n-2}g_{t_n}(X_k)\cdot(Q_{t_n}1)(X_{n-2})\right]\\ &= \mathbb{E}\left[\prod_{k=0}^{n-3}g_{t_n}(X_k)\cdot\mathbb{E}[g_{t_n}(X_{n-2})\cdot(Q_{t_n}1)(X_{n-2})|\mathcal{F}_{n-3}]\right]\\ &= \mathbb{E}\left[\prod_{k=0}^{n-3}g_{t_n}(X_k)\cdot(Q_{t_n}^21)(X_{n-3})\right]\\ &=\\ &\ \vdots\\ &= \mathbb{E}[(Q_{t_n}^n1)(X_0)]=\lambda_{t_n}^n\,\langle\nu_{t_n},1\rangle\,\langle\nu,\varphi_{t_n}\rangle+\mathcal{O}(\rho^n)\\ &\to \lim_{n\to\infty}\lambda_{t_n}^n\quad\text{mit }n\to\infty\end{aligned}$$

da $\nu_{t_n}\to\nu_0$, $\varphi_{t_n}\to\varphi_0=1$ und $\langle\nu_0,1\rangle=1$. Schließlich ist

$$\begin{aligned}\lim_{n\to\infty}\lambda_{t_n}^n &= \lim_{n\to\infty}\left(1+\frac{1}{2}t_n^2\lambda_0''+\mathcal{O}(t_n^3)\right)^n\\ &= \lim_{n\to\infty}\left(1-\frac{t^2\sigma^2}{2\,n}+\mathcal{O}(n^{-\frac{3}{2}})\right)^n\\ &= \exp\left(-\frac{\sigma^2t^2}{2}\right)\end{aligned}$$

□

Auf ganz ähnliche Weise kann man auch einen schon in Bemerkung 5.12 angesprochenen zentralen Grenzwertsatz für Partialsummenprozesse in dynamischen Systemen beweisen. Details findet man zum Beispiel in Kapitel 8.5 von [BG97].

7 Dynamische Störungstheorie für Markov-Operatoren, die eine Lasota-Yorke oder Doeblin-Fortet-Ungleichung erfüllen

Der Beweis des zentralen Grenzwertsatzes beruhte auf einem Satz der differenzierbaren Störungstheorie für isolierte einfache Eigenwerte linearer Operatoren. Die wichtigste Voraussetzung dieses Satzes, nämlich dass die Familie $t \mapsto Q_t$ differenzierbar ist, konnten wir nachweisen, weil alle Q_t *auf dem selben dynamischen System* beruhen.

Eine andere Situation liegt vor, wenn wenn wir das System etwas stören und zeigen möchten, dass dann auch z.B. die eindeutige invariante Dichte nur wenig gestört wird. Wir werden sehen, dass solche Störungen in der Regel nicht differenzierbar sind, und man benötigt eine andere Störungstheorie, die mit schwächeren Vorausetzungen auskommt (und dann natürlich auch etwas schwächere Resultate liefert).

Zunächst soll die Schwierigkeit illustriert werden. Sei $X = [0, 1]$, $T: X \to X$ eine stückweise expandierende Abbildung wie in Definition 3.8. $\tilde{T}$ sei eine weitere solche Abbildung, die sehr ähnlich zu T ist, aber sich etwas von ihr unterscheidet: Es gebe einen Punkt $x_0 \in X$ mit $\varepsilon := \tilde{T}x_0 - Tx_0 > 0$. Der Einfachheit halber setzen wir voraus, dass $1 \leqslant T'(x_0), \tilde{T}'(x_0) \leqslant S$ für ein $S > 0$. Sei $\delta > 0$ so klein, dass $Tx_0 < T(x_0 + \delta) < \tilde{T}x_0$ und $I := [x_0, x_0 + \delta]$. Dann ist $\mathrm{Var}(1_I) = 2$, aber $P_T 1_I$ und $P_{\tilde{T}} 1_I$ haben disjunkte Trägerintervalle $J := [Tx_0, T(x_0 + \delta)]$ und $\tilde{J} := [\tilde{T}x_0, \tilde{T}(x_0 + \delta)]$ und beide haben auf ihrem jeweiligen Trägerintervall mindestens den Wert $\frac{1}{S}$. Daher ist

$$\mathrm{Var}((P_T - P_{\tilde{T}})1_I \geqslant \frac{2}{S} + \frac{2}{S} = \frac{4}{S}$$

egal wie nahe T und $\tilde{T}$ zueinander sind. Also kann die Zuordnung $T \mapsto P_{\tilde{T}}$ noch nicht einmal stetig sein.

Einen Ausweg aus diesem Dilemma bildet die folgende Norm auf dem Raum der stetigen linearen Operatoren $P: \mathrm{BV}(X) \to \mathrm{BV}(X)$:

$$|||P||| := \sup \{ \|Pf\|_1 : \|f\|_{\mathrm{BV}} \leqslant 1 \}, \tag{7.1}$$

oder entsprechend auf dem Raum der Markov-Operatoren $Q: L(M) \to L(M)$:

$$|||Q||| := \sup \{ \|\varphi\|_\infty : \|\varphi\|_{\mathrm{Lip}} \leqslant 1 \}.$$

Wir werden uns von nun an nur auf Operatoren P: $\mathrm{BV}(X) \to \mathrm{BV}(X)$ konzentrieren; für die Markov-Operatoren auf $L(M)$ kann man aber ganz analog vorgehen. Details, insbesondere Beweise, die hier nicht durchgeführt werden (können), findet man in [Bal00, Abschnitt 3.3] oder [BG97, Kapitel 11].

7.1 Kleine Störungen in der $|||.|||$-Norm

Seien T und $\tilde{T}$ zwei stückweise expandierende Abbildungen auf dem Intervall X. Wir definieren einen Abstandsbegriff zwischen den Graphen von T und $\tilde{T}$, der sich an der Skorokhod-Metrik für den Abstand von Zufallspfaden mit Sprüngen orientiert. Im folgenden sei m das Lebesgue-Maß auf X.

Definition 7.1. *Für zwei stückweise expandierende Abbildungen $T, \tilde{T}: X \to X$ sei*

$$\begin{aligned} d(T, \tilde{T}) \;:=\; & \inf \{ \varepsilon > 0 : \exists A \subseteq X \text{ und } \exists \text{ Diffeomorphismus } \sigma: X \to X \text{ so dass } m(A) > 1 - \varepsilon, \\ & T_{|A} = \tilde{T} \circ \sigma_{|A} \text{ und } \forall x \in A : |\sigma(x) - x| < \varepsilon, \left|\frac{1}{\sigma'(x)} - 1\right| < \varepsilon \} \end{aligned}$$

So wie hier definiert ist d nicht symmetrisch, aber das spielt für unsere Zwecke keine Rolle. Mit diesem Abstandsbegriff kann man nämlich zeigen:

Lemma 7.2. [BG97, Lemma 11.2.1]

$$|||P_T - P_{\tilde{T}}||| \leqslant 14 \cdot d(T, \tilde{T})$$

Beispiel 7.3. Sei $X = [-\frac{1}{2}, \frac{1}{2}]$, und für $1 < a \leqslant 2$ sei $T_a: X \to X$, $T_a(x) := \frac{a-1}{2} - a|x|$. Die T_a heißen *Zeltabbildungen*. Sie haben Steigung $\pm a$. Seien $1 < a < \tilde{a} \leqslant 2$. Setze $b = (1 - \frac{a}{\tilde{a}})$ und wähle $A = [-\frac{1}{2}, -b] \cup [b, \frac{1}{2}]$. Dann ist $m(A) = 1 - 2b \geqslant 1 - 2(\tilde{a} - a)$. Definierfe nun $\sigma(x) = \frac{a}{\tilde{a}}(x + \frac{1}{2}) - \frac{1}{2}$ für $x \in [-\frac{1}{2}, -b]$, $\sigma(x) = \frac{a}{\tilde{a}}(x - \frac{1}{2}) + \frac{1}{2}$, und setze σ auf dem Intervall $[-b, b]$ zu einem Diffeomorphismus von X fort. Für $x \in A$ gilt dann: $T_{\tilde{a}}(\sigma(x)) = \frac{\tilde{a}-1}{2} - |a x + \frac{\tilde{a}-a}{2}| = T_a(x)$ (Fallunterscheidung!), $|\sigma(x) - x| \leqslant \tilde{a} - a$ und $|\frac{1}{\sigma'(x)} - 1| \leqslant \tilde{a} - a$, so dass $d(T_a, T_{\tilde{a}}) \leqslant 2(\tilde{a} - a)$ und daher $|||P_{T_a} - P_{T_{\tilde{a}}}||| \leqslant 28(\tilde{a} - a)$.

Ganz ähnlich kann man kleine stochastische Störungen behandeln. Dann betrachtet man eine Markovkette $(X_n)_{n \geqslant 0}$, die durch $X_{n+1} = T(X_n) + \xi_{n+1}$ definiert ist, wobei die ξ_n unabhängig identsich verteilt sind mit $|\xi_n| \leqslant \varepsilon$ und glatter Verteilungsdichte. Dann gilt folgendes Lemma:

Lemma 7.4. [BG97, Theorem 1.3.1] *Es gibt eine Konstante $C > 0$, so dass*

$$|||P_T - P_\varepsilon||| \leqslant C\varepsilon$$

wenn P_ε den Markov-Operator der Markovkette bezeichnet.

7.2 Störungssätze

In diesem Abschnitt betrachten wir eine Situation wie in Abschnitten 5.1 und 5.2: Sei $(\mathcal{B}, \|.\|)$ ein Banach-Raum und sei $|.|$ eine weitere Norm auf $\mathcal{B}$, $|f| \leqslant \|f\|$ für alle $f \in \mathcal{B}$. (Es kann auch nur eine Semi-Norm sein.) $(Q_\varepsilon)_{|\varepsilon| < \varepsilon_0}$ sei eine Familie linearer Operatoren $Q_\varepsilon: \mathcal{B} \to \mathcal{B}$ mit folgenden Eigenschaften:

i. $Q_\varepsilon\{f \in \mathcal{B}: \|f\| \leqslant 1\}$ ist total beschränkt (d.h. relativ kompakt) in $(\mathcal{B}, |.|)$,

ii. $|Q_\varepsilon| \leqslant 1$.

iii. Es gibt ein $k \in \mathbb{N}$ und reelle Konstanten $r, R > 0$, so dass $r < 1$ und dass für alle $f \in \mathcal{B}$ und alle $|\varepsilon| < \varepsilon_0$ gilt

$$\|Q_\varepsilon^k f\| \leqslant r^k \|f\| + R|f|. \tag{7.2}$$

Aus Satz 5.9 wissen wir dann, dass alle Q_ε quasikompakt sind.

In Verallgemeinerung des Vorgehens aus dem vorherigen Abschnitt definieren wir für einen Operator $P: (\mathcal{B}, \|.\|) \to (\mathcal{B}, |.|)$ die Operatornorm

$$|||P||| = \sup\{|Pf|: f \in \mathcal{B}, \|f\| \leqslant 1\}.$$

Für $|\lambda| > r$, $|\varepsilon| < \varepsilon_0$ und $\delta > 0$ sei

$$\Pi_\varepsilon^{(\lambda,\delta)} := \frac{1}{2\pi i} \int_{|z-\lambda|=\delta} (z - Q_\varepsilon)^{-1} \, dz.$$

Das ist eine Spektralprojektion, die im Fall eines isolierten Eigenwertes λ von Q_ε für jedes hinreichend kleine $\delta > 0$ gerade auf $\bigcup_{j=1}^\infty \ker((\lambda - Q_\varepsilon)^j)$ projiziert. (δ is hinreichend klein, wenn $\sigma(Q_\varepsilon) \cap \{z: |z - \lambda| \leqslant \delta\} = \{\lambda\}$.) Der folgende Satz stammt aus [KL99], siehe auch [Bal00]. Teil c) wurde schon in [Kel82] gezeigt.

Satz 7.5. [Störungssatz] *Unter den obigen Voraussetzungen gilt für jeden isolierten Eigenwert λ von Q_0 mit $|\lambda| > r$, jedes $\alpha \in (r, |\lambda|)$ und jedes $\delta \in (0, r - \alpha)$ mit $\{z: |z - \lambda| \leqslant \delta\} \cap \sigma(Q_0) = \{\lambda\}$: Sei*

$$\eta := \frac{\log\frac{\alpha}{r}}{\log\frac{1}{r}}$$

Dann gilt für ε hinreichend nahe bei 0:

a) Es gibt ein $K_1 = K_1(\delta, r) > 0$ so dass $|||\Pi_\varepsilon^{(\lambda,\delta)} - \Pi_0^{(\lambda,\delta)}||| \leqslant K_1\ |||Q_\varepsilon - Q_0|||^\eta$

b) $\mathrm{Rang}(\Pi_\varepsilon^{(\lambda,\delta)}) = \mathrm{Rang}(\Pi_0^{(\lambda,\delta)})$

c) *Ist* $|\lambda| = 1$*, so kann die obere Schranke in a) durch* $|||Q_\varepsilon - Q_0||| \cdot \log \frac{1}{|||Q_\varepsilon - Q_0|||}$ *ersetzt werden.*

Beweis. Hier sollen nur ein paar Anmerkungen zum Beweis gegeben werden:
A) Die Abschätzung in Teil a) folgt direkt aus der Ungleichung

$$|||(z - Q_\varepsilon)^{-1} - (z - Q_0)^{-1}||| \leqslant \mathrm{const} \cdot \left(\|(z - Q_0)^{-1}\| + \|(z - Q_0)^{-1}\|^2 \right) \cdot |||Q_\varepsilon - Q_0|||,$$

und diese Ungleichung lässt sich vollkommen elementar beweisen, d.h. durch flexible und vielfache Kombination von Ungleichung (7.2) mit der Dreieckungleichung.
B) Hier ist der Beweis von Teil c) im Fall, dass $\lambda = 1$ der einzige periphere Eigenwert ist, dass es ein einfacher Eigenwert ist und dass $\nu_\varepsilon = \nu_0$ (wie im Fall der Frobenius-Perron Operatoren). In diesem Fall ist $\Pi_0^{(1,\delta)} = \varphi_0 \otimes \nu_0$. Sei zunächst

$$\Phi_\varepsilon := \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} Q_\varepsilon^k$$

Die Existenz der Φ_ε folgt leicht aus Ungleichung (7.2), der Quasikompaktheit und der Voraussetzung $|Q_\varepsilon| \leqslant 1$; ebenso die Existenz einer Konstanten $C > 0$ so dass $\|\Phi_\varepsilon\| \leqslant C$. Beachte, dass $\Phi_0 = \Pi_0^{(1,\delta)} = \varphi_0 \otimes \nu_0$ und schreibe $Q_0 = \Phi_0 + N_0$. Dann ist

$$(1 - N_0)(\Phi_0 - 1)\Phi_\varepsilon \;=\; (\Phi_0 - 1 + N_0)\Phi_\varepsilon = (Q_0 - 1)\Phi_\varepsilon = (Q_0 - Q_\varepsilon)\Phi_\varepsilon\,,$$

also

$$(\Phi_0 - 1)\Phi_\varepsilon = (1 - N_0)^{-1}(Q_0 - Q_\varepsilon)\Phi_\varepsilon = \left(\sum_{n=0}^{L-1} N_0^n + N_0^L (1 - N_0)^{-1} \right) (Q_0 - Q_\varepsilon)\Phi_\varepsilon$$

und daher

$$\begin{aligned} |||\Phi_0\Phi_\varepsilon - \Phi_\varepsilon||| &\leqslant \sum_{n=0}^{L-1} |N_0^n| \cdot |||Q_0 - Q_\varepsilon||| \cdot \|\Phi_\varepsilon\| + \rho(N_0)^L \cdot \frac{\mathrm{const}}{1 - \rho(N_0)} \cdot (\|Q_0\| + \|Q_\varepsilon\|) \cdot \|\Phi_\varepsilon\| \\ &\leqslant \mathrm{const} \cdot \left(L \cdot |||Q_0 - Q_\varepsilon||| + \rho(N_0)^L \right) \end{aligned}$$

und für $L \approx \frac{\log |||Q_0 - Q_\varepsilon|||}{\log \rho(N_0)}$ erhält man

$$|||\Phi_0\Phi_\varepsilon - \Phi_\varepsilon||| \leqslant \mathrm{const} \cdot |||Q_\varepsilon - Q_0||| \cdot \log \frac{1}{|||Q_\varepsilon - Q_0|||}\,.$$

Bleibt zu beachten, dass $\Phi_0\Phi_\varepsilon = \langle \nu_0, \varphi_\varepsilon \rangle \cdot \varphi_0 \otimes \nu_\varepsilon = \varphi_0 \otimes \nu_0 = \Phi_0$ unter den speziellen oben gemachten Annahmen. □

Literaturverzeichnis

[Bal00] V Baladi. *Positive Transfer Operators and Decay of Correlations*. World Scientific, 2000.

[BG97] A Boyarsky and P Góra. *Laws of Chaos - Invariant Measures and Dynamical Systems in One Dimension*. Birkhäuser, 1997.

[DS57] N Dunford and J T Schwartz. *Linear Operators, Part I: General Theory*. Wiley, 1957.

[HH01] H Hennion and L Hervé. *Limit Theorems for Markov Chains and Stochastic Properties of Dynamical Systems by Quasi-compactness*, volume 1766 of *Lecture Notes in Mathematics*. Springer, 2001.

[Kel82] G Keller. Stochastic stability in some chaotic dynamical systems. *Monatshefte Math.*, 94:313–333, 1982.

[KL99] G Keller and C Liverani. Stability of the spectrum for transfer operators. *Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4)*, XXVIII:141–152, 1999.

[LM85] A Lasota and M C Mackey. *Probabilistic Properties of Deterministic Systems*. Cambridge University Press, 1985.

[LY73] A Lasota and J A Yorke. On the existence of invariant measures for piecewise monotonic transformations. *Transactions Amer. Math. Soc.*, 186:481–488, 1973.

[Nat75] I P Natanson. *Theorie der Funktionen einer reellen Veränderlichen*. H. Deutsch, 1975.

[Rud70] W Rudin. *Real and complex Analysis*. McGraw-Hill, 1970.

[Ryc83] M R Rychlik. Bounded variation and invariant measures. *Studia Mathematica*, LXXVI:69–80, 1983.

[Sau05] T Sauer. Kettenbrüche. Vorlesungsskript (http://www.uni-giessen.de/tomas.sauer/Skripten/Kettenbrueche.pdf), 2005.

[TM50] C T Ionescu Tulcea and G Marinescu. Théorème ergodique pour les classes d'opérations non complètement continues. *Ann. of Math.*, 52:140–147, 1950.