

Diskretwertige stochastische Vektorräume

Grundlagen, Ergodentheorie und Darstellungen

Inaugural-Dissertation
zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultät
der Universität zu Köln

vorgelegt von
Alexander Schönhuth
aus Bad Mergentheim

Druckerei Schönhuth, Bad Mergentheim

2007

Berichterstatter: Prof. Dr. U. Faigle
Prof. Dr. H. Jäger

Tag der mündlichen Prüfung: 31.05.2006

Kurzzusammenfassung

Inhalt dieser Arbeit sind eine Ausarbeitung eines neuen auf Vektorräumen basierten Formalismus zur Untersuchung stochastischer Prozesse sowie, darauf aufbauend, Dimensionsbegriffe, die zur Definition neuer oder wenig analysierter Prozessklassen führen. Diese Prozessklassen enthalten die weit- hin beliebte Klasse der Hidden Markov Prozesse, die intuitiv leicht zugänglich sind, sich rein mathematischen Untersuchungen gegenüber jedoch unzugänglich erweisen.

Wir zeigen in dieser Arbeit, dass mit Hilfe des neuen Formalismus deutlich einfachere Beweise für bereits bestehende Erkenntnisse erhalten werden können und fügen eine Reihe neuer Ergebnisse hinzu. Wir zeigen darüber hinaus Wege zum konkreten Einsatz der der neuen Klassen entstammenden Prozesse für Lernalgorithmen auf.

Abstract

In this thesis we develop a new framework, within which the analysis of stochastic processes based on vector spaces is made possible. Building on it, we outline conceptions of dimension for stochastic processes and examine the resulting classes of finite dimensional processes. These classes contain the prevalent class of Hidden Markov Models, which are successfully employed in statistical modeling and statistical pattern recognition. Although intuitively easy to perceive their mathematical analysis proves rather difficult.

We thus show that new insights can be gained from the vantage point of our framework. We give simpler proofs for existing results and add new theorems. Moreover, we point out how one may use finite dimensional processes within practical applications by providing hints for the construction of learning algorithms.

Inhaltsverzeichnis

Vorwort	vii
Kapitel 1. Einleitung	1
1.1. Motivation und Zielsetzung	1
1.2. Geschichte und verwandte Arbeiten	3
1.3. Kapitelübersicht	7
Teil 1. Stochastische Vektorräume	11
Kapitel 2. Stochastische Vektorräume - Grundbegriffe	13
2.1. Wörter, Zylindermengen und (σ) -Algebren	14
2.2. Stochastische Vektorräume und (Prä)-Maße	18
2.3. Normen, Konvergenz und signierte Maße	27
2.4. Ereignisvektoren	46
Kapitel 3. Operatoren und Dimension	53
3.1. Prognoseoperatoren	54
3.2. Dimension signierter Maßvektoren	59
3.3. Basen und Charakteristische Systeme	76
Kapitel 4. Evolutionstheorie	85
4.1. Motivation: Dynamische Systeme	86
4.2. Stationarität	87
4.3. Evolutionssummierbarkeit	103
4.4. Ergodizität	115
Kapitel 5. Die ergodische Zerlegung	129
5.1. Stationäre Vektoren	129
5.2. Der allgemeine Fall	134
5.3. Klassifikation von endlich-dimensionalen Prozessen	137
Kapitel 6. Entropieraten	141
6.1. Einführung: Entropie	141
6.2. Entropieraten stationärer Vektoren	142
6.3. Entropieraten von ES-Vektoren	145
Kapitel 7. Prognosegraphen	153
7.1. Einführung	153
7.2. Ergodische Zerlegung von Prognosegraphen	160
Kapitel 8. Quanten-Prognose-Modelle	167
8.1. Der Zustandsraum	167
8.2. Schatten-Normen	168

8.3. Quanten-Prognose-Modelle	170
8.4. Quanten-Markov-Ketten	173
8.5. Beispiele von Quanten-Markov-Ketten	175
Kapitel 9. Zusammenfassung und Ausblick	179
9.1. Zusammenfassung	179
9.2. Ausblick und offene Probleme	180
Teil 2. Lineare Algebra: Stabilitätstheorie, Kegel und Komponentensummen	183
Kapitel 10. Stabilitätstheorie	185
10.1. Grundlagen und Notationen	185
10.2. Stabile Matrizen	186
10.3. Reelle Stabilitätstheorie	187
10.4. Cesaro-Mittel und Stabilität	192
Kapitel 11. Folgenrang und Folgenstabilität	199
11.1. Folgenrang	199
11.2. Folgenstabilität	206
Kapitel 12. Komponentensummen und Kegel	209
12.1. Komponentensummenerhaltende Matrizen	209
12.2. Invarianz von Kegeln und Komponentensummen, Stabilität	212
Teil 3. Appendix	217
Anhang A. Vier technische Lemmata für (signierte) Maße	219
Anhang B. Beweis von Satz 4.41	221
B.1. Konvergenzformen von Zufallsvariablen	222
B.2. Absolutstetigkeit und der Satz von Radon-Nikodym	223
B.3. Der Raum $L_1(P)$	224
B.4. Formen der Konvergenz von Maßen	226
B.5. Positive Kontraktionen	227
B.6. Beweis des Satzes	228
Anhang C. Jordan-Normal-Formen	229
C.1. Grundlagen	229
C.2. Jordan-Normal-Formen	230
C.3. Zwei technische Lemmata	236
Anhang D. Korollare des Satzes von Jacobi	241
Literaturverzeichnis	243
Index	249
Nachwort und Danksagung	255
Erklärung	259

Vorwort

*The whole of science is nothing more than
a reformulation of everyday thinking.*

Albert Einstein: *Physics and Reality*

Die folgenden rund 250 Seiten sind der offizielle Ausdruck einer mathematischen Doktorarbeit. Tatsächlich scheint mir, im Sinne des obigen Zitats, die eigentliche “Doktorarbeit” woanders stattgefunden zu haben. In unzähligen Diskussionen, in Vorträgen und natürlich vor allem in zahllosen bewussten und unbewussten Gedankengängen zu den verschiedensten Tages- und auch Nachtzeiten, denen ein Schriftstück unmöglich gerecht werden kann. Dennoch hoffe ich, so viel wie möglich davon transportiert und kein lebloses Formelwerk abgeliefert zu haben.

Im Übrigen ist dies einer der ersten Beiträge zur Mathematik des 21. Jahrhunderts. Doch auch wenn man wahrscheinlich hin und wieder gerade an den Grundfesten rütteln sollte, um eine Wissenschaft am Leben zu erhalten, habe ich mit dieser Arbeit nichts dazu beigetragen. Alle Variablen entstammen Mengen, die, gemäß den Gepflogenheiten des 20. Jahrhunderts, genau definiert werden und dem Zermelo-Fraenkel-Cantorschen Axiomensystem gerecht werden. Auch wenn in neuerer Zeit “neoorthodoxe” Stimmen laut werden (siehe z.B. [Jay06], Appendix B: Formalities and Mathematical Style), dieses System würde der computerisierten Welt von heute nicht mehr ganz gerecht und z.B. das Positivitätsproblem, das im Kontext dieser Arbeit aufgeworfen wird, gerade Ausdruck dieses Makels ist.

Alexander Schönhuth
Köln, den 27.03.2006

KAPITEL 1

Einleitung

Diskretwertige stochastische Prozesse, das heißt von Gesetzen des Zufalls gesteuerte Symbolquellen, sind in weiten Teilen der Mathematik und in den ihr interdisziplinär angeschlossenen Wissenschaften wichtige Gegenstände der Forschung. Ist man nicht in der Lage, den vermuteten deterministischen Generationsmechanismus einer Klasse von Symbolsequenzen (als evidente Beispiele wären hier z.B. die Aminosäureabfolgen von Proteinen oder die zeitliche Entwicklung börsennotierter Werte zu nennen) zu erkennen, so können meist nichtsdestotrotz Gesetzmäßigkeiten ausgemacht werden, die dem Reich des Zufalls zuzuordnen sind. Versteht man diese, so ist es immerhin möglich, vorhandene Sequenzen zu klassifizieren und Vorhersagen bezüglich neu entstehender Sequenzen zu treffen.

Eben die diskretwertigen stochastischen Prozesse sind nun Gegenstand dieser Arbeit. Sie präsentiert im Kern einen neuen Formalismus, mit ihnen umzugehen und zu rechnen. Das neue Kalkül stützt sich dabei auf eine einfache und elementare Erkenntnis. Ist Σ die (endliche) Menge von Symbolen, die zu erzeugen die Prozesse in der Lage sind, und ist Σ^* die Menge der "Wörter", die mit diesen Symbolen gebildet werden können, so betrachten wir den Raum der reellwertigen Funktionen über dieser Menge der Wörter

$$l_{\Sigma} = \mathbb{R}^{\Sigma^*} = \{g : \Sigma^* \longrightarrow \mathbb{R}\}.$$

Auf diesem lässt sich ein linearer Operator einführen, dessen Fixpunkte sich als stochastische Objekte entpuppen. Die stochastischen Prozesse mit Werten in Σ lassen sich nun gerade mit den positiven, normalisierten Vektoren dieses Untervektorraums der Fixpunkte identifizieren. Das Studium solcher und verwandter Untervektorräume und der darin enthaltenen stochastischen Prozesse ist der Inhalt dieser Arbeit.

1.1. Motivation und Zielsetzung

Ursprung und Ausgangspunkt war die Erforschung einer Klasse stochastischer Prozesse, die auf einem Anfang der 60er Jahre geschaffenen Dimensionsbegriff beruht und sich als Klasse der "endlich-dimensionalen" Prozesse vorgestellt hat. Der Dimensionsbegriff lässt sich mit dem hier vorgestellten Formalismus besonders klar erfassen und so sind die endlich-dimensionalen Prozesse ein zentraler Bestandteil dieser Arbeit. Dabei bezeichnen wir allgemein ein Element $g \in l_{\Sigma}$ als endlich-dimensional, falls der Spaltenraum (und damit auch der Zeilenraum) der Matrix

$$\mathcal{P}_g := (g(\bar{a}\bar{b}))_{\bar{a}, \bar{b} \in \Sigma^*} \in \mathbb{R}^{\Sigma^* \times \Sigma^*}$$

endlich-dimensional ist, wobei für zwei Wörter $\bar{a} = a_1 \dots a_t \in \Sigma^t$, $\bar{b} = b_1 \dots b_s \in \Sigma^s$ mit $\bar{a}\bar{b} = a_1 \dots a_t b_1 \dots b_s$ gerade die Verkettung der beiden Wörter gemeint sei. Endlich-dimensionale Prozesse sind dann gerade diejenigen, deren assoziierte Vektoren endlich-dimensional sind.

1.1.1. Motivation: Endlich-dimensionale Prozesse und Hidden-Markov-Prozesse. Das Interesse an den endlich-dimensionalen Prozessen hat gute Gründe. Zum einen erlauben sie endliche Parametrisierungen, können also mit Hilfe einer endlichen Menge von Kennzahlen komplett beschrieben werden, was für einen beliebigen stochastischen Prozess nicht klar, aber nun einmal Grundvoraussetzung für praktische Anwendungen ist. Zweitens beinhalten sie als (echte) Unterklasse die Hidden-Markov-Prozesse, eine Klasse von Prozessen, die sich aufgrund ihrer intuitiven Formulierung in weiten Anwenderkreisen sehr beliebt gemacht haben [Rab89], [Jel97], [EAM95], [MZ97].

Hidden-Markov-Prozesse sind, wahrscheinlich gerade wegen ihrer anschaulichen Formulierung, mathematisch betrachtet eher störrische Objekte. Ihre in der Terminologie der Mathematik unflexible Handhabung hat deshalb eine Reihe von Nachteilen zur Folge. Nützliche Erkenntnisse hinsichtlich einer Reihe von geläufigen Begriffen aus der sie umgebenden Welt der stochastischen Prozesse, der Ergodentheorie und der Informationstheorie bleiben verwehrt bzw. benötigen umständliche und komplizierte Beweise. Dieses Problem hat die Klasse der endlich-dimensionalen Prozesse nicht. Was an Anschaulichkeit verloren geht, wird durch mathematische Flexibilität aufgewogen. Diese Arbeit wird zeigen, dass etliche Erkenntnisse der Ergodentheorie mit einfachsten Mitteln erhalten werden können. Erkenntnisse, die so für Hidden-Markov-Prozesse bislang nicht bzw. nur mit ungleich viel mehr Aufwand erhältlich waren.

1.1.2. Motivation: Lernalgorithmen. Eben diese Problematik der Hidden Markov Modelle ist auch der Grund für die Schwierigkeiten bei der Konstruktion von geeigneten Lernalgorithmen. Algorithmen, die Hidden-Markov-Modelle aus einer Menge von ihnen erzeugter Sequenzen lernen (rekonstruieren), benötigen entweder massive Vorgaben wie die Kenntnis der Topologie [Bil98], [BP66], [Rab89], beschränken sich auf (meist wirklichkeitsfremde) Zusatzeigenschaften wie Stationarität [Sch02] oder sind extrem komplex und langsam [SO93]. Die mathematisch einfache Formulierung der endlich-dimensionalen Prozesse verspricht auch hier Abhilfe, wie die Arbeiten von Herbert Jäger und seiner Arbeitsgruppen zeigen [Jae00], [Obe02], [Kre03], [JZK⁺05]. Um sich von der allerdings auch dort benötigten Zusatzvoraussetzung der Stationarität lösen zu können, ist die theoretische Kenntnis nichtstationärer Vertreter endlich-dimensionaler Prozesse von Nöten.

1.1.3. Zielsetzung. Ziel dieser Arbeit ist es, endlich-dimensionale Prozesse einem gründlichen Studium bezüglich der fundamentalen Begriffe aus der Ergodentheorie zu unterziehen, wiederum mit dem Fernziel, Lernalgorithmen für die spezielle Klasse "hochgradig nicht-stationärer" (wobei dieser

Begriff einer genauen Definition noch entbehrte) endlich-dimensionaler Prozesse zu konstruieren. Da die Ergodentheorie eine Brücke schlägt zwischen den Kennzahlen eines stochastischen Prozesses einerseits und den Schätzwerten von Größen, die “entlang” von Sequenzen des Prozesses erhalten werden, sind diesbezügliche Ergebnisse für den Bau von fundierten Lernalgorithmen notwendig. Darüber hinaus beinhaltet eine Ergodentheorie endlich-dimensionaler Prozesse einfache Beweisvarianten für Fakten, die so für Hidden-Markov-Prozesse nicht oder nur kompliziert zu erhalten sind.

Dieses Studium nun kann mit Hilfe des oben angerissenen, neuen Modellansatzes effektiv betrieben werden. Etliche diesbezügliche Ergebnisse belegen den Nutzen des neuen Modells einerseits, und schaffen ein solides ergodentheoretisches Fundament für den Bau von Lernalgorithmen auf der anderen Seite. Darüberhinaus ergeben sich durch den neuen Blickwinkel auf die diskretwertigen stochastischen Prozesse etliche interessante Fragestellungen und auch Lösungsansätze für die sie umgebenden Theorien. Ein besonderes Beispiel hierfür mag das informationstheoretische Kapitel 6 über die Existenz von Entropieraten sein.

1.2. Geschichte und verwandte Arbeiten

Um einen Einblick in das Umfeld dieser Arbeit zu vermitteln und ihre Leistung besser einordnen zu können, wollen wir einen kurzen Abriss der Historie der ihr nahestehenden Konzepte geben.

1.2.1. Endlich-dimensionale Prozesse. Die Geschichte der endlich-dimensionalen stochastischen Prozesse ist kurz und schnell erzählt. Zwar ist der Begriff der Dimension eines stochastischen Prozesses nicht sonderlich neu, aber in der einschlägigen Literatur der stochastischen Prozesse oder verwandter Gebiete niemals ausführlich und explizit von Interesse gewesen. Als Vater des Begriffes kommt sicherlich Heller mit seiner Arbeit aus dem Jahre 1965 [Hel65] in Betracht. Die durch den Dimensionsbegriff geschaffene Klasse der endlich-dimensionalen Prozesse bekommt von ihm den Namen “finitary processes”.

Die Erforschung dieser Prozessklasse muss in einen direkten Zusammenhang mit der Erforschung des Problems der Identifikation von Hidden-Markov-Modellen, auf Englisch “identifiability problem”, gebracht werden. Die Fragestellung dieses Problems ist es, unter welchen Bedingungen sich zwei Hidden-Markov-Modelle als äquivalent erweisen, d.h. im Grunde denselben stochastischen Prozess kodieren. Eine Reihe von Arbeiten, denen auch Hellers Werk zuzuordnen ist, beschäftigte sich mit diesem Problem, das von Blackwell und Koopmans [BK57] aufgeworfen wurde. Gilbert [Gil59] zum Beispiel griff das Problem auf und erweiterte die nur für Spezialfälle angegebenen Lösungen aus [BK57] um einige allgemeinere Fälle. Ebenso warf er die Frage auf, wie beliebige stationäre Prozesse als Hidden-Markov-Modelle erkannt werden könnten. Diese Frage wurde in den Arbeiten von Dharmadhikari [Dha63a], [Dha63b], [Dha65] (beinahe erfolgreich) aufgegriffen. Heller schließlich löste dieses Problem [Hel65], wenn auch seine Lösung

einem Stochastiker formal kompliziert erscheinen muss (Heller ist homologischer Algebraiker). Darauf aufbauend haben Fox und Rubin in den Arbeiten [Gil59], [Dha63a], [Dha63b], [Dha65] zitierte Fragestellungen auf Alphabete nicht-endlicher Kardinalitäten verallgemeinert [FR68], [FR69], [FR70]. Auch Erickson [Eri70] hat auf die Schriften von Dharmadhikari sowie Fox und Rubin aufbauende Resultate erzielt, diesmal Verallgemeinerungen der Annahme diskreter Zeitschritte. Eine Zusammenfassung der Arbeiten kann auch in [Ros71] nachgelesen werden. In all diesen Arbeiten wird der Begriff der endlich-dimensionalen Prozesse bereits, wenn nicht ausdrücklich, so doch implizit verwendet. Rund 35 Jahre später schließlich haben Ito, Amari und Kobayashi das Identifikationsproblem in voller Allgemeinheit gelöst [IAK92]. Sie konstruieren einen Algorithmus zur Überprüfung der Äquivalenz zweier endlich-dimensionaler Prozesse, und damit auch der Äquivalenz zweier Hidden-Markov-Modelle, die sich in Hellers Schrift als echte Unterklasse der “finitary processes” erwiesen haben. Die Präsentation des Problems und seiner Lösung wirkt jedoch undurchsichtig, und tatsächlich ist eine saubere Definition des involvierten Dimensionsbegriffs nicht zu erkennen. Parallel zu dieser Arbeit hat jedoch Balasubramanian [Bal93] ebenfalls eine Lösung des Problems zusammen mit einer präziseren Definition des Dimensionsbegriffs und einem polynomiellen Algorithmus (im Gegensatz zur exponentiellen Variante in [IAK92]) vorgestellt.

Im Rahmen dieser letzten beiden Arbeiten werden die endlich-dimensionalen Prozesse unter dem Namen “Linearly Dependent Processes” (LDPs) geführt, ein deutlicher Hinweis darauf, dass bei der Formulierung dieser Prozessklasse linear algebraische Begriffe eine Rolle spielen. Es ist zu bemerken, dass die Hellersche Charakterisierung der Hidden-Markov-Modelle (HMMs) innerhalb der involvierten Vektorräume sehr unpraktikabel ist. Dies mag die Probleme der HMMs bezüglich mathematischer Analysen erklären helfen.

Nach Erfüllung dieser ersten Schuldigkeit wurden die LDPs für weitere vier Jahre zu den Akten gelegt. 1997 sind sie unabhängig voneinander in zwei Arbeitsgruppen zu neuem Leben erweckt worden, einmal unter dem Namen “Generalized Hidden Markov Models” (verallgemeinertes Hidden-Markov-Modell) (GHMMs) [Upp97] im Umfang einer Doktorarbeit an der Universität Berkeley, und - ungleich viel ausführlicher - als “Observable Operator Models” (OOMs) [Jae97a], [Jae97b] in Arbeiten von Herbert Jäger in Bonn und Bremen. Die Arbeiten beider Gruppen hatten das Bestreben, neue Lernalgorithmen für endlichwertige, diskretezeitige stochastische Prozesse zu entwickeln. Wie bereits angedeutet, bezieht auch dieses Problem seine Motivation aus einem Problemkreis der Hidden-Markov-Modelle. (siehe den Unterabschnitt 1.1.2). Hier versprechen die als OOMs formulierten endlich-dimensionalen Prozesse Abhilfe; in den Jägerschen Arbeitsgruppen ist eine Klasse von Algorithmen entstanden, die bessere Ergebnisse hervorbringen und weniger Vorgaben benötigen als die in den HMM-Gemeinden üblicherweise zur Anwendung kommenden. Siehe hierzu die Arbeiten [Jae00], [Obe02], [Kre03] und [JZK⁺05].

Als letzter Beitrag zur Theorie des Dimensionsbegriffes bzw. des linear algebraischen Rahmenwerks sind noch die Ergebnisse der Arbeitsgruppe Singh [JS04], [LSS01] zu erwähnen. Hier wird der Begriff der “predictive states” betont, der auch in unserem Modell auf einfache Art und Weise wiederentdeckt werden kann, um optimale Entscheidungen autonomer Agenten treffen zu können.

1.2.2. Hidden-Markov-Modelle. Die Motivation aller bisherigen Bestrebungen der Erforschung endlich-dimensionaler Prozesse war, gewissen Fehlerhaftigkeiten der Hidden-Markov-Modelle beizukommen bzw. Aussagen über diese zu erzielen. Hidden-Markov-Modelle [Rab89], [Jel97], [EAM95], [MZ97], die Motivation aller sich auf endlich-dimensionalen Prozesse beziehenden Arbeiten, sind eine Klasse statistischer Modelle, die sich in weiten Bereichen und vielzähligen Anwendungen bewährt haben, von denen zunächst einmal die Spracherkennung [Rab89],[Jel97] zu zitieren ist. Die Biologie kann inzwischen als zweites großes Betätigungsfeld der HMMs genannt werden. Zu nennen sind die Analyse von Genexpressionszeitreihen, ein Gebiet, in dem sich der Autor selbst gut auskennt [SSS03], [SSS04], [SCSS05], aber auch und vor allem die Proteinklassifikation [DEKM98a], [PSS⁺02] und das Auffinden von Genen [RKTH00], [BK97]. Die Modellierung von Finanzzeitreihen [Wic01], [Kna00], [KSSW02] soll ebenfalls erwähnt werden. Darüberhinaus existiert eine große Anzahl unveröffentlichter Anwendungen in der Industrie.

HMMs kommen berechtigterweise dann zum Tragen, wenn davon ausgegangen werden darf, gewisse Sequenzdatenmuster ließen sich mit Hilfe von Grundzuständen erklären, die das zu modellierende System durchläuft (siehe Abbildung 1). Diese Sicht der Dinge ist intuitiv, und vor allem deshalb sind die HMMs einer breiten Masse von Anwendern zugänglich. Der Vorteil von variabel zu durchlaufenden (für den Anwender nicht sichtbaren, daher “versteckten”) Grundzuständen ist, dass solcherlei Modelle Datenmuster vor allem qualitativ, bis zu einem gewissen Grade zeitunabhängig, beschreiben. Dies ist ganz offensichtlich von Vorteil, handelt es sich bei den Daten z.B. um Sprachsignale - Wörter oder auch nur Silben können hastig oder gedehnt und überbetont gesprochen werden.

Die Prominenz der HMMs erklärt sich weiter mit der Existenz effizienter Algorithmen für die Lösung zentraler Probleme, die im Zusammenhang mit der Formulierung der HMMs aufkommen (die sogenannten “drei fundamentalen Probleme” der HMMs, [BP66], [BPSW70], [Bil98], [Rab89], [Jel97], [EAM95]). Ihre intuitive Formulierung jedoch scheint, wie bereits mehrfach angedeutet, mathematische Schwierigkeiten zu bereiten. Siehe [EM02] für eine äußerst gründliche und umfassende Übersicht theoretischer Erkenntnisse.

Algorithmen, die in gewissem Sinne hochgradig nicht-stationäre stochastische Prozesse aus einer meist größeren Menge daraus generierter Sequenzen relativ kurzer Länge (im Bereich 10-1000 Symbole) möglichst ohne Vorgaben lernen, sind eine Aufgabe, die der HMM-Gemeinde traditionell schwer fällt

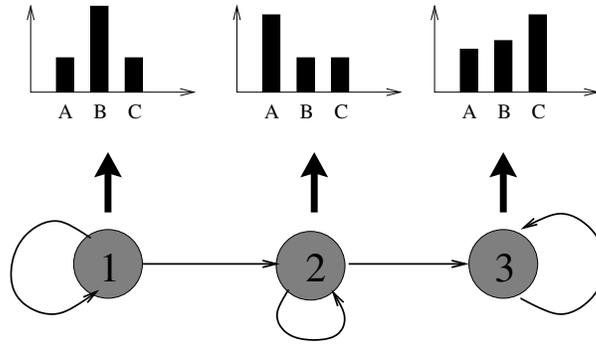


ABBILDUNG 1. Ein sogenanntes “Links-Rechts” Hidden-Markov-Modell mit drei versteckten Zuständen über dem Alphabet $\{A, B, C\}$.

(siehe Unterabschnitt 1.1.2). Kann die Topologie jedoch als bekannt angenommen werden (eine massive Vorgabe), so leistet z.B. der Baum-Welch-Variante des EM-Algorithmus gute Dienste [BP66], [BPSW70], [Bi98]. Wir verweisen wiederum auf die Spracherkennung und unter dem Stichwort “Profile HMM’s” [DEKM98a] auch auf die Proteinklassifizierung, wo der Einsatz nichtstationärer HMM’s inzwischen den Goldstandard darstellen.

1.2.3. Ergodentheorie und Dynamische Systeme. Ein letzter, stark mit dieser Arbeit verbandelter Themenkreis ist die Ergodentheorie dynamischer Systeme. Im weitesten Sinne beschäftigt sich die Ergodentheorie mit durch Gruppen oder Halbgruppen induzierten messbaren Transformationen von Maßräumen. Im allgemeinen Fall agiert also eine (Halb-)Gruppe (\mathcal{T}, \circ) von messbaren Transformationen $T \in \mathcal{T}$ auf einem Maßraum (Ω, \mathcal{B}, m) (d.h. Ω eine Menge mit einer σ -Algebra \mathcal{B} , einem Maß $m : \mathcal{B} \rightarrow \mathbb{R}^+$ und $\mathcal{T} \ni T : \Omega \rightarrow \Omega$ Abbildungen, so dass für $B \in \mathcal{B} : T^{-1}(B) \in \mathcal{B}$, mit \circ ist die Verknüpfung von Abbildungen gemeint) und die Ergodentheorie studiert die Effekte solcher Aktionen. Als ihre Väter können von Neumann, Birkhoff und Koopman genannt werden, als Entstehungszeitraum die 30er Jahre des zwanzigsten Jahrhunderts. Ihre ursprüngliche Motivation ist in einer Arbeit von Boltzmann zu finden; er beschäftigte sich mit Problemen aus der statistischen Mechanik, bei denen Erwartungswerte mit zeitlichen Mitteln übereinstimmen.

Dynamische Systeme sind Maßräume (Ω, \mathcal{B}, m) zusammen mit einer messbaren Transformation $T : \Omega \rightarrow \Omega$. Untersucht man die Effekte der Halbgruppe von Transformationen $(T^k : \Omega \rightarrow \Omega \mid k \in \mathbb{N})$, so betreibt man Ergodentheorie. Dynamische Systeme sind daher unweigerlich in einen Zusammenhang mit der Ergodentheorie zu stellen. Diskretzeitige stochastische Prozesse mit Werten in einem endlichen Symbolvorrat Σ können auf natürliche Art und Weise mit dynamischen Systemen assoziiert werden (siehe die Kapitel 2 und 4 für eine gründliche Aufklärung der Zusammenhänge). In der Sprache der stochastischen Prozesse beschäftigt sich die Ergodentheorie

dann mit stochastischen Größen, die “entlang von Sequenzen” gemessen werden.

Die Ergodentheorie ist inzwischen ein weites Feld. Etliche Arbeitsgruppen (siehe z.B. [IfMS]) und zum Teil dicke Bücher [Bil65], [PY98], [Sin94], [Kre85], [Gra01] (um nur einige zu nennen) zeugen davon. Bearbeitet werden sehr anwendungsnahe Probleme bis hin zu sehr abstrakten Fragestellungen. Zu unserem Zweck reichten einige (zentrale) Begriffe aus der Theorie der dynamischen Systeme, und als dieser Arbeit sehr nahe Referenzen können [Shi96], [PY98] und [Gra01] genannt werden. Einzig der etwas involviertere Beweis von Satz 4.41 benötigt einige Begriffe mehr (die meist [Kre85] entnommen worden sind) und wir haben ihn daher in den Anhang (siehe Appendix B) verschoben.

1.3. Kapitelübersicht

Die Arbeit ist in zwei Teile gegliedert. Das Studium der oben erklärten Vektorräume zeigte schon bald, dass gewisse Teildisziplinen der linearen Algebra von großem Nutzen sein würden. So war die Arbeit phasenweise eng an entsprechende Ergebnisse gekoppelt, die meist der Stabilitätstheorie, wie sie von Brayton und Tong [BT79] dargestellt wird, zuzuschreiben sind. Dementsprechend enthält der erste und, für sich genommen, deutlich gewichtigere Teil die eigentliche, oben versprochene Theorie der stochastischen Vektorräume. Der zweite Teil mag, obwohl kürzer und theoretisch eindeutig dem überaus klassischen Gebiet der linearen Algebra zuzuordnen, insofern von Interesse sein, als dass seine Ergebnisse in dieser Zusammenstellung in den klassischen Büchern nicht anzutreffen sind, oder ganz verschwiegen werden. Hat der Autor an dieser Stelle Räder neu erfunden, so fährt das auf ihnen aufsitzende Vehikel mit den neu entworfenen Varianten deutlich besser.

1.3.1. Erster Teil: Stochastische Vektorräume. Dieser Hauptteil der Arbeit enthält eine vollständige Beschreibung des neuen Modells und der darauf basierenden Ergebnisse. Im Einzelnen stellen sich die Kapitel dieses Teils wie folgt dar:

Kapitel 2 enthält die grundlegenden Definitionen unseres Modells zusammen mit einer parallel entwickelten maßtheoretischen Interpretation.

Wir präsentieren daher in den ersten beiden Abschnitten die fundamentale Definition des stochastischen Operators und darauf aufbauend die Vektorräume, die Gegenstand unserer Betrachtungen sind. Wir tun dies Hand in Hand mit den Begriffen der Maßtheorie, die in unserem Modell wiedererkannt werden können.

Im dritten Abschnitt führen wir eine Norm ein mit der Folge, dass wir die Elemente der dadurch induzierten Untervektorräume unter der zusätzlichen Hilfe des dort ebenfalls entwickelten Zerlegungssatzes mit bekannten Objekten der Maßtheorie identifizieren können. Der durch die Norm induzierte Konvergenzbegriff wird einer gründlichen Prüfung unterzogen und mit den einschlägigen Konvergenzformen der Maßtheorie in Abgleich gebracht werden.

Das Kapitel schließt mit einem Abschnitt, der den Begriff des Ereignisvektors ins Spiel bringt und mit der Formulierung eines Analogons des klassischen Approximationssatzes endet, ein wichtiger Baustein für spätere Arbeiten.

In Kapitel 3 führen wir zunächst den wesentlichen Begriff der Prognoseoperatoren ein und kümmern uns um deren elementare Eigenschaften.

Im zweiten Abschnitt sehen wir die Folgen dieser Definition: Neben der in dieser Arbeit üblichen wahrscheinlichkeitstheoretischen Interpretation sind dies vor allem zwei Dimensionsbegriffe (deren einer mit dem in der Einleitung bereits vielfach angesprochenen identifiziert werden kann) und die dadurch entstehenden Klassen stochastischer Prozesse endlicher Dimension. Zur Illustration dieser Begriffe rollen wir eine aus dem Themenkreis des Identifikationsproblems stammende Fragestellung der 60er Jahre erneut auf und erhalten so eine deutlich elementarere Lösung derselben. Wir bieten darüber hinaus eine konzeptionell verblüffend einfache, wenn auch ineffiziente Lösung des Identifikationsproblems selbst an.

Das Kapitel schließt mit der Untersuchung von Jäger zwecks Bau von Lernalgorithmen ins Leben gerufener sogenannter charakteristischer Basen. Wir verallgemeinern den Satz von Jäger auf nicht notwendigerweise positive endlich-dimensionale Vektoren und erhalten durch einen trickreichen und technischen Beweis ein abgeschwächtes Resultat für den Fall beliebiger (unendlich-dimensionaler) Vektoren.

Kapitel 4 schließlich kann als das Herzstück dieser Arbeit bezeichnet werden. Hier werden, in Abgleich mit ergodentheoretischen Objekten, Begriffe eingeführt, die sich mit der “Evolution”, d.h. mit den Auswirkungen der Aktionen der im vorigen Kapitel eingeführten Prognoseoperatoren und insbesondere des Evolutionsoperators auf unseren Vektorräumen beschäftigt.

Zunächst werden Fixpunkte des Evolutionsoperators (die als stationäre Vektoren bezeichnet werden) untersucht und in einen Zusammenhang mit den sie umgebenden Begriffen unseres Modells gestellt. Das Zusammenspiel von stationären Vektoren und “invarianten” Ereignissen wird untersucht mit dem Ergebnis zweier wichtiger und vielzitiertem Lemmata. Die arbeitsintensivste Leistung dieses Abschnitts ist jedoch eindeutig das Beispiel eines unendlich-dimensionalen, stationären, stochastischen Vektors, dessen Existenz lange bezweifelt wurde.

Im Anschluss wird der Begriff der Evolutionssummierbarkeit in den Mittelpunkt gerückt. Dieser Begriff erweist sich, auf stochastische, d.h. Wahrscheinlichkeitsmaße kodierende Vektoren angewendet, als äquivalent zum Begriff der asymptotischen Stationarität im Mittel. Der Beweis dieses Zusammenhangs jedoch hat mit unserem Modell wenig zu tun und erfordert den Einsatz etlicher zusätzlicher Begriffe der Ergodentheorie. Wir haben ihn daher in den Anhang (Teil B) verschoben. Nichtsdestotrotz ist dieser erstaunliche Satz eine umfassende Rechtfertigung dieser Arbeit, da sich mit seiner Hilfe weite Teile der Theorien wie sie z.B. in [Gra01],[Gra90] entwickelt werden, in unserem Modell abbilden lassen. Im Übrigen erweisen sich Vektoren von endlicher Evolutionsdimension immer als evolutionssummierbar, ein

Resultat, das die Ergebnisse des zweiten Teils der Arbeit wesentlich nutzt. Das nicht-triviale Beispiel eines evolutionssummierbaren Prozesses von unendlicher Evolutionsdimension beschließt diesen Abschnitt.

Es folgt ein Abschnitt, der den in der Ergodentheorie zentralen Begriff der Ergodizität in unser Modell einführt. Ergodizität lässt sich für evolutionssummierbare Vektoren mit Hilfe unserer Begriffe äußerst prägnant ausdrücken, ein wichtiges Ergebnis dieser Arbeit.

Als eine der Hauptarbeiten dieser Dissertation erklären wir in Kapitel 5 schließlich eine ergodische Zerlegung für endlich-dimensionale stochastische Vektoren. Wir erarbeiten sie zunächst für stationäre Vektoren und verallgemeinern sie auf alle endlich-dimensionalen Vektoren. Sie baut auf den geleisteten Vorarbeiten auf und hat eine anschauliche geometrische Interpretation. Wendet man die ergodische Zerlegung auf die in Kapitel 7 entwickelten graphischen Darstellungen an, so erhält man eine Aussage über deren Zusammenhangskomponenten. Der letzte Abschnitt dieses Kapitels mag der Ausgangspunkt für eine "Markov"-Theorie klassifizierender Natur für endlich-dimensionale stochastische Prozesse sein.

Kapitel 6 wiederum enthält eine Anwendung aus der Informationstheorie und dreht sich im Wesentlichen um den Begriff der Entropierate eines stochastischen Prozesses. Im Falle stationärer Prozesse belegt man deren Existenz mit Hilfe eines kurzen und elementaren Beweises. Im Falle im Mittel asymptotisch stationärer Prozesse kann deren Existenz zwar als Korollar des ebenso legendären wie aufwändigen, in vielzähligen Varianten erhältlichen Satzes von Shannon-McMillan-Breiman aufgezeigt werden, eine elementare Variante existierte bislang jedoch nicht. Wir schließen diese Lücke mit diesem Kapitel.

Kapitel 7 und 8 beschäftigten sich mit zwei neuen Varianten der Darstellung für endlich-dimensionale stochastische Prozesse, eine graphischer Natur, die andere mit der Motivation, statistische Phänomene aus der Quantenmechanik abbilden zu können.

Kapitel 7 befasst sich mit den graphischen Modellen, Prognosegraphen genannt. Wir beweisen, dass die durch Prognosegraphen induzierten stochastischen Prozesse gerade den endlich-dimensionalen Prozessen entsprechen. Darüberhinaus setzen wir die Anzahl der irreduziblen Komponenten eines Prognosegraphen in Beziehung zur Anzahl der ergodischen Komponenten des induzierten stochastischen Prozesses.

In Kapitel 8 definieren wir die quantenmechanisch inspirierten Quanten-Prognose-Modelle. Wir zeigen wiederum, dass die dazu assoziierten statistischen Phänomene gerade den endlich-dimensionalen Prozessen entsprechen. Für die wichtige Unterklasse der Quanten-Markov-Modelle erzielen wir eine Verschärfung des für endlich-dimensionale Prozesse erhältlichen Ergodensatzes. Die Modellierung von Quanten-Irrfahrten auf Graphen als Quanten-Markov-Modelle bringt daher eine Verschärfung der dort bislang erhaltenen Resultate mit sich.

Kapitel 9 schließlich enthält eine Zusammenfassung der in diesem Hauptteil der Arbeit präsentierten Ergebnisse sowie einen Ausblick auf eine Menge interessanter Folgearbeiten.

1.3.2. Zweiter Teil: Stabilitätstheorie linearer Operatoren und verwandte Konzepte. Wie oben bereits erwähnt, sind die meisten in diesem Teil niedergeschriebenen Ergebnisse der linearen Algebra zuzuordnen. Sie sind höchstwahrscheinlich ausschließlich in der hier gelieferten Komposition und Formulierung als "neu" zu empfinden. Nichtsdestotrotz sind dem Autor Quellen für die z.B. in Kapitel 11 aufgezeigten Zusammenhänge unbekannt.

Kapitel 10 kümmert sich vor allem um die Übertragung der komplexen Stabilitätstheorie von Brayton und Tong auf den Fall reeller Vektorräume, eine für das Kriterium der invarianten Umgebungen nicht-triviale Aufgabe.

Im sich anschließenden Abschnitt bereichern wir sowohl die komplexe als auch die reelle Theorie um das Kriterium der Summierbarkeit. Der Zusammenhang zwischen Summierbarkeit und Eigenwertkriterium für den komplexen Fall ist jedoch bereits bekannt. Neu ist hier auf jeden Fall die vollständige Auflistung aller Kriterien in einem Atemzug.

In Kapitel 11 bringen wir die Begriffe des Folgenrangs und der Folgenstabilität ins Spiel. Wir stellen fest, dass der Folgenrang eines linearen Operators an seinem Minimalpolynom abgelesen werden kann. Dementsprechend charakterisieren wir die Folgenstabilität.

Kapitel 12 kümmert sich um weitere äquivalente Formulierungen von Stabilität für Klassen von linearen Operatoren, die spezielle Klassen von affinen Hyperebenen invariant lassen. Die Charakterisierung so beschaffener stabiler Operatoren gelingt dann unter Einsatz des Begriffs eines Kegels, eine geometrische Anschauung, die sich in der Theorie der stochastischen Vektorräume, wie z.B. bei der Entwicklung der ergodischen Zerlegung (Abschnitt 5) auszahlt.

Teil 1

Stochastische Vektorräume

Stochastische Vektorräume - Grundbegriffe

Dieses einleitende Kapitel des Hauptteils unserer Arbeit stellt die grundlegenden Definitionen und Bezeichnungen unseres Modells zur Verfügung. Das Modell besteht im Wesentlichen aus Teilräumen des Vektorraums der beschränkten Folgen über der Menge der abzählbar vielen Wörter mit Buchstaben aus einem endlichen Alphabet (die mit der Menge der beschränkten Folgen über den reellen Zahlen identifiziert werden kann), Normen und Operatoren, die auf diesen Teilräumen agieren. Es kann als solches frei von jeglicher Interpretation behandelt werden und ist daher auch weitestgehend ohne Begriffe anderer Disziplinen verständlich.

Nichtsdestotrotz ist die Grundmotivation der Modellbildung das Erzielen von Ergebnissen, die der Stochastik und/oder verwandten Disziplinen zuzuordnen sind. Das Modell hat demnach eine **wahrscheinlichkeitstheoretische Interpretation** und, um dem Leser die Inspiration und das zugrundeliegende, strukturelle Denkmuster dieses Textes nicht zu verheimlichen, wollen wir diese Interpretation in einem Atemzug mit aufführen. Wir werden also in der Folge unsere Begriffe vorstellen und im Anschluss meist deren Interpretation im Sinne der Maß- und Wahrscheinlichkeitstheorie.

Da dies wiederum ein Sichvertrautmachen mit maßtheoretischen Konzepten erfordert (wir führen die grundlegenden Begriffe in der Folge auf) und dies dem Leser lästig erscheinen mag, geben wir nun einen Fahrplan zur **interpretationsfreien Lektüre** an. Wer also auf (seines Empfindens überflüssige) Wahrscheinlichkeitstheorie verzichten mag, der starte mit der vollständigen Lektüre von Abschnitt 2.1 (die Definition einer σ -Algebra bleibt einem nicht erspart) und fahre zunächst mit Unterabschnitt 2.2.1 fort, der die ersten Definitionen unseres Modells enthält. Es reicht dann, sich an die Definition eines endlichen, signierten Maßes zu erinnern (siehe Definition 2.12 in Unterabschnitt 2.2.2). Die Unterabschnitte 2.2.3 sowie 2.2.4 können daraufhin vollständig übersprungen werden. Die Definitionen des Schiebeoperators 2.29/2.30 in Unterabschnitt 2.2.5 jedoch sollten gelesen werden.

In Abschnitt 2.3 sollten daraufhin die Unterabschnitte 2.3.1 und 2.3.2 vollständig gelesen werden. In Unterabschnitt 2.3.3 genügt es, den zentralen Zerlegungssatz 2.51 zu studieren, dem sich das wichtige Korollar 2.52 sowie Definition 2.54 anschließen. Daraufhin ist Definition 2.56 maßgeblich für das weitere Verständnis des Texts. Unterabschnitt 2.3.4 kann daraufhin wiederum übersprungen werden. Aus dem Abschnitt über Ereignisvektoren 2.4 sollte man sich den Approximationssatz für Vektoren (Unterabschnitt 2.4.2, Satz 2.71) behalten - für das Verständnis desselben sind die Definitionen des

einführenden Unterabschnitts 2.4.1 erforderlich.

Bei uns bezeichnet \mathbb{N} , falls nicht anders geschrieben, die Menge der natürlichen Zahlen einschließlich der Null. \mathbb{R} ist die Menge der reellen Zahlen und \mathbb{R}^+ die Menge der positiven Zahlen einschließlich der Null (im Falle der Schreibweise $\epsilon \in \mathbb{R}^+$ meinen wir jedoch nur reelle Zahlen größer Null). Ist A eine Menge, so bezeichnen wir mit $\mathcal{P}(A)$ oder auch 2^A ihre Potenzmenge und mit $\text{card}(A)$ oder auch, wenn das zu keinen Missverständnissen führt, einfach mit $|A|$ ihre Kardinalität. Wir benutzen die in den meisten Büchern üblichen Bezeichnungen für Mengenoperationen, z.B. $\complement A = \Omega \setminus A$ für A Teilmenge einer Grundmenge Ω und $A \triangle B = (A \setminus B) \cup (B \setminus A)$, die symmetrische Differenz zweier Teilmengen A, B einer gemeinsamen Grundmenge Ω .

2.1. Wörter, Zylindermengen und (σ) -Algebren

2.1.1. Buchstaben, Wörter und Sequenzen. In dieser Arbeit sei Σ immer eine endliche Menge, d.h. $\text{card}(\Sigma) < \infty$. Diese Menge kann in unserer Arbeit als Symbol- oder Buchstabenvorrat interpretiert werden und wird daher **Alphabet** genannt werden. Ihre Elemente bezeichnen wir dementsprechend als **Buchstaben** oder **Symbole**. Wir schreiben weiter

$$\Sigma^t = \underbrace{\Sigma \times \dots \times \Sigma}_{t\text{-mal}} = \{(a_1, \dots, a_t) \mid a_i \in \Sigma \forall i = 1, \dots, t\}$$

und bezeichnen dies als die Menge der **t -Wörter**. Wir schreiben meist

$$a_1 \dots a_t := (a_1, \dots, a_t) \in \Sigma^t$$

für ein t -Wort. Darauf aufbauend seien

$$\begin{aligned} \Sigma^+ &:= \bigcup_{t \geq 1} \Sigma^t \\ \Sigma^* &:= \bigcup_{t \geq 0} \Sigma^t, \end{aligned}$$

wobei

$$\Sigma^0 := \{\square\}$$

und \square als **leeres Wort** (über Σ) aufgefasst werden kann. Die beiden obigen Mengen können als die Mengen aller **Wörter** über Σ (einmal ohne, einmal mit leerem Wort) aufgefasst werden. Im Allgemeinen schreiben wir Wörter beliebiger Länge als \bar{a} ($\in \Sigma^*$). Wir schreiben

$$|\bar{a}| := t,$$

falls $\bar{a} \in \Sigma^t$. Wir bemerken an dieser Stelle, da wir Wörter meist als von einer Zufallsquelle in zeitlicher Abfolge ausgespuckten Sequenz auffassen, dass wir Wortlängen in Analogie zu einer zeitlichen Entwicklung meist mit kleinen Buchstaben aus dem den Buchstaben t umgebenden Buchstaben (s, t, u) bezeichnen. Für eine Teilmenge $A \subset \Sigma^*$ von Wörtern schreiben wir

$$|A| := \sup_{\bar{a} \in A} |\bar{a}|,$$

also z.B. $|A| = t$ für $A \subset \Sigma^t$. Ist $a \in \Sigma$, so schreiben wir

$$a^t := \underbrace{a \cdot \dots \cdot a}_{t\text{-mal}}$$

für das Wort der Länge t , das nur aus a 's besteht.

Wir bezeichnen im weiteren Verlauf des Texts (mit Ausnahme der aus der Maßtheorie stammenden Definitionen) mit Ω_Σ die Menge aller Folgen über Σ , also

$$\Omega_\Sigma := \Sigma^\mathbb{N}.$$

Wir sprechen von Ω_Σ auch von der Menge der **Sequenzen** über Σ . Eine einzelne Sequenz bezeichnen wir meist mit ω ($\in \Omega_\Sigma$). Wir schreiben dann

$$\omega_1 \dots \omega_n \dots := \omega(0) \dots \omega(n-1) \dots \in \Sigma^\mathbb{N},$$

wobei die Umindizierung zu keinen Verwirrungen führen sollte.

BEMERKUNG 2.1. Σ^* ist abzählbar, jedoch ist Ω_Σ überabzählbar ($\text{card}(\Omega_\Sigma) = \text{card}(\mathbb{R})$).

Für zwei Wörter $\bar{a} = a_1 \dots a_t \in \Sigma^t$, $\bar{b} = b_1 \dots b_s \in \Sigma^s$ schreiben wir

$$\bar{a} \leq \bar{b} \iff t \leq s \wedge a_1 \dots a_t = b_1 \dots b_t,$$

mit anderen Worten, falls \bar{b} eine Fortsetzung von \bar{a} ist oder, andersherum gesagt, \bar{a} ein Präfix von \bar{b} . Analog meinen wir mit $\bar{a} \leq \omega$ für ein Wort $\bar{a} = a_1 \dots a_t$ und eine Sequenz $\omega = \omega_1 \dots \omega_n \dots$, dass \bar{a} ein Präfix von ω ist, also $\bar{a} = \omega_1 \dots \omega_n$ gilt. Sind $s, t \in \mathbb{N}$, so dass $s \geq t$, und ist $\bar{a} \in \Sigma^t$ ein Wort der Länge t , so schreiben wir auch

$$\Sigma_{\bar{a}}^s := \{\bar{b} \in \Sigma^s \mid \bar{a} \leq \bar{b}\}$$

für die Menge der s -Wörter, die mit \bar{a} anfangen, wobei formal $\Sigma_{\bar{a}}^s := \Sigma^s$ und analog

$$\Sigma_{\bar{a}}^* := \{\bar{b} \in \Sigma^* \mid \bar{a} \leq \bar{b}\}$$

für die Menge aller Wörter, die mit \bar{a} anfangen, wobei wiederum formal $\Sigma_{\bar{a}}^* := \Sigma^*$ gesetzt sei. Ebenso seien

$$\Omega_{\bar{a}} := \{\omega \in \Omega_\Sigma \mid \bar{a} \leq \omega\}$$

und natürlich wieder $\Omega_{\square} := \Omega_\Sigma$.

2.1.2. Zylindermengen und (σ -)Algebren. Wir führen im Folgenden Begriffe ein, die als ‘‘Folklore’’ der Maß- und Wahrscheinlichkeitstheorie gelten und in praktisch allen Büchern darüber wiedergefunden werden können, so z.B. [Bau91], [Gra01], [Doo53], [Shi96]. Wir passen sie hiermit an unseren Sprachrahmen an.

Wir starten mit den sogenannten Zylindermengen, die mit Teilmengen von Wörtern identifiziert werden können und später eine große Rolle spielen.

DEFINITION 2.2. *Wir nennen eine Teilmenge von Sequenzen*

$$M \subset \Omega_\Sigma$$

eine **Zylindermenge**, falls

$$\exists t \in \mathbb{N}, \exists A \subset \Sigma^t : M = \bigcup_{\bar{a} \in A} \Omega_{\bar{a}}.$$

Ein solches A wird auch als **definierende Grundmenge (von Wörtern)** der Zylindermenge bezeichnet und wir schreiben $Z(A) \subset \Omega_\Sigma$ für die durch $A \subset \Sigma^n$ induzierte Zylindermenge.

Die definierenden Grundmengen von Zylindermengen sind keineswegs eindeutig. Tatsächlich gilt:

PROPOSITION 2.3. *Seien $n, m \in \mathbb{N}, n \leq m$ und $A \subset \Sigma^n, B \subset \Sigma^m$ zwei Mengen von Wörtern. Dann*

$$Z(A) = Z(B) \iff B = \bigcup_{\bar{a} \in A} \Sigma_{\bar{a}}^m.$$

BEWEIS. Das ist offensichtlich. \square

Wir setzen nun

$$\mathcal{F}_\Sigma := \{A \in \mathcal{P}(\Omega_\Sigma) \mid A \text{ Zylindermenge}\},$$

ein System von Teilmengen von Ω_Σ . Dieses System hat gewisse Eigenschaften, die als die definierenden Eigenschaften einer *Algebra* erkannt werden können, ein Begriff aus der Maßtheorie. Wir führen daher an dieser Stelle die Begriffe eines Rings und einer (σ) -Algebra ein.

DEFINITION 2.4. *Ist Ω eine Menge, so heißt ein System von Teilmengen $\mathcal{F} \subset \mathcal{P}(\Omega)$ ein **Ring**, falls*

$$\emptyset \in \mathcal{F}; \tag{2.1}$$

$$A, B \in \mathcal{F} \implies A \setminus B \in \mathcal{F}; \tag{2.2}$$

$$A, B \in \mathcal{F} \implies A \cup B \in \mathcal{F}. \tag{2.3}$$

Gilt zusätzlich

$$\Omega \in \mathcal{F}, \tag{2.4}$$

so sprechen wir von einer **Algebra** (in Ω). Ein System von Teilmengen \mathcal{B} nennen wir **σ -Algebra** (in Ω), falls

$$\Omega \in \mathcal{B}; \tag{2.5}$$

$$A \in \mathcal{B} \implies \complement A \in \mathcal{B}. \tag{2.6}$$

Ist weiter $(A_n)_{n \in \mathbb{N}}$ eine Folge von Teilmengen $A_n \in \mathcal{P}(\Omega)$, so

$$\forall n \in \mathbb{N} : A_n \in \mathcal{B} \implies \bigcup_{n \in \mathbb{N}} A_n \in \mathcal{B}. \tag{2.7}$$

Ein Tupel (Ω, \mathcal{B}) , bestehend aus einer Menge Ω und einer σ -Algebra \mathcal{B} wird **Messraum** genannt. Elemente von \mathcal{B} werden dann **messbare Mengen** genannt.

Zunächst wollen wir bemerken, dass der Unterschied zwischen einer Algebra und einer σ -Algebra gar nicht so kompliziert ist, wie die Definition zunächst vermuten lässt.

BEMERKUNG 2.5. Die Eigenschaften 2.1, 2.2, 2.3 und 2.4 sind äquivalent zu den Eigenschaften 2.5, 2.6 und 2.3. Der Unterschied zwischen einer Algebra und einer σ -Algebra ist also gerade die Verschärfung der Forderung der Zugehörigkeit von endlichen Vereinigungen auf die Zugehörigkeit von abzählbaren Vereinigungen.

BEWEIS. Das folgt mit Hilfe der Gleichung

$$A \setminus B = A \cap \mathbb{C}B = \mathbb{C}(A \cup B).$$

□

Wir stellen nun ganz offiziell fest:

PROPOSITION 2.6. Das System der Zylindermengen \mathcal{F}_Σ ist eine Algebra in Ω_Σ .

Das ist allgemein bekannt (siehe z.B. [Gra01],[Shi96]). Beweise dafür werden im Allgemeinen nicht einmal mehr angegeben, so auch nicht in [Gra01]. Deshalb hier den an unsere Sprache angepassten

BEWEIS. Nimmt man die leere Menge als definierende Grundmenge, so folgt $\emptyset \in \mathcal{F}_\Sigma$. Weiter seien $A \in \Sigma^s$ bzw. $B \in \Sigma^t$ die definierenden Grundmengen zweier Zylindermengen M bzw. N . Sei $u = \max\{s, t\}$. Dann lassen sich A und B als Teilmengen von Σ^u auffassen (ersetze A, B durch $\cup_{\bar{a} \in A} \Sigma_{\bar{a}}^u, \cup_{\bar{a} \in B} \Sigma_{\bar{a}}^u$). In diesem Sinne sind dann $A \cup B$ bzw. $A \setminus B$ definierende Grundmengen von $M \cup N$ bzw. $M \setminus N$, d.h. auch diese sind Zylindermengen. Zuguterletzt ist $\{\square\}$ definierende Grundmenge von Ω_Σ selbst. □

In der Stochastik sind jedoch fast ausschließlich σ -Algebren gefragt; sie sind nämlich gerade die Mengensysteme, die eine konsistente Axiomatik der Wahrscheinlichkeitstheorie bedienen. Der Name “Messraum” für ein Tupel (Ω, \mathcal{B}) deutet an, dass die durch die σ -Algebra festgelegten Mengen durch noch geeignet zu definierende Maßfunktionen “gemessen” werden können. Oftmals jedoch, wie z.B. hier, sind die sich aufdrängenden Mengensysteme keine σ -Algebren. Die folgende Definition hilft, diesem Dilemma zu entkommen.

DEFINITION 2.7. Da der Durchschnitt einer (beliebigen) Familie von σ -Algebren wieder eine σ -Algebra ergibt und die Potenzmenge $\mathcal{P}(\Omega)$ einer Menge Ω selbst stets eine σ -Algebra ist, ist der folgende Begriff sinnvoll. Ist $\mathcal{F} \in \mathcal{P}(\Omega)$ ein beliebiges System von Teilmengen, so sei

$$\sigma(\mathcal{F}) = \bigcap_{\mathcal{F} \subset \mathcal{B}, \mathcal{B} \sigma\text{-Algebra}} \mathcal{B} \quad (2.8)$$

die von \mathcal{F} erzeugte σ -Algebra. Dies ist, im Sinne der Inklusion von Teilmengensystemen, die kleinste σ -Algebra, die \mathcal{F} enthält.

Da wir später Wahrscheinlichkeitstheorie auf den von einer stochastischen Quelle erzeugten Wörtern betreiben wollen, betrachten wir meist

$$\mathcal{B}_\Sigma := \sigma(\mathcal{F}_\Sigma),$$

die vom System der Zylindermengen erzeugte σ -Algebra. Später eingeführte Maßfunktionen beziehen sich dann auf den Messraum

$$(\Omega_\Sigma = \Sigma^{\mathbb{N}}, \mathcal{B}_\Sigma = \sigma(\mathcal{F}_\Sigma)).$$

2.2. Stochastische Vektorräume und (Prä)-Maße

Wir wollen im weiteren Verlauf dieser Arbeit stochastische Prozesse als Elemente eines Vektorraums auffassen lernen. Wir wechseln daher zunächst das Thema, und präsentieren die Vektorräume, für die wir uns interessieren. Gewisse Elemente derselben können wir dann mit stochastischen Prozessen identifizieren.

Wir benötigen dazu eine erste eigenständige Definition, nämlich die des stochastischen Operators. Der stochastische Operator ist ein linearer Operator, der auf dem Vektorraum der beschränkten, reellwertigen Funktionen über Σ^* agiert. Diesen Vektorraum bezeichnen wir mit l_∞^Σ . Wir setzen also

$$l_\infty^\Sigma := \{g : \Sigma^* \rightarrow \mathbb{R} \mid \sup_{\bar{a} \in \Sigma^*} g(\bar{a}) < \infty\} \subset \mathbb{R}^{\Sigma^*}$$

und bemerken, dass wir l_∞^Σ wegen $\text{card}(\Sigma^*) = \text{card}(\mathbb{N})$ mit dem Vektorraum der beschränkten, reellen Folgen identifizieren können. Elemente von l_∞^Σ bezeichnen wir meist mit den Buchstaben g oder h und l_∞^Σ bezeichnen wir der Einfachheit halber auch mit l_∞ , falls die Wahl des Alphabets klar ist, oder es auf eine Auswahl desselben nicht ankommt.

2.2.1. Der stochastische Operator und stochastische Vektorräume. Wir kommen nun zu unserer ersten Definition, deren Motivation im Anschluss durch den Abgleich mit maßtheoretischen Begriffen offenbar werden wird.

DEFINITION 2.8. *Wir nennen den linearen Operator*

$$\sigma : l_\infty \longrightarrow l_\infty,$$

die durch die Vorschrift

$$(\sigma g)(\bar{c}) := \sum_{a \in \Sigma} g(\bar{c}a)$$

definiert wird, den **stochastischen Operator**.

Mehrfache Anwendung des stochastischen Operators führt auf

PROPOSITION 2.9. *Ist $t \in \mathbb{N}$ und $\bar{c} \in \Sigma^*$, so gilt*

$$(\sigma^t g)(\bar{c}) = \sum_{\bar{a} \in \Sigma^t} g(\bar{c}\bar{a}).$$

BEWEIS. Folgt mit vollständiger Induktion. Der Induktionsanfang ist die Definition und der Induktionsschritt ist trivial. \square

Der stochastische Operator induziert einen Untervektorraum von l_∞ , für dessen Elemente und Untervektorräume wir uns im weiteren Verlauf dieser Arbeit ausschließlich interessieren.

DEFINITION/BEMERKUNG 2.10. *Die Menge der Fixpunkte des stochastischen Operators σ , also der Elemente von $g \in l_\infty$ mit $\sigma g = g$ bilden einen Untervektorraum l_σ von l_∞ . Unterräume $V \subset l_\sigma$ bezeichnen wir auch als **stochastische Vektorräume** (über Σ , falls das eine Rolle spielt).*

BEWEIS. Das ist trivial: Seien g, h zwei Fixpunkte von σ . Dann folgt die Behauptung aus der Linearität von σ :

$$\sigma(\alpha g + \beta h) = \alpha \sigma g + \beta \sigma h = \alpha g + \beta h.$$

\square

Ist die Auswahl eines Alphabets Σ von Bedeutung, so schreiben wir genauer l_σ^Σ anstelle von l_σ .

Die Elemente von l_σ bekommen nun noch für Maßtheoretiker suggestive Namen. Im folgenden Abschnitt werden wir dann die Ursache der Namensgebung aufklären.

DEFINITION 2.11. *Wir nennen einen Vektor $g \in l_\sigma$ einen **signierten Maßvektor**. Gilt zusätzlich*

$$\forall \bar{c} \in \Sigma^* : g(\bar{c}) \geq 0, \tag{2.9}$$

*so sprechen wir von einem **Maßvektor** und darüberhinaus von einem **stochastischen Vektor**, falls*

$$g(\square) = 1.$$

2.2.2. (Prä)-Maße. Wie versprochen, folgt hier die Begründung der obigen Bezeichnungen. Dazu erinnern wir zunächst an weitere fundamentale Begriffe der Maßtheorie.

DEFINITION 2.12. *Ist \mathcal{F} ein Ring, so nennen wir eine Funktion*

$$P : \mathcal{F} \longrightarrow \mathbb{R} \cup \{-\infty, \infty\}$$

ein **signiertes Prämaß**, falls

$$P(\emptyset) = 0, \quad (2.10)$$

und für jede Folge $(A_n)_{n \in \mathbb{N}}$ paarweise disjunkter Mengen aus \mathcal{F} , so dass $\bigcup_n A_n \in \mathcal{F}$

$$P\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \sum_{n \in \mathbb{N}} P(A_n), \quad (2.11)$$

insbesondere existiere der Limes auf der rechten Seite. Im Falle $P(\mathcal{F}) \subset \mathbb{R}^+ \cup \{\infty\}$ nennen wir P ein **Prämaß** auf \mathcal{F} . Im Falle $P(\mathcal{F}) \subset \mathbb{R}$ bzw. $P(\mathcal{F}) \subset \mathbb{R}^+$ nennen wir P ein **endliches, signiertes Prämaß** bzw. **endliches Prämaß**. Ist \mathcal{F} eine σ -Algebra, so ersetzen wir in den obigen Fällen jeweils Prämaß durch **Maß**. Ein Maß P mit der Eigenschaft $P(\Omega) = 1$ wird **Wahrscheinlichkeitsmaß** genannt. Es folgt leicht, dass ein Wahrscheinlichkeitsmaß ein endliches Maß ist. Ein Tripel (Ω, \mathcal{B}, P) bestehend aus einem Messraum (Ω, \mathcal{B}) und einem Maß bzw. einem Wahrscheinlichkeitsmaß P wird **Maßraum** bzw. **Wahrscheinlichkeitsraum** genannt.

Ist nun $g \in l_\sigma$, $A \subset \Sigma^t$ eine Menge von Wörtern und $Z(A) \subset \mathcal{F}_\Sigma$ die dadurch induzierte Zylindermenge, so definieren wir

$$g(Z(A)) := g(A) := \sum_{\bar{a} \in A} g(\bar{a})$$

und erhalten damit eine reellwertige Funktion auf \mathcal{F}_Σ , denn es gilt der folgende Satz, der einschließlich unterstützender Lemmata für Maßvektoren g bekannt ist (siehe wiederum [Gra01]). Wir passen ihn hier unserem Sprachrahmen an und verallgemeinern ihn auf endliche, signierte Prämaße. Da die Beweise für die speziell die Zylindermengen betreffenden Aussagen in Büchern nicht zu finden sind, entwickeln wir sie an dieser Stelle ausführlich.

SATZ 2.13. *Die soeben definierte Funktion g auf \mathcal{F}_Σ ist wohldefiniert, d.h. sie ist unabhängig von der Auswahl einer definierenden Grundmenge. Darüberhinaus ist g ein endliches, signiertes Prämaß auf \mathcal{F}_Σ .*

BEWEIS. Seien $m, n \in \mathbb{N}$, $m \geq n$ und $A \subset \Sigma^n, B \subset \Sigma^m$ zwei Mengen von Wörtern, so dass $Z(A) = Z(B)$ in \mathcal{F}_Σ . Nach Proposition 2.3 gilt dann

$$B = \bigcup_{\bar{a} \in A} \Sigma_{\bar{a}}^m,$$

B ist also gerade die Menge aller m -Wort-Fortsetzungen von Wörtern aus A . Mit Proposition 2.9 folgt

$$\begin{aligned} g(A) &= \sum_{\bar{a} \in A} g(\bar{a}) \stackrel{\sigma g = g}{=} \sum_{\bar{a} \in A} (\sigma^{m-n} g)(\bar{a}) \\ &\stackrel{P.2.9}{=} \sum_{\bar{a} \in A} \sum_{\bar{c} \in \Sigma^{m-n}} g(\bar{a}\bar{c}) = \sum_{\bar{a} \in A} g(\Sigma_{\bar{a}}^m) \\ &= g\left(\bigcup_{\bar{a} \in A} \Sigma_{\bar{a}}^m\right) = g(B), \end{aligned}$$

und damit die Wohldefiniertheit von g . Die Endlichkeit von g ist klar nach Definition. Weiter ist $g(\emptyset) = 0$, da $g(\emptyset)$ eine leere Summe ist. Wir betrachten

nun zwei Lemmata:

LEMMA 2.14. *Sei $(F_n)_{n \in \mathbb{N}}$ eine Folge von Zylindermengen $F_n \in \mathcal{F}_\Sigma$ mit $F_{n+1} \subset F_n$ für alle n und $\bigcap_{n \in \mathbb{N}} F_n = \emptyset$. Dann gibt es ein $N_0 \in \mathbb{N}$, so dass*

$$\forall n \geq N_0 : F_n = \emptyset.$$

BEWEIS. Siehe dafür auch [Shi96], Lemma I.1.1 (b). Dort werden bekannte Sätze aus der Topologie zu Hilfe genommen. Wir geben einen elementaren Beweis.

Wir nehmen das Gegenteil an, also eine Folge von nichtleeren Zylindermengen F_n mit leerem Schnitt und $F_{n+1} \subset F_n$ für alle n . Wir konstruieren nun eine Folge von Wörtern \bar{a}_n , so dass jeweils $\bar{a}_n \leq \bar{a}_{n+1}$, $t_n := |\bar{a}_n| \rightarrow \infty$ und schließlich $\Sigma_{\bar{a}_n}^* \subset F_n$. Setzt man $\omega_k := (\bar{a}_n)_k$ für $t_n \geq k$ (das ist wohldefiniert wegen $(\bar{a}_n)_k = (\bar{a}_m)_k$ für $m > n$), so erhält man

$$\omega \in \bigcap_{n \in \mathbb{N}} \Sigma_{\bar{a}_n}^* \subset \bigcap_{n \in \mathbb{N}} F_n,$$

und damit einen Widerspruch.

Wir konstruieren nun diese Folge von Wörtern rekursiv. Wir suchen eine Folge von definierenden Grundmengen von Wörtern $(A_n)_{n \in \mathbb{N}}$ für die F_n , so dass für $t_n := |A_n|$ jeweils $t_n < t_{n+1}$ gilt, da die t_n beliebig groß gewählt werden können. Wegen $F_{n+1} \subset F_n$ sind die Wörter aus A_{n+1} dann Fortsetzungen der Wörter aus A_n . Wir betrachten nun die Funktion

$$\begin{aligned} f : \Sigma^* &\longrightarrow \mathbb{N} \cup \{-\infty, \infty\} \\ \bar{a} &\longmapsto \sup\{n \in \mathbb{N} \mid \exists \bar{b} \in A_n : \bar{a} \leq \bar{b}\}, \end{aligned}$$

die einem Wort \bar{a} den Index $f(\bar{a}) \in \mathbb{N}$ zuordnet, so dass in $A_{f(\bar{a})}$ die längstmögliche, in einem der A_n enthaltene Fortsetzung von \bar{a} zu finden ist. Im Falle $f(\bar{a}) = \infty$ gibt es in jedem A_n mit $t_n \geq |\bar{a}|$ ein Wort, das \bar{a} fortsetzt (und $f(\bar{a}) = -\infty$ trifft auf den Fall zu, dass in keinem der A_n eine Verlängerung und damit auch \bar{a} selbst nicht zu finden ist). Wir setzen nun für alle $n \in \mathbb{N}$

$$A_n^\infty := \{\bar{a} \in A_n \mid f(\bar{a}) = \infty\}$$

und stellen fest, dass die A_n^∞ für alle n nicht-leer sind, da die Annahme $f(\bar{a}) \leq K_n$ für alle (endlich vielen) Wörter einer Menge A_n zu einem Widerspruch führt, da die nicht-leere Menge $A_{n+K_{n+1}}$ ausschließlich Fortsetzungen von Wörtern aus A_n enthält. Wir stellen weiter fest, dass

$$\forall n \in \mathbb{N} \forall \bar{a} \in A_n^\infty \exists \bar{b} \in A_{n+1}^\infty : \bar{a} \leq \bar{b}, \quad (2.12)$$

denn die Annahme, es gäbe ein $n \in \mathbb{N}$ und ein $\bar{a} = a_1 \dots a_{t_n} \in A_n^\infty$, so dass für alle $\bar{b} \in A_{n+1}$ mit $\bar{a} \leq \bar{b}$: $f(\bar{b}) \leq \infty$ erbringt (es gibt nur endlich viele solcher \bar{b} 's) wiederum ein $K_0 \in \mathbb{N}$, so dass für alle diese \bar{b} gilt, dass $f(\bar{b}) \leq K_0$. Dies führt auf einen Widerspruch, denn wegen $f(\bar{a}) = \infty$ findet man ein $K_1 > K_0$ und eine Fortsetzung $\bar{b} = b_1 \dots b_{t_n} \dots b_{t_{n+1}} \dots b_{t_{K_0}} \dots b_{t_{K_1}} \in A_{K_1}$ von \bar{a} . Dann aber $b_1 \dots b_{t_{n+1}} \in A_{n+1}$ wegen $F_{K_1} \subset F_{n+1}$, ein Widerspruch!

Nun ist die Konstruktion der geforderten Folge von Wörtern einfach. Für \bar{a}_0 nehmen wir ein beliebiges Wort aus A_0^∞ . Ist \bar{a}_n dann konstruiert, so wählen wir für \bar{a}_{n+1} gemäß Eigenschaft (2.12) ein Wort aus A_{n+1}^∞ , das \bar{a}_n fortsetzt. \square

LEMMA 2.15. *Ist $P : \mathcal{F} \rightarrow \mathbb{R}$ eine Funktion auf einem Ring \mathcal{F} mit den Eigenschaften $P(\emptyset) = 0$ und $P(\cup_{i=1}^n F_i) = \sum_{i=1}^n P(F_i)$ für endlich viele paarweise disjunkte $F_i \in \mathcal{F}$, so sind die beiden folgenden Aussagen äquivalent:*

- (a) *P ist ein endliches, signiertes Prämaß.*
- (b) *Für jede Folge $(F_n)_{n \in \mathbb{N}}$ von Mengen aus \mathcal{F} mit $F_{n+1} \subset F_n$ für alle n und $\cap_n F_n = \emptyset$ gilt*

$$\lim_{n \rightarrow \infty} P(F_n) = 0.$$

BEWEIS. Der Beweis verläuft analog zu dem von Satz 3.2 aus [Bau90], S.12 ff. Dort wird die Aussage jedoch nur für endliche Prämaße P bewiesen. Bei Überprüfung der Argumente zeigt sich jedoch, dass die Positivität von P an keiner Stelle von Belang ist. \square

Weiter im Beweis von Satz 2.13: Die Voraussetzungen von Lemma 2.15 sind für g nach Definition trivialerweise erfüllt. Eigenschaft (b) wiederum ist wegen Lemma 2.14 erfüllt, woraus die Behauptung des Satzes folgt. \square

2.2.3. Fortsetzung von Prämaßen zu Maßen. Die Maßtheorie interessiert sich nun für Fortsetzungen von Prämaßen auf Ringen zu Maßen auf den von den Ringen erzeugten σ -Algebren. Ergebnis dieser Untersuchungen sind die folgenden Sätze.

SATZ 2.16 (Fortsetzungssatz). *Jedes Prämaß P auf einem Ring \mathcal{F} lässt sich auf mindestens eine Weise zu einem Maß \tilde{P} auf die von \mathcal{F} erzeugte σ -Algebra $\sigma(\mathcal{F})$ fortsetzen.*

BEWEIS. Siehe [Bau90], S. 23, 5.1. \square

Für endliche Prämaße auf Algebren erhält man obendrein eine Eindeutigkeitsaussage.

SATZ 2.17 (Eindeutigkeitsatz). *Jedes endliche Prämaß P auf einer Algebra \mathcal{F} kann auf genau eine Weise zu einem Maß \tilde{P} auf $\sigma(\mathcal{F})$ fortgesetzt werden.*

BEWEIS. Siehe [Bau90], S. 28, 5.6 (beachte, dass endliche Prämaße auf Algebren immer auch “ σ -endlich” sind). \square

Ist \mathcal{F} ein Mengensystem, so kann man im Allgemeinen keine Konstruktionsregel der Elemente von $\sigma(\mathcal{F})$ angeben (im Stile von: “Elemente von $\sigma(\mathcal{F})$ sind abzählbare Vereinigungen von Elementen von \mathcal{F} ” o.ä.). Deshalb sind Aussagen über Eigenschaften der Elemente von $\sigma(\mathcal{F})$ und darauf definierter Funktionen oftmals schwierig zu erzielen. Ist \mathcal{F} jedoch eine Algebra, so kann

man sich auf den folgenden Satz berufen.

SATZ 2.18 (Approximationssatz). *Sei \mathcal{F} eine Algebra und P ein endliches Maß auf $\sigma(\mathcal{F})$. Dann existiert zu jedem $A \in \sigma(\mathcal{F})$ eine Folge $(C_n)_{n \in \mathbb{N}}$ in \mathcal{F} , so dass*

$$\lim_{n \rightarrow \infty} P(A \triangle C_n) = 0. \quad (2.13)$$

BEWEIS. Siehe [Bau90], S.29, 5.7. □

All dies hat für unsere Begriffe Konsequenzen, die nun die Namensgebung motiviert. Die Maßvektoren bzw. die stochastischen Vektoren stehen nämlich in einer 1-1 Beziehung zu den endlichen Maßen bzw. den Wahrscheinlichkeitsmaßen auf dem vorhin entwickelten Messraum $(\Omega_\Sigma, \mathcal{B}_\Sigma)$. Das erläutert der folgende

SATZ 2.19. *Sei $g \in l_\sigma$ ein Maßvektor. Dann wird durch g ein auf eindeutige Weise festgelegtes endliches Maß m_g auf $(\Omega_\Sigma, \mathcal{B}_\Sigma)$ definiert. Ist umgekehrt ein endliches Maß m auf $(\Omega_\Sigma, \mathcal{B}_\Sigma)$ gegeben, so wird dadurch auf eindeutige Weise ein Maßvektor g_m definiert. Darüberhinaus gilt*

$$m_{g_m} = m \quad \text{und} \quad g_{m_g} = g.$$

BEWEIS. Da g nach Satz 2.13 ein endliches Prämaß auf der Algebra der Zylindermengen \mathcal{F}_Σ induziert, lässt sich g nach Satz 2.17 zu einem eindeutig bestimmten Maß m_g auf $\mathcal{B}_\Sigma = \sigma(\mathcal{F}_\Sigma)$ fortsetzen. Ist umgekehrt ein Maß m auf $(\Omega_\Sigma, \mathcal{B}_\Sigma)$ gegeben, so wird dadurch eine Funktion g_m auf der Menge der Wörter Σ^* definiert, indem man $g_m(\bar{a}) := m(Z(\{\bar{a}\}))$ setzt. Einfache Rechnungen zeigen dann, dass es sich bei g_m um einen Maßvektor handelt: $g_m \geq 0$ ist klar wegen $m \geq 0$ und

$$\begin{aligned} \sigma g_m(\bar{a}) &= \sum_{a \in \Sigma} g_m(\bar{a}a) = \sum_{a \in \Sigma} m(Z\{\bar{a}a\}) \\ &= m(Z(\Sigma^{\lvert \bar{a} \rvert + 1})) \stackrel{P.2.3}{=} m(Z(\{\bar{a}\})) = g_m(\bar{a}) \end{aligned}$$

bestätigt $\sigma g_m = g_m$. Ist $Z(A)$ eine Zylindermenge mit definierender Grundmenge, so

$$\begin{aligned} m_{g_m}(Z(A)) &= g_m(A) = \sum_{\bar{a} \in A} g_m(\bar{a}) = \sum_{\bar{a} \in A} m(Z(\{\bar{a}\})) \\ &= m(Z(A)) \end{aligned}$$

wegen $\cup_{\bar{a} \in A} Z(\{\bar{a}\}) = Z(A)$ und der definierenden Eigenschaften eines Maßes. Da Maße, die auf den Zylindermengen übereinstimmen, aufgrund des Eindeutigkeitssatzes 2.18 gleich sind, folgt $m_{g_m} = m$. Weiter folgt $g_{m_g} = g$ aus

$$g_{m_g}(\bar{a}) = m_g(Z(\{\bar{a}\})) = g(\bar{a})$$

was den Beweis komplettiert. □

BEMERKUNG 2.20. *Im weiteren Verlauf bezeichnen wir m_g einfach ebenfalls mit g , falls es auf den Unterschied nicht ankommt. Wir benutzen optional*

die Bezeichnung P_g für m_g im Falle durch g induzierter Wahrscheinlichkeitsverteilungen, also im Falle stochastischer Vektoren.

KOROLLAR 2.21. *Die stochastischen Vektoren stehen in einer 1-1-Beziehung mit den Wahrscheinlichkeitsmaßen auf $(\Omega_\Sigma, \mathcal{B}_\Sigma)$.*

BEWEIS. Ist ein stochastischer Vektor g gegeben, so ist m_g mit Satz 2.19 wegen $g(\Omega_\Sigma) = g(\square) = 1$ ein Wahrscheinlichkeitsmaß. Umgekehrt folgt offensichtlich $g_m(\square) = 1$ für ein Wahrscheinlichkeitsmaß m . \square

2.2.4. Messbare Abbildungen und stochastische Prozesse. Unser eigentliches Interesse ist die Betrachtung von endlichwertigen, diskretzeitigen stochastischen Prozessen. Dieser Abschnitt dient dazu, eine Brücke zwischen den bisher entwickelten Begriffen und der Theorie der stochastischen Prozesse zu schlagen, ohne die gesamte Theorie der stochastischen Prozesse aufzurollen. Nebenbei wollen wir den Schiebeoperator einführen, der eine Verbindung zwischen der Theorie der stochastischen Prozesse und der Theorie der dynamischen Systeme herstellt. Dies erfolgt so sparsam wie möglich. Für ausführlichere Abhandlungen der Theorie der stochastischen Prozesse verweisen wir auf Standardwerke wie z.B. [Doo53], [Bil95] und [Bau91]. Zur Theorie der dynamischen Systeme verweisen wir zunächst auf spätere Kapitel. Zur Vorbereitung benötigen wir zunächst weitere elementare Begriffe der Maßtheorie.

DEFINITION 2.22. *Eine Abbildung*

$$f : \Omega_1 \longrightarrow \Omega_2$$

zwischen Messräumen $(\Omega_1, \mathcal{B}_1)$, $(\Omega_2, \mathcal{B}_2)$, für die

$$\forall B \in \mathcal{B}_2 : f^{-1}(B) \in \mathcal{B}_1 \quad (2.14)$$

gilt, wird **messbar** genannt.

Messbare Abbildungen, deren Definitionsmenge ein Maßraum ist, induzieren ein Maß auf dem Bildraum:

DEFINITION/BEMERKUNG 2.23. *Ist (Ω, \mathcal{B}, P) ein Maßraum und*

$$f : (\Omega, \mathcal{B}, P) \rightarrow (\Omega_0, \mathcal{A})$$

eine messbare Abbildung in einen Messraum (Ω_0, \mathcal{A}) so induziert f auf \mathcal{A} ein Maß \tilde{P} durch die Vorschrift ($A \in \mathcal{A}$)

$$\tilde{P}(A) := P(f^{-1}(A)). \quad (2.15)$$

Dabei ist \tilde{P} offensichtlich ein Wahrscheinlichkeitsmaß, falls auch P eins ist. Im Falle eines Wahrscheinlichkeitsraums (Ω, \mathcal{B}, P) wird eine solche Abbildung f dann **Zufallsvariable** (mit Werten in Ω_0) genannt.

Da, wie schon oben erwähnt, die Elemente einer σ -Algebra schwer zu charakterisieren sind, kann man auch messbare Abbildungen im Allgemeinen nur schwer erkennen. Erfreulicherweise reicht es bei gegebenem Erzeugendensystem, die Messbarkeit nur darauf zu überprüfen.

SATZ 2.24. *Seien (Ω, \mathcal{B}) und (Ω', \mathcal{B}') Messräume, so dass $\mathcal{B}' = \sigma(\mathcal{F}')$ für ein beliebiges Mengensystem $\mathcal{F}' \subset \mathcal{P}(\Omega')$. Dann ist eine Abbildung $T : \Omega \rightarrow \Omega'$ genau dann messbar, falls*

$$\forall F' \in \mathcal{F}' : T^{-1}(F') \in \mathcal{B}.$$

Mit anderen Worten reicht es, Messbarkeit von Abbildungen auf den Elementen eines Erzeugendensystems zu überprüfen.

BEWEIS. Siehe [Bau90], S.40, Satz 7.2. □

Wir zitieren nun ohne weitere Umschweife die Definition eines (endlichwertigen, diskretzeitigen) stochastischen Prozesses.

DEFINITION 2.25. *Ein **diskretzeitiger, endlichwertiger stochastischer Prozess** ist ein Quadrupel $(\Omega, \mathcal{B}, P, (X_t)_{t \in \mathbb{N}})$, so dass (Ω, \mathcal{B}, P) ein Wahrscheinlichkeitsraum ist und $(X_t)_{t \in \mathbb{N}}$ eine Folge von Zufallsvariablen, die alle ihre Werte in derselben endlichen Menge Σ annehmen (also, um genau zu sein $X_t : (\Omega, \mathcal{B}, P) \rightarrow (\Sigma, \mathcal{P}(\Sigma))$ messbare Abbildungen).*

Wir bemerken, dass ein solcher stochastischer Prozess Wahrscheinlichkeitsmaße P_t auf $(\Sigma^t, \mathcal{P}(\Sigma^t))$ für $t \in \mathbb{N}$ induziert, indem man für $\bar{a} = a_1 \dots a_t \in \Sigma^t$

$$P_t(\bar{a}) := P_t(X_0 = a_1, \dots, X_{t-1} = a_t) := P(X_0^{-1}(\{a_1\}) \cap \dots \cap X_{t-1}^{-1}(\{a_t\}))$$

und für $A \subset \Sigma^t$ weiter $P_t(A) := \sum_{\bar{a} \in A} P_t(\bar{a})$ setzt. Mit Hilfe dieser Wahrscheinlichkeitsmaße kann man nun zwei stochastische Prozesse als im Wesentlichen gleich identifizieren:

DEFINITION 2.26. *Zwei stochastische Prozesse*

$$(\Omega_1, \mathcal{B}_1, P_1, (X_t)_{t \in \mathbb{N}}) \text{ und } (\Omega_2, \mathcal{B}_2, P_2, (Y_t)_{t \in \mathbb{N}})$$

*mit Werten in derselben endlichen Menge Σ werden **äquivalent** genannt, falls die gemäß obiger Beschreibung induzierten Wahrscheinlichkeitsmaße $(P_1)_t$ und $(P_2)_t$ für alle $t \in \mathbb{N}$ übereinstimmen.*

Der einem stochastischen Prozess zugrundeliegende Wahrscheinlichkeitsraum (Ω, \mathcal{B}, P) kann nun bei gegebenem Σ auf kanonische Art und Weise eingeführt werden. Eine Folgerung des Satzes von Kolmogorov ([Bau91], S. 307, Satz 35.3) ist nämlich, dass es zu jedem stochastischen Prozess einen zu diesem äquivalenten Prozess gibt, der über den kanonisch konstruierten Wahrscheinlichkeitsraum definiert wird.

Ohne an dieser Stelle die Einzelheiten der Beweisführung des Satzes von Kolmogorov zu zitieren (dafür siehe z.B. die einleitenden Text zur Konstruktion stochastischer Prozesse in [Bau91], S. 303 ff. oder [Doo53], S. 46 ff.)

bemerken wir, dass der kanonisch gegebene Wahrscheinlichkeitsraum eines diskretzeitigen stochastischen Prozesses mit Werten in einem endlichen Alphabet Σ gerade gegeben ist durch

$$(\Omega_\Sigma = \Sigma^\mathbb{N}, \mathcal{B}_\Sigma = \sigma(\mathcal{F}_\Sigma), \tilde{P}),$$

wobei \tilde{P} das eindeutig bestimmte Wahrscheinlichkeitsmaß ist, für das bei gegebenem Wort $\bar{a} \in \Sigma^t$ gilt, dass

$$g_P(\bar{a}) := P_t(\bar{a}) = P_t(X_0 = \bar{a}_1, \dots, X_{t-1} = \bar{a}_t).$$

(Bemerke, dass g_P einen stochastischen Vektor definiert, der nach oben Gesagtem zu einem eindeutigen Wahrscheinlichkeitsmaß auf \mathcal{B}_Σ fortgesetzt werden kann.)

Ist umgekehrt ein Wahrscheinlichkeitsmaß P auf $(\Omega_\Sigma, \mathcal{B}_\Sigma)$ gegeben (mit anderen Worten $(\Omega_\Sigma, \mathcal{B}_\Sigma, P)$ ein Wahrscheinlichkeitsraum), so erhält man für $t \in \mathbb{N}$ Zufallsvariablen X_t und damit einen diskretzeitigen, endlichwertigen stochastischen Prozess durch die Vorschrift

$$\begin{array}{ccc} X_t : & \Omega_\Sigma & \longrightarrow & \Sigma \\ & \omega = \omega_1 \omega_2 \dots \omega_t \dots & \mapsto & \omega(t) = \omega_{t+1}. \end{array}$$

wobei man die Messbarkeit der X_t unter Zuhilfenahme von Satz 2.24 erhält. Kurz gesagt, und dies ist die *Quintessenz* dieses Abschnitts:

SATZ 2.27. *Diskretzeitige stochastische Prozesse $(X_t)_{t \in \mathbb{N}}$ mit Werten in einem endlichen Alphabet Σ entsprechen gerade (bis auf Äquivalenz) den Wahrscheinlichkeitsmaßen P auf dem soeben erklärten Messraum $(\Omega_\Sigma, \mathcal{B}_\Sigma)$.*

KOROLLAR 2.28. *Ein stochastischer Vektor g legt einen (bis auf Äquivalenz) eindeutig bestimmten, diskretzeitigen stochastischen Prozess $(X_t)_{t \in \mathbb{N}}$ mit Werten in Σ fest und umgekehrt.*

2.2.5. Der Schiebeoperator. Der Schiebeoperator ist eine messbare Abbildung, die die Theorie der stochastischen Prozesse mit der der dynamischen Systeme verbindet. Mit dieser Verbindung erhält man die Möglichkeit, stochastische Prozesse bezüglich asymptotischer Eigenschaften “entlang von Sequenzen” zu untersuchen. An dieser Stelle führen wir jedoch nur den Begriff des Schiebeoperators selbst ein. Elementare Begriffe der Theorie der dynamischen Systeme werden an anderen, passenderen Stellen eingeführt werden.

DEFINITION 2.29. *Wir nennen die Abbildung*

$$\begin{array}{ccc} T_\Omega : & \Omega_\Sigma & \longrightarrow & \Omega_\Sigma \\ & \omega = (\omega_1, \omega_2, \omega_3, \dots) & \mapsto & (\omega_2, \omega_3, \dots), \end{array}$$

*also $(T_\Omega \omega)(k) := \omega(k+1)$, in Worten T_Ω die Abbildung, die eine Sequenz um eine Stelle nach links verschiebt, den **Schiebeoperator**.*

DEFINITION/BEMERKUNG 2.30. *Wir schreiben analog für ein Wort $\bar{a} := a_1 \dots a_t \in \Sigma^t$*

$$T_\Omega(\bar{a}) := a_2 \dots a_t \in \Sigma^{t-1}$$

Ist daher $A \in \Sigma^t$ eine Menge von t -Wörtern, so

$$T_{\Omega}^{-n}A = \{a_1 \dots a_n a_{n+1} \dots a_{n+t} \in \Sigma^{n+t} \mid a_{n+1} \dots a_{n+t} \in A\}.$$

Es folgt allgemein für $n \in \mathbb{N}$ und Mengen $A \subset \Sigma^*$

$$|A| = t \implies |T_{\Omega}^{-n}A| = t + n.$$

PROPOSITION 2.31. *Der Schiebeoperator ist messbar.*

BEWEIS. Ist $Z(A) \in \mathcal{F}_{\Sigma}$ eine Zylindermenge mit definierender Grundmenge $A \subset \Sigma^t$, so ist auch $T_{\Omega}^{-n}(Z(A)) \in \mathcal{F}_{\Sigma}$ eine Zylindermenge (mit definierender Grundmenge $T_{\Omega}^{-n}A \subset \Sigma^{t+n}$ eine Zylindermenge für $n \in \mathbb{N}$). Der Schiebeoperator ist daher gemäß Satz 2.24 eine messbare Abbildung. \square

BEMERKUNG 2.32. *Ist T_{Ω} der Schiebeoperator und P ein Wahrscheinlichkeitsmaß auf $(\Omega_{\Sigma}, \mathcal{B}_{\Sigma})$, so ist auch $P \circ T_{\Omega}^{-k}$ ein Wahrscheinlichkeitsmaß auf $(\Omega_{\Sigma}, \mathcal{B}_{\Sigma})$ (siehe Bem./Def. 2.23). Identifizieren wir den Wahrscheinlichkeitsraum $(\Omega_{\Sigma}, \mathcal{B}_{\Sigma}, P)$ mit einem stochastischen Prozess $(X_t)_{t \in \mathbb{N}}$, so ist der durch $(\Omega_{\Sigma}, \mathcal{B}_{\Sigma}, P \circ T_{\Omega}^{-k})$ induzierte stochastische Prozess $(Y_t)_{t \in \mathbb{N}}$ gerade gegeben durch*

$$Y_t = X_{t+k}.$$

2.3. Normen, Konvergenz und signierte Maße

Beim Studium des letzten Abschnitts mag die Frage aufgekommen sein, ob sich beliebige Elemente aus l_{σ} , also bezüglich des stochastischen Operators invariante Elemente mit (eindeutig bestimmten) signierten Maßen identifizieren lassen. Tatsächlich ist kein zu den obigen Fortsetzungssätzen allgemeineres Ergebnis für den Fall signierter Prämaße bekannt. (Zumindest ist dem Autor keines bekannt, und eine einfache Verallgemeinerung der obigen Sätze ist nicht möglich, da die Positivität der Prämaße an entscheidenden Stellen ausgenutzt wird). Dieser Abschnitt wird einen Untervektorraum von l_{σ} präsentieren, für dessen Elemente eindeutige Fortsetzungen möglich sind. Spätere Arbeit mit diesem Untervektorraum garantiert demnach das Handeln mit wohlbegründeten Objekten.

Die Einführung dieses Untervektorraums hat einen zweiten guten Grund. Die später entwickelte Ergodentheorie interessiert sich vor allem für die Konvergenz gewisser Messwerte. Um diese Aussagen erzielen zu können, benötigen wir einen Konvergenzbegriff, der die ergodentheoretischen Aussagen garantiert. Dies genau leistet gerade die Norm "der totalen Variation", die wir in diesem Abschnitt einführen werden. Elemente aus l_{σ} , die endlich bezüglich dieser Norm sind, bilden dann gerade den soeben erwähnten Untervektorraum, also die Grundmenge der meisten unserer Betrachtungen.

2.3.1. Motivation: Wortweise Konvergenz. Wir starten diesen Abschnitt, indem wir die eigentliche Motivation der Einführung einer Norm anreißen: die Schaffung eines sinnvollen Konvergenzbegriffs. Einen genauen Abgleich mit den in der Maßtheorie geläufigen Konvergenzbegriffen geben wir später. Dabei werden wir diese Begriffe dann auch auf den Fall signierter Maße ausdehnen, ein in der Maßtheorie nicht behandeltes Thema. Ohne Einführung einer Norm liegt zunächst der folgende Konvergenzbegriff auf der Hand:

DEFINITION 2.33. *Wir sagen, dass eine Folge von Vektoren $g_n \in l_\infty, n \in \mathbb{N}$ **punktweise** oder auch **wortweise** gegen einen Vektor $g \in l_\infty$ **konvergiert**, falls*

$$\forall \bar{a} \in \Sigma^* : \lim_{n \rightarrow \infty} g_n(\bar{a}) = g(\bar{a}).$$

Zunächst einmal hat wortweise Konvergenz eine schöne Konsequenz: Konvergiert eine Folge stochastischer Vektoren wortweise, so ist der resultierende Grenzvektor wieder ein stochastischer Vektor. Allgemeiner gilt

SATZ 2.34. *Sei $(g_n)_{n \in \mathbb{N}}$ eine Folge (signierter) Maßvektoren (stochastischer Vektoren), die punktweise gegen einen Vektor g konvergieren. Dann ist auch g ein (signierter) Maßvektor (stochastischer Vektor).*

BEWEIS. Für signierte Maßvektoren haben wir $\sigma g = g$ zu zeigen. Dies folgt aus elementaren Limesregeln der Analysis:

$$\begin{aligned} \sigma g(\bar{a}) &= \sum_{a \in \Sigma} g(\bar{a}a) = \sum_{a \in \Sigma} \lim_{n \rightarrow \infty} g_n(\bar{a}a) \\ &= \lim_{n \rightarrow \infty} \sum_{a \in \Sigma} g_n(\bar{a}a) \stackrel{\sigma g_n = g_n}{=} \lim_{n \rightarrow \infty} g_n(\bar{a}) = g(\bar{a}). \end{aligned} \quad (2.16)$$

Im Falle von Maßvektoren haben wir noch $g(\bar{a}) \in [0, 1]$ für alle $\bar{a} \in \Sigma^*$ und für stochastische Vektoren zusätzlich $g(\square) = 1$ zu zeigen. Das ist aber offensichtlich. \square

Das soeben zitierte Ergebnis verführt zu der Annahme, dass wortweise Konvergenz auch Konvergenz der zu den Vektoren assoziierten Maße auf allen Elementen der σ -Algebra \mathcal{B}_Σ nach sich zieht. Tatsächlich ist dies nicht der Fall. Das folgende Beispiel wird dies dokumentieren:

BEISPIEL 2.35. *Sei $\Sigma = \{a, b\}$ und*

$$g_n : \Sigma^* \longrightarrow \mathbb{R} \quad \bar{a} \longmapsto \begin{cases} 1 & \bar{a} \leq a^n \\ \frac{1}{2^{|\bar{a}|-n}} & a^n \leq \bar{a} \\ 0 & \text{sonst.} \end{cases} \quad (2.17)$$

In Worten beschreiben die zu den g_n assoziierten Maße m_{g_n} eine Gleichverteilung über alle Sequenzen, die mit n a 's anfangen. Sei weiter

$$\begin{aligned} g : \Sigma^* &\longrightarrow \mathbb{R} \\ \bar{a} &\longmapsto \begin{cases} 1 & \bar{a} = a^n \text{ für ein } n \\ 0 & \text{sonst.} \end{cases} \end{aligned} \quad (2.18)$$

Dann konvergieren die g_n offensichtlich punktweise gegen g . Sei weiter

$$B = \{aaaaaa\dots\} = \bigcap_m \Omega_{a^m} \in \mathcal{B}_\Sigma$$

das Element der Sigma-Algebra, das nur die Sequenz enthält, die nur aus a 's besteht. Dann gilt

$$m_{g_n}(B) \stackrel{(*)}{=} \lim_{m \rightarrow \infty} g_n(\bigcap_{k \leq m} \Omega_{a^k}) = \lim_{m \rightarrow \infty} \frac{1}{2^{m-n}} = 0$$

((*) nach Konstruktion der eindeutigen Fortsetzung m_{g_n} von g_n , siehe den Beweis von 2.16 in [Bau90]), aber (mit derselben Argumentation)

$$m_g(B) = 1.$$

Demnach konvergieren die m_{g_n} nicht auf allen Elementen von \mathcal{B}_Σ gegen m_g .

Um einen Konvergenzbegriff zu schaffen, der die Konvergenz der durch stochastische Vektoren induzierten Wahrscheinlichkeitsmaße auf allen Elementen der σ -Algebra nach sich zieht (das ist später in der Ergodentheorie wünschenswert), muss mehr gearbeitet werden. Wir beginnen mit der Definition eines Untervektorraums von l_∞ , dessen Elemente sich, wie einleitend bemerkt, gerade mit den endlichen, signierten Maßen identifizieren lassen können (diesen Zusammenhang beleuchten wir später). Auf diesem Untervektorraum lässt sich eine Norm einführen, die die gewünschte Konvergenz nach sich zieht.

2.3.2. Die Norm der totalen Variation. Um es noch in einmal kurz und klar zu sagen: Normen haben viele nützliche Konsequenzen, deren wichtigste meist die Schaffung einer Topologie ist. Damit einhergehend kommen dann geläufige Begriffe wie Konvergenz und Stetigkeit ins Spiel. Deshalb:

DEFINITION 2.36. *Wir setzen*

$$l_{TV} := \{g \in \mathbb{R}^{\Sigma^*} \mid \sup_{t \in \mathbb{N}} \sum_{\bar{a} \in \Sigma^t} |g(\bar{a})| < \infty\}.$$

Weiter definieren wir

$$\begin{aligned} \|\cdot\|_{TV} : l_{TV} &\longrightarrow \mathbb{R}^+ \\ g &\longmapsto \sup_{t \in \mathbb{N}} \sum_{\bar{a} \in \Sigma^t} |g(\bar{a})|, \end{aligned} \quad (2.19)$$

und nennen diese Abbildung die **Norm der totalen Variation**. Den Vektorraum l_{TV} nennen wir entsprechend **Vektorraum der totalen Variation**.

Der Name “Norm der totalen Variation” ist in jeder Hinsicht gerechtfertigt. Zunächst einmal versichern wir:

PROPOSITION 2.37. $(l_{TV}, \|\cdot\|_{TV})$ ist ein normierter Raum.

BEWEIS. Nach Definition ist $\|\cdot\|_{TV}(l_{TV}) \subset \mathbb{R}^+$. Ist $\lambda \in \mathbb{R}$ und $g \in l_{TV}$, so

$$\begin{aligned} \|\lambda g\|_{TV} &= \sup_{k \in \mathbb{N}} \sum_{\bar{a} \in \Sigma^k} |\lambda g(\bar{a})| = \sup_{k \in \mathbb{N}} \sum_{\bar{a} \in \Sigma^k} |\lambda| |g(\bar{a})| \\ &= \sup_{k \in \mathbb{N}} |\lambda| \sum_{\bar{a} \in \Sigma^k} |g(\bar{a})| = |\lambda| \sup_{k \in \mathbb{N}} \sum_{\bar{a} \in \Sigma^k} |g(\bar{a})| = |\lambda| \|g\|_{TV}, \end{aligned}$$

wobei wir die Tatsache $\sup\{ts \mid s \in S\} = t \cdot \sup S$ für eine Menge $S \subset \mathbb{R}^+$ und $t \in \mathbb{R}^+$ ausgenutzt haben. Sind nun $g, h \in l_{TV}$, so

$$\begin{aligned} \|g + h\|_{TV} &= \sup_{k \in \mathbb{N}} \sum_{\bar{a} \in \Sigma^k} |g(\bar{a}) + h(\bar{a})| \stackrel{(2.20)}{\leq} \sup_{k \in \mathbb{N}} \sum_{\bar{a} \in \Sigma^k} |g(\bar{a})| + |h(\bar{a})| \\ &\stackrel{(2.21)}{\leq} \sup_{(n,m) \in \mathbb{N}^2} \left(\sum_{\bar{a} \in \Sigma^n} |g(\bar{a})| + \sum_{\bar{a} \in \Sigma^m} |h(\bar{a})| \right) \\ &\stackrel{(2.22)}{=} \sup_{n \in \mathbb{N}} \sum_{\bar{a} \in \Sigma^n} |g(\bar{a})| + \sup_{m \in \mathbb{N}} \sum_{\bar{a} \in \Sigma^m} |h(\bar{a})| = \|g\|_{TV} + \|h\|_{TV}, \end{aligned}$$

Der Reihe nach haben wir die folgenden Eigenschaften für zwei Mengen $S, T \subset \mathbb{R}^+$ ausgenutzt:

$$\forall s \in S \exists t \in T : s \leq t \Rightarrow \sup S \leq \sup T \quad (2.20)$$

$$S \subset T \Rightarrow \sup S \leq \sup T \quad (2.21)$$

$$\sup\{s + t \mid s \in S, t \in T\} = \sup S + \sup T. \quad (2.22)$$

Gilt $\|g\|_{TV} = 0$, so für alle t

$$\sum_{\bar{a} \in \Sigma^t} |g(\bar{a})| = 0 \implies \forall \bar{a} \in \Sigma^t : |g(\bar{a})| = 0.$$

□

Damit wäre der erste Teil des Namens “Norm der totalen Variation” überprüft. Die Bezeichnung “totale Variation” wird später ihre Berechtigung erfahren. Siehe dazu Satz 2.63 sowie Bemerkung 2.62, die unsere Konvergenzbegriffe in einen Zusammenhang mit den in der Gemeinde der Maßtheoretiker üblichen Konvergenzbegriffen stellt.

Wir fahren damit fort, Eigenschaften von l_{TV} aufzulisten. Wir stellen als Erstes fest, dass l_{TV} echt in l_∞ enthalten ist.

BEMERKUNG 2.38. Bezeichnet man mit

$$\|g\|_\infty := \sup_{\bar{a} \in \Sigma^*} |g(\bar{a})|$$

die übliche (Supremums-)Norm auf dem Raum der beschränkten Folgen l_∞ über Σ^* , so gilt

$$\forall g \in \mathcal{G} : \|g\|_\infty \leq \|g\|_{TV}.$$

Weiter gilt

$$l_{TV} \subsetneq l_\infty,$$

wie man mit Hilfe der Funktion g mit $g(\bar{a}) = 1$ für alle $\bar{a} \in \Sigma^*$ einsieht.

Wir definieren nun

DEFINITION 2.39. Wir sagen, dass eine Folge von Vektoren $(g_n)_{n \in \mathbb{N}}$ (**in Norm**) gegen einen Vektor $g \in l_{TV}$ **konvergiert**, falls

$$\lim_{n \rightarrow \infty} \|g_n - g\|_{TV} = 0,$$

also die g_n bezüglich der von $\|\cdot\|_{TV}$ induzierten Topologie konvergieren. Der Einfachheit halber meinen wir Konvergenz in Norm, wenn wir

$$\lim_{n \rightarrow \infty} g_n = g$$

schreiben.

Wir stellen weiter fest, dass der eben definierte lineare Unterraum l_{TV} ein Banachraum ist. Dies ist im Folgenden an einigen Stellen von höchstem Interesse, da sich so Konvergenzbeweise einfacher führen lassen.

SATZ 2.40. Der normierte Raum $(l_{TV}, \|\cdot\|_{TV})$ ist vollständig.

BEWEIS. Sei $(g_n)_{n \in \mathbb{N}}$ eine Cauchy-Folge aus l_{TV} . Für $g \in l_{TV}$ setzen wir

$$g_t = g|_{\Sigma^t} \in \mathbb{R}^{\Sigma^t},$$

welche sich mit Vektoren endlich-dimensionaler, reellwertiger Vektorräume identifizieren lassen. Wegen

$$\|(g_n)_t - (g_m)_t\|_1 = \sum_{\bar{a} \in \Sigma^t} |(g_n - g_m)(\bar{a})| \leq \|g_n - g_m\|_{TV}$$

und der Vollständigkeit endlich-dimensionaler, normierter Vektorräume, konvergieren die induzierten endlich-dimensionalen Folgen gegen Elemente \bar{g}_t . Wir definieren einen Vektor $\bar{g} \in \mathbb{R}^{\Sigma^*}$ durch die Vorschrift $\bar{g}(\bar{a}) := \bar{g}_t(\bar{a})$ für $|\bar{a}| = t$. Sei nun $\epsilon \in \mathbb{R}^+$. Dann gibt es ein $N_0 \in \mathbb{N}$, so dass

$$\forall n, m \geq N_0 : \|g_n - g_m\|_{TV} < \frac{\epsilon}{2},$$

woraus

$$\forall n, m \geq N_0, \forall t \in \mathbb{N} : \|(g_n)_t - (g_m)_t\|_1 < \frac{\epsilon}{2}$$

folgt. Lässt man $m \rightarrow \infty$ erhält man somit

$$\forall n \geq N_0, \forall t \in \mathbb{N} : \|(g_n)_t - (\bar{g})_t\|_1 \leq \frac{\epsilon}{2} < \epsilon \quad \Rightarrow \quad \|g_n - \bar{g}\|_{TV} < \epsilon.$$

Also konvergieren die g_n gegen den Vektor \bar{g} . Ist nun $\epsilon = 1$, so gibt es ein N_1 , so dass

$$\forall n \geq N_1 : \|g_n - \bar{g}\|_{TV} < 1.$$

Es folgt

$$\begin{aligned} \forall t \in \mathbb{N} \forall n \geq N_1 : \sum_{\bar{a} \in \Sigma^t} |\bar{g}(\bar{a})| - \sum_{\bar{a} \in \Sigma^t} |g_n(\bar{a})| \\ \leq \sum_{\bar{a} \in \Sigma^t} |\bar{g}(\bar{a}) - g_n(\bar{a})| \\ \leq \|g_n - \bar{g}\|_{TV} < 1, \end{aligned} \quad (2.23)$$

woraus $\sup_t \sum_{\bar{a} \in \Sigma^t} |\bar{g}(\bar{a})| \leq \|g_n\|_{TV} + 1$ und damit $\bar{g} \in l_{TV}$ folgt. \square

Für konkrete Berechnungen der Norm der totalen Variation ist die folgende Proposition hilfreich.

PROPOSITION 2.41. *Ist $g \in l_\sigma \cap l_{TV}$, so gilt*

$$\|g\|_{TV} = \lim_{t \rightarrow \infty} \sum_{\bar{a} \in \Sigma^t} |g(\bar{a})|.$$

Ist g ein Maßvektor, so $g \in l_{TV}$ und

$$\|g\|_{TV} = g(\square).$$

BEWEIS. Der erste Teil folgt aus

$$\begin{aligned} \sum_{\bar{a} \in \Sigma^t} |g(\bar{a})| &\stackrel{\sigma g = g}{=} \sum_{\bar{a} \in \Sigma^t} \left| \sum_{a \in \Sigma} g(\bar{a}a) \right| \\ &\leq \sum_{\bar{a} \in \Sigma^t} \sum_{a \in \Sigma} |g(\bar{a}a)| = \sum_{\bar{a} \in \Sigma^{t+1}} |g(\bar{a})|. \end{aligned}$$

Der zweite folgt aus

$$\sum_{\bar{a} \in \Sigma^t} |g(\bar{a})| \stackrel{g \geq 0}{=} \sum_{\bar{a} \in \Sigma^t} g(\bar{a}) \stackrel{\sigma^t g = g}{=} g(\square).$$

\square

Eine interessante und später ungemein nützliche Beobachtung ist nun, dass sich in endlich-dimensionalen Unterräumen von l_{SM} die beiden Konvergenzmodi nichts tun. In diesem Fall impliziert die wortweise Konvergenz nämlich schon Konvergenz in Norm:

SATZ 2.42. *Sei $V \subset l_{TV}$ ein endlich-dimensionaler Unterraum von l_{TV} und $(g_n)_{n \in \mathbb{N}}$ eine Folge in V , die wortweise gegen $g \in l_{TV}$ konvergiert. Dann konvergieren die g_n auch in Norm gegen g , insbesondere ist $g \in V$.*

BEWEIS. Wir wählen zunächst eine Basis $(h_j)_{j \in \{1, \dots, m\}}$ von V und identifizieren über die Koordinatendarstellung

$$\begin{aligned} \Phi : \quad V &\longrightarrow \mathbb{R}^m \\ \sum_j \alpha_j h_j &\longmapsto (\alpha_1, \dots, \alpha_m) \end{aligned}$$

V mit dem \mathbb{R}^m und setzen

$$\Phi(g_n) =: (\alpha_{n1}, \dots, \alpha_{nm}) =: \alpha_n^T$$

sowie

$$\Phi(g) =: (\beta_1, \dots, \beta_m) =: \beta^T.$$

Wir bemerken weiter, dass via der Koordinatendarstellung auf dem \mathbb{R}^m durch die TV -Norm ebenfalls eine Norm induziert wird. Da in einem endlich-dimensionalen reellwertigen Vektorraum alle Normen äquivalent sind, reicht es demnach zu zeigen, dass

$$\forall j \in \{1, \dots, m\} : \lim_{n \rightarrow \infty} \alpha_{nj} = \beta_j.$$

Wir bestimmen nun mit dem nachstehenden Lemma 2.43 m Sequenzen $\bar{a}_i, i = 1, \dots, m$, so dass

$$A = \begin{pmatrix} h_1(\bar{a}_1) & \cdots & h_m(\bar{a}_1) \\ \vdots & \ddots & \vdots \\ h_1(\bar{a}_m) & \cdots & h_m(\bar{a}_m) \end{pmatrix} \in \mathbb{R}^{m \times m}$$

eine reguläre Matrix ist. Wir schreiben weiter

$$x_g = (g(\bar{a}_1), \dots, g(\bar{a}_m))^T.$$

Dann gilt zunächst $A\beta = x_g$ und weiter

$$A\alpha_n \xrightarrow[n \rightarrow \infty]{} x_g \quad A^{-1} \xrightarrow{\text{stetig}} \quad \alpha_n \xrightarrow[n \rightarrow \infty]{} A^{-1}x_g = \beta.$$

□

Das sich nun anschließende, im obigen Beweis bereits benutzte Lemma wirkt auf den ersten Blick trivial und wird auch später noch (im Abschnitt über Basen und charakteristische Systeme 3.3) vielfach zur Anwendung kommen:

LEMMA 2.43. *Sei $(g_i)_{i \in I}, I \subset \mathbb{N}$ eine abzählbare Familie von Vektoren aus l_∞ . Dann sind die beiden folgenden Aussagen äquivalent:*

- (i) *Die g_i sind linear unabhängig.*
- (ii) *Es gibt eine Familie von Wörtern $(\bar{a}_i)_{i \in I}$, so dass die Matrizen für alle $i \in I$*

$$A_i = \begin{pmatrix} g_1(\bar{a}_1) & \cdots & g_i(\bar{a}_1) \\ \vdots & \ddots & \vdots \\ g_1(\bar{a}_n) & \cdots & g_i(\bar{a}_i) \end{pmatrix}$$

regulär sind.

BEWEIS. Die Richtung (ii) \Rightarrow (i) ist einfach, denn wir haben nach Definition der linearen Unabhängigkeit die lineare Unabhängigkeit endlicher Teilfamilien der (g_i) zu zeigen. Ist eine endliche Teilfamilie gegeben, so ist A_N für den größten in dieser Teilfamilie auftretenden Index N regulär und die lineare Unabhängigkeit der Teilfamilie folgt.

Für die andere Richtung identifizieren wir I über entsprechende Bijektionen mit entweder einer endlichen Teilmenge $\{1, \dots, n\}$ von \mathbb{N} oder aber mit $\mathbb{N} \setminus \{0\}$. Wir zeigen die Behauptung dann in beiden Fällen mit Induktion über i . Im Falle $i = 1$ finden wir wegen $g_1 \neq 0$ ein Wort \bar{a}_0 mit $0 \neq g_1(\bar{a}_1) = \det(g_1(\bar{a}_1))$.

Für den Induktionsschritt sei $i \geq 2$. Dann können wir die Existenz einer regulären Matrix

$$A_{i-1} := \begin{pmatrix} g_1(\bar{a}_1) & \cdots & g_{i-1}(\bar{a}_1) \\ \vdots & \ddots & \vdots \\ g_1(\bar{a}_{i-1}) & \cdots & g_{i-1}(\bar{a}_{i-1}) \end{pmatrix}$$

annehmen. Damit finden wir ein eindeutig bestimmtes $\lambda = (\lambda_1, \dots, \lambda_{i-1})^T$, so dass

$$A_{i-1}\lambda = (g_1(\bar{a}_1), \dots, g_{i-1}(\bar{a}_{i-1}))^T \Leftrightarrow \lambda = A_{i-1}^{-1}(g_1(\bar{a}_1), \dots, g_{i-1}(\bar{a}_{i-1}))^T.$$

Ist nun $\bar{a}_i \in \Sigma^* \setminus \{\bar{a}_0, \dots, \bar{a}_{i-1}\}$ und die Matrix

$$A_{\bar{a}_i} := \begin{pmatrix} & & & g_i(\bar{a}_1) \\ & & & \vdots \\ & A_{i-1} & & g_i(\bar{a}_{i-1}) \\ g_1(\bar{a}_i) & \cdots & g_{i-1}(\bar{a}_i) & g_i(\bar{a}_i) \end{pmatrix}$$

singulär, so gilt aufgrund der Eindeutigkeit von λ auch

$$g_i(\bar{a}_i) = \sum_{j=1}^{i-1} \lambda_j g_j(\bar{a}_i).$$

Daher muss es ein Wort \bar{a}_i geben, so dass $A_{\bar{a}_i}$ regulär ist, da sonst

$$g_i = \sum_{j=1}^{i-1} \lambda_j g_j$$

folgte, ein Widerspruch! □

2.3.3. Signierte Maße und der Zerlegungssatz. Wir kommen nun zur ersten Rechtfertigung der Schaffung von l_{TV} . Tatsächlich können wir nämlich die signierten Maßvektoren, die in l_{TV} enthalten sind, eindeutig mit signierten Maßen identifizieren. Die folgenden Ergebnisse zeigen den Weg zu dieser Erkenntnis. Wir starten mit den fundamentalen Ergebnissen über signierte Maße, die in Büchern im Allgemeinen etwas stiefmütterlich abgehandelt werden, offensichtlich, weil negativ belegte Mengen intuitiv nicht eingeordnet werden können. Tatsächlich stellen sich zumindest endliche, signierte Maße als Differenzen von endlichen Maßen heraus. Das reicht den meisten Maßtheoretikern.

SATZ 2.44 (Hahn-Zerlegung). *Sei P ein endliches, signiertes Maß auf einer σ -Algebra \mathcal{B} über einer Menge Ω . Dann existieren zwei Mengen Ω^+, Ω^- , so dass $\Omega = \Omega^+ \dot{\cup} \Omega^-$ und*

$$\begin{aligned} P(B) &\geq 0, & B &\in (\Omega^+ \cap \mathcal{B}) \\ P(B) &\leq 0, & B &\in (\Omega^- \cap \mathcal{B}). \end{aligned}$$

Die Mengen Ω^+, Ω^- nennt man eine **Hahn-Zerlegung** von P .

BEWEIS. Siehe [Bau90], Seite 123. \square

KOROLLAR 2.45. *Jedes endliche, signierte Maß P auf einer σ -Algebra \mathcal{B} ist die Differenz zweier endlicher Maße.*

BEWEIS. Definiere endliche Maße P^+, P^- durch $P^+(A) := P(A \cap \Omega^+)$ und $P^-(A) := -P(A \cap \Omega^-)$, dann $P = P^+ - P^-$. \square

KOROLLAR 2.46 (Jordan-Zerlegung). *Sei P ein endliches, signiertes Maß auf einer σ -Algebra \mathcal{B} . Dann gibt es eindeutig bestimmte Maße P^+, P^- , so dass $P = P^+ - P^-$ und für jede andere Darstellung $P = P_1 - P_2$ mit Maßen P_1, P_2 gilt: $P_1 = P^+ + \delta$ und $P_2 = P^- + \delta$ für ein Maß $\delta \neq 0$. Die Maße P^+, P^- werden **Jordan-Zerlegung** von P genannt.*

BEWEIS. Folgt man, indem man den Beweis des Satzes ([Bau90], S. 123) über die Hahn-Zerlegung durchgeht. \square

BEMERKUNG 2.47. *Ist P ein endliches, signiertes Maß und $\Omega^+ \dot{\cup} \Omega^-$ eine Hahn-Zerlegung zu P , so ist die Jordan-Zerlegung gerade gegeben durch*

$$P^+(A) := P(A \cap \Omega^+) \text{ und } P^-(A) := P(A \cap \Omega^-).$$

Das ist eine direkte Konsequenz des Beweises zum Satz über die Jordan-Zerlegung.

Die Jordan-Zerlegung gibt Anlass für eine weitere Definition, die den “Betrag” eines endlichen, signierten Maßes erklärt.

DEFINITION 2.48. *Sei P ein endliches, signiertes Maß und $P = P^+ - P^-$ seine Jordan-Zerlegung. Dann nennen wir P^+ die **positive Variation** von P und P^- die **negative Variation** von P . Setzen wir weiter*

$$|P| := P^+ + P^-,$$

*so nennen wir dies **totale Variation** von P .*

Wie bereits zu Beginn des Abschnitts erwähnt, stehen für signierte Prämaße im Allgemeinen keine zu den für Prämaßen erhältlichen analogen Fortsetzungssätze zur Verfügung. Eine hinreichende Bedingung wäre in unserem Fall die Zerlegbarkeit der signierten Prämaße in Differenzen von Prämaßen, wie mit dem sich anschließenden Lemma klar wird.

LEMMA 2.49. *Sei*

$$g = \sum_{i=1}^n \alpha_i g_i$$

ein Maßvektor, der als Linearkombination von Maßvektoren g_i entstanden ist. Dann gilt

$$\forall B \in \mathcal{B}_\Sigma : \quad m_g(B) = \sum_{i=1}^n \alpha_i m_{g_i}(B).$$

BEWEIS. Das ist natürlich klar für Zylindermengen B . Für allgemeines B bemerken wir, dass durch die Vorschrift

$$f(B) := \sum_{i=1}^n \alpha_i m_{g_i}(B)$$

ein endliches, signiertes Maß auf $(\Omega_\Sigma, \mathcal{B}_\Sigma)$ festgelegt wird, das auf den Zylindermengen mit m_g übereinstimmt. Die Annahme eines $B \in \mathcal{B}_\Sigma$ mit $f(B) < 0$ führt mit dem Approximationssatz für Maße 2.18 auf eine Folge von Zylindermengen (B_n) mit $|f|(B \Delta B_n) \rightarrow 0$, wobei $|f|$ die totale Variation von f sei (siehe Definition 2.48). Das wiederum führt mit Lemma A.2 auf eine Zylindermenge B_{n_0} mit

$$0 > f(B_{n_0}) = g(B_{n_0}) \geq 0,$$

ein Widerspruch. Demnach ist f ein Maß. Mit Hilfe von Satz 2.19 sieht man daher ein, dass $f = m_g$. \square

KOROLLAR 2.50. *Ist g ein signierter Maßvektor, der als Differenz zweier Maßvektoren g_1, g_2 entstanden ist, also*

$$g = g_1 - g_2,$$

so definieren wir

$$\forall B \in \mathcal{B}_\Sigma : m_g(B) := m_{g_1}(B) - m_{g_2}(B).$$

Diese Definition ist wohlbegründet, denn im Falle von

$$g_1 - g_2 = g = g_3 - g_4$$

für Maßvektoren g_1, g_2, g_3, g_4 gilt wegen $g_1 + g_4 = g_2 + g_3$ nach Lemma 2.49

$$m_{g_1}(B) + m_{g_4}(B) = m_{(g_1+g_4)}(B) = m_{(g_2+g_3)}(B) = m_{g_2}(B) + m_{g_3}(B)$$

und damit natürlich $m_{g_1}(B) - m_{g_2}(B) = m_{g_3}(B) - m_{g_4}(B)$. In diesem Sinne schreiben wir für solche Vektoren g auch einfach g anstelle von m_g , falls es auf den Unterschied nicht ankommt.

Wir sehen damit ein, dass durch signierte Maßvektoren, die als Differenz zweier Maßvektoren entstehen, ein eindeutig festgelegtes signiertes Maß m_g gegeben ist. Diese Zerlegung funktioniert nun im Falle signierter Maßvektoren aus l_{TV} :

SATZ 2.51 (Zerlegungssatz). *Sei $g \in l_{TV}$ ein signierter Maßvektor. Dann gibt es eindeutig bestimmte Maßvektoren $g^+, g^- \in l_{TV}$, so dass*

$$g = g^+ - g^-$$

und

$$\|g^+\|_{TV} + \|g^-\|_{TV} = \|g\|_{TV}.$$

BEWEIS. Wir werden die Vektoren g^+, g^- explizit konstruieren. Wir setzen zunächst für alle $\bar{a} \in \Sigma^*$ und $t \geq |\bar{a}|$

$$\delta_{t,\bar{a}}^+ := \sum_{\substack{\bar{b} \in \Sigma_{\bar{a}}^t \\ g(\bar{b}) \geq 0}} |g(\bar{b})| = \sum_{\substack{\bar{b} \in \Sigma_{\bar{a}}^t \\ g(\bar{b}) \geq 0}} g(\bar{b}) \quad (2.24)$$

und analog

$$\delta_{t,\bar{a}}^- := \sum_{\substack{\bar{b} \in \Sigma_{\bar{a}}^t \\ g(\bar{b}) < 0}} |g(\bar{b})| = \sum_{\substack{\bar{b} \in \Sigma_{\bar{a}}^t \\ g(\bar{b}) < 0}} -g(\bar{b}). \quad (2.25)$$

Wir bemerken zunächst, dass für alle $\bar{a} \in \Sigma^*$ und $t \geq |\bar{a}|$

$$\delta_{t,\bar{a}}^{+,-} \leq \delta_{t+1,\bar{a}}^{+,-} \leq \|g\|_{TV}, \quad (2.26)$$

wobei die erste Ungleichung aus

$$g(\bar{b}) \stackrel{\sigma g = g}{=} \sum_{a \in \Sigma} g(\bar{b}a) \leq \sum_{\substack{a \in \Sigma \\ g(\bar{b}a) \geq 0}} g(\bar{b}a)$$

für $\bar{b} \in \Sigma_{\bar{a}}^t$ mit $g(\bar{b}) \geq 0$ bzw. einer analogen Rechnung für $g(\bar{b}) < 0$ folgt, und die zweite direkt aus der Definition der Norm der totalen Variation. Damit bilden die $\delta_{t,\bar{a}}^+$ und die $\delta_{t,\bar{a}}^-$ monoton wachsende, beschränkte Folgen in t und wir setzen

$$g^+(\bar{a}) := \lim_{t \rightarrow \infty} \delta_{t,\bar{a}}^+, \quad (2.27)$$

$$g^-(\bar{a}) := \lim_{t \rightarrow \infty} \delta_{t,\bar{a}}^-, \quad (2.28)$$

und zeigen, dass so Vektoren mit den gewünschten Eigenschaften definiert sind. Weiter folgt wegen

$$\begin{aligned} \sum_{a \in \Sigma} \delta_{t,\bar{a}a}^+ &= \sum_{a \in \Sigma} \sum_{\substack{\bar{b} \in \Sigma_{\bar{a}a}^t \\ g(\bar{b}) \geq 0}} g(\bar{b}) \\ &\stackrel{\sigma g = g}{=} \sum_{\substack{\bar{b} \in \Sigma_{\bar{a}}^t \\ g(\bar{b}) \geq 0}} g(\bar{b}) = \delta_{t,\bar{a}}^+ \end{aligned} \quad (2.29)$$

für $t \geq |\bar{a}| + 1$ und damit

$$\begin{aligned} (\sigma g^+)(\bar{a}) &= \sum_{a \in \Sigma} g^+(\bar{a}a) = \sum_{a \in \Sigma} \lim_{t \rightarrow \infty} \delta_{t,\bar{a}a}^+ \\ &= \lim_{t \rightarrow \infty} \sum_{a \in \Sigma} \delta_{t,\bar{a}a}^+ \stackrel{(2.29)}{=} \lim_{t \rightarrow \infty} \delta_{t,\bar{a}}^+ = g^+(\bar{a}) \end{aligned}$$

$\sigma g^+ = g^+$. Analog schließt man auf $\sigma g^- = g^-$. Nach Definition gilt offensichtlich $g^+, g^- \geq 0$. Damit sind g^+, g^- Maßvektoren. Beachten wir noch

$$\delta_{t,\square}^+ + \delta_{t,\square}^- = \sum_{\bar{a} \in \Sigma^t} |g(\bar{a})|, \quad (2.30)$$

so

$$\begin{aligned}
\|g^+\|_{TV} + \|g^-\|_{TV} &\stackrel{P.2.41}{=} g^+(\square) + g^-(\square) = \lim_{t \rightarrow \infty} \delta_{t,\square}^+ + \lim_{t \rightarrow \infty} \delta_{t,\square}^- \\
&= \lim_{t \rightarrow \infty} (\delta_{t,\square}^+ + \delta_{t,\square}^-) \stackrel{(2.30)}{=} \lim_{t \rightarrow \infty} \sum_{\bar{a} \in \Sigma^t} |g(\bar{a})| \quad (2.31) \\
&\stackrel{P.2.41}{=} \|g\|_{TV}.
\end{aligned}$$

Es bleibt die Eindeutigkeit zu beweisen. Hierzu nehmen wir die Existenz zweier zu g^+, g^- verschiedener Maßvektoren g_1, g_2 mit $g = g_1 - g_2$ und $\|g\|_{TV} = \|g_1\|_{TV} + \|g_2\|_{TV}$ an. Wir nehmen ohne Einschränkung (die anderen Fälle folgen völlig analog) die Existenz eines $\bar{a} \in \Sigma^*$ mit $g_1(\bar{a}) < g^+(\bar{a})$ an. Nach Definition von g^+ gibt es ein $t \geq |\bar{a}|$, so dass für $A := \{\bar{a} \in \Sigma_{\bar{a}}^t \mid g(\bar{a})\}$ gilt, dass $g(A) > g_1(\bar{a})$. Damit folgt, da g_1 ein Maßvektor ist,

$$g_1(A) \leq g_1(\Sigma_{\bar{a}}^t) = g_1(\bar{a}) < g(A) = g_1(A) - g_2(A) \leq g_1(A),$$

ein Widerspruch! □

KOROLLAR 2.52. *Jedes durch einen signierten Maßvektor aus $l_\sigma \cap l_{TV}$ induzierte signierte Prämaß lässt sich zu einem eindeutig bestimmten endlichen, signierten Maß auf $(\Omega_\Sigma, \mathcal{B}_\Sigma)$ fortsetzen. Die endlichen, signierten Maße auf $(\Omega_\Sigma, \mathcal{B}_\Sigma)$ stehen damit in einer 1-1-Beziehung zu den Elementen aus $l_\sigma \cap l_{TV}$, d.h. es gilt*

$$m_{g_m} = g \quad \text{und} \quad g_{m_g} = g.$$

BEWEIS. Ist $g \in l_\sigma \cap l_{TV}$, so sei $g = g^+ - g^-$ die Zerlegung gemäß Satz 2.51. Dann erhalten wir ein endliches, signiertes Maß m_g durch die Vorschrift

$$\forall B \in \mathcal{B}_\Sigma : \quad m_g(B) := m_{g^+}(B) - m_{g^-}(B),$$

wobei m_{g^+}, m_{g^-} die eindeutig bestimmten Maße seien, die die durch g^+, g^- induzierten Prämaße fortsetzen. Ist m^* ein beliebiges endliches, signiertes Maß, so dass die Einschränkung von m^* auf die Zylindermengen den Vektor g induziert, so schreiben wir $m^* = m_1 - m_2$ gemäß der Hahn-Zerlegung und erhalten mit Satz 2.19 $g = g_{m_1} - g_{m_2}$. Damit aber mit Korollar 2.50 und wiederum Satz 2.19

$$m_g = m_{g_{m_1}} - m_{g_{m_2}} = m_1 - m_2 = m^*.$$

Umgekehrt erhalten wir bei gegebenem endlichen, signierten Maß m über die Hahn-Zerlegung $m = m_1 - m_2$ (Satz 2.44 einen dazu assoziierten signierten Maßvektor $g_m := g_{m_1} - g_{m_2} \in l_{TV}$. Mit Hilfe von Satz 2.19 und Korollar 2.50 stellen wir dann noch fest, dass die Beziehungen $m_{g_m} = m$ und $g_{m_g} = g$ gelten, was den Beweis vervollständigt. □

BEMERKUNG 2.53. *Aus den Eigenschaften der Jordan-Zerlegung und ihrer Eindeutigkeit kann man nun folgern, dass die zu g^+ und g^- assoziierten Maße m_{g^+}, m_{g^-} gerade der Jordan-Zerlegung von m_g entsprechen. Es gilt also*

$$m_{g^+} = m_g^+ \quad \text{und} \quad m_{g^-} = m_g^-.$$

Darüberhinaus man mit dem Satz über die Hahn-Zerlegung die Existenz zweier Mengen $\Omega^+, \Omega^- \in \mathcal{B}_\Sigma$ mit $\Omega_\Sigma = \Omega^+ \dot{\cup} \Omega^-$ folgern, so dass für beliebiges $B \in \mathcal{B}_\Sigma$

$$m_{g^+}(B) = m_{g^+}(B \cap \Omega^+) \text{ und } m_{g^-}(B) = m_{g^-}(B \cap \Omega^-)$$

gelten. Offensichtlich dann auch

$$\|g^{+,-}\|_{TV} \stackrel{P.2.41}{=} g^{+,-}(\square) = m_{g^{+,-}}(\Omega_\Sigma) = m_g(\Omega^{+,-}).$$

Diese Bemerkung motiviert nun eine weitere Definition, die Namen für die Elemente des Zerlegungssatzes stiftet:

DEFINITION 2.54. Ist $g \in l_{TV} \cap l_\sigma$ und sind g^+, g^- die eindeutig bestimmten Maßvektoren des Zerlegungssatzes, für die

$$g = g^+ - g^-$$

und

$$\|g\|_{TV} = \|g^+\|_{TV} + \|g^-\|_{TV}$$

gilt, so nennen wir (im Einklang mit Bemerkung 2.53) g^+ den **positiven Variationsvektor**, g^- den **negativen Variationsvektor** und $|g| := g^+ + g^-$ den **totalen Variationsvektor** von g .

Ein Satz, der den Umgang mit der Jordan-Zerlegung in Bezug auf unsere Norm illustriert, beschließt diesen Abschnitt.

SATZ 2.55. Sei $g \in l_{SM}$ und $(g_n)_{n \in \mathbb{N}}$ eine Folge von Vektoren $g_n \in l_{SM}$. Dann

$$\lim_{n \rightarrow \infty} g_n = g \iff \lim_{n \rightarrow \infty} (g_n)^+ = g^+ \text{ und } \lim_{n \rightarrow \infty} (g_n)^- = g^-,$$

wobei jeweils Konvergenz in TV-Norm gemeint sei.

BEWEIS. “ \Leftarrow ” ist offensichtlich. Für “ \Rightarrow ” bemerken wir zunächst, dass es wegen

$$\begin{aligned} \|(g_n)^- - g^-\|_{TV} &= \|-(g_n - (g_n)^+) + (g - g^+)\|_{TV} \\ &\leq \|g - g_n\|_{TV} + \|(g_n)^+ - g^+\|_{TV} \end{aligned}$$

reicht, $\lim_{n \rightarrow \infty} \|(g_n)^+ - g^+\|_{TV} = 0$ zu zeigen. Sei $\epsilon \in \mathbb{R}^+$. Dann wählen wir ein $N_0 \in \mathbb{N}$, so dass für alle $n \geq N_0$

$$\|g_n - g\|_{TV} < \frac{\epsilon}{3}.$$

Sei nun $t \in \mathbb{N}$. Wir definieren zunächst für alle $n \in \mathbb{N}$

$$A_{n,s,\bar{a}} := \{\bar{b} \in \Sigma_{\bar{a}}^s \mid g_n(\bar{b}) \geq 0\} \quad \text{sowie} \quad A_{s,\bar{a}} := \{\bar{b} \in \Sigma_{\bar{a}}^s \mid g(\bar{b}) \geq 0\}.$$

Wir betrachten den Beweis des Zerlegungssatzes 2.51 und stellen zur Wiederholung fest:

$$(g_n)^+(\bar{a}) = \lim_{s \rightarrow \infty} g_n(A_{n,s,\bar{a}}) \quad \text{und ebenso} \quad g^+(\bar{a}) = \lim_{s \rightarrow \infty} g(A_{s,\bar{a}}).$$

Wir definieren weiter

$$\begin{aligned} B_{n,s,\bar{a}} &:= A_{n,s,\bar{a}} \setminus A_{s,\bar{a}}, \\ C_{n,s,\bar{a}} &:= A_{s,\bar{a}} \setminus A_{n,s,\bar{a}}, \\ D_{n,s,\bar{a}} &:= A_{s,n,\bar{a}} \cap A_{s,\bar{a}} \end{aligned}$$

und bemerken, dass $g_n(C_{n,s,\bar{a}}) \leq 0$ und $g(C_{n,s,\bar{a}}) \geq 0$ sowie umgekehrt $g(B_{n,s,\bar{a}}) \leq 0$ und $g_n(B_{n,s,\bar{a}}) \geq 0$, aber sowohl $g_n(D_{n,s,\bar{a}}) \geq 0$ als auch $g(D_{n,s,\bar{a}}) \geq 0$. Daher gilt für $n \geq N_0$, $t \in \mathbb{N}$ und $s \geq t$:

$$\begin{aligned} \sum_{\bar{a} \in \Sigma^t} g_n(B_{n,s,\bar{a}}) &\leq \sum_{\bar{a} \in \Sigma^t} g_n(B_{n,s,\bar{a}}) \underbrace{- g(B_{n,s,\bar{a}})}_{\geq 0} = \sum_{\bar{a} \in \Sigma^t} |g_n(B_{n,s,\bar{a}}) - g(B_{n,s,\bar{a}})| \\ &= \sum_{\bar{a} \in \Sigma^t} \left| \sum_{\bar{b} \in B_{n,s,\bar{a}}} g_n(\bar{b}) - \sum_{\bar{b} \in B_{n,s,\bar{a}}} g(\bar{b}) \right| \leq \sum_{\bar{a} \in \Sigma^t} \sum_{\bar{b} \in B_{n,s,\bar{a}}} |g_n(\bar{b}) - g(\bar{b})| \\ &\leq \sum_{\bar{b} \in \Sigma^s} |g_n(\bar{b}) - g(\bar{b})| \leq \|g_n - g\| < \frac{\epsilon}{3} \end{aligned} \tag{2.32}$$

sowie analog

$$\sum_{\bar{a} \in \Sigma^t} g(C_{n,s,\bar{a}}) < \frac{\epsilon}{3}. \tag{2.33}$$

Deshalb für alle $n \geq N_0$, $t \in \mathbb{N}$ und $s \geq t$

$$\begin{aligned} \sum_{\bar{a} \in \Sigma^t} |g_n(A_{n,s,\bar{a}}) - g(A_{s,\bar{a}})| &= \sum_{\bar{a} \in \Sigma^t} |g_n(B_{n,s,\bar{a}}) + (g_n - g)(D_{n,s,\bar{a}}) - g(C_{n,s,\bar{a}})| \\ &\leq \sum_{\bar{a} \in \Sigma^t} |g_n(B_{n,s,\bar{a}})| + \sum_{\bar{a} \in \Sigma^t} |(g_n - g)(D_{n,s,\bar{a}})| + \sum_{\bar{a} \in \Sigma^t} |-g(C_{n,s,\bar{a}})| \\ &\leq \sum_{\bar{a} \in \Sigma^t} g_n(B_{n,s,\bar{a}}) + \sum_{\bar{a} \in \Sigma^t} \sum_{\bar{b} \in D_{n,s,\bar{a}}} |g_n(\bar{b}) - g(\bar{b})| + \sum_{\bar{a} \in \Sigma^t} g(C_{n,s,\bar{a}}) \\ &\stackrel{(2.32), (2.33)}{<} \frac{\epsilon}{3} + \|g_n - g\|_{TV} + \frac{\epsilon}{3} < \epsilon. \end{aligned} \tag{2.34}$$

Daher folgt für $n \geq N_0$ und $t \in \mathbb{N}$

$$\begin{aligned} \sum_{\bar{a} \in \Sigma^t} |(g_n)^+(\bar{a}) - g^+(\bar{a})| &= \sum_{\bar{a} \in \Sigma^t} \left| \lim_{s \rightarrow \infty} g_n(A_{n,s,\bar{a}}) - \lim_{s \rightarrow \infty} g(A_{s,\bar{a}}) \right| \\ &= \lim_{s \rightarrow \infty} \sum_{\bar{a} \in \Sigma^t} |g_n(A_{n,s,\bar{a}}) - g(A_{s,\bar{a}})| \stackrel{(2.34)}{\leq} \epsilon \end{aligned}$$

und somit auch

$$\|(g_n)^+ - g^+\|_{TV} \leq \epsilon.$$

□

Mit Korollar 2.52 definieren wir nun den Vektorraum, der Grundlage der meisten unserer Ergebnisse sein wird:

DEFINITION 2.56. *Wir setzen*

$$l_{SM} := l_\sigma \cap l_{TV}$$

und bezeichnen diese Menge als **Vektorraum der endlichen, signierten Maßvektoren**. Wir schreiben m_g für das gemäß Korollar 2.52 durch g induzierte, eindeutig bestimmte, signierte, endliche Maß auf $(\Omega_\Sigma, \mathcal{B}_\Sigma)$.

Wir schließen der Vollständigkeit halber mit dem Beispiel eines Fixpunkts g des stochastischen Operators σ , der nicht in l_{TV} zu finden ist.

BEISPIEL 2.57. *Wir setzen*

$$g : \Sigma^* \longrightarrow \mathbb{R}$$

$$\bar{a} \mapsto \begin{cases} 0 & \bar{a} = \square \\ g(\bar{b}) + 1 & \bar{a} = \bar{b}a \text{ für ein } \bar{b} \in \Sigma^* \\ g(\bar{b}) - g(\bar{b}a) & \bar{a} = \bar{b}b \text{ für ein } \bar{b} \in \Sigma^* \end{cases}$$

und sehen direkt ein, dass $\sigma g = g$ gilt. Offensichtlich gilt

$$g(a^t) = t,$$

woraus $\sup_{t \in \mathbb{N}} \sum_{\bar{a} \in \Sigma^t} |g(\bar{a})| = \infty$ und damit $g \notin l_{TV}$ folgt.

2.3.4. Konvergenz von Maßen. Dieser Abschnitt liefert nun den - zu Beginn des Abschnitts bereits angerissenen - Abgleich mit den in der Maßtheorie bekannten Konvergenzbegriffen. Wir starten daher mit zwei aus der Maßtheorie stammenden Definitionen (siehe z.B. [JR97]).

DEFINITION 2.58. *Sei $(P_n)_{n \in \mathbb{N}}$ eine Folge von Maßen auf einem Messraum (Ω, \mathcal{B}) . Wir sagen, dass die P_n **stark** gegen ein Maß P^* konvergieren, falls*

$$\forall B \in \mathcal{B} : \lim_{n \rightarrow \infty} P_n(B) = P^*(B).$$

Wir reichen in einem Atemzug eine weitere Form der Konvergenz von Maßen nach:

DEFINITION 2.59. *Sei $(P_n)_{n \in \mathbb{N}}$ eine Folge von Maßen auf einem Messraum (Ω, \mathcal{B}) . Wir sagen, dass die P_n **schwach Skorokhod** gegen ein Maß P^* konvergieren, falls*

$$\forall \epsilon \in \mathbb{R}^+ \exists N_0 \in \mathbb{N} : \forall n \geq N_0 \sup_{B \in \mathcal{B}} |P_n(B) - P^*(B)| < \epsilon,$$

d.h. falls gleichmäßige Konvergenz auf allen Elementen der σ -Algebra \mathcal{B} stattfindet.

BEMERKUNG 2.60. *Wir weisen daraufhin, dass wir diese Konvergenzbegriffe verallgemeinert haben. In der Literatur werden sie nur für Folgen von Wahrscheinlichkeitsmaßen formuliert (siehe [JR97]).*

BEMERKUNG 2.61. Die Verwendung des Ausdrucks “schwach” im Begriff “schwach Skorokhod” rührt vom prominenten Satz über das “schwache Gesetz der großen Zahlen” her und kann sich mit einer äquivalenten Beschreibung dieser Konvergenzform, die in Appendix B aufgeführt wird, erklärt werden.

BEMERKUNG 2.62. Beide Formen, die starke und die schwache Skorokhod Konvergenz finden gemeinhin wenig Beachtung in der Literatur. Viel wichtiger und deshalb ungleich ausführlicher behandelt ist die sogenannte “schwache Konvergenz” von Maßen, die sich auf von Topologien erzeugten σ -Algebren definieren lässt. Eine Folge von Maßen konvergiert dann “schwach”, falls alle Integralfolgen von konvergenten Folgen stetiger Funktionen konvergieren. Der Satz von Portmanteau führt dann verschiedene Charakterisierungen für diese Sorte Konvergenz auf. Dafür siehe z.B. [Bil99]. In der Ergodentheorie jedoch ist die “starke” Konvergenz spezieller Folgen von Maßen, wie bereits erwähnt, von zentralem Interesse.

Offensichtlich impliziert “schwache Skorokhod” Konvergenz die “starke” Konvergenz. Was haben diese Formen der Konvergenz nun mit unseren Begriffen zu tun? Der folgende Satz wird diese Zusammenhänge aufzeigen, und wie oben versprochen, die Namensgebung “Norm der totalen Variation” begründen. Tatsächlich wird in der Literatur (siehe z.B. [JR97], [Bil95], [Bil99]) die “Metrik der totalen Variation” d_{TV} , gegeben zwei Maße P und Q auf einem Messraum (Ω, \mathcal{B}) wie folgt definiert:

$$d_{TV}(P, Q) := \sup_{B \in \mathcal{B}} |P(B) - Q(B)|.$$

Wir zeigen, dass die durch unsere Norm geschaffene Topologie zur der dieser Metrik äquivalent ist. Eine offensichtliche Folgerung ist, dass Konvergenz in Norm zu “schwacher Skorokhod” Konvergenz äquivalent ist.

SATZ 2.63. Seien $g, h \in l_{SM}$. Dann gilt

$$\sup_{B \in \mathcal{B}_\Sigma} |m_g(B) - m_h(B)| \leq \|g - h\|_{TV} \leq 2 \sup_{B \in \mathcal{B}_\Sigma} |m_g(B) - m_h(B)|. \quad (2.35)$$

BEWEIS. Wir zeigen zunächst

$$\sup_{B \in \mathcal{B}_\Sigma} |m_g(B) - m_h(B)| \leq \|g - h\|_{TV}.$$

Sei dazu zunächst $Z(B) \in \mathcal{B}_\Sigma$ eine Zylindermenge, also $B \subset \Sigma^t$ für ein t . Es folgt

$$\begin{aligned} |m_g(B) - m_h(B)| &= \left| \sum_{\bar{a} \in B} g(\bar{a}) - \sum_{\bar{a} \in B} h(\bar{a}) \right| \\ &\leq \sum_{\bar{a} \in B} |g(\bar{a}) - h(\bar{a})| \\ &\leq \sum_{\bar{a} \in \Sigma^t} |g(\bar{a}) - h(\bar{a})| \leq \|g - h\|_{TV}. \end{aligned} \quad (2.36)$$

Sei nun $B \in \mathcal{B}_\Sigma$ beliebig. Dann gibt es gemäß Satz 2.18 eine Folge von Zylindermengen $(B_k)_{k \in \mathbb{N}}$ mit

$$\lim_{k \rightarrow \infty} |(m_g - m_h)(B_k \triangle B)| = 0.$$

Mit Lemma A.2 folgt dann $\lim_{k \rightarrow \infty} (m_g - m_h)(B_k) = (m_g - m_h)(B)$ und somit weiter wegen Korollar 2.50

$$\lim_{k \rightarrow \infty} |m_g(B_k) - m_h(B_k)| = |m_g(B) - m_h(B)|.$$

Wegen $|m_g(B_k) - m_h(B_k)| \leq \|g - h\|_{TV}$ ist damit auch $|m_g(B) - m_h(B)| \leq \|g - h\|_{TV}$.

Sei nun $\epsilon \in \mathbb{R}^+$. Für die andere Richtung haben wir die Existenz einer Menge $B \in \mathcal{B}_\Sigma$ zu zeigen, so dass

$$|m_g(B) - m_h(B)| > \frac{1}{2} \|g - h\|_{TV} - \epsilon.$$

Wir finden zunächst ein $t \in \mathbb{N}$, so dass

$$\sum_{\bar{a} \in \Sigma^t} |g(\bar{a}) - h(\bar{a})| > \|g - h\|_{TV} - 2\epsilon.$$

Wir setzen nun

$$\begin{aligned} A_1 &:= \{\bar{a} \in \Sigma^t \mid g(\bar{a}) > h(\bar{a})\}, \\ A_2 &:= \{\bar{a} \in \Sigma^t \mid h(\bar{a}) > g(\bar{a})\} \end{aligned}$$

und folgern

$$\begin{aligned} & (m_g(A_1) - m_h(A_1)) + (m_h(A_2) - m_g(A_2)) \\ &= \sum_{\bar{a} \in A_1} (g(\bar{a}) - h(\bar{a})) + \sum_{\bar{a} \in A_2} (h(\bar{a}) - g(\bar{a})) \\ &= \sum_{\bar{a} \in \Sigma^t} |g(\bar{a}) - h(\bar{a})| > \|g - h\|_{TV} - 2\epsilon. \end{aligned} \tag{2.37}$$

Damit folgt für ein $i \in \{1, 2\}$

$$|m_g(A_i) - m_h(A_i)| > \frac{1}{2} \|g - h\|_{TV} - \epsilon. \tag{2.38}$$

□

BEMERKUNG 2.64. Sind $g, h \in l_{SM}$ zwei stochastische Vektoren, so gilt sogar

$$\|g - h\|_{TV} = 2 \sup_{B \in \mathcal{B}_\Sigma} |m_g(B) - m_h(B)|$$

wegen

$$\begin{aligned}
2|m_g(B) - m_h(B)| &= |m_g(B) - m_h(B)| + |(1 - m_g(B)) - (1 - m_h(B))| \\
&= |m_g(B) - m_h(B)| + |m_g(\mathbb{C}B) - m_h(\mathbb{C}B)| \\
&= \left| \sum_{\bar{a} \in B} g(\bar{a}) - \sum_{\bar{a} \in B} h(\bar{a}) \right| + \left| \sum_{\bar{a} \in \mathbb{C}B} g(\bar{a}) - \sum_{\bar{a} \in \mathbb{C}B} h(\bar{a}) \right| \\
&\leq \sum_{\bar{a} \in B} |g(\bar{a}) - h(\bar{a})| + \sum_{\bar{a} \in \mathbb{C}B} |g(\bar{a}) - h(\bar{a})| \\
&= \sum_{\bar{a} \in \Sigma^t} |g(\bar{a}) - h(\bar{a})| \leq \|g - h\|_{TV}.
\end{aligned}$$

für Zylindermengen $Z(B)$ und demselben Approximationsargument wie oben.

Weiter ergibt sich unmittelbar:

KOROLLAR 2.65. *Seien $(g_n)_{n \in \mathbb{N}}, g \in l_{SM}$ endliche, signierte Maßvektoren. Dann sind die beiden folgenden Aussagen äquivalent:*

(a) *Die Folge der g_n konvergiert in Norm gegen g , d.h.*

$$\lim_{n \rightarrow \infty} \|g_n - g\|_{TV} = 0.$$

(b) *Die Folge m_{g_n} der zu g_n assoziierten Maße konvergiert "schwach Skorokhod" gegen das Maß m_g .*

Beide Aussagen implizieren dann, dass die Folge der zu g_n assoziierten Maße m_{g_n} "stark" gegen das Maß m_g konvergieren.

Jedoch impliziert "starke" Konvergenz nicht die "schwache Skorokhod" Konvergenz. Hier ist ein Gegenbeispiel, das durch das in [JR97] aufgeführte Gegenbeispiel (dort wird ein anderer zugrundeliegender Messraum angegeben) inspiriert ist.

BEISPIEL 2.66. *Wir setzen $\Sigma := \{a, b\}$ und betrachten die folgende Folge von stochastischen Vektoren $(g_n)_{n \in \mathbb{N}}$:*

$$\begin{aligned}
g_n : \{a, b\}^* &\longrightarrow [0, 1] \\
\bar{a} &\longmapsto \begin{cases} \frac{1}{2^{|\bar{a}|}} & t < n \\ \frac{1}{2^{|\bar{a}|-1}} & t \geq n \text{ und } \bar{a}(n) = a \\ 0 & \text{sonst.} \end{cases}
\end{aligned}$$

In Worten beschreibt g_n eine Gleichverteilung über alle Sequenzen, die an n -ter Stelle ein a haben. Wir betrachten nun noch

$$\begin{aligned}
\bar{g} : \{a, b\}^* &\longrightarrow [0, 1] \\
\bar{a} &\longmapsto \frac{1}{2^{|\bar{a}|}},
\end{aligned}$$

also den Vektor, der eine Gleichverteilung über alle Sequenzen ausdrückt. Es ist unmittelbar klar, dass die Folge der g_n wortweise gegen \bar{g} konvergiert. Es

gilt aber

$$\begin{aligned}
\sum_{\bar{a} \in \{a,b\}^n} |g_n(\bar{a}) - \bar{g}(\bar{a})| &= \sum_{\substack{\bar{a} \in \{a,b\}^n \\ \bar{a}(n)=a}} |g_n(\bar{a}) - \bar{g}(\bar{a})| + \sum_{\substack{\bar{a} \in \{a,b\}^n \\ \bar{a}(n)=b}} |g_n(\bar{a}) - \bar{g}(\bar{a})| \\
&= \sum_{\substack{\bar{a} \in \{a,b\}^n \\ \bar{a}(n)=a}} \left| \frac{1}{2^{|\bar{a}|-1}} - \frac{1}{2^{|\bar{a}|}} \right| + \sum_{\substack{\bar{a} \in \{a,b\}^n \\ \bar{a}(n)=b}} \left| \frac{1}{2^{|\bar{a}|}} \right| \\
&= \text{card}(\{a,b\}^{n-1}) \frac{1}{2^{|\bar{a}|}} + \text{card}(\{a,b\}^{n-1}) \frac{1}{2^{|\bar{a}|}} \\
&= 2(2^{n-1} \frac{1}{2^n}) = 1,
\end{aligned}$$

was Konvergenz in Norm ausschließt. Die "starke" Konvergenz kann jedoch folgendermaßen gezeigt werden. Wir bemerken zunächst, dass für alle $B \in \mathcal{B}_\Sigma$ und alle $n \in \mathbb{N}$ wegen $m_{g_n}(T_\Omega^{-n}Z(b)) = 0$ ($T_\Omega^{-n}Z(b)$) sind alle Wörter, die an der n -ten Stelle ein b haben):

$$\begin{aligned}
|m_{g_n}(B) - m_{\bar{g}}(B)| &= |m_{g_n}(T_\Omega^{-n}(Z(a)) \cap B) \\
&\quad - m_{\bar{g}}(T_\Omega^{-n}(Z(a)) \cap B) - m_{\bar{g}}(T_\Omega^{-n}(Z(b)) \cap B)| \\
&\leq |m_{g_n}(T_\Omega^{-n}(Z(a)) \cap B) - m_{\bar{g}}(T_\Omega^{-n}(Z(a)) \cap B)| \\
&\quad + |m_{\bar{g}}(T_\Omega^{-n}(Z(b)) \cap B)| \\
&\leq |m_{\bar{g}}(T_\Omega^{-n}(Z(a)) \cap B)| + |m_{\bar{g}}(T_\Omega^{-n}(Z(b)) \cap B)| \\
&= m_{\bar{g}}(B),
\end{aligned}$$

woraus

$$m_{g_n}(B) \leq 2m_{\bar{g}}(B) \quad (2.39)$$

folgt. Sei nun $B \in \mathcal{B}_\Sigma$ und $\epsilon \in \mathbb{R}^+$. Zunächst gibt es nach dem Approximationssatz 2.18 eine Folge von Zylindermengen B_k mit $\lim_{k \rightarrow \infty} m_{\bar{g}}(B_k \triangle B) = 0$ und wegen Lemma A.3 somit eine Zylindermenge B_K mit

$$m_{\bar{g}}(B_K \triangle B) < \frac{\epsilon}{3}.$$

Aus (2.39) folgt dann

$$m_{g_n}(B_K \triangle B) < \frac{2\epsilon}{3}. \quad (2.40)$$

Sei $A \subset \Sigma^N$ eine definierende Grundmenge von B_K , also $B_K = Z(A)$. Dann folgt nach Definition der g_n

$$\forall n \geq N+1: \quad g_n(Z(A)) = \bar{g}(Z(A)), \quad (2.41)$$

da die g_n für $n \geq N+1$ mit \bar{g} auf Wörtern der Länge kleiner gleich N übereinstimmen. Schließlich folgt für alle $n \geq N+1$:

$$\begin{aligned}
|m_{g_n}(B) - m_{\bar{g}}(B)| &\leq |m_{g_n}(B) - m_{g_n}(B_K) + m_{g_n}(B_K) - m_{\bar{g}}(B)| \\
&\stackrel{(2.41)}{=} |m_{g_n}(B) - m_{g_n}(B_K) + m_{\bar{g}}(B_K) - m_{\bar{g}}(B)| \\
&\leq |m_{g_n}(B) - m_{g_n}(B_K)| + |m_{\bar{g}}(B_K) - m_{\bar{g}}(B)| \\
&\stackrel{L.A.2}{\leq} m_{g_n}(B \triangle B_K) + m_{\bar{g}}(B \triangle B_K) \\
&\stackrel{(2.39),(2.40)}{\leq} \frac{2\epsilon}{3} + \frac{\epsilon}{3} = \epsilon.
\end{aligned}$$

2.4. Ereignisvektoren

Dieser Abschnitt befasst sich mit signierten Maßvektoren, die durch Bedingungen durch Ereignisse aus anderen signierten Maßvektoren entstehen und ist eher technischer Natur. Die hier erzielten Ergebnisse sind jedoch wesentliche Bausteine für die später entwickelte Ergodentheorie. Sie können aber auch ohne spätere Folgeerscheinungen als elementare Erweiterung der bis dato entwickelten Grundbegriffe aufgefasst werden. Insbesondere der Approximationsatz für Ereignisvektoren mag - in Analogie zum Approximationsatz 2.18 - von Interesse sein.

2.4.1. Ereignis- und Zylindervektoren. Zur Einführung der Definitionen erinnern wir an den Zerlegungssatz 2.51 bzw. Definition 2.54 und die Konsequenzen. Jeder Vektor $g \in l_{SM}$ lässt sich demnach als Differenz zweier Maßvektoren g^+, g^- , positiver bzw. negativer Variationsvektor genannt, schreiben:

$$g = g^+ - g^-$$

und wir schreiben $|g| := g^+ + g^-$ und nennen dies totalen Variationsvektor von g . Es gilt dann

$$m_{|g|} = |m_g|,$$

wobei $|m_g|$ die totale Variation (siehe Definition 2.48) des endlichen, signierten Maßes m_g bezeichne.

Wir verwechseln weiterhin konsequent Vektoren g mit den dazu assoziierten Maßen m_g und schreiben somit z.B.

$$|g|(B) := m_{|g|}(B) = |m_g|(B)$$

für alle $B \in \mathcal{B}_\Sigma$.

DEFINITION 2.67. Sei $g \in l_{SM}$ ein endlicher, signierter Maßvektor und sei B ein Ereignis aus \mathcal{B}_Σ mit $|m_g|(B) > 0$. Dann setzen wir

$$\begin{aligned} g^B : \Sigma^* &\longrightarrow \mathbb{R} \\ \bar{a} &\mapsto \frac{|g|_{TV}}{|g|(B)} g(Z(\{\bar{a}\}) \cap B) \end{aligned}$$

und weiter $g^B = 0$ für $|g|(B) = 0$. Im Falle eines stochastischen Vektors g beschreibt g^B also das durch g induzierte Wahrscheinlichkeitsmaß gegeben das Ereignis B . Wir nennen allgemein g^B den durch g und B induzierten **Ereignisvektor**. Ist B eine Zylindermenge, so sprechen wir auch vom durch den von g und B induzierten **Zylindervektor**.

Im Folgenden schreiben wir der Einfachheit halber für Wörter $\bar{b} \in \Sigma^*$ und Mengen von t -Wörtern $B \subset \Sigma^t$

$$g^{\bar{b}} := g^{Z(\{\bar{b}\})}, g^B := g^{Z(B)}$$

für die dazu assoziierten Zylindervektoren. Im Falle, dass ein Wort oder eine Teilmenge von t -Wörtern als Element von \mathcal{B}_Σ aufgefasst werden muss,

so meinen wir damit die dazu assoziierten Zylindermengen. Wir schreiben demnach z.B.

$$g(\bar{a} \cap B)$$

und meinen

$$g(Z(\{\bar{a}\}) \cap B).$$

Die obige Definition mag auf den ersten Blick kompliziert erscheinen. Die folgende Proposition wird diese Definition motivieren und den Sinn der Vorfaktoren $\frac{\|g\|_{TV}}{|g|(B)}$ ersichtlich machen.

PROPOSITION 2.68. *Ist g ein signierter Maßvektor bzw. Maßvektor bzw. stochastischer Vektor, so ist auch g^B ein signierter Maßvektor bzw. Maßvektor bzw. stochastischer Vektor. Ist B ein Ereignis mit $|g|(B) > 0$, so gilt*

$$\|g^B\|_{TV} = \|g\|_{TV}.$$

BEWEIS. Aus

$$\begin{aligned} (\sigma g^B)(\bar{a}) &= \sum_{a \in \Sigma} g^B(\bar{a}a) = \sum_{a \in \Sigma} g(\bar{a}a \cap B) \\ &= g(\bar{a} \cap B) = g^B(\bar{a}) \end{aligned}$$

folgt zunächst $\sigma g^B = g^B$. Aus $g \geq 0$ folgt $g^B \geq 0$, da m_g ein Maß ist und ist zusätzlich $g(\square) = 1$, so wegen $\|g\|_{TV} \stackrel{P.2.41}{=} g(\square) = 1$ und $g = |g|$

$$g^B(\square) = (\|g\|_{TV}/|g|(B))g(\square \cap B) = \frac{1}{|g|(B)}g(B) = 1.$$

Es bleibt, $\|g^B\|_{TV} = \|g\|_{TV}$ zu zeigen. Dazu wählen wir gemäß Hahn-Zerlegung 2.44 zwei Mengen Ω^+, Ω^- , so dass $\Omega_\Sigma = \Omega^+ \dot{\cup} \Omega^-$ und $g(A \cap \Omega^+) \geq 0, g(A \cap \Omega^-) \leq 0$ für alle $A \in \mathcal{B}_\Sigma$. Offensichtlich ist diese Zerlegung dann auch eine Hahn-Zerlegung für g^B , also $g^B(A \cap \Omega^+) \geq 0$ und $g^B(A \cap \Omega^-) \leq 0$. Gemäß der Definition der Jordan-Zerlegung und den Erkenntnissen aus Bemerkung 2.53 gilt daher für die Zerlegung $g^B = (g^B)^+ - (g^B)^-$ gemäß Zerlegungssatz 2.51:

$$\begin{aligned} (g^B)^{+,-}(\bar{a}) &= g^B(\bar{a} \cap \Omega^{+,-}) = \frac{\|g\|_{TV}}{|g|(B)}g(\bar{a} \cap B \cap \Omega^{+,-}) \\ &= \frac{\|g\|_{TV}}{|g|(B)}g^{+,-}(\bar{a} \cap B), \end{aligned}$$

mit anderen Worten $(g^B)^+ = (g^+)^B$ sowie $(g^B)^- = (g^-)^B$. Weiter gilt

$$\begin{aligned} \|(g^{+,-})^B\|_{TV} &\stackrel{P.2.41}{=} (g^{+,-})^B(\square) \\ &= \frac{\|g^{+,-}\|_{TV}}{g^{+,-}(B)}m_{g^{+,-}}(B \cap \square) = \|g^{+,-}\|_{TV}. \end{aligned}$$

Deshalb

$$\begin{aligned} \|g^B\|_{TV} &= \|(g^B)^+\|_{TV} + \|(g^B)^-\|_{TV} = \|(g^+)^B\|_{TV} + \|(g^-)^B\|_{TV} \\ &= \|g^+\|_{TV} + \|g^-\|_{TV} = \|g\|_{TV}. \end{aligned}$$

□

Aufgrund der einfacheren Handhabbarkeit von Zylindermengen im Vergleich mit beliebigen Ereignissen sind Zylindervektoren ebenfalls einfacher als beliebige Ereignisvektoren zu begreifen. Für Maßvektoren gestalten sich die Zylindervektoren besonders übersichtlich:

BEISPIEL 2.69. Ist $g \in l_{SM}, g \geq 0$ ein Maßvektor $\bar{b} \in \Sigma^t$ ein t -Wort mit $|g|(\bar{b}) = g(\bar{b}) > 0$, so gilt für $\bar{a} \in \Sigma^*$

$$g^{\bar{b}}(\bar{a}) = \begin{cases} \|g\|_{TV} & \bar{a} \leq \bar{b} \\ \|g\|_{TV} \frac{g(\bar{a})}{g(\bar{b})} & \bar{b} \leq \bar{a} \\ 0 & \text{sonst.} \end{cases}$$

Ist $B \subset \Sigma^t$ eine Teilmenge von t -Wörtern mit $g(B) > 0$, so gilt

$$\begin{aligned} g^B(\bar{a}) &= \frac{\|g\|_{TV}}{g(B)} g(\bar{a} \cap B) = \frac{\|g\|_{TV}}{g(B)} g(Z(\{\bar{a}\}) \cap (\bigcup_{\bar{b} \in B} Z(\{\bar{b}\}))) \\ &= \frac{\|g\|_{TV}}{g(B)} g(\bigcup_{\bar{b} \in B} (Z(\{\bar{a}\}) \cap Z(\{\bar{b}\}))) \\ &= \frac{\|g\|_{TV}}{g(B)} \sum_{\bar{b} \in B} g(Z(\{\bar{a}\}) \cap Z(\{\bar{b}\})) \\ &= \begin{cases} \frac{\|g\|_{TV}}{g(B)} \sum_{\bar{b} \in B, \bar{a} \leq \bar{b}} g(\bar{b}) & |\bar{a}| \leq t, \\ \frac{\|g\|_{TV}}{g(B)} g(\bar{a}) & |\bar{a}| > t, \exists \bar{b} \in B : \bar{b} \leq \bar{a}, \\ 0 & |\bar{a}| > t, \forall \bar{b} \in B : \bar{b} \not\leq \bar{a}. \end{cases} \end{aligned}$$

Der folgende Satz zeigt, dass sich Konvergenz in Norm auf Ereignisvektoren überträgt.

SATZ 2.70. Sei $(g_n)_{n \in \mathbb{N}}, g_n \in l_{SM}$ eine Folge endlicher, signierter Maßvektoren, die in Norm gegen einen Vektor $g \in l_{SM}$ konvergiert. Sei weiter $B \in \mathcal{B}_\Sigma$ ein Ereignis, so dass $|g|(B) > 0$. Dann

$$\lim_{n \rightarrow \infty} \|g_n^B - g^B\|_{TV} = 0.$$

BEWEIS. Ohne Einschränkung ($\lim_{n \rightarrow \infty} |g_n|(B) = |g|(B)$, siehe Satz 2.55 und Korollar 2.65) sei $|g_n|(B) > 0$ für alle n . Sei $\epsilon \in \mathbb{R}^+$. Zunächst gibt es wegen (Satz 2.55) $\lim_{n \rightarrow \infty} \frac{\|g_n\|_{TV}}{|g_n|(B)} = \frac{\|g\|_{TV}}{|g|(B)}$ ein $C > 0$, so dass

$$\forall n \in \mathbb{N} : \frac{\|g_n\|_{TV}}{|g_n|(B)} \leq C.$$

Weiter gibt es aufgrund von Satz 2.63 ein $N_0 \in \mathbb{N}$, so dass (beachte, dass wegen $|g|(B) > 0$ auch $\|g\|_{TV} > 0$)

$$\forall n \geq N_0 : \sup_{A \in \mathcal{B}_\Sigma} |g_n(A) - g(A)| < \frac{\epsilon}{2(\|g\|_{TV} + C)} \quad (2.42)$$

und weiter ein $N_1 \in \mathbb{N}$, so dass

$$\forall n \geq N_1 : \quad \left| \frac{\|g_n\|_{TV}}{|g_n|(B)} - \frac{\|g\|_{TV}}{|g|(B)} \right| < \frac{\epsilon}{2(\|g\|_{TV} + C)}. \quad (2.43)$$

Deshalb für $n \geq \max\{N_0, N_1\}$ und $A \in \mathcal{B}_\Sigma$ (setze $b_n := \frac{\|g_n\|_{TV}}{|g_n|(B)}$ und $b := \frac{\|g\|_{TV}}{|g|(B)}$)

$$\begin{aligned} |g_n^B(A) - g^B(A)| &= |b_n g_n(A \cap B) - b g(B \cap A)| \\ &= |b_n g_n(A \cap B) - b_n g(A \cap B) + b_n g(A \cap B) - b g(A \cap B)| \\ &\leq |b_n g_n(A \cap B) - b_n g(A \cap B)| + |b_n g(A \cap B) - b g(A \cap B)| \\ &= |b_n| |g_n(A \cap B) - g(A \cap B)| + |g(A \cap B)| |b_n - b| \\ &\stackrel{(2.42), (2.43)}{<} (|b_n| + |g(A \cap B)|) \frac{\epsilon}{2(\|g\|_{TV} + C)} \leq \frac{\epsilon}{2}, \end{aligned} \quad (2.44)$$

woraus mit Satz 2.63 $\|g_n^B - g^B\|_{TV} \leq \epsilon$ und damit die Behauptung folgt. \square

2.4.2. Approximation von Ereignisvektoren durch Zylindervektoren. Das Beispiel des letzten Abschnitts zeigt, dass Zylindervektoren einfach handhabbare Objekte sind. Beliebige Ereignisvektoren hingegen können sich als deutlich störrischer erweisen. Dieses Problem ist analog zu dem der Beschreibung beliebiger Ereignisse selbst (siehe dazu den Kommentar vor dem Approximationssatz 2.18). Wir streben daher nach einem Satz, der uns die Approximation beliebiger Ereignisvektoren durch Zylindervektoren garantiert. Um genau zu sein, schwebt uns folgendes Resultat vor:

SATZ 2.71 (Approximationssatz für Vektoren). *Sei $g \in l_{SM}$ und $A \in \mathcal{B}_\Sigma$. Dann gibt es eine Folge von Zylindermengen $(A_n)_{n \in \mathbb{N}}$, so dass*

$$\lim_{n \rightarrow \infty} \|g^{A_n} - g^A\|_{TV} = 0.$$

Wir werden im Folgenden den obigen Satz als Korollar des folgenden zentralen Lemmas erhalten. Wir bemerken vorab, dass die Aussage dieses Lemmas mehr als nötig bereitstellt. Zum Beweis von Satz 2.71 ist tatsächlich weniger erforderlich.

LEMMA 2.72. *Sei $g \in l_{SM}$ und A, B zwei Ereignisse mit $|g|(A), |g|(B) > 0$. Bezeichnen wir mit $A \triangle B := (A \setminus B) \cup (B \setminus A)$ die symmetrische Differenz der beiden Ereignisse, dann gilt:*

$$\|g^A - g^B\|_{TV} \leq \frac{2\|g\|_{TV}^2}{|g|(A)|g|(B)} |g|(A \triangle B).$$

Wir zeigen zunächst das Resultat für A, B Zylindermengen.

LEMMA 2.73. Sei $g \in l_{SM}$. Seien weiter zwei Zylindermengen A und B mit $|g|(A), |g|(B) > 0$ gegeben. Dann gilt:

$$\|g^A - g^B\|_{TV} \leq \frac{2\|g\|_{TV}^2}{|g|(A)|g|(B)} |g|(A \triangle B).$$

BEWEIS. Wir identifizieren A und B mit definierenden Grundmengen $A, B \subset \Sigma^n$ (ohne Einschränkung wählen wir dabei jeweils Mengen von n -Wörtern). Sei $m \geq n$. Durch Übergang zu $\Sigma_A^m := \cup_{\bar{a} \in A} \Sigma_{\bar{a}}^m, \Sigma_B^m := \cup_{\bar{a} \in B} \Sigma_{\bar{a}}^m$ fassen wir A und B ebenso als Teilmengen von Σ^m auf. Es gilt dann (siehe Beispiel 2.69) $g^A(\bar{a}) = g^B(\bar{a}) = 0$ für $\bar{a} \in \Sigma^m \setminus (A \cup B)$. Deshalb

$$\sum_{\bar{a} \in \Sigma^m} |g^A(\bar{a}) - g^B(\bar{a})| = \sum_{\bar{a} \in A \cup B} |g^A(\bar{a}) - g^B(\bar{a})|.$$

Wegen $A \cup B = (A \triangle B) \dot{\cup} (A \cap B)$ untersuchen wir

$$\begin{aligned} \sum_{\bar{a} \in (A \cap B) \subset \Sigma^m} |g^A(\bar{a}) - g^B(\bar{a})| &= \sum_{\bar{a} \in (A \cap B) \subset \Sigma^m} \left| \frac{\|g\|_{TV}}{|g|(A)} g(\bar{a}) - \frac{\|g\|_{TV}}{|g|(B)} g(\bar{a}) \right| \\ &= \|g\|_{TV} \sum_{\bar{a} \in (A \cap B)} |g(\bar{a})| \left| \frac{1}{|g|(A)} - \frac{1}{|g|(B)} \right| \\ &\stackrel{L.A.2}{=} \|g\|_{TV} \sum_{\bar{a} \in (A \cap B)} |g(\bar{a})| \frac{||g|(A) - |g|(B)|}{|g|(A)|g|(B)} \\ &\leq \|g\|_{TV} \sum_{\bar{a} \in A \cap B} |g(\bar{a})| \frac{|g|(A \triangle B)}{|g|(A)|g|(B)} \\ &\leq \frac{\|g\|_{TV}^2}{|g|(A)|g|(B)} |g|(A \triangle B), \end{aligned} \tag{2.45}$$

als auch

$$\begin{aligned} \sum_{\bar{a} \in (A \triangle B) \subset \Sigma^m} |g^A(\bar{a}) - g^B(\bar{a})| &= \sum_{\bar{a} \in (A \setminus B)} \left| \frac{\|g\|_{TV}}{|g|(A)} g(\bar{a}) \right| + \sum_{\bar{a} \in (B \setminus A)} \left| \frac{\|g\|_{TV}}{|g|(B)} g(\bar{a}) \right| \\ &\stackrel{|g(\bar{a})| \leq |g|(\bar{a})}{\leq} \frac{\|g\|_{TV}}{|g|(A)} |g|(A \setminus B) + \frac{\|g\|_{TV}}{|g|(B)} |g|(B \setminus A) \\ &= \frac{\|g\|_{TV} |g|(B)}{|g|(A)|g|(B)} |g|(A \setminus B) + \frac{\|g\|_{TV} |g|(A)}{|g|(A)|g|(B)} |g|(B \setminus A) \\ &= \frac{\|g\|_{TV}^2}{|g|(A)|g|(B)} |g|(A \setminus B) + \frac{\|g\|_{TV}^2}{|g|(A)|g|(B)} |g|(B \setminus A) \\ &= \frac{\|g\|_{TV}^2}{|g|(A)|g|(B)} |g|(A \triangle B). \end{aligned} \tag{2.46}$$

Setzt man diese beiden Abschätzungen zusammen, so folgt die Behauptung aufgrund $\|g^A - g^B\| = \lim_{m \rightarrow \infty} \sum_{\bar{a} \in \Sigma^m} |g^A(\bar{a}) - g^B(\bar{a})|$, siehe Proposition 2.41. \square

Mit Hilfe dieses Lemmas zeigen wir nun:

LEMMA 2.74. *Sei $g \in l_{SM}$ und A ein Ereignis mit $|g|(A) > 0$. Sei A_n eine Folge von Zylindermengen mit*

$$|g|(A_n \triangle A) \xrightarrow{n \rightarrow \infty} 0.$$

Dann folgt

$$\|g^{A_n} - g^A\|_{TV} \xrightarrow{n \rightarrow \infty} 0.$$

BEWEIS. Zunächst folgt

$$\begin{aligned} \|g^{A_n} - g^{A_m}\|_{TV} &\stackrel{L.2.73}{\leq} \frac{2\|g\|_{TV}^2}{|g|(A_n)|g|(A_m)} |g|(A_n \triangle A_m) \\ &\stackrel{L.A.3}{\leq} \frac{2\|g\|_{TV}^2}{|g|(A_n)|g|(A_m)} (|g|(A_n \triangle A) + |g|(A \triangle A_m)) \xrightarrow{m,n \rightarrow \infty} 0. \end{aligned} \quad (2.47)$$

Daraus folgt wiederum, dass die g^{A_n} eine Cauchy-Folge in l_{TV} bilden. Wegen der Vollständigkeit von l_{TV} (Satz 2.40) konvergieren die g^{A_n} also gegen einen Vektor $\bar{g} \in l_{TV}$. Sei nun $\bar{a} \in \Sigma^*$. Mit Hilfe der Lemmata A.1 und A.2 folgern wir

$$|g(\bar{a} \cap A_n) - g(\bar{a} \cap A)| \stackrel{L.A.2}{\leq} |g|((\bar{a} \cap A_n) \triangle (\bar{a} \cap A)) \stackrel{L.A.1}{\leq} |g|(A_n \triangle A)$$

und damit $\lim_{n \rightarrow \infty} g(\bar{a} \cap A_n) = g(\bar{a} \cap A)$. Da wegen Lemma A.2 und Satz 2.55 auch $\lim_{n \rightarrow \infty} |g|(A_n) = |g|(A)$ gilt, folgt

$$g^{A_n}(\bar{a}) = \frac{\|g\|_{TV}}{|g|(A_n)} g(\bar{a} \cap A_n) \xrightarrow{n \rightarrow \infty} \frac{\|g\|_{TV}}{|g|(A)} g(\bar{a} \cap A) = g^A(\bar{a}),$$

die g^{A_n} konvergieren also punktweise gegen g^A . Daher folgt $\bar{g} = g^A$ und damit die Behauptung. \square

An dieser Stelle sind wir nun in der Lage, einen **Beweis für Satz 2.71** liefern zu können: Wir finden die in Lemma 2.74 geforderte Folge mit Hilfe des Approximationssatzes 2.18 und erhalten die Aussage. \square

Der Vollständigkeit und Allgemeinheit halber geben wir nichtsdestotrotz nun noch den

Beweis von Lemma 2.72: Mit dem Approximationssatz 2.18 finden wir zunächst Folgen von Zylindermengen A_n, B_n , so dass

$$\lim_{n \rightarrow \infty} |g|(A_n \triangle A) = 0 \text{ und } \lim_{n \rightarrow \infty} |g|(B_n \triangle B) = 0.$$

Mit Lemma A.4 folgt

$$\lim_{n \rightarrow \infty} |g|(A_n \triangle B_n) = |g|(A \triangle B) \quad (2.48)$$

und mit Lemma A.2

$$\lim_{n \rightarrow \infty} |g|(A_n)|g|(B_n) = |g|(A)|g|(B). \quad (2.49)$$

Sei nun $\epsilon \in \mathbb{R}^+$. Wegen (2.48), (2.49) findet man nun ein $N_1 \in \mathbb{N}$, so dass

$$\forall n \geq N_1 : \left| \frac{2\|g\|_{TV}^2}{|g|(A_n)|g|(B_n)} |g|(A_n \triangle B_n) - \frac{2\|g\|_{TV}^2}{|g|(A)|g|(B)} |g|(A \triangle B) \right| < \frac{\epsilon}{3}. \quad (2.50)$$

Wegen Lemma 2.74 konvergieren die g^{A_n}, g^{B_n} nun gegen g^A, g^B . Wir finden daher weiter ein $N_2 \in \mathbb{N}$, so dass

$$\forall n \geq N_2 : \|g^A - g^{A_n}\|_{TV} < \frac{\epsilon}{3}, \quad (2.51)$$

und ein $N_3 \in \mathbb{N}$, so dass

$$\forall n \geq N_3 : \|g^B - g^{B_n}\|_{TV} < \frac{\epsilon}{3}. \quad (2.52)$$

Insgesamt erhält man für $n \geq \max\{N_1, N_2, N_3\}$:

$$\begin{aligned} \|g^A - g^B\|_{TV} &= \|g^A - g^{A_n} + g^{A_n} - g^{B_n} + g^{B_n} - g^B\|_{TV} \\ &\leq \|g^{A_n} - g^{B_n}\|_{TV} + \|g^{A_n} - g^A\|_{TV} + \|g^{B_n} - g^B\|_{TV} \\ &\stackrel{L.2.73,(2.51),(2.52)}{\leq} \frac{2\|g\|_{TV}^2}{|g|(A_n)|g|(B_n)} |g|(A_n \triangle B_n) + \frac{\epsilon}{3} + \frac{\epsilon}{3} \\ &\stackrel{(2.50)}{\leq} \frac{2\|g\|_{TV}^2}{|g|(A)|g|(B)} |g|(A \triangle B) + \frac{\epsilon}{3} + 2\frac{\epsilon}{3} \\ &= \frac{2\|g\|_{TV}^2}{|g|(A)|g|(B)} |g|(A \triangle B) + \epsilon, \end{aligned} \quad (2.53)$$

und damit die Behauptung. \square

Als Korollar erhalten wir dann eine Verschärfung von Lemma 2.74:

KOROLLAR 2.75. *Sei $g \in l_{SM}$ und A ein Ereignis mit $|g(A)| > 0$. Sei A_n eine Folge beliebiger Ereignisse mit*

$$|g|(A_n \triangle A) \xrightarrow{n \rightarrow \infty} 0.$$

Dann folgt

$$\|g^{A_n} - g^A\|_{TV} \xrightarrow{n \rightarrow \infty} 0.$$

BEWEIS. Der Beweis verläuft analog zu dem von Lemma 2.74, wenn man sich jeweils an den Stellen, die Lemma 2.73 benötigen, auf Lemma 2.72 bezieht. \square

BEMERKUNG 2.76. *Wir wollen schlussendlich bemerken, dass man die einen beliebigen Ereignisvektor g^A approximierenden Zylindervektoren g^{A_n} aus Satz 2.71 immer so wählen kann, dass für die definierenden Grundmengen A_n jeweils $|A_n| = n$ gilt.*

Operatoren und Dimension

*Verbringe nicht die Zeit
mit der Suche nach einem Hindernis.
Vielleicht ist keines da.*

Franz Kafka

Nachdem im vorhergehenden Kapitel der Vektorraum, der Gegenstand unserer Betrachtungen sein wird, samt seiner maßtheoretischen Interpretation vorgestellt worden ist, beschäftigt sich dieses Kapitel mit Operatoren, die auf diesem Vektorraum agieren. Diese Operatoren kann man wiederum wahrscheinlichkeitstheoretisch interpretieren und sie führen auf die Definition der für diese Arbeit wichtigen Dimensionsbegriffe. Prominente Vertreter für demgemäß endlich-dimensionale Vektoren sind stochastische Vektoren, die als Hidden Markov Prozesse interpretiert werden können.

Wir wollen an dieser Stelle bemerken, dass Heller in seiner Arbeit [Hel65] den Nutzen von Vektorräumen zur Beschreibung von stochastischen Prozessen anzweifelt, da eine “natürliche” Basis für die gegebenen Vektorräume nicht ersichtlich ist und betrachtet stochastische Prozesse als Moduln über Polynomalgebren. Mit seiner Kritik bezieht er sich auf die drei Arbeiten von Dharmadhikari ([Dha63a],[Dha63b],[Dha65]), in denen ein erster Versuch unternommen wird, stochastische Prozesse als Elemente von Vektorräumen zu betrachten. Dharmadhikari klammert sich jedoch an die durch einzelne Prozesse induzierten Vektorräume und untersucht Koordinatendarstellungen derselben. Heller’s Kritik an diesem Ansatz scheint daher durchaus berechtigt. Die Einführung von Operatoren auf diesen Vektorräumen ist schwierig - sie werden zunächst in Abhängigkeit einer wie auch immer gewählten Basis definiert und benötigen in der Folge einen Beweis, dass sie von der Auswahl derselben unabhängig sind. Wir betrachten die von Dharmadhikari eingeführten Vektorräume als Untervektorräume der im vorigen Kapitel beschriebenen, allgemeineren Vektorräume. Auf diesen lassen sich die von Dharmadhikari beschriebenen und von Heller als unschön abgetanen Operatoren *natürlich* definieren. Wir plädieren daher für unseren Ansatz als einen dem von Heller gegenüber **Elementareren** (da nur einfache Begriffe wie Folgen und Vektorräume benötigt werden) und als einen dem von Dharmadhikari gegenüber **Kanonischeren**.

3.1. Prognoseoperatoren

In diesem Abschnitt werden nun die Operatoren eingeführt, die für die Bildung der Dimensionsbegriffe nötig sind. Prognoseoperatoren sind, wie bereits oben erwähnt, nichts wesentlich Neues. Zusätzlich zu den oben erwähnten Texten werden sie in englischsprachigen Texten von Jäger unter dem Namen “Observable Operators” geführt ([Jae97a], [Jae00], [JZK⁺05]). Die hier gegebene Definition hat, wie ebenfalls oben erwähnt, nicht das Problem, den Nachweis der Unabhängigkeit von der Auswahl einer Basis führen zu müssen. Darüberhinaus können sie so in einfache Zusammenhänge mit der Norm der totalen Variation gestellt werden. So werden sie sich z.B. als stetige Operatoren erweisen.

3.1.1. Definitionen und Notationen. Wir starten mit der Auflistung der maßgeblichen Definitionen.

DEFINITION 3.1. Sei $a \in \Sigma$. Dann nennen wir den linearen Operator

$$\tau_{\bar{a}} : l_{\infty} \longrightarrow l_{\infty},$$

der durch die Vorschrift

$$(\tau_{\bar{a}}g)(\bar{c}) := g(\bar{a}\bar{c})$$

definiert wird, den **Prognoseoperator zum Wort** oder auch **Geschichte** \bar{a} . Wir nennen desweiteren den linearen Operator

$$\mu : l_{\infty} \longrightarrow l_{\infty},$$

der durch die Vorschrift

$$(\mu g)(\bar{c}) := \sum_{a \in \Sigma} g(a\bar{c})$$

festgelegt wird, den **Evolutionoperator**.

Diese Definition lässt sich dann auch auf Mengen von Wörtern verallgemeinern:

DEFINITION 3.2. Wir verallgemeinern diese Definition von Wörtern auf beliebige endliche Mengen gleichlanger Wörter, wobei wir für eine solche Menge $A \subset \Sigma^t$

$$\begin{array}{ccc} \tau_A : l_{\infty} & \longrightarrow & l_{\infty} \\ g & \mapsto & \sum_{\bar{a} \in A} \tau_{\bar{a}}g \end{array}$$

definieren. Offensichtlich

$$\tau_A g(\bar{b}) = \sum_{\bar{a} \in A} g(\bar{a}\bar{b}).$$

Wir bezeichnen diese Operatoren analog als **Prognoseoperatoren zur Geschichte** A .

Die Prognoseoperatoren und der Evolutionoperator stehen natürlich in einem Zusammenhang:

PROPOSITION 3.3. *Ist $g \in l_\infty$, so gilt*

$$\mu(g) = \sum_{a \in \Sigma} \tau_a(g).$$

Iterierte Anwendung von μ ergibt

$$\mu^t(g) = \sum_{\bar{a} \in \Sigma^t} \tau_{\bar{a}}(g).$$

BEWEIS. Der zweite Teil folgt mit multinomialer Entwicklung. \square

PROPOSITION 3.4. *Ist $g \in l_\sigma$ ein (signierter) Maßvektor, so auch $\tau_a g$ für alle $a \in \Sigma$ und μg ein (signierter) Maßvektor. Ist g ein stochastischer Vektor, so auch μg .*

BEWEIS. Sei $a \in \Sigma$ und $\bar{a} \in \Sigma^*$. Dann

$$\sigma \tau_a g(\bar{a}) = \sum_{b \in \Sigma} \tau_a(\bar{a}b) = \sum_{b \in \Sigma} g(a\bar{a}) \stackrel{\sigma g = g}{=} g(a\bar{a}) = \tau_a g(\bar{a}).$$

Mit Proposition 3.3 folgt dasselbe für μ . Ist $g \geq 0$, so nach Definition offensichtlich auch $\tau_a g$ und μg . Gilt $g(\square) = 1$ und $\sigma g = g$, so $\mu g(\square) = \sum_{a \in \Sigma} g(a) \stackrel{\sigma g = g}{=} g(\square) = 1$, also ist μg ein stochastischer Vektor, falls es g ist. \square

Bevor wir uns mit den zentralen Ergebnissen beschäftigen, notieren wir noch eine später hilfreiche Proposition:

PROPOSITION 3.5. *Seien $h, g \in l_\infty$ zwei Vektoren. Dann gilt $g = h$ genau dann, wenn*

$$g(\square) = h(\square) \quad \wedge \quad \forall a \in \Sigma : \tau_a h = \tau_a g. \quad (3.1)$$

BEWEIS. Die erste Richtung ist trivial. Gelte nun $g \neq h$. Das heißt, es gibt ein Wort \bar{a} , so dass $g(\bar{a}) \neq h(\bar{a})$. Im Falle $\bar{a} = \square$ sind wir fertig. Andernfalls gilt $\tau_a h \neq \tau_a g$ für den Anfangsbuchstaben von \bar{a} . \square

3.1.2. Stetigkeit und Stabilität. Hauptergebnis dieses Abschnitts ist, dass die Prognoseoperatoren stetig bezüglich der TV -Norm sind.

SATZ 3.6. *Es gilt $\tau_a(l_{TV}) \subset l_{TV}$ für alle $a \in \Sigma$ sowie $\mu(l_{TV}) \subset l_{TV}$. Die linearen Operatoren*

$$\tau_a : (l_{TV}, \|\cdot\|_{TV}) \rightarrow (l_{TV}, \|\cdot\|_{TV})$$

sowie

$$\mu : (l_{TV}, \|\cdot\|_{TV}) \rightarrow (l_{TV}, \|\cdot\|_{TV})$$

sind stetig und es gilt für alle $a \in \Sigma$

$$\|\tau_a\|_{TV} = \|\mu\|_{TV} = 1.$$

BEWEIS. Sei $a \in \Sigma$. Aus den Rechnungen

$$\sum_{\bar{a} \in \Sigma^t} |\tau_a g(\bar{a})| = \sum_{\bar{a} \in \Sigma^t} |g(a\bar{a})| \leq \sum_{a \in \Sigma} \sum_{\bar{a} \in \Sigma^t} |g(a\bar{a})| = \sum_{\bar{a} \in \Sigma^{t+1}} |g(\bar{a})| \leq \|g\|_{TV} \quad (3.2)$$

und analog

$$\sum_{\bar{a} \in \Sigma^t} |\mu g(\bar{a})| = \sum_{\bar{a} \in \Sigma^t} \left| \sum_{a \in \Sigma} g(a\bar{a}) \right| \leq \sum_{a \in \Sigma} \sum_{\bar{a} \in \Sigma^t} |g(a\bar{a})| = \sum_{\bar{a} \in \Sigma^{t+1}} |g(\bar{a})| \leq \|g\|_{TV} \quad (3.3)$$

folgen $\tau(l_{TV}) \subset l_{TV}$ und $\mu(l_{TV}) \subset l_{TV}$ sowie darüberhinaus $\|\tau_a\| \leq 1$ und $\|\mu\| \leq 1$. Setzt man jeweils für jedes $a \in \Sigma$

$$f_a : \Sigma^* \longrightarrow \mathbb{R} \\ \bar{a} \longmapsto \begin{cases} 1 & \bar{a} = a^t \text{ für ein } t \\ 0 & \text{sonst,} \end{cases}$$

so gilt jeweils $\|\tau_a f_a\|_{TV} = \|f_a\|_{TV}$ und auch $\|\mu f_a\|_{TV} = \|f_a\|_{TV}$. Es folgen $\|\tau_a\| = 1$ und $\|\mu\| = 1$ und damit die Behauptung, da die Stetigkeit linearer Operatoren äquivalent zur Endlichkeit der Operatornorm ist. \square

Eine unmittelbare und wichtige Folgerung ist zunächst:

KOROLLAR 3.7. *Für alle $a \in \Sigma$ gilt $\tau_a(l_{SM}) \subset l_{SM}$ und außerdem $\mu(l_{SM}) \subset l_{SM}$. Ist g ein stochastischer Vektor, so dann offensichtlich auch μg .*

Als weitere Folgerung dieses Satzes ein zweiter, der sich in der Ergodentheorie von großem Nutzen erweisen wird.

SATZ 3.8. *Die Prognoseoperatoren*

$$\tau_a : l_{TV} \longrightarrow l_{TV}$$

und der Evolutionsoperator

$$\mu : l_{TV} \longrightarrow l_{TV}$$

sind stabil.

BEWEIS. Wir haben $\|\mu^t\|_{TV} \leq K$ für ein $K \in \mathbb{R}^+$ zu zeigen. Das folgt mit Satz 3.6 aus

$$\|\mu^t\|_{TV} \leq \|\mu\|_{TV}^t = 1.$$

Analog erweisen sich die Prognoseoperatoren als stabil. \square

BEMERKUNG 3.9. *Hingegen ist*

$$\mu : (l_\infty, \|\cdot\|_\infty) \longrightarrow (l_\infty, \|\cdot\|_\infty)$$

zwar noch stetig, aber nicht mehr stabil. Für $g \in l_\infty$ mit $g(\bar{a}) = 1$ für alle $\bar{a} \in \Sigma$ gilt nämlich

$$\mu^k g(\bar{a}) = \text{card}(\Sigma)^k$$

für alle $\bar{a} \in \Sigma$ und somit

$$\|\mu^k g\|_\infty \xrightarrow{k \rightarrow \infty} \infty,$$

woraus die Instabilität von μ auf l_∞ folgt. Für beliebiges g gilt jedoch

$$|\mu g(\bar{a})| = \left| \sum_{a \in \Sigma} g(a\bar{a}) \right| \leq \sum_{a \in \Sigma} |g(a\bar{a})| \leq \text{card}(\Sigma) \cdot \|g\|_\infty,$$

woraus mit dem obigen Beispiel $\|\mu\|_\infty = \text{card}(\Sigma) < \infty$ und damit die Stetigkeit von μ folgt.

3.1.3. Prognosevektoren und Evolution stochastischer Prozesse. Wir reichen abschließend eine wahrscheinlichkeitstheoretische Motivation der Definition der Prognoseoperatoren nach. Durch Prognoseoperationen modifizierte stochastische Vektoren können als der ursprünglich gegebene stochastische Prozess unter neuen Bedingungen interpretiert werden können.

Diese Interpretation wirft die Frage nach dem Zusammenhang mit den Ereignisvektoren auf, die ja für stochastische Prozesse ebenfalls gewisse Bedingungen kodieren. Auch diesen Zusammenhang klären wir auf.

DEFINITION 3.10. Ist $g \in l_{SM}$ ein endlicher, signierter Maßvektor und $A \subset \Sigma^t$ eine Menge von t -Wörtern, so setzen wir

$$g_A = \begin{cases} \frac{\|g\|_{TV}}{|g|(A)} \tau_{Ag} & |g|(A) > 0 \\ 0 & \text{sonst.} \end{cases}$$

und nennen diesen Vektor **Prognosevektor zur Geschichte A** .

Für Wörter $\bar{a} \in \Sigma^*$ schreiben wir der Einfachheit halber auch

$$g_{\bar{a}} := g_{\{\bar{a}\}}$$

Das folgende Beispiel hilft dann, die Motivation der Prognosevektoren zu verstehen.

BEISPIEL 3.11. Sei $g \in l_{SM}$ ein stochastischer Vektor und $\bar{a} = a_1 \dots a_t \in \Sigma^t$. Dann gilt

$$g(\bar{a}) = g(a_1) \prod_{s=1}^{t-1} g_{a_1 \dots a_s}(a_{s+1}).$$

Das folgt, da $\|g\|_{TV} = g(\square) = 1$ (siehe Prop. 2.41), da g ein stochastischer Vektor ist, aus

$$g_{a_1 \dots a_s}(a_{s+1}) = \frac{1}{g(a_1 \dots a_s)} \tau_{a_1 \dots a_s} g(a_{s+1}) = \frac{g(a_1 \dots a_s a_{s+1})}{g(a_1 \dots a_s)}.$$

Das nun folgende Lemma klärt den oben angesprochenen Zusammenhang auf:

LEMMA 3.12. *Ist $g \in l_{SM}$ und $A \subset \Sigma^t$ eine Menge von t -Wörtern, so*

$$g_A = \mu^t g^A = \tau_A g^A. \quad (3.4)$$

BEWEIS. Sei $A = \{\bar{a}_1, \dots, \bar{a}_k\} \subset \Sigma^t$. Der Fall $|g|(A) = 0$ ist trivial. Sei daher $|g|(A) > 0$ und $\bar{b} \in \Sigma^*$, so gilt

$$\begin{aligned} \mu^t g^A(\bar{b}) &= \sum_{\bar{a} \in \Sigma^t} g^A(\bar{a}\bar{b}) \stackrel{g^A(\bar{a})=0, \bar{a} \notin A}{=} \sum_{\bar{a} \in A} g^A(\bar{a}\bar{b}) \\ &= \tau_A g^A(\bar{b}) \\ &= \frac{\|g\|_{TV}}{|g|(A)} \sum_{\bar{a} \in A} g(\bar{a}\bar{b}) = \frac{\|g\|_{TV}}{|g|(A)} \tau_A g(\bar{b}), \end{aligned} \quad (3.5)$$

woraus die Behauptungen folgen. \square

Die Bilder von endlichen, signierten Maßvektoren unter der Anwendung von Prognoseoperatoren erzielt man also durch die einfache Evolution von Zylindervektoren.

Die folgende Bemerkung zeigt, wie die Prognosevektoren stochastischer Vektoren nun in der Sprache der stochastischen Prozesse interpretiert werden können.

BEMERKUNG 3.13. *Sei $g \in l_{SM}$ ein stochastischer Vektor (also $g \geq 0$ und $g(\square) = 1$) und $(X_t)_{t \in \mathbb{N}}$ der dazu assoziierte stochastische Prozess. Sei weiter $A \subset \Sigma^s$ eine Menge von s -Wörtern. Dann ist der zu A assoziierte Prognosevektor g_A ebenfalls ein stochastischer Vektor gegeben und für den g_A zugeordneten stochastischen Prozess $(Y_t)_{t \in \mathbb{N}}$ gilt:*

$$P(Y_t = a) = P(X_{t+s} = a \mid X_0 X_1 \dots X_{s-1} \in A).$$

Die Verteilungen des stochastischen Prozesses (Y_t) entsprechen also denen von (X_{t+s}) , falls als bekannt angenommen werden darf, die ersten s Zufallsvariablen $X_0 \dots X_{s-1}$ hätten ein Wort aus A erzeugt.

Im Falle, dass $A = \Sigma^t$ für ein $t \in \mathbb{N}$ ergibt sich eine spezielle Situation, die mit Hilfe des in Abschnitt 2.2.5 eingeführten Schiebeoperators charakterisiert werden kann, wie aus der folgenden Proposition ersichtlich wird.

PROPOSITION 3.14. *Ist $g \in l_{SM}$ und $T_\Omega : \Omega_\Sigma \rightarrow \Omega_\Sigma$ der in Abschnitt 2.2.5 definierte Schiebeoperator, so gilt*

$$m_{\mu^t g} = m_g \circ T_\Omega^{-t}.$$

BEWEIS. Wegen der Eindeutigkeit der endlichen, signierten Maße gemäß Korollar 2.52 reicht es die Identität auf Zylindermengen $Z(A)$, $A \subset \Sigma^s$ zu überprüfen. Dort gilt

$$\begin{aligned} m_{\mu^t g}(Z(A)) &= \mu^t g(A) = \sum_{\bar{b} \in \Sigma^t} \sum_{\bar{a} \in A} g(\bar{b}\bar{a}) \\ &\stackrel{B./D.2.30}{=} g(T_\Omega^{-t} A) = m_g(Z(T_\Omega^{-t} A)) = m_g(T_\Omega^{-t} Z(A)), \end{aligned}$$

wobei die letzten beiden Gleichungen offensichtliche Folgerungen der Definition des Schiebeoperators sind. \square

BEMERKUNG 3.15. *Ist $g \in l_{SM}$ ein stochastischer Vektor und $(X_t)_{t \in \mathbb{N}}$ der dazu assoziierte stochastische Prozess, so gilt nach der vorangehenden Bemerkung, dass der Prognosevektor mit Geschichte Σ^s einen stochastischen Prozess $(Y_t)_{t \in \mathbb{N}}$ induziert, für den*

$$Y_t = X_{t+s}$$

gilt. Für den entsprechenden Prognosevektor g_{Σ^s} gilt nach Definition

$$g_{\Sigma^s} = \mu^s g^{\Sigma^s} = \mu^s g^{\Omega_\Sigma} = \mu^s g.$$

Wir bemerken, dass dies mit Bemerkung 2.32 und Proposition 3.14 übereinstimmt.

3.2. Dimension signierter Maßvektoren

Dieser Abschnitt führt schließlich die zentralen Begriffe dieser Arbeit ein. In einfachen Worten ist die Dimension eines signierten Maßvektors die des von seinen Prognosevektoren aufgespannten Untervektorraums von l_{SM} . Endlich-dimensionale stochastische Vektoren können daher mit stochastischen Prozessen identifiziert werden, für die unter der Annahme von Vorwissen nichts wesentlich Neues erwartet werden muss - der Prozess kann in jedem Zustand als Linearkombination endlich vieler "Basisprozesse" beschrieben werden. Die Definition des Begriffs rührt von der Erkenntnis her, dass Hidden Markov Prozesse eben solche Prozesse sind.

3.2.1. Definitionen und Einführung. Wir beginnen ohne Umschweife mit den Definitionen der für diesen Abschnitt sinnstiftenden Objekte. Wir motivieren sie mit einfachen Beispielen von stochastischen Prozessen und illustrieren sie desweiteren mit Querverbindungen zur Stabilitätstheorie des ersten Teils.

DEFINITION 3.16. *Sei $g \in l_\infty$. Dann nennen wir*

$$\mathcal{V}_g := \text{span}\{\tau_{\bar{a}}(g) \mid \bar{a} \in \Sigma^*\}$$

den Prognoseraum des Vektors g und

$$\mathcal{E}_g := \text{span}\{\mu^t(g) \mid t \in \mathbb{N}\}$$

den Evolutionsraum des Vektors g . Wir setzen weiter

$$\dim g := \dim \mathcal{V}_g$$

und nennen diese Zahl die (stochastische) Dimension des Vektors g sowie

$$\text{edim } g := \dim \mathcal{E}_g$$

die (stochastische) Evolutionsdimension des Vektors g . Ist g ein stochastischer Vektor und $(X_t)_{t \in \mathbb{N}}$ der durch g gegebene stochastische Prozess, so setzen wir

$$\dim(X_t) := \dim(g) \tag{3.6}$$

$$\text{edim}(X_t) := \text{edim}(g). \tag{3.7}$$

Wir nennen dies dann die **Dimension** bzw. **Evolutionsdimension** des stochastischen Prozesses (X_t) .

Wir motivieren den Dimensionsbegriff mit einem kleinen Beispiel einer geläufigen Klasse von stochastischen Prozessen, den identisch und unabhängig verteilten stochastischen Prozessen. Dabei heißt ein stochastischer Prozess $(X_t)_{t \in \mathbb{N}}$ **identisch und unabhängig verteilt**, falls es ein Wahrscheinlichkeitsmaß P auf $(\Sigma, \mathcal{P}(\Sigma))$ gibt, so dass für die durch den Prozess induzierten Verteilungen P_t auf $(\Sigma^t, \mathcal{P}(\Sigma^t))$ (siehe Abschnitt 2.2.3) gilt:

$$P_t(\bar{a} = a_1 \dots a_t) = \prod_{s=1}^t P(a_s).$$

Dann erhält man:

PROPOSITION 3.17. Für einen stochastischen Prozess $(X_t)_{t \in \mathbb{N}}$ gilt

$$\dim(X_t) = 1 \Leftrightarrow (X_t) \text{ ist unabhängig und identisch verteilt.} \quad (3.8)$$

BEWEIS. Sei jeweils g der zu (X_t) assoziierte stochastische Vektor.

Gilt $\dim(X_t) = 1$, so für alle $\bar{a} \in \Sigma^*$

$$g_{\bar{a}} = g.$$

Sei nun $\bar{b} = b_0 \dots b_t \in \Sigma^*$. Dann (siehe Beispiel 3.11)

$$g(\bar{b}) = g(b_0)g_{b_0}(b_1)g_{b_0b_1}(b_2) \cdot \dots \cdot g_{b_0 \dots b_{t-1}}(b_t) = \prod_{s=0}^t g(b_s),$$

was zeigt, dass (X_t) unabhängig und identisch verteilt ist. Ist umgekehrt (X_t) unabhängig und identisch verteilt, so für alle $\bar{b} = b_0 \dots b_t$

$$g(\bar{b}) = \prod_{s=0}^t g(b_s).$$

Damit für alle $\bar{a}, \bar{b} \in \Sigma^*$

$$\begin{aligned} g_{\bar{a}}(\bar{b}) &= \frac{1}{g(\bar{a})} g(\bar{a}\bar{b} = a_1 \dots a_s b_1 \dots b_t) \\ &= \frac{\prod_{r=1}^s g(a_r) \prod_{u=1}^t g(b_u)}{\prod_{r=1}^s g(a_r)} = \prod_{u=1}^t g(b_u) = g(\bar{b}), \end{aligned}$$

woraus die Behauptung folgt. \square

BEMERKUNG 3.18. Für einen stochastischen Vektor g ist es bisweilen hilfreich, sich die Dimension als den Rang des Spaltenraums $\mathcal{V}_P = \mathcal{V}_g$ der abzählbar unendlichen Matrix

$$\mathcal{P} := (g(\bar{b}\bar{a}))_{\bar{a}, \bar{b} \in \Sigma^*} = (\tau_{\bar{b}} g(\bar{a}))_{\bar{a}, \bar{b} \in \Sigma^*} \in \mathbb{R}^{\mathbb{N} \times \mathbb{N}}$$

vorzustellen bzw. die Evolutionsdimension als den Rang des Spaltenraums $\mathcal{E}_P = \mathcal{E}_g$ der Matrix

$$\mathcal{P}_\mu := \left(\sum_{|\bar{b}|=t} g(\bar{b}\bar{a}) \right)_{\bar{a} \in \Sigma^*, t \in \mathbb{N}} = (\mu^t g(\bar{a}))_{\bar{a} \in \Sigma^*, t \in \mathbb{N}} \in \mathbb{R}^{\mathbb{N} \times \mathbb{N}}.$$

Mit Hilfe von Lemma 2.43 sieht man schnell ein, dass Spalten- und Zeilenrang der obigen Matrizen identisch sein müssen.

Wir machen für spätere Zwecke noch die offensichtliche

BEMERKUNG 3.19. Gilt $f \in \mathcal{V}_g$, so auch $\mathcal{V}_f \subset \mathcal{V}_g$.

SATZ 3.20. Sei g ein Vektor mit $\text{edim}(g) = m < \infty$. Dann ist die Familie von Vektoren $(g_t := \mu^t g)_{t=0, \dots, m-1}$ linear unabhängig. Gilt weiter $g_m = \mu^m g = \sum_{t=0}^{m-1} c_t g_t$, so ist das Polynom

$$p_g(X) := X^m - c_{m-1}X^{m-1} - \dots - c_1X - c_0$$

das Minimalpolynom von $\mu : \mathcal{E}_g \rightarrow \mathcal{E}_g$. Wegen $\deg p_g = m$ ist dies natürlich auch das charakteristische Polynom von μ .

BEWEIS. Der erste Teil ist leicht einzusehen: Nehmen wir an, dass $g_{m-1} = \sum_{t=0}^{m-2} \alpha_t g_t$, so folgt wegen

$$g_m = \mu^m g = \mu g_{m-1} = \sum_{t=0}^{m-2} \alpha_t \mu g_t = \sum_{t=0}^{m-2} \alpha_t g_{t+1}$$

induktiv auch für alle $t \geq m$, dass $g_t \in \text{span}\{g_t \mid t = 0, \dots, m-2\}$. Für den zweiten Teil bemerken wir, dass die Vektoren $g_t, t = 0, \dots, m-1$ linear unabhängig sind, und damit (siehe Definition 11.2)

$$\text{fdim}_\mu(g_0) = m.$$

Deshalb gilt nach Satz 11.3 gerade, dass der Grad des Minimalpolynoms maximal ist, womit klar ist, dass sich Minimalpolynom und charakteristisches Polynom entsprechen. Es bleibt zu zeigen, dass

$$\forall t = 0, \dots, m-1 : p_g(\mu)(g_t) = 0,$$

denn daraus folgt $p_g(\mu) = 0$, da $p_g(\mu)$ auf einer Basis verschwindet. Dies ist aber einfach einzusehen:

$$p_g(\mu)(g_0) = g_m - c_{m-1}g_{m-1} - \dots - c_1g_1 - c_0g_0 = 0$$

nach Definition der c_t . Für $t \geq 1$ gilt weiter

$$p_g(\mu)(g_t) = p_g(\mu)(\mu^t g_0) = \mu^t(p_g(\mu)(g_0)) = \mu^t 0 = 0.$$

□

KOROLLAR 3.21. Sei $g \in l_{TV}$ mit $\text{edim}(g) < \infty$. Dann ist

$$\mu : \mathcal{E}_g \rightarrow \mathcal{E}_g$$

folgenstabil.

BEWEIS. Aufgrund Satz 11.8 (Folgenstabilität - Äquivalenz) hat man zu zeigen, dass μ stabil ist, und dass das Minimalpolynom von μ zum charakteristischen Polynom äquivalent ist (beachte, dass Eigenschaft (ii) von Satz 11.8 dazu gerade äquivalent ist). Dass μ stabil ist, sagt Satz 3.8. Die zweite Eigenschaft folgt aus Satz 3.20. \square

KOROLLAR 3.22. *Sei g ein stochastischer Vektor mit $\text{edim}(g) = m < \infty$. Sei μ der Evolutionsoperator. Dann gilt*

$$\dim (\text{Eig}(\mu; 1) \cap \mathcal{E}_g) = 1$$

und wegen $\mathcal{E}_g \subset \mathcal{V}_g$ somit

$$\dim (\text{Eig}(\mu; 1) \cap \mathcal{V}_g) \geq 1.$$

BEWEIS. Da die Eigenräume eines linearen Operators, für den das Minimalpolynom maximalen Grad hat, höchstens Dimension 1 haben, folgt mit Satz 3.20

$$\dim \text{Eig}(\mu; 1) \leq 1.$$

Seien nun $g_t := \mu^t g$ und $p_g(X) = X^m + c_{m-1}X^{m-1} + \dots + c_1X + c_0$ das charakteristische Polynom wie in Satz 3.20. Wegen

$$\sum_{t=0}^{m-1} c_t = \sum_{t=0}^{m-1} c_t g_t(\square) = g_m(\square) = 1$$

folgt $p_g(1) = 0$ und damit die Behauptung. \square

3.2.2. Kegel. Die Prognoseoperatoren endlich-dimensionaler stochastischer Vektoren lassen einen echten Kegel in \mathcal{V}_g invariant. Der Nutzen dieser Tatsache wird später ganz besonders an zwei Stellen wichtig werden. Zum einen bildet sie den Grundstein für die Lösung eines mit dem "Identifikationsproblem" verwandten Problem, welches wiederum die Motivation der Dimensionsbegriffe darstellt. Zum anderen lässt sich ein polyedrischer Unterkegel dieses Kegels finden, dem eine besondere Bedeutung zugesprochen werden kann - er ist die zugrundeliegende Struktur der im folgenden Kapitel besprochenen "Ergodische Zerlegung" endlich-dimensionaler stochastischer Vektoren.

Zum Verständnis dieses Kapitels benötigen wir ebenden Begriff des Kegels und ein paar Folgedefinitionen, die wir nun bereitstellen.

In diesem Abschnitt bezeichnet

$$\text{conv}(B) := \{x \in V \mid x = \sum_{i=1}^n \alpha_i x_i, x_i \in B, \sum_i \alpha_i = 1, \alpha_i \geq 0\}$$

zunächst den konvexen Abschluss einer Menge $B \subset V$ eines \mathbb{R} -Vektorraums. Eine Menge B bezeichnen wir als konvex, falls $\text{conv}(B) = B$.

DEFINITION 3.23. *Sei V ein \mathbb{R} -Vektorraum. Eine Menge $K \subset V$ heißt **konvexer Kegel**, falls die folgenden Bedingungen erfüllt sind:*

(i) Für alle $\alpha \in \mathbb{R}^+, v \in K$:

$$\alpha v \in K. \quad (3.9)$$

(ii) Für alle $v_1, v_2 \in K$:

$$v_1 + v_2 \in K. \quad (3.10)$$

BEMERKUNG 3.24. *Offensichtlich ist ein konvexer Kegel K eine konvexe Menge, mit anderen Worten gilt*

$$\text{conv}(K) = K.$$

Ist $K \subset V$ ein Kegel in einem \mathbb{R} -Vektorraum V , so setzen wir

$$-K := \{x \in V \mid -x \in K\}.$$

Ist $\|\cdot\|$ die durch das Standardskalarprodukt induzierte Norm (also $\|\cdot\| = \|\cdot\|_2$, die euklidische Norm), so für $v_0 \in V$: $U_d(v_0) := \{v \in V \mid \|v - v_0\| < d\}$ und weiter für eine beliebige Menge $B \subset V$

$$\text{int}(B) := \{x \in B \mid \exists \epsilon \in \mathbb{R}^+ : U_\epsilon(x) \subset B\},$$

also $\text{int}(B)$ das Innere von B bezüglich der durch das Standardskalarprodukt induzierten Topologie von V (also nicht bezüglich der durch das Skalarprodukt auf B induzierten Teilraumtopologie).

DEFINITION 3.25. *Ein konvexer Kegel heißt **spitz**, falls*

$$K \cap -K = \{0\}, \quad (3.11)$$

reproduzierend, falls

$$\text{span}\{x - y \mid x, y \in K\} = V, \quad (3.12)$$

und **solide**, falls

$$\text{int}(K) \neq \emptyset. \quad (3.13)$$

BEMERKUNG 3.26. *Die Eigenschaften 3.25 und 3.25 sind äquivalent. Ein Kegel ist also genau dann reproduzierend, falls er solide ist.*

BEWEIS. Siehe z.B. [Ber70], Lemma 2.1. □

DEFINITION 3.27. *Ein konvexer Kegel heißt **echt**, falls er abgeschlossen, spitz und solide (bzw. reproduzierend) ist.*

Ausgangspunkt für die nachfolgende Theorie ist die Einsicht, dass es für endlich-dimensionale stochastische Vektoren immer eine Basis des Prognose-raums gibt, bezüglich derer alle Prognosevektoren eine positive Darstellung haben. Dies liefert dann Aufschluss über die Menge der Prognosevektoren: Ihre konvexe Hülle bildet nämlich einen echten Kegel.

LEMMA 3.28. Sei g ein Maßvektor mit $\dim(g) = m < \infty$. Dann gibt es eine Basis $(\mathfrak{g}_j), j = 1, \dots, m$, so dass

$$\forall \bar{a} \in \Sigma^* : g_{\bar{a}} = \sum_{j=1}^m \alpha_j \mathfrak{g}_j \implies \forall j : \alpha_j \in [0, 1].$$

Mit anderen Worten, es gibt eine Basis $(\mathfrak{g}_j), j = 1, \dots, m$, bezüglich derer alle Prognosevektoren eine positive Darstellung mit Koeffizienten zwischen 0 und 1 haben.

BEWEIS. Wir finden zunächst m Wörter $\bar{b}_j, j = 1, \dots, m$, so dass die $(g_j := g_{\bar{b}_j})_{j=1, \dots, m}$ eine linear unabhängige Familie bilden. Mit Hilfe von Lemma 2.43 finden wir weiter m Wörter $\bar{a}_i, i = 1, \dots, m$, so dass die Matrix

$$A = \begin{pmatrix} g_1(\bar{a}_1) & \dots & g_m(\bar{a}_1) \\ \vdots & \ddots & \vdots \\ g_1(\bar{a}_m) & \dots & g_m(\bar{a}_m) \end{pmatrix}$$

regulär ist. Wir setzen nun für $k \in \{1, \dots, m\}$ und e_k die kanonischen Basisvektoren des \mathbb{R}^m zunächst $(\alpha_{k,1}, \dots, \alpha_{k,m}) := (A^{-1}e_k)^T$ und weiter

$$\mathfrak{g}_k = \sum_{j=1}^m \alpha_{k,j} g_j$$

und behaupten, damit eine Basis von geforderter Qualität gefunden zu haben. Dass wir eine Basis erhalten haben, ist klar. Weiterhin stellen wir fest:

$$\mathfrak{g}_k(\bar{a}_i) = \sum_{j=1}^m \alpha_{k,j} g_j(\bar{a}_i) = \delta_{ki} \quad (3.14)$$

nach Definition der $\alpha_{k,j}$. Sei nun $\bar{a} \in \Sigma^*$ und $g_{\bar{a}} = \sum_{k=1}^m \beta_k \mathfrak{g}_k$ die Darstellung des zugehörigen Prognosevektors über den \mathfrak{g}_k . Dann folgt

$$\forall i \in \{1, \dots, m\} : \beta_i = \sum_{k=1}^m \beta_k \delta_{ki} \stackrel{(3.14)}{=} \sum_{k=1}^m \beta_k \mathfrak{g}_k(\bar{a}_i) = g_{\bar{a}}(\bar{a}_i) \geq 0,$$

also die Behauptung. \square

Wir präzisieren nun die bereits oben angerissene Aussage, dass die Prognosevektoren in gewisser Weise einen echten Kegel bilden.

SATZ 3.29. Sei g ein endlich-dimensionaler stochastischer Vektor. Sei K_g der Abschluss von

$$C_g = \text{cone}(\tau_{\bar{a}}g \mid \bar{a} \in \Sigma^*),$$

des von den Vektoren $\{\tau_{\bar{a}}g \mid \bar{a} \in \Sigma^*\}$ erzeugten Kegels in \mathcal{V}_g . K_g ist ein echter Kegel, der von allen observablen Operatoren $\tau_{\bar{a}}$ und demnach auch von μ invariant gelassen wird. Wir nennen K_g den **Prognosekegel** von g .

BEWEIS. Der Abschluss eines konvexen Kegels ist offensichtlich wiederum ein konvexer Kegel. Die Abgeschlossenheit und die Volldimensionalität der Menge K_g sind klar nach Definition. Die Spitzheit folgt aus dem vorangehenden Lemma 3.28, da Spitzheit unter Basistransformationen erhalten bleibt und ein Kegel, der im positiven Orthanten enthalten ist, offensichtlich spitz

ist. Dass K_g von den observablen Operatoren invariant gelassen wird, ist ebenfalls klar nach Definition von K_g . \square

KOROLLAR 3.30. *Sei g ein endlich-dimensionaler, stochastischer Vektor und $g_1 \in K_g$ ein stochastischer Vektor, der im oben beschriebenen echten Kegel liegt. Dann gilt*

$$K_{g_1} \subset K_g.$$

BEWEIS. Das ist offensichtlich, da K_g ja von allen Prognoseoperatoren invariant gelassen wird, nun aber wiederum $K_{g_1} = \overline{\text{cone}(\tau_{\bar{a}}g_1, \bar{a} \in \Sigma^*)}$ gilt. \square

BEMERKUNG 3.31. *Sei g ein endlich-dimensionaler stochastischer Vektor und $0 \neq f \in \mathcal{V}_g$ mit $f(\square) = 0$. Dann gilt*

$$f \notin K_g.$$

BEWEIS. Das folgt, indem man bemerkt, dass für die Elemente g_i der Basis aus Lemma 3.28 $g_i(\square) = 1$ gilt, und alle Elemente des Kegels Konvexkombinationen der g_i sind. \square

BEMERKUNG 3.32. *Sei g ein endlich-dimensionaler, stochastischer Vektor und $g^* \in K_g$ ein Vektor aus dem Prognosekegel mit $g^*(\square) = 1$. Dann ist g^* ein stochastischer Vektor.*

BEWEIS. Dass $\sigma g^* = g^*$, ist offensichtlich, da $\mathcal{V}_g \subset l_\sigma$. Dass $0 \leq g^* \leq 1$, folgt für $g^* \in \text{Int}(K_g)$ daraus, dass sich g^* nach Konstruktion von K_g als Konvexkombination stochastischer Vektoren schreiben lässt. Gilt $g^* \in \delta K_g$, so findet man eine Folge stochastischer Vektoren, die gegen g^* konvergiert. Damit ist g^* nach Korollar 2.52 offensichtlich ebenfalls ein stochastischer Vektor. \square

3.2.3. Markov-Vektoren. Dieser Abschnitt nun macht den ersten Schritt hin zur historischen Motivation der Dimensionsbegriffe, nämlich der Tatsache, dass stochastische Vektoren, die Hidden-Markov-Prozesse kodieren, endlich-dimensional sind. Wir starten daher mit der Definition eines Markov-Prozesses und gehen im nächsten Abschnitt mit Hilfe sogenannter Alphabetwechsel zu Hidden-Markov-Prozessen über.

Da in diesem und dem folgenden Abschnitt die Auswahl des Alphabets sehr wohl eine Rolle spielt, schreiben wir hier meist

$$l_\Sigma := l_\infty^\Sigma,$$

um das zugrundeliegende Alphabet zu benennen und so zwischen verschiedenen Alphabeten unterscheiden zu können.

Darüberhinaus sei im weiteren Verlauf mit einer stochastischen Matrix eine Matrix mit nicht-negativen Einträgen gemeint, deren Spaltensummen 1 ergeben. In diesem Sinne nennen wir einen Vektor $w \in \mathbb{R}^n$ stochastisch, falls er, als Spaltenvektor aufgefasst, eine stochastische Matrix ist. Wir machen

auf den Unterschied zwischen stochastischen Vektoren $w \in \mathbb{R}^n$ und stochastischen Vektoren $g \in l_\Sigma$ aufmerksam.

Wir wollen zuguterletzt schon an dieser Stelle darauf hinweisen, dass zu den nachfolgenden Sätzen in den meisten Fällen verwandte Resultate in den Schriften von Gilbert, Dharmadikari und Heller [Gil59], [Dha63a], [Dha63b], [Dha65], [Hel65] zu finden sind. Diese sind jedoch, wie wir bereits in der Einleitung argumentiert haben, mit unserem Modell nun elementarer formuliert. Wie gesagt bestreitet Heller den auf einfacheren Vektorräumen basierenden Ansatz. Wir beweisen hiermit das Gegenteil.

DEFINITION 3.33. *Wir nennen einen Vektor $g \in l_\Sigma$ einen **Markov-Vektor**, falls es einen stochastischen Vektor $w_g \in \mathbb{R}^\Sigma$ und eine stochastische Matrix $M \in \mathbb{R}^{\Sigma \times \Sigma}$ gibt, so dass*

$$\forall \bar{a} = a_1 \dots a_t \in \Sigma^+ : \quad g(\bar{a}) = w_g(a_1) \prod_{i=2}^t M(a_i, a_{i-1}) \quad (3.15)$$

und darüberhinaus $g(\square) = 1$.

PROPOSITION 3.34. *Ein Markov-Vektor ist ein stochastischer Vektor.*

BEWEIS. Die Aussagen $g(\square) = 1$ und $g \geq 0$ sind klar nach Definition. Weiter gilt

$$(\sigma g)(\square) = \sum_{a \in \Sigma} g(a) = \sum_{a \in \Sigma} w_g(a) \stackrel{w_g \text{ stoch.}}{=} 1 = g(\square)$$

und für $\bar{a} = a_1 \dots a_t \in \Sigma^t$

$$\begin{aligned} (\sigma g)(\bar{a}) &= \sum_{a \in \Sigma} g(\bar{a}a) = \sum_{a \in \Sigma} g(\bar{a}a) = \sum_{a \in \Sigma} w_g(a_1) \prod_{i=2}^t M(a_i, a_{i-1}) M(a, a_t) \\ &= w_g(a_1) \prod_{i=2}^t M(a_i, a_{i-1}) \underbrace{\sum_{a \in \Sigma} M(a, a_t)}_{=1, M \text{ st.}} = g(\bar{a}), \end{aligned}$$

woraus zusammen $\sigma g = g$ folgt. □

Natürlich sind die durch Markov-Vektoren induzierten stochastischen Prozesse gerade die Markov-Ketten. Das Interessante an den Markov-Vektoren ist nun gerade, dass sie im Sinne unserer obigen Definition endlich-dimensional sind.

PROPOSITION 3.35. *Ist g ein Markov-Vektor, so*

$$\dim g \leq \text{card}(\Sigma). \quad (3.16)$$

BEWEIS. Für $a \in \Sigma$ definieren wir Vektoren $f_a \in l_\Sigma$ durch die Vorschrift ($\bar{a} = a_1 \dots a_t \in \Sigma^+$)

$$f_a(\bar{a}) := M(a_1, a) \prod_{i=2}^t M(a_i, a_{i-1})$$

und $f_a(\square) = 1$. Wir bemerken, dass die f_a Markov-Vektoren (nimm dasselbe M wie für g und setze w_{f_a} den zu a gehörenden Spaltenvektor von M) und daher stochastische Vektoren sind. Weiter gilt für $\bar{b} = b_1 \dots b_s$, so dass $g(\bar{b}) > 0$ und für $\bar{a} = a_1 \dots a_t$

$$(\tau_{\bar{b}}g)(\bar{a}) = g(\bar{b}\bar{a}) = g(\bar{b})M(a_1, b_t) \prod_{i=1}^t M(a_i, a_{i-1}) = g(\bar{b})f_{b_t}(\bar{a}),$$

also $\tau_{\bar{b}}g \in \text{span}\{f_{b_t}\}$. Im Falle, dass der Spaltenrang von M kleiner als $\text{card}(\Sigma)$, M also nicht regulär ist, folgt $\dim \text{span}\{f_a \mid a \in \Sigma\} \leq \text{card}(\Sigma) - 1$ und daher

$$\begin{aligned} \dim \mathcal{V}_g &\leq \dim \text{span}\{f_a \mid a \in \Sigma\} + \dim \text{span}\{g\} \\ &\leq \text{card}(\Sigma) - 1 + 1 = \text{card}(\Sigma). \end{aligned}$$

Andernfalls (bei Regularität von M) liegt w_g im Spaltenraum von M und daher $w_g(b) = \sum_{a \in \Sigma} \alpha_a M(b, a)$ für geeignete von b unabhängige α_a . Es folgt

$$g = \sum_{a \in \Sigma} \alpha_a f_a$$

und daher

$$\dim \mathcal{V}_g = \dim \text{span}\{f_a \mid a \in \Sigma\} = \text{card}(\Sigma). \quad \square$$

Aus dem Beweis wird dabei sofort ersichtlich, dass die Menge $\{g_{\bar{a}} \mid \bar{a} \in \Sigma^*\}$ für einen Markov-Vektor endlich ist. Das führt direkt auf folgende wichtige Bemerkung. Dabei bezeichnen wir einen Kegel $K \subset V$ in einem \mathbb{R} -Vektorraum als *polyedrisch*, falls er ein endliches Erzeugendensystem besitzt.

KOROLLAR 3.36. *Ist g ein Markov-Vektor, so ist der dazu assoziierte Prognosekegel K_g polyedrisch.*

BEWEIS. Mit Hilfe des Beweises von Proposition 3.35 sehen wir ein, dass der Prognosekegel gerade von den endlich vielen f_a und möglicherweise g selbst erzeugt wird. \square

Wir geben nun eine Charakterisierung von Markov-Vektoren an, die mit Objekten unseres Modells beschrieben werden kann:

SATZ 3.37. *Ein stochastischer Vektor g ist genau dann ein Markov-Vektor, falls*

$$\forall a \in \Sigma : \quad \dim \tau_a(\mathcal{V}_g) \leq 1. \quad (3.17)$$

BEWEIS. “ \implies ” ist eine direkte Konsequenz der Beweisführung von Proposition 3.35. Sei umgekehrt ein stochastischer Vektor mit der Eigenschaft 3.17 gegeben. Sei dann jeweils f_a ein Erzeuger von $\text{span } \tau_a(\mathcal{V}_g)$, wobei entweder f_a als stochastischer Vektor gewählt werden kann oder $\tau_a(\mathcal{V}_g) = \{0\}$ ist. Wir definieren dann weiter einen (offensichtlich stochastischen) Vektor $w_g \in \mathbb{R}^\Sigma$ durch die Vorschrift

$$w_g(a) := g(a)$$

für alle $a \in \Sigma$ und eine (offensichtlich stochastische) Matrix $M \in \mathbb{R}^{\Sigma \times \Sigma}$ durch die Vorschrift

$$M(a, b) := \tau_a(f_b)(\square) = f_b(a).$$

Wir zeigen nun für $\bar{a} = a_1 \dots a_t \in \Sigma^t$

$$g(\bar{a}) = w_g(a_1) \prod_{i=2}^t M(a_i, a_{i-1})$$

und damit die Behauptung mit vollständiger Induktion über t . Dabei gilt die Behauptung für $t = 1$ nach Definition von w_g offensichtlich. Sei nun $t \geq 1$ und $\bar{a} = a_1 \dots a_t a_{t+1} \in \Sigma^{t+1}$. Wegen

$$g_{a_1 \dots a_t} = \frac{1}{g(a_1 \dots a_t)} \tau_{a_1 \dots a_t} g = \frac{1}{g(a_1 \dots a_t)} \tau_{a_t} \tau_{a_1 \dots a_{t-1}} g = f_{a_t} \quad (3.18)$$

im Falle von $g(a_1 \dots a_t) > 0$ (denn dann ist $g_{a_1 \dots a_t}$ ein stochastischer Vektor, der in $\tau_{a_t}(\mathcal{V}_g)$ liegt) gilt

$$\begin{aligned} g(\bar{a}) &= \tau_{a_1 \dots a_{t+1}} g(\square) = \tau_{a_{t+1}} \tau_{a_1 \dots a_t} g(\square) \\ &= \tau_{a_1 \dots a_t} g(a_{t+1}) \stackrel{(3.18)}{=} g(a_1 \dots a_t) f_{a_t}(a_{t+1}) \\ &\stackrel{(IV)}{=} w_g(a_1) \left(\prod_{i=2}^t M(a_i, a_{i-1}) \right) f_{a_t}(a_{t+1}) \\ &= w_g(a_1) \left(\prod_{i=2}^t M(a_i, a_{i-1}) \right) M(a_{t+1}, a_t) \end{aligned}$$

und die Behauptung folgt. \square

3.2.4. Alphabetwechsel und Hidden-Markov-Vektoren. Wie mehrfach in dieser Arbeit erwähnt, ist die Motivation der Einführung des Dimensionsbegriffs ein Ableger des Identifikationsproblems gewesen: Wann ist ein stochastischer Prozess ein Hidden-Markov-Prozess oder mit den Worten von Gilbert, Dharmadikhari und Heller zu sprechen, wann ist ein stochastischer Prozess die Funktion eines Markov-Prozesses? Im Rahmen der Arbeiten [Gil59], [Dha63a], [Dha63b], [Dha65] und [Hel65] ist diese Frage mit Hilfe des Dimensionsbegriffs neu charakterisiert worden. Wir beschreiben im Folgenden diese Charakterisierung in unserem Sprachrahmen und geben einen Beweis dieser Charakterisierung. Dieser ist weitestgehend durch die Erkenntnisse in [Hel65] inspiriert, vermeidet aber die dort verwendeten Spezialbegriffe aus der Algebra.

Wir betrachten einführend die folgende Situation: Seien zwei Alphabete Σ_0, Σ gegeben und eine Abbildung

$$\phi : \Sigma_0 \longrightarrow \Sigma,$$

wobei $\text{card}(\Sigma) \leq \text{card}(\Sigma_0)$ angenommen werden darf, da wir uns zukünftig nur für den Bildraum von ϕ interessieren. Wir setzen ϕ auf die offensichtliche Art und Weise zu einer Abbildung

$$\phi : \Sigma_0^* \longrightarrow \Sigma^*$$

($\phi(\square) := \square$) fort. ϕ induziert dann durch die Vorschrift

$$(\hat{\phi}g)(\bar{a}) := \sum_{\bar{q} \in \phi^{-1}(\bar{a})} g(\bar{q})$$

für $\bar{a} = a_1 \dots a_t \in \Sigma^t$ eine lineare Abbildung

$$\hat{\phi} : l_{\Sigma_0} \longrightarrow l_{\Sigma}.$$

Wir geben dem Ganzen Namen:

DEFINITION 3.38. *Wir nennen das ϕ von oben einen **Alphabetwechsel** und $\hat{\phi}$ den dazu assoziierten **Wechseloperator**.*

Nun die formale Definition eines Hidden-Markov-Vektors:

DEFINITION 3.39. *Wir nennen $g \in l_{\Sigma}$ einen **Hidden-Markov-Vektor**, falls es einen Alphabetwechsel $\phi : \Sigma_0 \rightarrow \Sigma$ und einen Markov Vektor $g_0 \in l_{\Sigma_0}$ gibt, so dass*

$$g = \hat{\phi}g_0.$$

Hidden-Markov-Vektoren sind nun endlich-dimensionale, stochastische Vektoren, wie mit Hilfe der beiden folgenden Propositionen klar wird. Es ist dann einfach zu zeigen, dass die zu ihnen assoziierten stochastischen Prozesse gerade die Hidden-Markov-Prozesse im klassischen Sinne sind. Hierbei beachte man, dass die Hidden-Markov-Prozesse den Funktionen von Markov-Prozessen entsprechen [BP66], [HJ99], [EM02].

PROPOSITION 3.40. *Mit den obigen Bezeichnungen gilt: Ist $g \in l_{\Sigma_0}$ ein stochastischer Vektor, so auch $\hat{\phi}g$.*

BEWEIS. Klar sind $g(\square) = 1$ und $g \geq 0$. Weiter gilt

$$\begin{aligned} \sigma \hat{\phi}g(\square) &= \sum_{a \in \Sigma} \hat{\phi}g(a) = \sum_{a \in \Sigma} \sum_{q \in \phi^{-1}(a)} g(q) \\ &= \sum_{q \in \Sigma_0} g(q) = 1 = \hat{\phi}g(\square) \end{aligned}$$

und für $\bar{a} = a_1 \dots a_t \in \Sigma^+$

$$\begin{aligned} (\sigma \hat{\phi} g)(\bar{a}) &= \sum_{a \in \Sigma} \hat{\phi} g(\bar{a}a) = \sum_{a \in \Sigma} \sum_{q_1 \dots q_t q_{t+1} \in \phi^{-1}(\bar{a}a)} g(q_1 \dots q_t q_{t+1}) \\ &= \sum_{\bar{q} \in \phi^{-1}(\bar{a})} \sum_{q \in \Sigma_0} g(\bar{q}q) = \sum_{\bar{q} \in \phi^{-1}(\bar{a})} g(\bar{q}) \\ &= \hat{\phi} g(\bar{a}), \end{aligned}$$

woraus zusammen $\sigma \hat{\phi} g = \hat{\phi} g$ folgt. \square

PROPOSITION 3.41. *Mit den obigen Bezeichnungen: Ist $g \in l_{\Sigma_0}$, so*

$$\dim \hat{\phi} g \leq \dim g.$$

BEWEIS. Wir betrachten dafür zunächst das folgende

LEMMA 3.42. *Mit den obigen Bezeichnungen kommutiert das Diagramm*

$$\begin{array}{ccc} l_{\Sigma_0} & \xrightarrow{\hat{\phi}} & l_{\Sigma} \\ \tau_{\phi^{-1}(a)} \downarrow & & \downarrow \tau_a \\ l_{\Sigma_0} & \xrightarrow{\hat{\phi}} & l_{\Sigma} \end{array}$$

für alle $a \in \Sigma$.

BEWEIS. Sei $g \in l_{\Sigma_0}$ und $\bar{a} \in \Sigma^*$. Dann

$$\begin{aligned} \tau_a \hat{\phi} g(\bar{a}) &= \hat{\phi} g(a\bar{a}) = \sum_{q\bar{q} \in f^{-1}(a\bar{a})} g(q\bar{q}) \\ &= \sum_{q \in \phi^{-1}(a)} \sum_{\bar{q} \in \phi^{-1}(\bar{a})} g(q\bar{q}) = \sum_{\bar{q} \in \phi^{-1}(\bar{a})} \sum_{q \in \phi^{-1}(a)} \tau_q g(\bar{q}) \\ &= \sum_{\bar{q} \in \phi^{-1}(\bar{a})} \tau_{\phi^{-1}(a)} g(\bar{q}) = \hat{\phi} \tau_{\phi^{-1}(a)} g(\bar{a}), \end{aligned}$$

woraus $\tau_a \circ \hat{\phi} = \hat{\phi} \circ \tau_{\phi^{-1}(a)}$ und damit die Behauptung folgt. \square

Beweis von Proposition 3.41: Das ist nun einfach, da mit Lemma 3.42

$$\mathcal{V}_{\hat{\phi} g} \subset \hat{\phi}(\mathcal{V}_g)$$

folgt und daher, da $\hat{\phi}$ linear ist, die Behauptung. \square

Aus dem obigen Satz 3.37 erhält man zusammen mit Lemma 3.42 noch das folgende

KOROLLAR 3.43. *Ist g ein Hidden-Markov-Vektor vermöge eines Alphabetwechsels $\phi : \Sigma_0 \rightarrow \Sigma$, so*

$$\forall a \in \Sigma : \dim \tau_a(\mathcal{V}_g) \leq \text{card}(\phi^{-1}(a)).$$

Es bleibt nun die Frage zu beantworten, wie man Hidden-Markov-Vektoren innerhalb der Klasse endlich-dimensionaler Prozesse charakterisieren für alle $a \in \Sigma$ kann, ohne Alphabetwechsel zu Hilfe nehmen zu müssen. Diese Charakterisierung ist anhand des Prognosekegels möglich und von Heller (unseres Erachtens umständlich [Hel65]) beschrieben worden. Wir bezeichnen mit

$$\mathcal{V}_g^+ := \{g \in \mathcal{V}_g \mid g \geq 0\}$$

die Menge der Maßvektoren in \mathcal{V}_g . Diese bildet offensichtlich einen von den Prognoseoperatoren invariant gelassenen Kegel. Die in [Hel65] erarbeitete Charakterisierung liest sich, in unser Modell übersetzt, dann so:

SATZ 3.44 (Heller, 1965). *Ein endlich-dimensionaler stochastischer Vektor $g \in l_\Sigma$ ist genau dann ein Hidden-Markov-Vektor, falls es einen Kegel $K \subset \mathcal{V}_g$ gibt, der die folgenden Eigenschaften erfüllt:*

- (a) $g \in K$,
- (b) $K \subset \mathcal{V}_g^+$,
- (c) $\forall a \in \Sigma: \tau_a(K) \subset K$,
- (d) K ist echt und polyedrisch.

BEWEIS. “ \implies ”: Ist $g = \hat{\phi}g_0$ ein Hidden-Markov-Vektor vermöge des Alphabetwechsels $\phi: \Sigma_0 \rightarrow \Sigma$, so ist

$$\hat{K} := \hat{\phi}(K_{g_0}) \cap \mathcal{V}_g$$

ein Kegel, der die oben genannten Anforderungen erfüllt. Mit Lemma 3.42 folgt zunächst die Invarianz von \hat{K} bezüglich der Prognoseoperatoren und weiter wegen $K_{g_0} \subset \hat{K}$ auch die Echtheit des Kegels. Weiter ist \hat{K} polyedrisch, da K_{g_0} als Prognosekegel eines Markov-Vektors polyedrisch ist (siehe Korollar 3.36), und die Eigenschaft, polyedrisch zu sein, unter linearen Abbildungen und Schnittbildung mit linearen Unterräumen erhalten bleibt. Offensichtlich auch $\hat{K} \subset \mathcal{V}_g^+$ nach Definition von $\hat{\phi}$.

“ \impliedby ”: Sei umgekehrt ein endlich-dimensionaler stochastischer Vektor g zusammen mit einem Kegel K gegeben, der die obigen Anforderungen erfüllt. Wir wählen ein endliches Erzeugendensystem von oBdA (da $K \subset \mathcal{V}_g^+$) stochastischen $\{h_1, \dots, h_m\}$ von K gemäß Eigenschaft (d), also

$$K = \text{cone}\{h_i \mid i = 1, \dots, m\}.$$

Wir konstruieren nun ein Alphabet Σ_0 , einen Alphabetwechsel $\phi: \Sigma_0 \rightarrow \Sigma$, und einen Markov-Vektor $g_0 \in l_{\Sigma_0}$, so dass $g = \hat{\phi}g_0$ ist. Wir starten mit dem Alphabet:

$$\Sigma_0 := \Sigma \times \{1, \dots, m\} = \{(a, i) \mid a \in \Sigma, i \in \{1, \dots, m\}\},$$

wobei m gerade die Anzahl der Erzeugenden h_i von K sei. Wir definieren für alle $i = 1, \dots, m$ Koeffizienten α_i , indem wir

$$g =: \sum_{i=1}^m \alpha_i h_i$$

gemäß Eigenschaft (a) als Konvexkombination der h_i schreiben und weiter für alle $(a, i, j) \in \Sigma \times \{1, \dots, m\}^2$ Koeffizienten α_{aij} , indem wir jeweils $\tau_a h_j$

$$\tau_a h_j =: \sum_{i=1}^m \alpha_{aij} h_i$$

nun gemäß Eigenschaft (c) als Konvexkombination der h_i schreiben. Wir definieren dann einen Vektor $w_{g_0} \in \mathbb{R}^{\Sigma_0}$ durch

$$w_{g_0}((a, i)) := \sum_{j=1}^m \alpha_j \alpha_{aij},$$

der wegen

$$\begin{aligned} \sum_{(a,i) \in \Sigma_0} w_{g_0}((a, i)) &= \sum_{a \in \Sigma} \sum_{i=1}^m \sum_{j=1}^m \alpha_j \alpha_{aij} = \sum_{a \in \Sigma} \sum_{i=1}^m \sum_{j=1}^m \alpha_j \alpha_{aij} \underbrace{h_i(\square)}_{=1} \\ &= \sum_{a \in \Sigma} \sum_{j=1}^m \alpha_j \sum_{i=1}^m \alpha_{aij} h_i(\square) = \sum_{a \in \Sigma} \sum_{j=1}^m \alpha_j \tau_a h_j(\square) \\ &= \sum_{j=1}^m \alpha_j h_j(a) = \sum_{a \in \Sigma} g(a) = 1 \end{aligned}$$

stochastisch ausfällt. Weiter definieren wir eine Matrix $M \in \mathbb{R}^{\Sigma_0 \times \Sigma_0}$ durch die Vorschrift

$$M((a, i), (b, j)) := \alpha_{aij},$$

welche sich wegen

$$\begin{aligned} \sum_{(a,i) \in \Sigma_0} M((a, i), (b, j)) &= \sum_{a \in \Sigma} \sum_{i=1}^m \alpha_{aij} = \sum_{a \in \Sigma} \sum_{i=1}^m \alpha_{aij} \underbrace{h_i(\square)}_{=1} \\ &= \sum_{a \in \Sigma} \tau_a h_j(\square) = \sum_{a \in \Sigma} h_j(a) = 1. \end{aligned}$$

als stochastisch erweist.

Wir definieren nun einen Alphabetwechsel $\phi : \Sigma_0 \rightarrow \Sigma$ durch

$$\phi((a, i)) := a,$$

also durch die Projektion von Σ_0 auf Σ . Wir behaupten dann, dass

$$g = \hat{\phi} g_0,$$

wobei g_0 der zu w_{g_0} und M assoziierte Markov-Vektor sei. Offensichtlich natürlich

$$g(\square) = 1 = \hat{\phi} g_0(\square).$$

Für $\bar{a} = a_1 \dots a_t \in \Sigma^t$ zeigen wir dann zunächst

$$\tau_{\bar{a}} g = \sum_{j=1}^m \alpha_j \sum_{i_1 \dots i_t \in \{1, \dots, m\}^t} \alpha_{a_1 i_1 j} \prod_{s=2}^t \alpha_{a_s i_s i_{s-1}} h_{i_t} \quad (3.19)$$

mit vollständiger Induktion über t . Der Induktionsanfang $t = 1$ folgt aus

$$\tau_a g = \tau_a \left(\sum_{j=1}^m \alpha_j h_j \right) = \sum_{j=1}^m \alpha_j (\tau_a h_j) = \sum_{j=1}^m \alpha_j \sum_{i=1}^m \alpha_{aij} h_i$$

gemäß der Definitionen für die verschiedenen α 's. Sei nun $t \geq 1$ und $\bar{a} = a_1 \dots a_t a_{t+1} \in \Sigma^{t+1}$. Dann

$$\begin{aligned} \tau_{\bar{a}} g &= \tau_{a_{t+1}} \tau_{a_1 \dots a_t} g \\ &\stackrel{IV}{=} \tau_{a_{t+1}} \sum_{j=1}^m \alpha_j \sum_{i_1 \dots i_t \in \{1, \dots, m\}^t} \alpha_{a_1 i_1 j} \prod_{s=2}^t \alpha_{a_s i_s i_{s-1}} h_{i_t} \\ &= \sum_{j=1}^m \alpha_j \sum_{i_1 \dots i_t \in \{1, \dots, m\}^t} \alpha_{a_1 i_1 j} \prod_{s=2}^t \alpha_{a_s i_s i_{s-1}} \tau_{a_{t+1}} h_{i_t} \\ &= \sum_{j=1}^m \alpha_j \sum_{i_1 \dots i_t \in \{1, \dots, m\}^t} \alpha_{a_1 i_1 j} \prod_{s=2}^t \alpha_{a_s i_s i_{s-1}} \sum_{k=1}^m \alpha_{a_{t+1} k i_t} h_k \\ &= \sum_{j=1}^m \alpha_j \sum_{i_1 \dots i_t i_{t+1} \in \{1, \dots, m\}^{t+1}} \alpha_{a_1 i_1 j} \prod_{s=2}^{t+1} \alpha_{a_s i_s i_{s-1}} h_{i_{t+1}}, \end{aligned}$$

womit (3.19) bewiesen wäre. Aus (3.19) folgt dann wiederum

$$\begin{aligned} g(a_1 \dots a_t) &= \tau_{a_1 \dots a_t} g(\square) \stackrel{(3.19)}{=} \sum_{j=1}^m \alpha_j \sum_{i_1 \dots i_t \in \{1, \dots, m\}^t} \alpha_{a_1 i_1 j} \prod_{s=2}^t \alpha_{a_s i_s i_{s-1}} \underbrace{h_{i_t}(\square)}_{=1} \\ &= \sum_{j=1}^m \alpha_j \sum_{i_1 \dots i_t \in \{1, \dots, m\}^t} \alpha_{a_1 i_1 j} \prod_{s=2}^t \alpha_{a_s i_s i_{s-1}} \end{aligned} \tag{3.20}$$

Für $\bar{a} = a_1 \dots a_t \in \Sigma^t$ erhalten wir damit

$$\begin{aligned} \hat{\phi} g_0(\bar{a}) &= \sum_{\bar{q} \in \phi^{-1}(\bar{a})} g_0(\bar{q}) \\ &= \sum_{i_1 \dots i_t \in \{1, \dots, m\}^t} g_0((a_1, i_1), \dots, (a_t, i_t)) \\ &= \sum_{i_1 \dots i_t \in \{1, \dots, m\}^t} w_g((a_1, i_1)) \prod_{j=2}^t M((a_j, i_j), (a_{j-1}, i_{j-1})) \\ &= \sum_{i_1 \dots i_t \in \{1, \dots, m\}^t} \sum_{j=1}^m \alpha_j \alpha_{a_1 i_1 j} \prod_{s=2}^t \alpha_{a_s i_s i_{s-1}} \\ &= \sum_{j=1}^m \alpha_j \sum_{i_1 \dots i_t \in \{1, \dots, m\}^t} \alpha_{a_1 i_1 j} \prod_{s=2}^t \alpha_{a_s i_s i_{s-1}} \\ &\stackrel{(3.20)}{=} g(\bar{a}). \end{aligned}$$

□

3.2.5. Eine einfache Lösung des Identifikationsproblems. Wir stellen zum Schluss dieses Abschnitts ein konzeptionell beinahe simples Verfahren zur Lösung des Identifikationsproblems dar. Das Identifikationsproblem beschäftigt sich mit der Frage, wann zwei Hidden-Markov-Modelle äquivalent im Sinne der von ihnen induzierten stochastischen Prozesse sind. Zwei Arbeiten beantworten diese Frage, die eine veröffentlicht mit einem nicht polynomiellen Algorithmus [IAK92], die andere unveröffentlicht mit einem polynomiellen Algorithmus [Bal93]. Beide Verfahren sind konzeptionell aufwändig. Wir liefern ein exponentielles, jedoch äußerst elementares Verfahren, das in endlicher Zeit entscheiden kann, in welchem Falle zwei endlich-dimensionale, signierte Maßvektoren gleich sind. Das Verfahren basiert auf dem folgenden, zentralen Satz.

SATZ 3.45 (Identifikationssatz). *Sei $1 \leq m \in \mathbb{N}$ und $g, h \in l_\sigma$ zwei signierte Maßvektoren mit $\dim g, \dim h \leq m$. Dann gilt*

$$\forall \bar{a} \in \Sigma^{2m-1} : g(\bar{a}) = h(\bar{a}) \iff g = h.$$

Anders ausgedrückt ist ein höchstens m -dimensionaler, signierter Maßvektor durch die Angabe aller Werte auf Wörtern der Länge $2m-1$ bereits eindeutig festgelegt. Zur Überprüfung der Gleichheit zweier Hidden-Markov-Vektoren brauchen wir also nur deren Verteilung auf Wörtern einer fixen Wortlänge zu überprüfen. Hierbei kann bei der üblichen Darstellungsform der Hidden-Markov-Modelle als Graphen zwei mal die Anzahl der versteckten Zustände des größeren Modells minus eins gewählt werden.

Der Beweis benötigt einige Vorbereitungen.

LEMMA 3.46. *Sei $g \in l_\infty$ ein endlich-dimensionaler Vektor mit $\dim g = m$. Dann gibt es eine Menge von Wörtern $\bar{a}_k, k = 0, \dots, m-1$ mit $|\bar{a}_k| \leq k$, so dass $(\tau_{\bar{a}_k} g)_{k=0, \dots, m-1}$ eine linear unabhängige Familie von Vektoren und damit eine Basis von \mathcal{V}_g bildet.*

BEWEIS. Wir konstruieren diese Basis rekursiv. Dazu starten wir mit

$$\bar{a}_0 = \square,$$

d.h. $\tau_{\bar{a}_0} g = g$. Wir schreiben

$$\mathcal{V}_g^k := \text{span}\{\tau_{\bar{c}} g \mid k \geq |\bar{c}|\}.$$

Haben wir die Wörter $\bar{a}_0, \dots, \bar{a}_k$ für $k < m-1$ bereits gefunden, so nehmen wir zunächst an, dass

$$\mathcal{V}_g^{k+1} \subset \mathcal{V}_g^k.$$

Daraus folgt dann wegen

$$\tau_a(\mathcal{V}_g^{k+1}) \subset \tau_a(\mathcal{V}_g^k) \subset \mathcal{V}_g^{k+1} \subset \mathcal{V}_g^k$$

und $\mathcal{V}_g^{k+2} = \text{span}\{\tau_a(\mathcal{V}_g^{k+1}) \mid a \in \Sigma\}$ induktiv auch $\mathcal{V}_g^m \subset \mathcal{V}_g^k$ für alle $m \geq k$ und daher wegen $\mathcal{V}_g = \text{span} \cup_{k \geq 0} \mathcal{V}_g^k$

$$\mathcal{V}_g \subset \mathcal{V}_g^k.$$

Somit können wir ein Wort \bar{a}_{k+1} mit $|\bar{a}_{k+1}| \leq k < k+1$ wählen, so dass $(\tau_{\bar{a}_i} g)_{i=0, \dots, k+1}$ eine linear unabhängige Familie ist. Im Falle

$$\mathcal{V}_g^k \subsetneq \mathcal{V}_g^{k+1}$$

können wir offensichtlich ein Wort $\bar{a}_{k+1} \in \Sigma^{k+1}$ bestimmen, so dass

$$(\tau_{\bar{a}_i} g)_{0 \leq i \leq k+1}$$

eine linear unabhängige Familie ist. \square

Darauf aufbauend erhalten wir die eigentliche Einsicht:

LEMMA 3.47. *Sei $g \in l_\infty$ mit $\dim g = m$. Dann gilt*

$$g \equiv 0 \iff \forall \bar{a}, |\bar{a}| \leq m-1 : g(\bar{a}) = 0.$$

BEWEIS. Die erste Richtung ist trivial. Für die Rückrichtung wählen wir gemäß Lemma 3.46 eine Basis $(\tau_{\bar{a}_i} g)_{0 \leq i \leq m-1}$ mit $|\bar{a}_i| \leq i$. Sei nun $\bar{b} \in \Sigma^*$ beliebig und $\tau_{\bar{b}} g = \sum_{i=0}^{m-1} \alpha_i \tau_{\bar{a}_i} g$ eine Darstellung über der Basis. Dann rechnen wir

$$g(\bar{b}) = \tau_{\bar{b}} g(\square) = \sum_{i=0}^{m-1} \alpha_i \tau_{\bar{a}_i} g(\square) = \sum_{i=0}^{m-1} \alpha_i \underbrace{g(\bar{a}_i)}_{=0, |\bar{a}_i| \leq i \leq m-1} = 0$$

und erhalten die Behauptung. \square

Weiter bemerken wir:

LEMMA 3.48. *Seien $g, h \in l_\infty$. Dann*

$$\dim(g+h) \leq \dim g + \dim h.$$

BEWEIS. Das folgt aus

$$\mathcal{V}_{(g+h)} \subset \mathcal{V}_g + \mathcal{V}_h$$

und allgemein $\dim(V+W) \leq \dim V + \dim W$ für Untervektorräume V, W eines gemeinsamen Obervektorraums. \square

Nun ist der **Beweis von Satz 3.45** einfach. Stimmen zwei Vektoren $g, h \in l_\sigma$ auf allen Wörtern der Länge $2m-1$ überein, so stimmen sie wegen $(\bar{a} \in \Sigma^*$ mit $|\bar{a}| \leq 2m-1)$

$$\begin{aligned} g(\bar{a}) \stackrel{\sigma g = g}{=} \sigma^{2m-1-|\bar{a}|} g(\bar{a}) &= \sum_{\bar{b} \in \Sigma^{2m-1-|\bar{a}|}} g(\underbrace{\bar{b}\bar{a}}_{\in \Sigma^{2m-1}}) \\ &\stackrel{g|_{\Sigma^{2m-1}} = h|_{\Sigma^{2m-1}}}{=} \sum_{\bar{b} \in \Sigma^{2m-1-|\bar{a}|}} h(\underbrace{\bar{b}\bar{a}}_{\in \Sigma^{2m-1}}) \\ &= \sigma^{2m-1-|\bar{a}|} h(\bar{a}) \stackrel{\sigma h = h}{=} h(\bar{a}) \end{aligned}$$

auf allen Wörtern der Länge kleiner gleich $2m-1$ überein. Daher $(g-h)(\bar{a}) = 0$ für alle Wörter mit $|\bar{a}| \leq 2m-1$ und wegen Lemma 3.48

$$\dim(g-h) \leq \dim g + \dim(-h) = \dim g + \dim h \leq 2m.$$

Deshalb wegen Lemma 3.47

$$g - h = 0$$

und wir haben die Behauptung des Satzes gezeigt (die andere Richtung ist trivial). \square

Wir wollen an dieser Stelle das im Vorwort erwähnte Positivitätsproblem formulieren. Dies, weil der soeben aufgeführte Satz vielleicht Inspiration für neue Ansätze zu seiner Lösung sein könnte.

PROBLEM 3.49. Sei $g \in l_\sigma$ ein endlich-dimensionaler, signierter Maßvektor. Gesucht ist ein Algorithmus, der anhand einer Koordinatendarstellung des Prognoseraums zusammen mit den zugehörigen Matrixdarstellungen der Prognoseoperatoren entscheidet, ob g ein Maßvektor ist, das heißt, $g \geq 0$.

3.3. Basen und Charakteristische Systeme

3.3.1. Einleitung und Definitionen. Will man reellwertige Koordinatendarstellungen des Prognoseraums eines endlich-dimensionalen Prozesses samt der dazugehörigen Parametrisierungen der Prognoseoperatoren studieren, stellt sich unmittelbar die Frage nach einer geeigneten Basis für den Prognoseraum. Diverse Kriterien können bei der Auswahl der Basis eine Rolle spielen. Möchte man zum Beispiel die Parameter einer der Parametrisierung des Prozesses aus einer (Menge von) von ihm erzeugten Sequenz(en) "lernen", so sollte sich die Basis gewissen numerischen Problemen gegenüber gutmütig verhalten. Der von Jäger entwickelte Lernalgorithmus ([Jae00],[JZK⁺05]) erfordert die Inversion einer von der Auswahl der Basis abhängigen Matrix. Bei geschickter Wahl der Basis ist die Kondition dieser Matrix günstig und eventuelle Schätzungsfehler pflanzen sich nur unerheblich fort. Ein anderes Kriterium ist, mit Hilfe einer geeigneten Basis die Koordinaten der Prognosevektoren als Wahrscheinlichkeiten interpretieren zu können. Dies impliziert gleichzeitig die Nichtnegativität dieser Koordinaten, eine weitere angenehme Eigenschaft einer solchen Basis.

Tatsächlich sind die oben beschriebenen Kriterien erfüllbar, was der weitere Verlauf dieses Abschnitts bezeugen wird. Ist der Prozess endlich-dimensional, so lässt sich eine Basis finden, die allen oben beschriebenen Anforderungen gerecht wird. Wir konzentrieren uns in diesem Abschnitt jedoch ausschließlich auf das Kriterium der Interpretierbarkeit der Koordinaten der Prognosevektoren als Wahrscheinlichkeiten. Die numerischen Konsequenzen der Wahl einer solchen Basis sind ausführlich in den Schriften von Jäger und seiner Mitarbeiter ([Jae00],[Obe02],[Kre03], [JZK⁺05]) beschrieben worden.

Bevor wir mit der entscheidenden Definition starten, wollen wir darauf hinweisen, dass wir Zylindermengen im Sinne der einleitenden Ergebnisse und Hinweise in Kapitel 2 meist mit ihren definierenden Grundmengen identifizieren.

DEFINITION 3.50. Sei $g \in l_{SM}$ ein endlicher, signierter Maßvektor. Eine Familie $(g_i)_{i \in I \subset \mathbb{N}}$ von linear unabhängigen Vektoren $g_i \in \mathcal{V}_g$ heisst **charakteristisches System** von g , falls es eine Partition

$$\Omega = \dot{\bigcup}_{i \in I} A_i$$

von Mengen $A_i \subset \mathcal{B}_\Sigma$ gibt, so dass die folgenden beiden Eigenschaften erfüllt sind:

(i)

$$\text{card}(I) = \dim g,$$

(ii)

$$m_{g_i}(A_j) = \delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j. \end{cases}$$

Die A_i nennt man dementsprechend **charakteristische Ereignisse**. Ersetzt man Eigenschaft (ii) durch

(iia)

$$m_{g_i}(A_i) = 1, \quad m_{g_i}(\complement A_i) = 0,$$

so sprechen wir von **quasi-charakteristischem System**. Bildet ein (quasi-) charakteristisches System $(g_i)_{i \in I}$ eine Basis von \mathcal{V}_g , so sprechen wir auch von einer **(quasi-) charakteristischen Basis**.

BEMERKUNG 3.51. Man bemerkt sofort, dass im Falle endlich-dimensionaler Prozesse g ein charakteristisches System $(g_i)_{i \in I}$ immer auch eine charakteristische Basis von \mathcal{V}_g bildet.

Gibt es eine charakteristische Basis für einen Vektor g , so können wir die diesbezüglichen Parameter der Prognosevektoren wie oben gefordert als Wahrscheinlichkeiten interpretieren.

PROPOSITION 3.52. Sei g ein stochastischer Vektor und (g_i, A_i) eine charakteristische Basis von g . Ist dann $h \in \mathcal{V}_g$ (also z.B. ein Prognosevektor) und

$$h = \sum_{i \in I} \alpha_i g_i$$

seine Darstellung auf der Basis (g_i) (nur endlich viele der α_i sind von Null verschieden), so gilt:

$$m_h(A_i) = \alpha_i.$$

Die α_i entsprechen also gerade den Messwerten von h für die A_i . Ist $h \geq 0$, so folgt dann $\alpha_i \geq 0$.

BEWEIS. Das folgt sofort aus der Definition eines charakteristischen Systems:

$$m_h(A_j) = \sum_{i \in I} \alpha_i m_{g_i}(A_j) = \alpha_j g_j(A_j) = \alpha_j.$$

Ist $h \geq 0$, so ist m_h ein Maß, und daher $m_h(A_j) \geq 0$. □

Die Frage nach der Existenz charakteristischer Systeme ist im Falle endlich-dimensionaler stochastischer Vektoren von Jäger positiv beantwortet worden ([Jae97b]). Im Falle eines unendlich-dimensionalen Vektors jedoch kann keine solche Aussage gemacht werden, hier existieren nur die vergleichsweise unbrauchbaren quasi-charakteristischen Systeme.

3.3.2. Der endlich-dimensionale Fall. Im Falle eines endlich-dimensionalen $g \in l_{SM}$ verallgemeinern wir den bereits erwähnten Satz von Jäger. Wir folgen dabei seinen Argumenten, müssen jedoch aufpassen, da in seinem Beweis von der Positivität der Vektoren an entscheidender Stelle Gebrauch gemacht wird.

SATZ 3.53. *Sei $g \in l_{SM}$ ein endlicher, signierter Maßvektor mit $\dim g < \infty$. Dann gibt es ein charakteristisches System $(g_i)_{i \in I}$ zu g .*

BEWEIS. Ist $g = 0$, so ist nichts zu zeigen. Ist $g \neq 0$, so sei $n := \dim g$ und und \bar{b}_1 ein Wort mit $g(\bar{b}_1) \neq 0$ (*). Dann ist der zugehörige Vektor $\tau_{\bar{b}_1} g$ von Null verschieden. Wir können somit eine Basis $(\tau_{\bar{b}_1}, \dots, \tau_{\bar{b}_n})$ von \mathcal{V}_g finden, die den Prognosevektor von \bar{b}_1 enthält. Mit Lemma 2.43 finden wir dann paarweise verschiedene Wörter $\bar{a}_1, \dots, \bar{a}_n$, so dass die Matrix

$$\begin{pmatrix} \tau_{\bar{b}_1} g(\bar{a}_1) & \dots & \tau_{\bar{b}_n} g(\bar{a}_1) \\ \vdots & \ddots & \vdots \\ \tau_{\bar{b}_1} g(\bar{a}_n) & \dots & \tau_{\bar{b}_n} g(\bar{a}_n) \end{pmatrix}$$

regulär ist. Wir bemerken, dass für zwei Wörter \bar{a}_i, \bar{a}_j die folgende Beziehung gilt:

$$\bar{a}_i \leq \bar{a}_j \iff Z(\{\bar{a}_j\}) \subset Z(\{\bar{a}_i\}). \quad (3.21)$$

Wir bemerken dann weiter, dass für alle $i = 1, \dots, n$

$$\bigcup_{j \neq i, \bar{a}_i \leq \bar{a}_j} Z(\{\bar{a}_j\}) \subsetneq Z(\{\bar{a}_i\}),$$

wobei aus 3.21 zunächst die Inklusionsbeziehung folgt, und die Annahme der Gleichheit der beiden Mengen auf die Gleichungen

$$\sum_{j \neq i, \bar{a}_i \leq \bar{a}_j} \tau_{\bar{b}_k} g(\bar{a}_j) = \sum_{j \neq i, \bar{a}_i \leq \bar{a}_j} \tau_{\bar{b}_k} g(Z(\{\bar{a}_j\})) = \tau_{\bar{b}_k} g(Z(\{\bar{a}_i\})) = \tau_{\bar{b}_k} g(\bar{a}_i)$$

für alle $k = 1, \dots, n$ und damit auf die lineare Abhängigkeit der entsprechenden Zeilen der obigen Matrix führen würde. Mit dem Wissen um diese Tatsache setzen wir für alle $i = 1, \dots, n$

$$\tilde{A}_k := Z(\{\bar{a}_i\}) \setminus \bigcup_{j \neq i, \bar{a}_i \leq \bar{a}_j} Z(\{\bar{a}_j\})$$

und bemerken, dass die \tilde{A}_k nicht-leer und paarweise disjunkt sind und dass die dadurch gegebene Matrix

$$M := \begin{pmatrix} \tau_{\bar{b}_1} g(\tilde{A}_1) & \dots & \tau_{\bar{b}_n} g(\tilde{A}_1) \\ \vdots & \ddots & \vdots \\ \tau_{\bar{b}_1} g(\tilde{A}_n) & \dots & \tau_{\bar{b}_n} g(\tilde{A}_n) \end{pmatrix}$$

ebenfalls regulär ist. Wir setzen zuguterletzt

$$A_0 := \Omega \setminus \left(\bigcup_{k=1}^n \tilde{A}_k \right)$$

und betrachten den Vektor $v_0 := (\tau_{\bar{b}_1} g(A_0), \dots, \tau_{\bar{b}_n} g(A_0))$, der wegen der Regularität von M im Zeilenraum von M zu finden ist. Sei

$$v_0 := \sum_{k=1}^n \alpha_k M_k,$$

wobei M_k die k -te Zeile von M sei, die entsprechende Darstellung. Wir bemerken, dass die Annahme $\forall k : \alpha_k = -1$ auf die Gleichung

$$g(\bar{b}_1) = \tau_{\bar{b}_1} g(\square) = \sum_{k=1}^n \tau_{\bar{b}_1} g(\tilde{A}_k) + \tau_{\bar{b}_1} g(A_0) = \sum_{k=1}^n \tau_{\bar{b}_1} g(\tilde{A}_k) - \tau_{\bar{b}_1} g(\tilde{A}_k) = 0,$$

föhrt, ein Widerspruch zur Wahl von \bar{b}_1 (siehe (*)). Wir suchen demnach ein k_0 , so dass $\alpha_{k_0} \neq -1$ und setzen

$$A_k := \begin{cases} \tilde{A}_k & k \neq k_0 \\ \tilde{A}_k \cup A_0 & k = k_0, \end{cases}$$

stellen dann fest, dass die A_k eine Partition von Ω bilden und dass die Matrix

$$N := \begin{pmatrix} \tau_{\bar{b}_1} g(A_1) & \dots & \tau_{\bar{b}_n} g(A_1) \\ \vdots & \ddots & \vdots \\ \tau_{\bar{b}_1} g(A_n) & \dots & \tau_{\bar{b}_n} g(A_n) \end{pmatrix}$$

wiederum regulär ist. Wir setzen nun zunächst

$$(\beta_{k1}, \dots, \beta_{kn})^T := N^{-1} e_k$$

(wobei $e_k := (0, \dots, 0, \underset{k}{1}, 0, \dots, 0)^T$) und weiter

$$g_k := \sum_{j=1}^n \beta_{kj} \tau_{\bar{b}_j} g$$

und stellen schlussendlich fest

$$g_k(A_i) = \sum_{j=1}^n \beta_{kj} \tau_{\bar{b}_j} g(A_i) = \begin{cases} 1 & i = k \\ 0 & i \neq k \end{cases}$$

nach Wahl der β_{kj} . Damit haben wir ein charakteristisches System bestimmt. \square

Wir bemerken zum Schluss, dass die charakteristischen Ereignisse endlich-dimensionaler Vektoren als Zylindermengen gewählt werden können.

3.3.3. Der unendlich-dimensionale Fall. Dieser Abschnitt hat über die oben besprochene Motivation hinaus eine weitere, speziellere: Für stationäre, unendlich-dimensionale, stochastische Vektoren kann es charakteristische Basen nicht geben. Der Beweis dazu (und natürlich die Definition von Stationarität) ist im Abschnitt über stationäre Vektoren zu finden, siehe Satz 4.20. Hätte man nun die Existenz charakteristischer Basen ganz allgemein beweisen können, wäre die Existenz eines stationären, unendlich-dimensionalen Prozesses ausgeschlossen gewesen. Dem Dimensionsbegriff wäre somit noch eine andere Bedeutung zuteil geworden. Tatsächlich gibt es aber stationäre, unendlich-dimensionale, stochastische Vektoren. Siehe dazu ebenfalls den Abschnitt über Stationarität (Satz 4.21).

Wie bereits angekündigt, muss im Falle eines unendlich-dimensionalen g 's ungleich viel mehr gearbeitet werden, um weniger zu erreichen. Genau gesagt erhält man unter einigen Mühen die Existenz quasi-charakteristischer Systeme, eine Aussage, deren Wert dem Autor noch lange nicht an Wünschenswertes heranreicht.

Bevor wir starten, wollen wir bemerken, dass wir in der Folge für beliebige $B \in \mathcal{B}_\Sigma$ (also nicht ausschließlich für Zylindermengen B)

$$\tau_{\bar{b}}g(B) \text{ anstelle von } m_{\tau_{\bar{b}}g}(B)$$

schreiben, wie bereits in früheren Kapiteln angekündigt.

SATZ 3.54. *Sei $g \in l_{SM}$ ein endlicher, signierter Maßvektor mit $\dim g = \infty$. Dann gibt es ein quasi-charakteristisches System $(g_i)_{i \in \mathbb{N}}$ zu g .*

BEWEIS. Wir machen die zu g assoziierte Matrix (siehe Bemerkung 3.18)

$$\mathcal{P}_g = (\tau_{\bar{b}}g(\bar{a}))_{\bar{a}, \bar{b} \in \Sigma^*} \in \mathbb{R}^{\mathbb{N} \times \mathbb{N}} \quad (3.22)$$

zum zentralen Objekt unserer Betrachtungen.

Wir können die Spaltenvektoren $(\mathcal{P}_g)^{\bar{b}}$ mit den $\tau_{\bar{b}}g$ identifizieren und schreiben darüberhinaus $f_{\bar{a}}$ für die Zeilenvektoren, z.B. $f_{\square}(\bar{b}) = \tau_{\bar{b}}g(\square) = g(\bar{b})$. Wegen $\dim g = \infty$ gilt unter Zuhilfenahme von Lemma 2.43 folglich

$$\dim \text{span}\{f_{\bar{a}} | \bar{a} \in \Sigma^*\} = \infty. \quad (3.23)$$

Idee hier ist nun der rekursive Aufbau einer Partition von Ω unter Zuhilfenahme einer geeigneten Sequenz von Wörtern \bar{a}_n , so dass die $f_{\bar{a}_n}$ eine linear unabhängige Familie von Vektoren bilden. Dazu betrachten wir das folgende

LEMMA 3.55. *Sei $(\bar{a}_n)_{n \in \mathbb{N}}$ eine Folge paarweise verschiedener Wörter über Σ . Dann liegt einer der beiden Fälle vor:*

(T1) *Es gibt eine Teilfolge (\bar{a}_{n_i}) , so dass*

$$\forall i, j \in \mathbb{N}^2, i \leq j : \quad \bar{a}_{n_i} \leq \bar{a}_{n_j},$$

mit anderen Worten, die Teilfolge ist ein sich "streng monoton" verlängerndes Wort oder

(T2) *es gibt eine Teilfolge (\bar{a}_{n_i}) , so dass*

$$\forall i, j \in \mathbb{N}^2, i \neq j: \quad \bar{a}_{n_k} \not\leq \bar{a}_{n_l} \wedge \bar{a}_{n_l} \not\leq \bar{a}_{n_k},$$

mit anderen Worten kein Teilfolgenglied ist Präfix eines anderen.

BEWEIS. Wir gehen davon aus, dass Fall (T1) nicht vorliegt. Dann haben wir die Existenz einer Teilfolge vom zweiten Typ zu zeigen. Diese finden wir rekursiv. Wir stellen fest, dass in unserer Folge jede induktiv geordnete Teilmenge der v_n eine obere Schranke hat. Mit dem Lemma von Zorn wählen wir ein maximales Element \bar{a}_{n_1} , also ein Wort, das durch andere Folgenglieder nicht mehr fortgesetzt werden kann. Haben wir bereits die Wörter $\bar{a}_{n_1}, \dots, \bar{a}_{n_m}$ gefunden, so betrachten wir die Teilfolge der Originalfolge, aus der alle (endlich vielen) Wörter gestrichen wurden, die Präfixe von $\bar{a}_{n_1}, \dots, \bar{a}_{n_m}$ waren, einschliesslich der $\bar{a}_{n_1}, \dots, \bar{a}_{n_m}$ selbst. In dieser Teilfolge finden wir wiederum ein maximales Element $\bar{a}_{n_{m+1}}$. Offensichtlich ist $\bar{a}_{n_{m+1}}$ nicht das Präfix eines der Wörter $\bar{a}_{n_1}, \dots, \bar{a}_{n_m}$. Da diese maximale Folgenglieder waren, können diese ebenfalls nicht Präfix von $\bar{a}_{n_{m+1}}$ sein. \square

Weiter im Beweis des Satzes: Wir betrachten nun eine Basis $(f_{\bar{a}_n})_{n \in \mathbb{N}}$ des Zeilenraums von \mathcal{P} , s.d. f_{\square} Element dieser Basis ist. Wir betrachten die Folge der zugrundeliegenden, paarweise verschiedenen Wörter \bar{a}_n und unterscheiden gemäß Lemma 3.55 zwei Fälle:

(T1): Sei $(\bar{c}_i := \bar{a}_{n_i})$ diese Teilfolge (also $\bar{c}_i \leq \bar{c}_j$ für $i \leq j$). Es sei $\bar{c}_0 := \square$ gewählt. Mit Lemma 2.43 finden wir eine Folge von Wörtern $\bar{b}_j, j \in \mathbb{N}$, so dass die Matrizen

$$C_n := (f_{\bar{c}_i}(\bar{b}_k))_{i,j=0,\dots,n} = \begin{pmatrix} \tau_{\bar{b}_0}g(\square) & \cdots & \tau_{\bar{b}_n}g(\square) \\ \vdots & \ddots & \vdots \\ \tau_{\bar{b}_0}g(\bar{c}_n) & \cdots & \tau_{\bar{b}_n}g(\bar{c}_n) \end{pmatrix} \in \mathbb{R}^{(n+1) \times (n+1)}$$

regulär sind für alle n . Für alle $i \in \mathbb{N}$ definieren wir Zylindermengen A_i durch die Vorschrift

$$A_i := Z(\{\bar{c}_i\}) \setminus Z(\{\bar{c}_{i+1}\}).$$

Dann gilt natürlich auch, dass die Matrizen

$$B_n := \begin{pmatrix} \tau_{\bar{b}_0}g(A_0) & \cdots & \tau_{\bar{b}_{n+1}}g(A_0) \\ \vdots & \ddots & \vdots \\ \tau_{\bar{b}_0}g(A_n) & \cdots & \tau_{\bar{b}_{n+1}}g(A_n) \\ \tau_{\bar{b}_0}g(\bar{c}_{n+1}) & \cdots & \tau_{\bar{b}_{n+1}}g(\bar{c}_{n+1}) \end{pmatrix} \in \mathbb{R}^{(n+2) \times (n+2)}$$

regulär sind. Weiter sind die A_i paarweise disjunkt und es gilt

$$\left(\bigcup_{i=0}^n A_i \right) \cup Z(\{\bar{c}_{n+1}\}) = \Omega_{\Sigma}$$

(beachte $\bar{c}_0 = \square$). Wir setzen nun $e_n := (0, \dots, 0, 1, 0)^T \in \mathbb{R}^{n+2}$,

$$(\alpha_{n0}, \alpha_{n1}, \dots, \alpha_{n,n+1}) := (B_n^{-1} e_n)^T$$

und weiter

$$g_n := \alpha_{n0} \tau_{\bar{b}_0} g + \dots + \alpha_{n,n+1} \tau_{\bar{b}_{n+1}} g.$$

Dann gilt

$$\begin{aligned} \forall i = 0, \dots, n-1 : \quad g_n(A_k) &= 0 \\ g_n(A_n) &= 1 \\ g_n(\Omega_\Sigma \setminus (\bigcup_{k=0}^n A_k)) &= g_n(\bar{c}_{n+1}) = 0. \end{aligned}$$

Hieraus folgert man einfach die lineare Unabhängigkeit der g_n sowie noch einfacher Eigenschaft (iii).

(T2): Wir konstruieren nun ein charakteristisches System nach der Grundidee des nachfolgenden Lemmas.

LEMMA 3.56. *Sei $g \in l_{SM}$ ein endlicher, signierter Maßvektor und sei eine Partition von Ω_Σ*

$$\Omega_\Sigma = \dot{\bigcup}_{i \geq 0} A_i,$$

mit dazu assoziierten Vektoren $h_i \in l_\infty^\Sigma, i \geq 0$ und einem Extravektor h_\square , die durch die Vorschrift

$$h_i(\bar{b}) := \tau_{\bar{b}}g(A_i) \text{ und } h_\square(\bar{b}) := \tau_{\bar{b}}g(\square)$$

festgelegt sind, gegeben. Ist dann die Menge dieser Vektoren

$$\{h_\square, h_i \mid i \geq 0\},$$

linear unabhängig, so gibt es eine Menge von Vektoren $g_i, i \geq 0$, so dass jeweils

$$g_i(A_i) = 1 \text{ und } g_i(\mathbb{C}A_i) = 0.$$

Mit anderen Worten gibt es ein charakteristisches System (g_i) , so dass die A_i die dazu passenden charakteristischen Ereignisse bilden.

BEWEIS. Wir konstruieren die zu den A_i gehörenden Vektoren rekursiv. Da die Menge $\{h_\square - h_0, h_0\}$ eine linear unabhängig ist, finden wir mit Lemma 2.43 Wörter \bar{b}_0, \bar{b}_1 , so dass die Matrix

$$B_0 := \begin{pmatrix} h_0(\bar{b}_0) & h_0(\bar{b}_1) \\ (h_\square - h_0)(\bar{b}_0) & (h_\square - h_0)(\bar{b}_1) \end{pmatrix} = \begin{pmatrix} \tau_{\bar{b}_0}g(A_0) & \tau_{\bar{b}_1}g(A_0) \\ \tau_{\bar{b}_0}g(\mathbb{C}A_0) & \tau_{\bar{b}_1}g(\mathbb{C}A_0) \end{pmatrix} \in \mathbb{R}^{2 \times 2}$$

regulär ist. Wir setzen $e_0 := (1, 0)^T$ und

$$(\alpha_{00}, \alpha_{01}) := (B_0^{-1}e_1)^T$$

und weiter

$$g_0 := \alpha_{00}g_{\bar{b}_0} + \alpha_{01}g_{\bar{b}_1}$$

und stellen fest, dass $g_0(A_0) = 1$ und $g_0(\mathbb{C}A_0) = 0$. Wiederum mit Lemma 2.43 finden wir dann für $n \geq 1$, da die Mengen $\{h_0, \dots, h_n, \tilde{h}_n := h_\square - h_0 -$

$\dots - h_n\}$ linear unabhängig sind, Wörter $\bar{b}_0, \bar{b}_1, \dots, \bar{b}_{n+1}$, so dass die Matrizen

$$B_n := \begin{pmatrix} h_0(\bar{b}_0) & \cdots & h_0(\bar{b}_{n+1}) \\ \vdots & \ddots & \vdots \\ h_n(\bar{b}_0) & \cdots & h_n(\bar{b}_{n+1}) \\ (\tilde{h}_n)(\bar{b}_0) & \cdots & (\tilde{h}_n)(\bar{b}_{n+1}) \end{pmatrix} \\ = \begin{pmatrix} \tau_{\bar{b}_0}g(A_0) & \cdots & \tau_{\bar{b}_{n+1}}g(A_0) \\ \vdots & \ddots & \vdots \\ \tau_{\bar{b}_0}g(A_n) & \cdots & \tau_{\bar{b}_{n+1}}g(A_n) \\ \tau_{\bar{b}_0}g(\Omega_\Sigma \setminus \cup_{i=0}^n A_i) & \cdots & \tau_{\bar{b}_{n+1}}g(\Omega_\Sigma \setminus \cup_{i=0}^n A_i) \end{pmatrix} \in \mathbb{R}^{(n+2) \times (n+2)}$$

regulär sind. Wir setzen nun $e_n := (0, \dots, 0, 1, 0) \in \mathbb{R}^{n+2}$ und

$$(\alpha_{n0}, \alpha_{n1}, \dots, \alpha_{n,n+1}) := (B_n^{-1}e_n)^T$$

und weiter

$$g_n := \alpha_{n0}\tau_{\bar{b}_0}g + \dots + \alpha_{n,n+1}\tau_{\bar{b}_{n+1}}g$$

und sehen sofort

$$\forall i = 0, \dots, n-1 : \quad \begin{aligned} g_n(A_k) &= 0 \\ g_n(A_n) &= 1 \\ g_n(\Omega \setminus (\cup_{i=0}^n A_k)) &= 0 \end{aligned}$$

Hieraus folgert man einfach die lineare Unabhängigkeit der g_n sowie trivialeweise Eigenschaft (iii) der Definition eines quasi-charakteristischen Systems. \square

Weiter im **Beweis von Satz 3.54**. In Anbetracht dieses Lemmas ist es nun unser Ziel, die dort geforderte Menge von Vektoren bereitzustellen.

Wir bezeichnen die entsprechend Fall **T2** aus Lemma 3.55 erhaltene Teilfolge mit $(\bar{c}_i := \bar{a}_{n_i})$ (also $\forall i, j : \bar{c}_i \not\leq \bar{c}_j, \bar{c}_j \not\leq \bar{c}_i$), wobei wir diese aus technischen Gründen mit dem Index 1 starten lassen. Wir bemerken weiter, dass in diesem Fall das Wort \square nicht in dieser Teilfolge auftreten kann. Wir setzen nun für $j \geq 1$

$$C_j := Z(\{\bar{c}_j\})$$

und

$$C_0 := \Omega_\Sigma \setminus \bigcup_{j \geq 1} C_j = \Omega_\Sigma \setminus \bigcup_{j \geq 1} Z(\{\bar{c}_j\})$$

als das Komplement aller so gewählten Mengen. Wir setzen weiter $f_i := f_{\bar{c}_i}$ für $i \geq 1$ und definieren f_0 durch die Vorschrift

$$f_0(\bar{b}) := \tau_{\bar{b}}g(C_0)$$

für $\bar{b} \in \Sigma^*$. Man kann f_0 auch als den punktweisen Limes der Folgenglieder $\tilde{f}_k := f_\square - \sum_{i=1}^k f_i$ auffassen. Es ist nicht klar, ob $f_0 \in \text{span}\{f_{\bar{a}} \mid \bar{a} \in \Sigma^*\}$.

Um die Situation aus Lemma 3.56 herzustellen, haben wir nun zwischen verschiedenen Fällen zu unterscheiden.

1. *Fall:* $\{f_{\square}, f_0, f_i \mid i \geq 1\}$ ist eine linear unabhängige Menge. In diesem Fall erhalten wir die gewünschte Situation, indem wir

$$\forall i \geq 0: \quad A_i := C_i \text{ und } h_i := f_i$$

und $h_{\square} := f_{\square}$ setzen.

2. *Fall:* $f_0 = \alpha f_{\square}$ für ein $\alpha \in \mathbb{R}$, also f_{\square} und f_0 linear abhängige Vektoren. Wir bemerken, dass $\{f_{\square}, f_i \mid i \geq 1\}$ nach Voraussetzung eine linear unabhängige Menge von Vektoren ist. Daher ist auch $\{f_{\square}, f_0 + f_1 = \alpha f_{\square} + f_1, f_i \mid i \geq 2\}$ eine linear unabhängige Menge von Vektoren. Wir setzen nun

$$A_0 := C_0 \cup C_1 \text{ und } \forall i \geq 1: \quad A_i := C_{i+1}$$

und dementsprechend $h_{\square} := f_{\square}, h_0 := f_0 + f_1, h_i := f_{i+1}, i \geq 1$ und stellen fest, die für Lemma 3.56 gewünschte Menge von Vektoren erhalten zu haben.

3. *Fall:* $f_0 \in \text{span}\{f_{\square}, f_i \mid i \geq 1\}$, aber f_{\square} und f_0 sind nicht linear abhängig. Wir suchen dann eine endliche Teilmenge $J \subset \mathbb{N}$ und schreiben

$$f_0 = \alpha_{\square} f_{\square} + \sum_{i \in J} \alpha_i f_i,$$

wobei die $\alpha_i \neq 0$ nach Voraussetzung dieses Falls. Dann wählen wir eine Bijektion $\phi: \mathbb{N} \rightarrow \mathbb{N} \setminus J$ und setzen

$$A_0 := C_0 \cup \bigcup_{i \in J} C_i \text{ und für } i \in \mathbb{N} \setminus J: \quad A_i := C_{\phi(i)}$$

und weiter $h_{\square} := f_{\square}, h_0 := f_0 + \sum_{i \in J} f_i$ und $h_i := f_{\phi(i)}$ für $i \geq 1$. Die Annahme $h_0 \in \text{span}\{h_{\square}, h_i \mid i \geq 1\}$ führte auf eine Gleichung $0 \neq \sum_{i \in J} \alpha_i f_i = \beta_{\square} + \sum_{i \geq 1} \beta_i f_i$, wobei nur endlich viele der β_i von Null verschieden wären. Diese Gleichung hieße wiederum aber, dass die Menge $\{f_{\square}, f_i \mid i \geq 1\}$ linear abhängig wäre, ein Widerspruch! Wir haben in den h_i somit eine für Lemma 3.56 passende Menge von Vektoren gefunden. \square

Wie schon oben erwähnt, der Beweis ist lang und mühselig, das Ergebnis scheint unbefriedigend. Eine Existenz echter charakteristischer Systeme scheint nur unter noch größeren Qualen erhalten werden zu können. Daher verlassen wir diesen schwer zerfurchten, trotz allem spärlich ertragreichen Acker und wenden uns erfreulicheren Dingen zu.

KAPITEL 4

Evolutionstheorie

*Everything happens to everybody
sooner or later if there is time enough.*

George Bernard Shaw

Die Veränderung von zufallsgesteuerten Prozessen im Wandel der Zeit ist ein Hauptthema etlicher Teildisziplinen der Wahrscheinlichkeitstheorie. Die Motivation solcher Studien liegt auf der Hand: Lassen sich mit Hilfe einer anfänglich gegebenen Parametrisierung stochastischer Objekte Rückschlüsse auf deren Verhalten im weiteren zeitlichen Verlauf schließen, so können Erwartungswerte stabil geschätzt werden und Volatilitäten zuverlässig angegeben werden. Dies hilft bei der Berechnung zufallsbedingter Risiken und kann in der Finanzwelt im wahrsten Sinne des Wortes Gold wert sein. Jedoch auch umgekehrt kann die Kenntnis des asymptotischen Verhaltens stochastischer Prozesse helfen, aus einzelnen Stichproben auf anfängliche Parametrisierungen rückzuschließen und damit stochastische Prozesse in ihrer Gesamtheit ein wenig in den Griff zu bekommen.

Mathematisches Zeugnis der Relevanz solcher Zusammenhänge ist zum Beispiel die Ergodentheorie, die sich in den dreißiger Jahren entwickelt hat und deren Ursprung in der Erforschung stochastischer Prozesse in der statistischen Mechanik, bei denen “zeitliche” Mittel mit Erwartungswerten übereinstimmen, zu sehen ist (siehe auch die Einleitung, Kapitel 1, Abschnitt 1.2.3). Die Ergodentheorie hat sich inzwischen von diesen Spezialitäten emanzipiert und hat im Laufe der Jahre wilde Blüten getrieben, so dass dicke Bücher dazu geschrieben werden konnten. Grundthema ist jedoch immer eine Art zeitliche Entwicklung von Messwerten und das Studium derselben.

Auch wir wollen unsere Theorie um ein dem Studium “zeitlicher” Entwicklung unserer Objekte gewidmeten Kapitel bereichern. Dabei drängt es sich aus der Definition heraus auf, das Verhalten unserer (signierten) Maßvektoren unter dem Einfluss des Evolutionsoperators bzw. der Prognoseoperatoren gründlich unter die Lupe zu nehmen. Das folgende Kapitel beschäftigt sich daher vornehmlich mit den Auswirkungen der iterierten Anwendung des Evolutionsoperators und zum Teil auch der Prognoseoperatoren und bezieht seinen Namen daher.

Wir machen darauf aufmerksam, dass die Motivation wiederum in der Theorie der dynamischen Systeme, wie sie z.B. [Gra01] oder in [PY98]

betrieben wird, gesehen werden kann. Wir führen, wie bereits bisher in diesem Text üblich, die Querverbindungen meist in einem Atemzug mit auf. Ausschließlich an unseren Objekten interessierte Leser überspringen diese Hinweise geflissentlich.

4.1. Motivation: Dynamische Systeme

Wie bereits vorab erwähnt, ist es wiederum nicht zwingend notwendig, aber doch möglicherweise hilfreich, die Verknüpfung unseres Modells mit den Begriffen der sie umgebenden Theorie, in diesem Fall der Ergodentheorie der dynamischen Systeme aufzuzeigen. Wie schon in der Einleitung berichtet, kann die Theorie der dynamischen Systeme als eine Teildisziplin der allgemeiner formulierbaren Ergodentheorie aufgefasst werden. Was das Studium dieses Teilbereichs der Ergodentheorie so populär macht, ist, dass stochastische Prozesse als dynamische Systeme aufgefasst werden können. Übersetzt man die Ergebnisse der Theorie der dynamischen Systeme in die Sprache der stochastischen Prozesse, so erhält man meist Aussagen über Statistiken, die anhand einzelner Sequenzen (also anhand einzelner Stichproben des stochastischen Prozesses) erhoben werden. Tatsächlich hat die Ergodentheorie einige starke Resultate unter vergleichsweise schwachen und in der Praxis oft zutreffenden Grundannahmen erzielt.

Neben dem rein theoretischen Nutzen einer Evolutions-/Ergodentheorie für unser Modell hat, wie schon an anderer Stelle dieser Arbeit erwähnt, eine solche Theorie auch einen ganz konkreten, praktischen Nutzen. Aufgrund der endlichen Parametrisierbarkeit endlich-dimensionaler Prozesse liegt die Vermutung nahe, dass diese Modelle aus von ihnen erzeugten Daten ermittelt werden können. Dies wiederum ist davon abhängig, wieviel Information aus Stichproben des stochastischen Prozesses zurückgewonnen werden kann - eine Frage, zu der eine Ergodentheorie dieser Modelle Auskunft geben kann.

Wir starten an dieser Stelle mit der allgemeinen Definition eines dynamischen Systems, nicht ohne den Leser, wie bereits gesagt, dazu zu ermutigen, diese einfach zu überspringen - sie ist zum Verständnis unseres Modells nicht notwendig, sondern dient ausschließlich dazu, es zu interpretieren.

DEFINITION 4.1. *Ein dynamisches System ist ein Quadrupel $(\Omega, \mathcal{B}, m, T)$, bestehend aus einem Maßraum (Ω, \mathcal{B}, m) und einer messbaren Abbildung T .*

In unserem speziellen Fall werden wir bei gegebenem Alphabet Σ immer dynamische Systeme der Struktur

$$(\Omega_\Sigma, \mathcal{B}_\Sigma, m, T_\Omega)$$

d.h. $(\Omega_\Sigma, \mathcal{B}_\Sigma)$ ist der von uns fast ausschließlich betrachtete Messraum mit einem Maß m und T_Ω ist der Schiebeoperator. Diese speziellen dynamischen Systeme geben Anlass zu einer Folgedefinition:

DEFINITION/BEMERKUNG 4.2. Ist (X_t) ein stochastischer Prozess, so ist das durch ihn **kanonisch induzierte dynamische System** gerade

$$(\Omega_\Sigma, \mathcal{B}_\Sigma, m, T_\Omega),$$

also $(\Omega_\Sigma, \mathcal{B}_\Sigma, m)$ der durch (X_t) kanonisch gegebene Wahrscheinlichkeitsraum (siehe Abschnitt 2.2.4) und T_Ω der Schiebeoperator (Abschnitt 2.2.5). In diesem Sinne beziehen sich die Begriffe der Welt der dynamischen Systeme (z.B. Stationarität) auf stochastische Prozesse.

Studiert man nun ein dynamisches System, so interessiert man sich meist für die induzierten Maße

$$m \circ T^{-k}, \quad k \in \mathbb{N}$$

und deren Gesetzmäßigkeiten. Betrachten wir Maße m , so praktisch ausschließlich, indem wir deren Maßvektoren g betrachten. Für die in Kapitel 2 etablierte 1-1-Beziehung zwischen Maßen auf $(\Omega_\Sigma, \mathcal{B}_\Sigma)$ und Maßvektoren $g \in l_\Sigma$ fällt dann auf (siehe dazu Proposition 3.14), dass

$$g \leftrightarrow m \iff \mu^k g \leftrightarrow m \circ T_\Omega^{-k},$$

mit anderen Worten ist Ergodentheorie in unserem Modell dann gerade das Studium der Auswirkungen des Evolutionsoperators auf Maßvektoren $g \in l_{SM}$.

4.2. Stationarität

In diesem Abschnitt definieren wir den Begriff eines stationären Vektors und zeigen den Zusammenhang mit dem entsprechenden Begriff aus der Welt der dynamischen Systeme auf. Stationäre Vektoren sind Fixpunkte des Evolutionsoperators, mit anderen Worten, Vektoren, die sich zeitlich nicht verändern wollen. Wir geben die Definition an und zeigen Zusammenhänge mit anderen Begriffen unseres Modells auf.

4.2.1. Einführung und Motivation. Wir starten geradewegs mit der Definition dieses Abschnitts.

DEFINITION 4.3. Wir nennen einen Vektor $g \in l_\infty$ **stationär**, falls

$$\mu g = g, \tag{4.1}$$

mit anderen Worten, falls er invariant bezüglich des Evolutionsoperators ist.

Zur Aufklärung der Begriffsbildung betrachten wir gleich im Anschluss eine weitere Definition, die diesmal der Welt der dynamischen Systeme entstammt.

DEFINITION 4.4. Ein dynamisches System $(\Omega, \mathcal{B}, m, T)$ wird **stationär (bezüglich T)** genannt, falls

$$\forall B \in \mathcal{B} : \quad m(T^{-1}B) = m(B).$$

Einen stochastischen Prozess mit Werten in einem endlichen Alphabet Σ bezeichnet man als **stationär**, falls das durch ihn induzierte, kanonisch gegebene dynamische System $(\Omega_\Sigma, \mathcal{B}_\Sigma, m, T_\Omega)$ stationär ist.

Wir werden den Begriff der Stationarität auch auf Maße anwenden, wobei dann klar sein muss, um welches dynamische System es sich dabei eigentlich handelt. Wird nichts anderes gesagt, so handelt es sich bei uns um die kanonisch durch den Schiebeoperator induzierten dynamischen Systeme, gegeben ein Maß auf dem Messraum $(\Omega_\Sigma, \mathcal{B}_\Sigma)$.

Um die Stationarität eines dynamischen Systems (oder eben eines stochastischen Prozesses, siehe Bem./Def. 4.2) zu bestätigen, reicht es, die definierende Eigenschaft auf den Elementen eines Erzeugendensystems der involvierten σ -Algebra festzustellen.

LEMMA 4.5. *Ein dynamisches System $(\Omega, \mathcal{B} = \sigma(\mathcal{F}), m, T)$ ist genau dann stationär, falls*

$$\forall F \in \mathcal{F} : \quad m(T^{-1}F) = m(F),$$

also falls es invariant bezüglich T auf den Ereignissen des erzeugenden Rings ist.

BEWEIS. Folgt mit Hilfe des Approximationssatzes für Maße 2.18, der ausführliche Beweis dazu findet sich in [Gra01], siehe S.162, Lemma 7.4.1. \square

Im Falle durch Maßvektoren g induzierter Maße m_g auf $(\Omega_\Sigma, \mathcal{B}_\Sigma)$ bedeutet das, dass wir Stationarität auf den Zylindermengen allein nachweisen müssen. Das hilft uns, den folgenden Satz nachweisen zu können, der gleichzeitig die Motivation der Namensgebung ist.

SATZ 4.6. *Für einen durch einen stochastischen Vektor g gegebenen stochastischen Prozess $(X_t)_{t \in \mathbb{N}}$ gilt:*

$$(X_t)_{t \in \mathbb{N}} \text{ stationär} \Leftrightarrow g \text{ stationär} \Leftrightarrow \text{edim}(X_t) = 1. \quad (4.2)$$

BEWEIS. Die zweite Äquivalenz ist klar nach Definition 3.16 von *edim*. Sei g ein stationärer, stochastischer Vektor. Dann gilt für Wörter $\bar{a} \in \Sigma^*$:

$$\begin{aligned} m_g(T_\Omega^{-1}Z(\{\bar{a}\})) &= \sum_{a \in \Sigma} m_g(Z(\{a\bar{a}\})) = \sum_{a \in \Sigma} g(a\bar{a}) \\ &= \mu g(\bar{a}) = g(\bar{a}) = m_g(Z(\{\bar{a}\})). \end{aligned}$$

Damit dann offensichtlich auch $m_g(T_\Omega^{-1}Z(B)) = m_g(Z(B))$ auf den Zylindermengen $Z(B)$ mit beliebigen Grundmengen von Wörtern B . Hieraus folgt die Stationarität von m_g aus Lemma 4.5. Ist umgekehrt g_m ein durch einen stationären, stochastischen Prozess m induzierter Vektor, so gilt

$$\begin{aligned} \mu g_m(\bar{a}) &= \sum_{a \in \Sigma} g_m(a\bar{a}) = \sum_{a \in \Sigma} m(Z(\{a\bar{a}\})) \\ &= m(T_\Omega^{-1}Z(\{\bar{a}\})) = m(Z(\{\bar{a}\})) = g_m(\bar{a}). \end{aligned}$$

□

Die stationären Vektoren als Fixpunkte des Evolutionsoperators bilden nun einen Untervektorraum von l_∞ . Dies führt uns auf folgende

DEFINITION 4.7. *Wir setzen*

$$l_\mu := \{g \in l_{SM} \mid \mu g = g\},$$

also den Vektorraum der stationären, endlichen, signierten Maßvektoren. Dies ist offensichtlich ein Untervektorraum von l_{SM} .

Das folgende Beispiel zeigt, dass im Allgemeinen ein Fixpunkt g von μ nicht aus l_{TV} sein muss, wie man zunächst vermuten könnte.

BEISPIEL 4.8. *Sei $\Sigma = \{a, b\}$. Wir definieren rekursiv*

$$g : \Sigma^* \longrightarrow \mathbb{R}$$

$$\bar{a} \longmapsto \begin{cases} 0 & \bar{a} = \square \\ g(\bar{b}) + \frac{1}{2^{|\bar{b}|+1}} & \bar{a} = \bar{b}a \\ g(\bar{b}) - \frac{1}{2^{|\bar{b}|+1}} & \bar{a} = \bar{b}b. \end{cases}$$

Dass $(\mu g)(\bar{a}) = g(\bar{a})$ gilt, folgt mit vollständiger Induktion über die Wortlänge $t := |\bar{a}|$:

$t = 0$:

$$(\mu g)(\square) = g(a) + g(b) = \frac{1}{2} - \frac{1}{2} = 0 = g(\square).$$

$t \rightarrow t + 1$: *Sei $\bar{a} \in \Sigma^{t+1}$. Dann $\bar{a} = \bar{b}x$ mit $\bar{b} \in \Sigma^t$. Es folgt*

$$\begin{aligned} (\mu g)(\bar{a}) &= g(a\bar{a}) + g(b\bar{a}) = g(a\bar{b}x) + g(b\bar{b}x) \\ &= \begin{cases} g(a\bar{b}) + g(b\bar{b}) + 2 \frac{1}{2^{|\bar{b}|+2}} \stackrel{IV}{=} g(\bar{b}) + \frac{1}{2^{|\bar{b}|+1}}, & x = a \\ g(a\bar{b}) + g(b\bar{b}) - 2 \frac{1}{2^{|\bar{b}|+2}} \stackrel{IV}{=} g(\bar{b}) - \frac{1}{2^{|\bar{b}|+1}}, & x = b \end{cases} \quad (4.3) \\ &= g(\bar{b}x) = g(\bar{a}). \end{aligned}$$

Weiter gilt

$$\forall t \geq 1 : \sum_{\bar{a} \in \Sigma^t} |g(\bar{a})| = 2^{t-1},$$

wie man wiederum mit vollständiger Induktion einsieht:

$t = 1$:

$$\sum_{\bar{a} \in \Sigma} |g(\bar{a})| = |g(a)| + |g(b)| = \frac{1}{2} + \frac{1}{2} = 1.$$

$t \rightarrow t + 1$: *Wir bemerken zunächst, dass offensichtlich $|g(\bar{a})| \geq \frac{1}{2^{|\bar{a}|}}$ und daher, dass $g(\bar{a})$, $g(\bar{a}) + \frac{1}{2^{|\bar{a}|}}$ und $g(\bar{a}) - \frac{1}{2^{|\bar{a}|}}$ alle das gleiche Vorzeichen haben, woraus wiederum*

$$|g(\bar{a}a)| + |g(\bar{a}b)| = |g(\bar{a}) + \frac{1}{2^{|\bar{a}|+1}}| + |g(\bar{a}) - \frac{1}{2^{|\bar{a}|+1}}| = 2|g(\bar{a})| \quad (4.4)$$

folgt. Damit

$$\begin{aligned} \sum_{\bar{a} \in \Sigma^{t+1}} |g(\bar{a})| &= \sum_{\bar{a} \in \Sigma^t} |g(\bar{a}a)| + |g(\bar{a}b)| \stackrel{(4.4)}{=} \sum_{\bar{a} \in \Sigma^t} 2|g(\bar{a})| \\ &\stackrel{IV}{=} 2 \cdot 2^{t-1} = 2^t. \end{aligned} \quad (4.5)$$

Es bleibt zu bemerken, dass es uns an dieser Stelle so auf die Schnelle kein Beispiel eines gemeinsamen Fixpunkts des stochastischen Operators σ und des Evolutionsoperators μ eingefallen ist, das nicht in l_{TV} zu finden ist (wir erinnern an Beispiel 2.57). Es bleibt somit eine offene Frage (?), ob bereits

$$\begin{aligned} l_\sigma \cap \{g \in l_\infty \mid \mu g = g\} &\stackrel{(?)}{=} l_\mu = l_{SM} \cap \{g \in l_\infty \mid \mu g = g\} \\ &= l_\sigma \cap l_{TV} \cap \{g \in l_\infty \mid \mu g = g\} \end{aligned}$$

gilt oder nicht. Wir lüften dieses Geheimnis an anderer Stelle (siehe Beispiel 4.28).

Der Raum l_μ ist nun abgeschlossen, wie die Folge der nachstehenden Proposition sein wird.

PROPOSITION 4.9. *Sei (g_n) eine Folge mit $\mu g_n = g_n$, die wortweise gegen einen Vektor g^* konvergiert. Dann auch $\mu g^* = g^*$.*

BEWEIS. Sei $\bar{a} \in \Sigma^*$, dann gilt

$$\begin{aligned} \mu g^*(\bar{a}) &= \sum_{a \in \Sigma} g^*(a\bar{a}) = \sum_{a \in \Sigma} \lim_{n \rightarrow \infty} g_n(a\bar{a}) \\ &= \lim_{n \rightarrow \infty} \sum_{a \in \Sigma} g_n(a\bar{a}) = \lim_{n \rightarrow \infty} \mu g_n(\bar{a}) \\ &= \lim_{n \rightarrow \infty} g_n(\bar{a}) = g^*(\bar{a}), \end{aligned}$$

also $\mu g^* = g^*$. □

KOROLLAR 4.10. *l_μ ist abgeschlossen.*

BEWEIS. Sei (g_n) eine Folge aus l_μ , die in Norm gegen einen Vektor g^* konvergiert. Dass $\sigma g^* = g^*$, folgt aus Satz 2.34 und $\mu g^* = g^*$ aus obiger Proposition 4.9. □

Wir beschließen diesen Abschnitt mit einem wichtigen und elementaren Lemma, das zeigt, dass die Variationsvektoren g^+, g^- (siehe Definition 2.54) stationärer Vektoren g ebenfalls stationär sind.

LEMMA 4.11. *Sei $g \in l_{SM}$. Dann*

$$g \in l_\mu \iff g^+, g^- \in l_\mu.$$

BEWEIS. Die Rückrichtung ist trivial. Für die andere Richtung betrachten wir zunächst die Gleichung

$$g = \mu g = \mu(g^+ - g^-) = \mu g^+ - \mu g^-.$$

und weiter die Rechnung

$$\begin{aligned} \|g\|_{TV} &= \|\mu g\|_{TV} = \|\mu(g^+ - g^-)\|_{TV} \\ &\leq \|\mu g^+\|_{TV} + \|\mu g^-\|_{TV} \\ &\stackrel{\text{Satz 3.6}}{\leq} \|g^+\|_{TV} + \|g^-\|_{TV} = \|g\|_{TV}, \end{aligned}$$

woraus $\|g\|_{TV} = \|\mu g^+\|_{TV} + \|\mu g^-\|_{TV}$ und damit wegen der Eindeutigkeit der Variationsvektoren, dass $\mu g^+ = g^+$, $\mu g^- = g^-$. \square

Nach diesen einführenden Illustrationen machen wir uns richtig an die Arbeit. Zunächst studieren wir die Auswirkungen von Stationarität auf Ereignisvektoren.

4.2.2. Stationarität und Ereignisse. Dieser Abschnitt enthält im Wesentlichen zwei kleine, aber um so wichtigere Erkenntnisse. Die Ereignisvektoren stationärer signierter Maßvektoren von Ereignissen, die invariant bezüglich des Schiebeoperators sind (siehe nachstehende Definition), sind wiederum stationäre Vektoren. Diese stationären Ereignisvektoren liegen nun im Abschluss des Prognoseraums. Für endlich-dimensionale stationäre Vektoren folgt, dass ein "invarianter" Ereignisvektor im Prognoseraum selbst liegt, was auf den ersten Blick nicht unbedingt einleuchtend ist, da Ereignisvektoren im Prognoseraum im Allgemeinen wenig verloren haben. Um die Ergebnisse verständlich zu machen benötigen wir zuerst noch eine weitere Definition aus dem Reich der dynamischen Systeme.

DEFINITION 4.12. Sei (Ω, \mathcal{B}) ein Messraum und $T : \Omega \rightarrow \Omega$ eine messbare Abbildung. Ein Ereignis $B \in \mathcal{B}$ wird **invariant (bezüglich T)** genannt, falls

$$T^{-1}B = B.$$

Das System der invarianten Ereignisse bildet nun wieder eine σ -Algebra, eine wichtige Erkenntnis für zahlreiche Ergebnisse aus der Ergodentheorie:

LEMMA 4.13. Die Menge $\mathcal{I} \subset \mathcal{B}$ aller invarianten Ereignisse ist eine σ -Algebra.

BEWEIS. Siehe [Gra01], Lemma 6.3.3, Seite 135. \square

Mit diesen Begriffen ausgerüstet setzen wir zunächst

$$\mathcal{I}_\Sigma = \{I \in \mathcal{B}_\Sigma \mid T_\Omega^{-1}I = I\},$$

also \mathcal{I}_Σ die Unter- σ -Algebra der bezüglich des Schiebeoperators invarianten Ereignisse. Diese wird später eine wichtige Rolle spielen, wie sich in diesem Unterabschnitt bereits andeuten wird.

LEMMA 4.14. Sei $g \in l_\mu$. Dann

$$g = 0 \iff \forall I \in \mathcal{I}_\Sigma : g(I) = 0.$$

BEWEIS. Die Hinrichtung ist trivial. Für die andere Richtung nehmen wir die Existenz eines stationären Vektors $0 \neq g \in l_{SM}$ an, so dass $g(I) = 0$ für alle $I \in \mathcal{I}_\Sigma$. Wegen Lemma 4.11 sind dann g^+, g^- stationäre Maßvektoren. Ohne Einschränkung sei $g^+ \neq 0$ (der Fall $g^- \neq 0$ wird analog abgehandelt). Sei weiter Ω^+, Ω^- die Hahn-Zerlegung von g (siehe Bemerkung 2.53). Es gilt also

$$g^-(\Omega^+) = 0 \text{ und } g^+(\Omega^+) = \|g^+\|_{TV} > 0.$$

Wir setzen nun

$$I^+ := \bigcap_{n \geq 0} \bigcup_{m \geq n} T_\Omega^{-m} \Omega^+ \subset \bigcup_{n \geq 0} T_\Omega^{-n} \Omega^+.$$

Man kann nun zeigen, dass I^+ invariant ist (siehe dafür [Gra01], S.132/133). Weiter gilt

$$\begin{aligned} g^-(I^+) &\leq g^-\left(\bigcup_{n \geq 0} T_\Omega^{-n} \Omega^+\right) \leq \sum_{n \geq 0} g^-(T_\Omega^{-n} \Omega^+) \\ &= \sum_{n \geq 0} \mu^n g^-(\Omega^+) \stackrel{L.4.11}{=} \sum_{n \geq 0} g^-(\Omega^+) = 0 \end{aligned}$$

sowie

$$g^+(I^+) \stackrel{(*)}{\geq} \limsup_{n \rightarrow \infty} g^+(T_\Omega^{-n} \Omega^+) = \limsup_{n \rightarrow \infty} \mu^n g^+(\Omega^+) \stackrel{L.4.11}{=} g^+(\Omega^+) > 0,$$

wobei hier (*) aus [Bau90], Korollar 15.3, S.93 (beachte, dass hier gerade die Ungleichung vertauscht werden muss, siehe den sich anschließenden Beweis) folgt. Damit

$$g(I^+) = g^+(I^+) - g^-(I^+) = g^+(I^+) > 0,$$

und wir haben eine invariante Menge gefunden, die von g nicht mit Null belegt wird, ein Widerspruch! \square

Damit folgt natürlich, dass die Gleichheit zweier stationärer Vektoren ausschließlich auf den invarianten Mengen überprüft werden muss. Diese Tatsache wird uns später sehr unter die Arme greifen.

Ein weiteres nützliches Ergebnis folgt im Anschluss. Die Ereignisvektoren stationärer Vektoren bezüglich invarianter Ereignisse sind wieder stationär. Darüberhinaus findet man solche Ereignisvektoren unweit der Prognoseräume ihrer Grundvektoren wieder.

LEMMA 4.15. *Sei $g \in l_{SM}$ stationär und $I \in \mathcal{I}_\Sigma$ ein bezüglich des Schiebepoperators invariantes Ereignis, also $T_\Omega^{-1}I = I$. Dann ist auch g^I ein stationärer Vektor, mit anderen Worten*

$$\mu g^I = g^I.$$

Darüberhinaus gilt

$$g^I \in \overline{V_g},$$

g^I liegt also im Abschluss des Prognoseraums.

BEWEIS. Der Fall $|g|(I) = 0$ ist trivial. Für $|g|(I) > 0$ sei $\bar{a} \in \Sigma^*$. Dann gilt

$$\begin{aligned}
\mu g^I(\bar{a}) &= \sum_{a \in \Sigma} g^I(a\bar{a}) = \sum_{a \in \Sigma} \frac{\|g\|_{TV}}{|g|(I)} g(a\bar{a} \cap I) \\
&= \frac{\|g\|_{TV}}{|g|(I)} \sum_{a \in \Sigma} g(a\bar{a} \cap T_\Omega^{-1}I) = \frac{\|g\|_{TV}}{|g|(I)} g(T_\Omega^{-1}\bar{a} \cap T_\Omega^{-1}I) \\
&= \frac{\|g\|_{TV}}{|g|(I)} g(T_\Omega^{-1}(\bar{a} \cap I)) = \frac{\|g\|_{TV}}{|g|(I)} \mu g(\bar{a} \cap I) \\
&\stackrel{\mu g = g}{=} \frac{\|g\|_{TV}}{|g|(I)} g(\bar{a} \cap I) = g^I(\bar{a}),
\end{aligned} \tag{4.6}$$

woraus der erste Teil der Behauptung folgt. Für den zweiten Teil beachten wir, dass wir mit Hilfe von Satz 2.71 eine Folge von Zylindervektoren g^{F_n} finden, die gegen g^I konvergieren, also

$$\lim_{n \rightarrow \infty} \|g^{F_n} - g^I\|_{TV} = 0.$$

Unter Anwendung von Lemma 3.12 folgt dann

$$\begin{aligned}
\|g_{F_n} - g^I\|_{TV} &\stackrel{L.3.12}{=} \|\mu^{F_n} g^{F_n} - \mu^{F_n} g^I\|_{TV} \\
&\leq \|\mu^{F_n}\| \cdot \|g^{F_n} - g^I\|_{TV} \stackrel{S.3.6}{\leq} \|g^{F_n} - g^I\|_{TV},
\end{aligned}$$

womit wir eine Folge von Vektoren aus dem Prognoseraum gefunden hätten, die gegen g^I konvergiert. \square

Oft konzentrieren wir uns auf endlich-dimensionale Vektoren. Da hier der Abschluss des Prognoseraums mit dem Prognoseraum selbst übereinstimmt, erhalten wir

KOROLLAR 4.16. *Sei $g \in l_{SM}$ endlich-dimensional und stationär und sei $I \in \mathcal{B}_\Sigma$ ein invariantes Ereignis, also $T_\Omega^{-1}I = I$. Dann gilt*

$$g^I \in \mathcal{V}_g. \tag{4.7}$$

BEWEIS. Folgt aus der Tatsache, dass endlich-dimensionale Unterräume eines topologischen Vektorraums abgeschlossen sind und eben Lemma 4.15. \square

Wir wollen abschließend bemerken, dass man für durch beliebige Ereignisse bedingte Vektoren, die von endlich-dimensionalen Vektoren herrühren, immerhin beweisen kann, dass gewisse "stationäre Grenzvektoren" derselben im Prognoseraum liegen. Siehe dazu den Abschnitt über asymptotische Stationarität für eine genaue Definition dieser Begriffe.

4.2.3. Stationarität und Kegel. Der für Maßvektoren bereits im letzten Kapitel vorgeführte (siehe Satz 3.29), von den Prognoseoperatoren invariant gelassene "Prognosekegel" hat für stationäre Vektoren g die Eigenschaft, g selbst in seinem Innern zu enthalten:

SATZ 4.17. Sei $g \in l_{SM}, g \geq 0$ ein endlich-dimensionaler stationärer Maßvektor. Sei K_g der Abschluss von

$$C_g = \text{cone}(\tau_{\bar{a}}g \mid \bar{a} \in \Sigma^*),$$

also der in Satz 3.29 entwickelte echte Kegel, der von allen Prognoseoperatoren $\tau_{\bar{a}}$ invariant gelassen wird. Dann gilt

$$g \in \text{Int}(K_g),$$

wobei für $A \subset \mathcal{V}_g$ mit $\text{Int}(A)$ das Innere bezüglich der auf \mathcal{V}_g durch $\|\cdot\|_{TV}$ induzierten Teilraumtopologie gemeint sei.

BEWEIS. Wir wählen zunächst (das ist erlaubt, da Echtheit Solidität impliziert) einen Vektor

$$h := \sum_{j=1}^m \alpha_j \tau_{\bar{a}_j} g \in \text{Int}(C_g) = \text{Int}(K_g),$$

und stellen ihn auf dem Erzeugendensystem mit positiven α_j für eine Auswahl von Wörtern \bar{a}_j dar. Natürlich ist auch

$$\bar{h} = \sum_{j=1}^m \alpha_j \tau_{\bar{A}_j} g \in C_g,$$

wobei $\bar{A}_j := \{\bar{b} \in \Sigma^{|\bar{a}_j|} \mid \bar{b} \neq \bar{a}_j\}$ das jeweilige Komplement des Wortes \bar{a}_j bezeichne. Es gilt also

$$\tau_{\bar{A}_j} g = \sum_{\bar{b} \in \Sigma^{|\bar{a}_j|} \setminus \{\bar{a}_j\}} \tau_{\bar{b}} g.$$

Weiter folgt

$$\begin{aligned} \frac{1}{2}(h + \bar{h}) &= \frac{1}{2} \sum_{j=1}^m \alpha_j \tau_{\bar{a}_j} g + \frac{1}{2} \sum_{j=1}^m \alpha_j \tau_{\bar{A}_j} g \\ &= \frac{1}{2} \sum_{j=1}^m \alpha_j (\tau_{\bar{a}_j} g + \tau_{\bar{A}_j} g) = \frac{1}{2} \sum_{j=1}^m \alpha_j \sum_{\bar{b} \in \Sigma^{|\bar{a}_j|}} \tau_{\bar{b}} g \quad (4.8) \\ &= \frac{1}{2} \sum_{j=1}^m \alpha_j \mu^{|\bar{a}_j|} g \stackrel{\mu g = g}{=} \frac{1}{2} \sum_{j=1}^m \alpha_j g = \left(\frac{1}{2} \sum_{j=1}^m \alpha_j\right) g. \end{aligned}$$

Die Behauptung folgt nun durch Anwendung des folgenden Lemmas, indem man dort $u = h$ und $v = \bar{h}$ setzt, womit man

$$\left(\frac{1}{2} \sum_{j=1}^m \alpha_j\right) g \in \text{Int}(K_g)$$

erhält und damit offensichtlich auch $g \in \text{Int}(K_g)$. \square

LEMMA 4.18. Sei $(V, \|\cdot\|)$ ein endlich-dimensionaler, normierter \mathbb{K} -Vektorraum (in diesem Sinne für $v_1 \in V: U_\epsilon(v_1) := \{v \in V \mid \|v - v_1\| < \epsilon\}$) und G eine konvexe Menge aus V . Seien $u, v \in V$ mit $u \in \text{Int}(G), v \in G$. Dann gilt

$$\forall \alpha \in]0, 1[: \quad \alpha u + (1 - \alpha)v \in \text{Int}(G).$$

BEWEIS. Wir finden zunächst ein $\epsilon \in \mathbb{R}^+$, so dass

$$U_\epsilon(u) \subset G.$$

Sei nun also $\alpha \in]0, 1]$ und $w := \alpha u + (1 - \alpha)v$. Wir behaupten

$$U_{\alpha\epsilon}(w) \subset G,$$

woraus die Behauptung folgt. Sei $z \in U_{\alpha\epsilon}(w)$. Dann gilt $u + \frac{1}{\alpha}(z - w) \in U_\epsilon(u)$:

$$\|u + \frac{1}{\alpha}(z - w) - u\| = \|\frac{1}{\alpha}(z - w)\| = \frac{1}{\alpha}\|z - w\| < \frac{1}{\alpha}\alpha\epsilon = \epsilon.$$

Weiter gilt

$$\alpha(u + \frac{1}{\alpha}(z - w)) + (1 - \alpha)v = \alpha u + (z - w) + (1 - \alpha)v = w + z - w = z,$$

womit die Behauptung wegen der Konvexität von G folgt. \square

Über die Eigenschaft, im Innern ihres Prognosekegels zu liegen hinaus, haben es endlich-dimensionale, stationäre Maßvektoren an sich, die anderen Maßvektoren im Prognosekegel in gewissem Sinne zu dominieren. Belegen die stationären Maßvektoren Ereignisse mit Null, so auch alle anderen Prognosevektoren.

SATZ 4.19. *Sei g ein stationärer Maßvektor und $g^* \in \mathcal{V}_g$ ein Maßvektor aus dem Prognoseraum. Dann $m_{g^*} \ll m_g$, d.h. per definitionem*

$$\forall B \in \mathcal{B}_\Sigma : m_g(B) = 0 \implies m_{g^*}(B) = 0.$$

(Eine solche Beziehung zwischen zwei Maßen hat einen Namen: siehe Definition B.6 im Anhang im Anhang).

BEWEIS. Sei $B \in \mathcal{B}_\Sigma$ und $g^* \in \mathcal{V}_g$ ein Maßvektor mit $m_{g^*}(B) > 0$. Nach Definition von \mathcal{V}_g finden wir unter Beachtung von Lemma 2.49 (g^* ist eine Linearkombination von Prognosevektoren) ein Wort $\bar{a} \in \Sigma^*$, so dass $m_{\tau_{\bar{a}}g}(B) > 0$. Wir identifizieren der einfacheren Schreibweise g mit m_g und erhalten (beachte wiederum Satz 2.49)

$$\begin{aligned} g(B) &= \mu^{|\bar{a}|}g(B) = \sum_{\bar{c} \in \Sigma^{|\bar{a}|}} \tau_{\bar{c}}g(B) \\ &\geq \tau_{\bar{a}}g(B) > 0, \end{aligned} \tag{4.9}$$

also die Behauptung. \square

4.2.4. Stationarität und Dimension. Die folgenden Ergebnisse könnten auch als Nachtrag zum Abschnitt über charakteristische Ereignisse aus dem letzten Kapitel betrachtet werden. Wie bereits erwähnt, dienten die dortigen Bemühungen der Ergründung des Zusammenhangs von Stationarität und Dimension. Der folgende Satz lieferte den Hinweis, dass eine Existenz charakteristischer Systeme für unendlich-dimensionale Vektoren die Existenz stationärer, unendlich-dimensionaler Vektoren ausschließt. Da die Existenz

charakteristischer Ereignisse für endlich-dimensionale Vektoren leicht bewiesen werden kann, bestand die Vermutung, dass es tatsächlich nur endlich-dimensionale stationäre Vektoren geben kann. Darüberhinaus nährte die (naive) Anschauung, dass stationäre, also in gewissem Sinne “evolutionsresistente” Vektoren sich auch unter Anwendung der Prognoseoperatoren nur in begrenztem Ausmaß verändern würden können, die Vermutung. Das Gegenteil ist jedoch der Fall, wie hier mit einem Beispiel belegt wird. Zunächst jedoch der angesprochene

SATZ 4.20. *Sei g ein stationärer Maßvektor mit einer charakteristischen Basis $(g_i)_{i \in I}$. Dann gilt*

$$\dim g < \infty.$$

BEWEIS. Wir nehmen an, es gäbe einen unendlich-dimensionalen stationären Maßvektor mit einer charakteristischen Basis $(g_i)_{i \in I}$, so dass $\text{card}(I) = \dim g$. Wegen $g_i \in \mathcal{V}_g$ finden wir Wörter $\bar{b}_j, j = 1, \dots, m$ und Koeffizienten α_{ij} , so dass

$$g_i := \sum_{j=1}^m \alpha_{ij} g_{\bar{b}_j}.$$

Wegen $g_i(A_i) = 1$ finden wir somit ein Wort $\bar{b} \in \Sigma^*$, so dass $\tau_{\bar{b}} g(A_i) > 0$. Mit Satz 4.19 folgt daher $g(A_i) > 0$. Andererseits finden wir eine endliche Teilmenge $J \subset I$ und Koeffizienten $\alpha_j, j \in J$, so dass

$$g = \sum_{j \in J} \alpha_j g_j.$$

Für $i \notin J$ folgt $g(A_i) = \sum_{j \in J} \alpha_j g_j(A_i) = 0$, ein Widerspruch! \square

So weit, so gut. Hier jedoch das (sich nicht eben aufdrängende) Beispiel eines unendlich-dimensionalen, stationären stochastischen Vektors.

SATZ 4.21. *Es gibt einen stationären, stochastischen Vektor g , so dass*

$$\dim g = \infty.$$

BEWEIS. Unser Alphabet sei $\Sigma = \{a, b\}$. Wir setzen zunächst für $\bar{c} := a_1 \dots a_t \in \Sigma^t$:

$$I_a(\bar{c}) := \#\{i \mid a_i = a\} \in \{0, \dots, t\}$$

$$I_b(\bar{c}) := \#\{i \mid a_i = b\} \in \{0, \dots, t\}$$

Es gilt offensichtlich $I_a(\bar{c}) + I_b(\bar{c}) = |\bar{c}| = t$. Damit definieren wir einen Vektor $g : \Sigma^* \rightarrow \mathbb{R}$ durch die Vorschrift

$$g(\bar{c}) := \sum_{i=0}^{I_b(\bar{c})} (-1)^i \binom{I_b(\bar{c})}{i} \frac{1}{I_a(\bar{c}) + 1 + i} \quad (4.10)$$

und haben zu zeigen, dass

- (a) g ist ein stochastischer Vektor, d.h. $\sigma g = g$, $g(\square) = 1$ und $g(\bar{c}) \geq 0$,
- (b) g ist stationär, d.h. $\mu g = g$, sowie
- (c) $\dim g = \infty$.

“ad (a)”: Es gilt

$$g(\square) = (-1)^0 \binom{0}{0} \frac{1}{0+0+1} = 1.$$

Die Gleichung $\sigma g = g$ sieht man durch die folgende Rechnung ein:

$$\begin{aligned} g(\bar{c}a) + g(\bar{c}b) &= \\ &= \sum_{i=0}^{I_b(\bar{c})} (-1)^i \binom{I_b(\bar{c})}{i} \frac{1}{I_a(\bar{c}) + 2 + i} + \sum_{i=0}^{I_b(\bar{c})+1} (-1)^i \binom{I_b(\bar{c})+1}{i} \frac{1}{I_a(\bar{c}) + 1 + i} \\ &= \sum_{i=1}^{I_b(\bar{c})+1} (-1)^{i-1} \binom{I_b(\bar{c})}{i-1} \frac{1}{I_a(\bar{c}) + 1 + i} \\ &\quad + \sum_{i=0}^{I_b(\bar{c})+1} (-1)^i \binom{I_b(\bar{c})+1}{i} \frac{1}{I_a(\bar{c}) + 1 + i} \\ &= \sum_{i=1}^{I_b(\bar{c})+1} (-1)^{i-1} \left(\binom{I_b(\bar{c})}{i-1} - \binom{I_b(\bar{c})+1}{i} \right) \frac{1}{I_a(\bar{c}) + 1 + i} \\ &\quad + (-1)^0 \binom{I_b(\bar{c})+1}{0} \frac{1}{n+1+0} \\ &= \sum_{i=1}^{I_b(\bar{c})+1} (-1)^{i-1} \left(- \binom{I_b(\bar{c})}{i} \right) \frac{1}{I_a(\bar{c}) + 1 + i} + (-1)^0 \binom{I_b(\bar{c})}{0} \frac{1}{n+1+0} \\ &= \sum_{i=0}^{I_b(\bar{c})} (-1)^i \binom{I_b(\bar{c})}{i} \frac{1}{I_a(\bar{c}) + 1 + i} \\ &= g(\bar{c}). \end{aligned}$$

Die Tatsache $\forall \bar{c} \in \Sigma^* : 1 \geq g(\bar{c}) \geq 0$ zeigen wir mit folgendem

LEMMA 4.22. *Seien $m, n \in \mathbb{N}$, $m \geq 0, n \geq 1$. Dann gilt*

$$\sum_{i=0}^m (-1)^i \binom{m}{i} \frac{1}{n+i} = \frac{(n-1)!m!}{(m+n)!}. \quad (4.11)$$

Beweis des Lemmas. Mit vollständiger Induktion über m . Wir setzen $\binom{l}{j} = 0$ für $j > l$.

$m = 0$:

$$(-1)^0 \binom{0}{0} \frac{1}{n} = \frac{(n-1)!}{n!}.$$

$m \rightarrow m + 1$:

$$\begin{aligned}
& \sum_{i=0}^{m+1} (-1)^i \binom{m+1}{i} \frac{1}{n+i} \\
&= \sum_{i=0}^m (-1)^i \binom{m}{i} \frac{1}{n+i} - \sum_{i=0}^{m+1} (-1)^i \binom{m}{i} \frac{1}{n+1+i} \\
&\stackrel{IV}{=} \frac{(n-1)!m!}{(m+n)!} - \frac{n!m!}{(m+n+1)!} \\
&= \frac{(n-1)!m!(m+n+1) - n!m!}{(m+n+1)!} \\
&= \frac{(n-1)!m!(m+n+1-n)}{(m+n+1)!} \\
&= \frac{(n-1)!(m+1)!}{(m+n+1)!},
\end{aligned}$$

wobei wir für die zweite Gleichung die Beziehung $\binom{m+1}{i} = \binom{m}{i} + \binom{m}{i-1}$ ausgenützt und eine Indexverschiebung vorgenommen haben. \square

Weiter im **Beweis des Satzes**. Wir erhalten somit

$$\begin{aligned}
g(\bar{c}) &= \sum_{i=0}^{I_b(\bar{c})} (-1)^i \binom{I_b(\bar{c})}{i} \frac{1}{I_a(\bar{c}) + 1 + i} \\
&= \frac{I_a(\bar{c})! I_b(\bar{c})!}{(I_b(\bar{c}) + I_a(\bar{c}) + 1)!} \in [0, 1]
\end{aligned} \tag{4.12}$$

und damit das Behauptete.

“ad (b)”: Das sieht man einfach wegen

$$g(x\bar{c}) = g(\bar{c}x) \tag{4.13}$$

für $x \in \{a, b\}$. Hieraus folgt wegen $\sigma(g) = g$

$$g(a\bar{c}) + g(b\bar{c}) = g(\bar{c}a) + g(\bar{c}b) = g(\bar{c}) \tag{4.14}$$

und damit $\mu g = g$.

“ad (c)”: Um die Unendlichkeit der Dimension des Vektors nachzuweisen, betrachten wir die folgenden quadratischen Matrizen:

$$A_p := \begin{pmatrix} g(\square) & \tau_a g(\square) & \tau_{aa} g(\square) & \dots & \tau_{a^{p-1}} g(\square) \\ g(a) & \tau_a g(a) & \tau_{aa} g(a) & \dots & \tau_{a^{p-1}} g(a) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ g(a^{p-1}) & \tau_a g(a^{p-1}) & \tau_{aa} g(a^{p-1}) & \dots & \tau_{a^{p-1}} g(a^{p-1}) \end{pmatrix} \in \mathbb{R}^{p \times p}$$

für beliebige Primzahlen p . Wegen $\tau_{a^t}g(a^s) = \frac{1}{t+s+1}$ nach Definition gilt also

$$A_p = \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{3} & \cdots & \frac{1}{p} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \cdots & \frac{1}{p+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{1}{p} & \frac{1}{p+1} & \frac{1}{p+2} & \cdots & \frac{1}{2p-1} \end{pmatrix} \in \mathbb{R}^{p \times p}.$$

Wir zeigen, dass $\det(A_p) \neq 0$, woraus die lineare Unabhängigkeit je p aufeinanderfolgender Vektoren $\tau_{a^{t-1}}g, t = 1, \dots, p$ folgt und weiter wegen der Existenz unendlich vieler Primzahlen die Behauptung. Wir berechnen die Determinante von A_p nach der Formel

$$\det(A_p) = \sum_{\rho \in S_p} \operatorname{sgn}(\rho) \prod_{i=1}^p a_{i\rho(i)} = \sum_{\rho \in S_p} \operatorname{sgn}(\rho) \prod_{i=1}^p \frac{1}{i + \rho(i) - 1},$$

wobei hier mit S_p die Menge der Bijektionen auf $\{1, \dots, p\}$ gemeint sei, und $\operatorname{sgn}(\rho) \in \{-1, 1\}$ das Signum einer Permutation $\rho \in S_p$ bezeichne. Wir betrachten die Permutation

$$\rho_0 : \begin{array}{ccc} \{1, \dots, p\} & \longrightarrow & \{1, \dots, p\} \\ i & \mapsto & p - i + 1, \end{array}$$

die die Indizes $\{1, \dots, p\}$ in umgekehrter Reihenfolge anordnet, und stellen fest, dass für das entsprechende Produkt der Elemente der Gegendiagonale von A_p gerade

$$\prod_{i=1}^p \frac{1}{i + \rho_0(i) - 1} = \frac{1}{p^p}$$

gilt. Damit

$$\begin{aligned} \det(A_p) &= \sum_{\rho \in S_p} \operatorname{sgn}(\rho) \prod_{i=1}^p \frac{1}{i + \rho(i) - 1} \\ &= \sum_{\rho \in S_p, \rho \neq \rho_0} \operatorname{sgn}(\rho) \prod_{i=1}^p \frac{1}{i + \rho(i) - 1} + \operatorname{sgn}(\rho_0) \frac{1}{p^p} \\ &=: \frac{m}{n} \pm \frac{1}{p^p}, \end{aligned}$$

wobei wir den ersten Summanden als vollständig gekürzten Bruch $\frac{m}{n}$ geschrieben haben. Wir stellen dann fest, dass p^p nicht n teilt, da keiner der Nenner der von der Gegendiagonale verschiedenen Elemente von A_p von p geteilt wird. Deshalb sieht man sofort, dass $\frac{m}{n} \neq \pm \frac{1}{p^p}$ und damit die Behauptung. \square

In der Sprache der Stochastiker, es gibt stationäre, stochastische Prozesse von unendlicher Dimension.

KOROLLAR 4.23. *Es gibt einen stochastischen Prozess $(X_t)_{t \in \mathbb{N}}$, für den gilt*

$$\dim(X_t) = \infty \quad \text{und} \quad e - \dim(X_t) = 1.$$

BEWEIS. Klar nach Definition und dem vorangehenden Satz 4.21. \square

BEMERKUNG 4.24. *An dieser Stelle sei gesagt, dass die Existenz eines unendlich-dimensionalen "ergodischen" (für eine genaue Definition des Begriffs siehe Abschnitt 4.4), stationären, stochastischen Vektors eine ungeklärte Frage ist. Fest steht, dass für den soeben im Beweis von Satz 4.21 aufgeführten Vektor für alle $\bar{a} \in \{a, b\}^+$ und $\bar{b} \in \{a, b\}^*$*

$$\begin{aligned}
\mu\tau_{\bar{a}}g(\bar{b}) &= \tau_{\bar{a}}g(a\bar{b}) + \tau_{\bar{a}}g(b\bar{b}) \\
&= g(\bar{a}a\bar{b}) + g(\bar{a}b\bar{b}) \\
&= \sum_{i=0}^{I_b(\bar{a}a\bar{b})} (-1)^i \binom{I_b(\bar{a}a\bar{b})}{i} \frac{1}{I_a(\bar{a}a\bar{b}) + 1 + i} \\
&\quad + \sum_{i=0}^{I_b(\bar{a}b\bar{b})} (-1)^i \binom{I_b(\bar{a}b\bar{b})}{i} \frac{1}{I_a(\bar{a}b\bar{b}) + 1 + i} \\
&= \sum_{i=0}^{I_b(a\bar{a}\bar{b})} (-1)^i \binom{I_b(a\bar{a}\bar{b})}{i} \frac{1}{I_a(a\bar{a}\bar{b}) + 1 + i} \\
&\quad + \sum_{i=0}^{I_b(b\bar{a}\bar{b})} (-1)^i \binom{I_b(b\bar{a}\bar{b})}{i} \frac{1}{I_a(b\bar{a}\bar{b}) + 1 + i} \\
&= g(a\bar{a}\bar{b}) + g(b\bar{a}\bar{b}) = \mu g(\bar{a}\bar{b}) \\
&= g(\bar{a}\bar{b}) = \tau_{\bar{a}}g(\bar{b})
\end{aligned}$$

gilt, mit anderen Worten, dass $\mathcal{V}_g \subset l_\mu$. Dieser Prozess ist demnach in gewissem Sinne "total unergodisch", wenn man sich an dieser Stelle provisorisch vorstellt, dass die Ergodizität eines Vektors durch die Eigenschaft charakterisiert ist, dass im Prognoseraum nur ein Eigenvektor von μ zum Eigenwert 1 zu finden ist.

4.2.5. Fortsetzung zu stationären Vektoren. Dieser Abschnitt enthält einige elementare Erkenntnisse über die Fortsetzung von Vektoren zu stationären Vektoren. Diese haben bei der Konstruktion des unendlich-dimensionalen stationären Vektors im vorangehenden Abschnitt geholfen und können unter Umständen bei der Konstruktion weiterer Beispiele dienen bzw. deren Existenz widerlegen helfen. In diesem Abschnitt immer

$$\Sigma = \{a, b\}$$

und weiter (zur Erinnerung)

$$\Sigma_a^n := \{\bar{a} = a_1 \dots a_n \mid a_1 = a\}$$

(analog sei Σ_b^n definiert). Die hier erzielten Ergebnisse lassen sich sicherlich leicht auf endliche Alphabete beliebiger Größe verallgemeinern.

Interessant mag an diesem Kapitel erscheinen, dass sich mit Hilfe seiner Ergebnisse Schätzungen zur Anzahl der frei wählbaren Parameter endlich-dimensionaler, stationärer Prozesse anstellen lassen könnten. Wie die nachfolgende Proposition zeigt, reicht es bereits, nur die Hälfte eines stationären Prozesses zu kennen, um ihn genau bestimmen zu können.

PROPOSITION 4.25. *Sei $g : \Sigma_a^* \rightarrow \mathbb{R}$ ein Vektor, für den $g(\bar{a}a) + g(\bar{a}b) = g(\bar{a})$ für alle $\bar{a} \in \Sigma_a^*$ gilt. Sei $C \in \mathbb{R}$. Dann lässt sich g eindeutig zu einem stationären Vektor*

$$\hat{g} : \Sigma^* \rightarrow \mathbb{R}$$

fortsetzen, für den $\sigma\hat{g} = \hat{g}$ und $\hat{g}(\square) = C$ gilt.

BEWEIS. Wir konstruieren g rekursiv über die Formeln

$$\hat{g}(\square) := C \tag{4.15}$$

$$\hat{g}(b\bar{a}) := \hat{g}(\bar{a}) - g(a\bar{a}). \tag{4.16}$$

Dass der so erhaltene Vektor stationär ist und g fortsetzt, ist klar. Die Eindeutigkeit und $\sigma(\hat{g}) = \hat{g}$ zeigen wir nun mit Induktion über Σ^t .

$t = 1$: Die Eindeutigkeit ist klar wegen $\hat{g}(b) = C - g(a)$. Wegen

$$\begin{aligned} \sigma\hat{g}(b) &= \hat{g}(ba) + \hat{g}(bb) \\ &= \hat{g}(a) - \hat{g}(aa) + \hat{g}(b) - \hat{g}(ab) \\ &= g(a) - g(aa) + (C - g(a)) - g(ab) \\ &\stackrel{\sigma g = g}{=} C - g(a) = C - \hat{g}(a) = \hat{g}(b) \end{aligned} \tag{4.17}$$

gilt $\sigma g|_{\Sigma} = g|_{\Sigma}$.

$t \rightarrow t + 1$: Nehmen wir an, \hat{g} sei auf Σ^{t+1} nicht eindeutig, so folgt aus der Eindeutigkeit über Σ^t , dass ein $\bar{a} \in \Sigma^t$ existiert, so dass $\hat{g}(b\bar{a}) \neq \hat{g}(\bar{a}) - \hat{g}(a\bar{a})$, was ein Widerspruch zur Stationarität von \hat{g} ist. Sei nun $\bar{a} \in \Sigma^t$. Dann folgt

$$\begin{aligned} \sigma\hat{g}(b\bar{a}) &= \hat{g}(b\bar{a}a) + \hat{g}(b\bar{a}b) \\ &= \hat{g}(\bar{a}a) - \hat{g}(a\bar{a}a) + \hat{g}(\bar{a}b) - \hat{g}(a\bar{a}b) \\ &\stackrel{IV, \sigma g = g}{=} \hat{g}(\bar{a}) - \hat{g}(a\bar{a}) \\ &= \hat{g}(b\bar{a}), \end{aligned} \tag{4.18}$$

was zu zeigen war. □

KOROLLAR 4.26. *Ein stationärer Maßvektor g über dem Alphabet $\Sigma = \{a, b\}$ ist durch die Angabe von $\tau_a g$ eindeutig bestimmt.*

KOROLLAR 4.27. *In einem stationären Prozess gilt*

$$g(b^t \bar{a}) = g(\bar{a}) - \sum_{i=0}^{t-1} g(ab^i \bar{a}).$$

BEWEIS. Vollständige Induktion zusammen mit der Konstruktionsregel aus Proposition 4.25. \square

Wir enträtseln an dieser Stelle das Geheimnis der Existenz eines gemeinsamen Fixpunkts des stochastischen Operators und des Evolutionsoperators, der nicht in l_{TV} enthalten:

BEISPIEL 4.28. Wir betrachten erneut den in Beispiel 2.57 definierten Vektor $g : \Sigma^* \rightarrow \mathbb{R}$ und seine Einschränkung $g|_{\Sigma_a^*}$. Natürlich erfüllt $g|_{\Sigma_a^*}$ die Anforderungen von Proposition 4.25 und lässt sich damit zu einem stationären Vektor \tilde{g} fortsetzen, so dass $\tilde{g}|_{\Sigma_a^*} = g|_{\Sigma_a^*}$. Weiterhin gilt

$$\tilde{g}(a^t) = \tilde{g}|_{\Sigma_a^*}(a^t) = g|_{\Sigma_a^*}(a^t) = g(a^t) = t,$$

woraus $\tilde{g} \notin l_{TV}$ folgt. Damit folgt

$$l_\mu \subsetneq l_\sigma \cap \{g : \Sigma^* \rightarrow \mathbb{R} \mid \mu g = g\}.$$

Natürlich lässt sich nicht jeder “halbe” Vektor zu einem stationären Maßvektor fortsetzen. Dazu bemerken wir, dass der in Proposition 4.25 resultierende Vektor \hat{g} nicht unbedingt ein Maßvektor sein muss, da nicht garantiert ist, dass $\hat{g} \geq 0$. Welchen Ansprüchen ein “halber” Vektor gerecht werden muss, um ein echter Maßvektor werden zu können, das sagt die folgende

PROPOSITION 4.29. Ein Vektor $g : \Sigma_a^* \rightarrow \mathbb{R}$ lässt sich genau dann zu einem stationären, stochastischen Vektor $\hat{g} : \Sigma^* \rightarrow \mathbb{R}$ fortsetzen, falls die folgenden Bedingungen erfüllt sind:

$$(i) \quad \forall \bar{a} \in \Sigma_a^* : g(\bar{a}a) + g(\bar{a}b) = g(\bar{a}), \quad (4.19)$$

$$(ii) \quad \forall \bar{a} \in \Sigma_a^* : 0 \leq g(\bar{a}) \leq 1, \quad (4.20)$$

$$(iii) \quad \forall \bar{a} \in \Sigma_a^* : \sum_{t \geq 0} g(ab^t \bar{a}) \leq g(\bar{a}). \quad (4.21)$$

BEWEIS. “ \Leftarrow ”: Es gelten (i), (ii) und (iii). Gemäß Proposition 4.25 lässt sich g auf eindeutige Art und Weise zu einem stationären Vektor mit $g(\square) = 1$ fortsetzen, den wir ebenfalls mit g bezeichnen und für den $\sigma g = g$. Es bleibt zu zeigen, dass dieser ein stochastischer Vektor ist.

Sei dazu $\bar{a} \in \Sigma_b^*$. Dann ist $0 \leq g(\bar{a}) \leq 1$ zu zeigen. Schreibt man $\bar{a} = b^t \bar{c}$ mit $\bar{c} \in \Sigma_a^*$, so folgt wegen Korollar 4.27 $g(\bar{a}) = g(\bar{c}) - \sum_{i=0}^{t-1} g(ab^i \bar{c})$, woraus mit (iii) die Behauptung folgt.

“ \Rightarrow ”: Dass (i) und (ii) notwendig sind, sieht man unmittelbar. (iii) folgt wiederum aus Korollar 4.27. \square

PROPOSITION 4.30. Sei g ein stationärer, stochastischer Vektor. Dann gilt:

$$\forall \bar{a} \in \Sigma^+ \setminus \{b^t \mid t \in \mathbb{N}\} : \sum_{t \geq 0} g(ab^t \bar{a}) = g(\bar{a}).$$

BEWEIS. Mit Korollar 4.27 ist

$$\lim_{t \rightarrow \infty} g(b^t \bar{a}) = 0$$

zu zeigen. Dies folgt aus

$$g(b^t \bar{a}) \leq g(b^t) - g(b^{t+|\bar{a}|})$$

und daraus, dass $(g(b^t))_{t \in \mathbb{N}}$ eine konvergente Folge ist. \square

4.3. Evolutionssummierbarkeit

In diesem Abschnitt führen wir den Begriff der Evolutionssummierbarkeit signierter Maßvektoren ein. Evolutionssummierbarkeit eines signierten Maßvektors bedeutet gerade, dass die Cesaro-Mittel unter der Anwendung des Evolutionsoperators konvergieren. Das bedeutet anschaulich z.B., dass Häufigkeiten von Wörtern eines sich im Laufe der Zeit entwickelnden evolutionssummierbaren, stochastischen Vektors konvergieren, eine für einen stochastischen Prozess günstige Eigenschaft, wie unmittelbar einsichtig ist.

Die Motivation dieses Begriffs liegt wiederum in der Theorie der dynamischen Systeme begraben. Tatsächlich erweisen sich evolutionssummierbare stochastische Vektoren gerade als die Pendanten der im Mittel asymptotisch stationären stochastischen Prozesse. Der Begriff der asymptotischen Stationarität im Mittel ist deshalb zentral, weil diese Eigenschaft äquivalent dazu ist, dass die Messwerte, die entlang von Sequenzen genommen werden, im Mittel konvergieren (das ist die Aussage des berühmten Ergodensatzes von Birkhoff, siehe Abschnitt 4.4, Satz 4.59. In einfachen Worten: Relative Häufigkeiten, die mit Hilfe von Auszählmethoden entlang einer Sequenz, die von einem AMS-Prozess herrührt, erhoben werden, konvergieren. Das ist natürlich wiederum notwendige Voraussetzung, will man auf Eigenschaften des stochastischen Prozesses aus einzelnen von ihm erzeugten Sequenzen rückschließen.

Eine der Hauptideen dieses Abschnitts ist nun, dass endlich-dimensionale Prozesse evolutionssummierbar sind. Der Beweis dazu ist nach den getanen Vorarbeiten ebenso kurz wie einfach. Bislang existierende Beweise, dass zum Beispiel Hidden Markov Prozesse asymptotisch stationär im Mittel sind, sind ungleich viel komplizierter, siehe dazu [EM02], [KR81].

4.3.1. Einführung. Wir starten mit den Definitionen, die Gegenstand der sich anschließenden Diskussionen sein werden.

DEFINITION 4.31. *Wir nennen einen Vektor $g \in l_{TV}$ **wort- oder punktweise evolutionssummierbar** oder kurz **PES**, falls*

$$\exists \bar{g} \in l_\infty, \forall \bar{a} \in \Sigma^* : \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} \mu^i g(\bar{a}) = \bar{g}(\bar{a}), \quad (4.22)$$

mit anderen Worten, falls die Folge $(\frac{1}{n} \sum_{i=0}^{n-1} \mu^i g)_{n \in \mathbb{N}}$ "wortweise" gegen einen Vektor \bar{g} konvergiert. Gilt darüberhinaus

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} \mu^i g = \bar{g} \iff \lim_{n \rightarrow \infty} \left\| \frac{1}{n} \sum_{i=0}^{n-1} \mu^i g - \bar{g} \right\|_{TV} = 0, \quad (4.23)$$

(also Konvergenz im Sinne der Norm der totalen Variation, es folgt natürlich $\bar{g} \in l_{TV}$) so nennen wir den Vektor g **evolutionssummierbar** oder kurz **ES**. Wir nennen Vektoren $g \in l_{TV}$ darüberhinaus **wortweise evolutionskonvergent** bzw. **evolutionskonvergent**, falls die Folgen $\mu^n g$ punktweise bzw. in Norm gegen einen Vektor \bar{g} konvergieren.

Natürlich sind (wortweise) evolutionskonvergente Vektoren auch (P)ES-Vektoren. Der Fall evolutionskonvergenter Vektoren entspricht in der Welt der stochastischen Prozesse gerade den asymptotisch stationären Prozessen und ist meist nicht von speziellem Belang. Die meisten Ergebnisse werden sich auf (P)ES-Vektoren beziehen. Wir starten also damit, einfache Eigenschaften von PES-Vektoren zu bemerken.

PROPOSITION 4.32. *Der Grenzvektor \bar{g} eines PES-Vektors ist ein stationärer Vektor, mit anderen Worten, der durch \bar{g} beschriebene stochastische Prozess ist stationär.*

BEWEIS. Die Behauptung folgt aus nachstehender Rechnung:

$$\begin{aligned} \mu \bar{g}(\bar{a}) &= \sum_{a \in \Sigma} \bar{g}(a\bar{a}) = \sum_{a \in \Sigma} \lim_{n \rightarrow \infty} g_n(a\bar{a}) \\ &= \lim_{n \rightarrow \infty} \sum_{a \in \Sigma} g_n(a\bar{a}) = \lim_{n \rightarrow \infty} \mu g_n(\bar{a}) \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} \mu^{i+1} g(\bar{a}) = \lim_{n \rightarrow \infty} g_n(\bar{a}) + \underbrace{\frac{1}{n} (\mu^n g(\bar{a}) - g(\bar{a}))}_{\rightarrow 0} \\ &= \bar{g}(\bar{a}). \end{aligned} \quad (4.24)$$

□

Das berechtigt zu folgender

DEFINITION 4.33. *Wir nennen den Limesvektor \bar{g} eines PES-Vektors g , also $\forall \bar{a} \in \Sigma^* : \bar{g}(\bar{a}) = \lim_{n \rightarrow \infty} 1/n \sum_{i=0}^{n-1} \mu^i g(\bar{a})$ auch den **stationären Grenzvektor** von g .*

Den Zusammenhang von ES-Vektoren und stationären Vektoren kann man noch in einem anderen Licht sehen. Dazu betrachten wir zunächst die folgende

DEFINITION 4.34. *Wir bezeichnen mit $l_{ES} \subset l_{SM}$ den **Vektorraum der ES-Vektoren** sind.*

Wir bemerken dann, dass ES-Vektoren durch eine lineare Abbildung auf den Raum der stationären Vektoren abgebildet werden kann. Diese lineare Abbildung wird sich als Projektion erweisen.

PROPOSITION 4.35. *Definiert man*

$$\mu_n := \frac{1}{n} \sum_{i=0}^{n-1} \mu^i,$$

so gilt natürlich, dass $\lim_{n \rightarrow \infty} \mu_n(g)$ existiert für alle $g \in l_{ES}$. Sei weiter $l_\mu \subset l_{ES}$ der lineare Unterraum der stationären Vektoren. Wir setzen dann

$$\begin{aligned} \bar{\mu} : l_{ES} &\longrightarrow l_\mu \\ g &\longmapsto \lim_{n \rightarrow \infty} \mu_n g. \end{aligned}$$

Das so definierte $\bar{\mu}$ ist eine lineare Abbildung, für die $\|\bar{\mu}\| \leq 1$ gilt. Wegen $\bar{\mu}^2 = \bar{\mu}$ ist diese lineare Abbildung eine Projektion.

BEWEIS. Für den Beweis betrachten wir nachstehende Folgerung aus dem Prinzip der gleichmäßigen Beschränktheit, die sich in [MV92] nachlesen lässt:

LEMMA 4.36. *Seien E ein Banachraum und $(A_n)_{n \in \mathbb{N}}$ eine Folge von stetigen, linearen Abbildungen $A_n : E \rightarrow E$. Es gelte*

- (a) $\sup_{n \in \mathbb{N}} \|A_n\| < \infty$,
- (b) $\forall x \in E : (A_n x)_{n \in \mathbb{N}}$ ist eine Cauchy-Folge.

Dann ist $(A_n x)_{n \in \mathbb{N}}$ konvergent für alle $x \in E$ und durch

$$\begin{aligned} A : E &\longrightarrow E \\ x &\longmapsto \lim_{n \rightarrow \infty} A_n x \end{aligned}$$

wird eine stetige, lineare Abbildung definiert mit $\|A\| \leq \sup_{n \in \mathbb{N}} \|A_n\|$.

BEWEIS. Vereinfachte Aussage von [MV92], Lemma 8.14. \square

Nun zum **Beweis von Proposition 4.35**: Wir setzen für A_n in Lemma 4.36 μ_n ein. Wegen

$$\sup_{n \in \mathbb{N}} \|\mu_n\| = \sup_{n \in \mathbb{N}} \left\| \frac{1}{n} \sum_{i=0}^{n-1} \mu^i \right\| \leq \sup_{n \in \mathbb{N}} \frac{1}{n} \sum_{i=0}^{n-1} \|\mu^i\| = 1$$

folgt damit das Behauptete. Für einen stationären Vektor g gilt dann offensichtlich $\bar{\mu}g = g$, woraus folgt, dass $\bar{\mu}$ eine Projektion ist. \square

Wir bemerken an dieser Stelle eine wichtige Kleinigkeit, die an späterer Stelle (in der Theorie der ergodischen Vektoren) zur Anwendung kommen wird.

BEMERKUNG 4.37. *Ist $I \in \mathcal{B}_\Sigma$ ein bezüglich des Schiebeoperators invariantes Ereignis, d.h. $T_\Omega^{-1}I = I$, so offensichtlich $m_g(I) = m_g(T^{-1}I) = m_{\mu g}(I)$. Deshalb auch*

$$m_g(I) = m_{\bar{\mu}g}(I)$$

für einen ES-Vektor g .

Der Raum der stationären Vektoren erwies sich als abgeschlossen, ebenso ist dies der Raum der ES-Vektoren, eine Erkenntnis, die sich später als hilfreich erweist.

SATZ 4.38. l_{ES} ist abgeschlossen.

BEWEIS. Sei $(g_k)_{k \in \mathbb{N}}$ eine Folge von ES-Vektoren, die gegen einen Vektor g^* konvergieren, mit anderen Worten

$$\lim_{k \rightarrow \infty} \|g_k - g^*\|_{TV} = 0. \quad (4.25)$$

Sei μ_n wie in Proposition 4.35 definiert. Die $\mu_n g_k$ konvergieren gleichmäßig in n gegen $\mu_n g^*$:

$$\|\mu_n g_k - \mu_n g^*\|_{TV} \leq \|\mu_n\|_{TV} \cdot \|g_k - g^*\|_{TV} \stackrel{Prop.4.35}{\leq} \|g_k - g^*\|_{TV}. \quad (4.26)$$

Wir betrachten nun die Folge $(\bar{\mu} g^k)_{k \in \mathbb{N}}$. Wegen

$$\|\bar{\mu} g_{k_1} - \bar{\mu} g_{k_2}\|_{TV} \leq \|\bar{\mu}\| \cdot \|g_{k_1} - g_{k_2}\|_{TV} \stackrel{Prop.4.35}{\leq} \|g_{k_1} - g_{k_2}\|_{TV} \quad (4.27)$$

ist $(\bar{\mu} g_k)$ eine Cauchy-Folge in l_μ , also einem abgeschlossenen Unterraum (siehe Korollar 4.10) eines Banachraums und konvergiert daher gegen einen Vektor $\bar{g}^* \in l_\mu$.

Wir zeigen nun, dass die Folge der $\mu_n g^*$ ebenfalls gegen \bar{g}^* konvergiert, womit wir am Ziel wären. Sei also $\epsilon \in \mathbb{R}^+$. Dann finden wir mit (4.26) ein $K_1 \in \mathbb{N}$, so dass

$$\forall n \in \mathbb{N} \forall k \geq K_1 : \quad \|\mu_n g_k - \mu_n g^*\| < \frac{\epsilon}{3}. \quad (4.28)$$

Wir finden mit Gleichung (4.27) weiter ein $K_2 \in \mathbb{N}$, so dass

$$\forall k \geq K_2 : \quad \|\bar{\mu} g_k - \bar{g}^*\|_{TV} < \frac{\epsilon}{3}. \quad (4.29)$$

Zuletzt finden wir für $K_0 := \max\{K_1, K_2\}$, da g_{K_0} ein ES-Vektor ist, ein $N_0 \in \mathbb{N}$, so dass

$$\forall n \geq N_0 : \quad \|\mu_n g_{K_0} - \bar{\mu} g_{K_0}\|_{TV} < \frac{\epsilon}{3}. \quad (4.30)$$

Insgesamt erhalten wir für alle $n \geq N_0$

$$\begin{aligned} \|\mu_n g^* - \bar{g}^*\|_{TV} &= \|\mu_n g^* - \mu_n g_{K_0} + \mu_n g_{K_0} - \bar{\mu} g_{K_0} + \bar{\mu} g_{K_0} - \bar{g}^*\|_{TV} \\ &\leq \|\mu_n g^* - \mu_n g_{K_0}\|_{TV} + \|\mu_n g_{K_0} - \bar{\mu} g_{K_0}\|_{TV} + \|\bar{\mu} g_{K_0} - \bar{g}^*\|_{TV} \\ &\stackrel{(4.28), (4.29), (4.30)}{<} \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon. \end{aligned} \quad (4.31)$$

□

4.3.2. Evolutionssummierbarkeit und asymptotische Stationarität im Mittel. Wir zeigen nun die Zusammenhänge unserer Definitionen mit den schon angesprochenen Begriffen der Theorie der dynamischen Systeme auf. Dies wird auf einen Vergleich der in Unterabschnitt 2.3.4 angegebenen Konvergenzbegriffe hinauslaufen und daher hier nicht im Detail

aufgerollt werden. Wir beginnen mit der folgenden, bereits erwähnten, fundamentalen Definition aus der Welt der dynamischen Systeme.

DEFINITION 4.39. *Ein dynamisches System $(\Omega, \mathcal{B}, m, T)$ wird **asymptotisch stationär im Mittel** oder **schlicht AMS** genannt, falls*

$$\forall B \in \mathcal{B} : \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} m(T^{-i}B) =: \bar{m}(B) \text{ existiert.}$$

DEFINITION/BEMERKUNG 4.40. *Nach dem Satz von Vitali-Hahn-Saks (siehe [Gra01], Lemma 6.2.2, Seiten 128-131) ist die so definierte Funktion \bar{m} auf der σ -Algebra \mathcal{B} ein Wahrscheinlichkeitsmaß. Man sieht auf einfache Art und Weise ein, dass es sich um ein bezüglich T stationäres Maß handelt. Wir nennen \bar{m} auch **stationäres Grenzmaß**.*

Der nachfolgende Satz zeigt nun, dass sich die Begriffe der Evolutionssummierbarkeit und der asymptotischen Stationarität im Mittel gerade entsprechen. Da die Eigenschaft der asymptotischen Stationarität im Mittel eine wichtige Grundlage für etliche weitere wichtige asymptotische Eigenschaften dynamischer Systeme bzw. stochastischer Prozesse ist, ist die Begriffsbildung der Evolutionssummierbarkeit damit natürlich umfassend gerechtfertigt.

SATZ 4.41. *Ist g Maßvektor, so sind die beiden folgenden Aussagen äquivalent:*

- (a) *g ist ein ES-Vektor.*
- (b) *Das zu g assoziierte dynamische System $(\Omega_\Sigma, \mathcal{B}_\Sigma, m_g, T_\Omega)$ ist AMS.*

Der BEWEIS allerdings benötigt eine Menge mehr Apparatur aus der Maßtheorie als der Text an dieser Stelle verträgt. Er ist daher in Appendix B zu finden. Wir wollen an dieser Stelle nur erwähnen, dass Aussage (a) gerade besagt, dass die involvierte Folge von Maßen schwach Skorokhod gegen das stationäre Grenzmaß konvergiert, wohingegen Aussage (b) starke Konvergenz ausdrückt. Die Implikation (a) \Rightarrow (b) ist daher trivial, siehe Korollar 2.65. \square

Eines steht jedoch fest: Es gibt PES-Vektoren, deren dynamische Systeme nicht AMS (und damit natürlich nicht ES) sind, wie das folgende Beispiel belegen wird. Dabei kann man den PES-Vektor sogar stochastisch wählen.

BEISPIEL 4.42. *Wir betrachten den stochastischen Vektor g des stochastischen Prozesses über dem Alphabet $\Sigma = \{a, b\}$, der mit Wahrscheinlichkeit 1 die Sequenz*

$$a \ b \ a \ bb \ a \ bbb \ a \ bbbb \ a \ bbbbb \ a \ \cdots \ a \ \underbrace{b \cdots b}_{n\text{-mal}} \ a \ \underbrace{b \cdots b}_{(n+1)\text{-mal}} \ \cdots$$

erzeugt. Wir zeigen, dass g PES ist, aber das induzierte Prozessmaß m_g nicht AMS ist. Dazu setzen wir

$$\bar{g} : \Sigma^* \longrightarrow \mathbb{R}$$

$$\bar{b} \longmapsto \begin{cases} 1 & \bar{b} = b^n \text{ für ein } n \\ 0 & \text{sonst} \end{cases}$$

und zeigen, dass \bar{g} der stationäre Grenzvektor des obenstehenden Prozesses ist, was intuitiv völlig klar ist, nichtsdestotrotz technische Schwierigkeiten mit sich bringt. Dazu sei $\bar{b} = b^n$ für ein $n \in \mathbb{N}$. Betrachten wir die Folge $(\mu^t g(\bar{b}))_{t \in \mathbb{N}}$, so ergibt sich folgendes Bild:

$$\underbrace{00 \cdots 00}_{\binom{n(n+3)}{2} - 1 \text{ mal}} \quad 1 \quad \underbrace{0 \cdots 0}_{n \text{ mal}} \quad 11 \quad \underbrace{0 \cdots 0}_{n \text{ mal}} \quad 111 \quad \underbrace{0 \cdots 0}_{n \text{ mal}} \quad 1111 \cdots,$$

woraus sofort klar wird, dass

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=0}^{n-1} \mu^t g(\bar{b}) = 1$$

ist. Daher folgt, dass g PES ist. Wir setzen weiter $B := \Omega_\Sigma \setminus \{bbbb\dots\}$, also B das Element der σ -Algebra, das alle Sequenzen enthält bis auf die, die nur aus b 's besteht. Dann gilt für alle $t \in \mathbb{N}$: $m_{\mu^t g}(B) = 1$, aber $m_{\bar{g}}(B) = 0$. Daher gilt für $g_n := \frac{1}{n} \sum_{t=0}^{n-1} \mu^t g$

$$\lim_{n \rightarrow \infty} m_{g_n}(B) = 1 \neq 0 = m_{\bar{g}}(B).$$

4.3.3. Asymptotische Stationarität und Dimension. Es ist einfach einzusehen, dass der zuletzt aufgeführte Prozess unendlich-dimensional, mehr noch von unendlicher Evolutionsdimension ist. Dass dies für einen Vektor, der nicht evolutionssummierbar ist, notwendigerweise der Fall ist, wird das folgende (zentrale) Ergebnis zeigen. Vektoren von endlicher Evolutionsdimension (und damit natürlich auch endlicher Dimension) sind immer evolutionssummierbar. Das zeigt der folgende Satz, der nach getaner Vorarbeit einfach zu beweisen ist.

SATZ 4.43. Sei $g \in l_{TV}$ mit $\text{edim}(g) < \infty$. Dann ist g ein ES-Vektor.

BEWEIS. Wegen Satz 3.8 ist auch

$$\mu : \mathcal{E}_g \longrightarrow \mathcal{E}_g,$$

also die Einschränkung von μ auf den Evolutionsraum von g stabil. Aus $\dim \mathcal{E}_g = \text{edim}(g) < \infty$ folgt mit Satz 10.14 und der Äquivalenz von Normen, dass die Folge $g_n = \frac{1}{n} \sum_{t=0}^{n-1} \mu^t g$ in Norm gegen \bar{g} konvergiert, mit anderen Worten, g ist ein ES-Vektor. \square

Der Vollständigkeit halber bleibt die Frage zu beantworten, ob es stochastische Prozesse gibt, die sowohl von unendlicher Evolutionsdimension als auch ES sind. Tatsächlich gibt es solche Prozesse, wie das folgende Beispiel demonstriert.

SATZ 4.44. *Es gibt einen stochastischen ES-Vektor g mit*

$$\text{edim } g = \infty.$$

BEWEIS. Sei $\Sigma = \{a, b\}$ und $p \in]0, 1[$. Wir betrachten dann den folgenden rekursiv definierten Vektor

$$g : \Sigma^* \longrightarrow \mathbb{R}$$

$$\bar{a} \longmapsto \begin{cases} 1 & \bar{a} = \square \\ p^{|\bar{b}|+1}g(\bar{b}) & \bar{a} = \bar{b}a \text{ für ein } \bar{b} \in \Sigma^* \\ (1 - p^{|\bar{b}|+1})g(\bar{b}) & \bar{a} = \bar{b}b \text{ für ein } \bar{b} \in \Sigma^* \end{cases}.$$

Es gilt z.B. $g(abab) = p(1 - p^2)p^3(1 - p^4)$. g ist offensichtlich ein stochastischer Vektor. Wir bemerken, dass g einen stochastischen Prozess $(X_t)_{t \in \mathbb{N}}$ mit Werten in $\Sigma = \{a, b\}$ kodiert, für den (für die folgenden Bezeichnungen siehe Definition 2.25 und den Folgetext)

$$P(X_t^{-1}(\{a\})) = p^{t+1}$$

und weiter

$$P(\bar{a} = a_1 \dots a_t) = P(X_0^{-1}(\{a_1\})) \times \dots \times P(X_{t-1}^{-1}(\{a_t\}))$$

gilt, d.h. die durch die X_t induzierten Verteilungen sind unabhängig. Wir behaupten, dass g die Anforderungen des Satzes erfüllt.

Zur Vorbereitung des Beweises behaupten wir, dass ($\bar{a} \in \Sigma^*$)

$$\mu^k g(\bar{a}) = \begin{cases} 1 & \bar{a} = \square \\ p^{|\bar{b}|+k+1}g(\bar{b}) & \bar{a} = \bar{b}a \text{ für ein } \bar{b} \in \Sigma^* \\ (1 - p^{|\bar{b}|+k+1})g(\bar{b}) & \bar{a} = \bar{b}b \text{ für ein } \bar{b} \in \Sigma^* \end{cases}$$

gilt. Wir zeigen dies mit vollständiger Induktion über $k \in \mathbb{N}$, wobei der Anfang ($k = 0$) mit der Definition von g bereits erledigt ist. Für den Induktionsschritt sei $k \geq 0$. Wir sehen zunächst ein, dass $\mu^{k+1}g(\square) = 1$, da $\mu^{k+1}g$ ein stochastischer Vektor ist (siehe dazu Korollar 3.7). Für $\bar{a} \in \Sigma^+ = \Sigma^* \setminus \{\square\}$, so dass $\bar{a} = \bar{b}x$ mit $\bar{b} \in \Sigma^*$ und $x \in \Sigma$ rechnen wir dann

$$\begin{aligned} \mu^{k+1}g(\bar{a}) &= \mu\mu^k g(\bar{b}x) = \mu^k(a\bar{b}x) + \mu^k(b\bar{b}x) \\ &\stackrel{IV}{=} \begin{cases} p^{|\bar{a}\bar{b}|+k+1}(\mu^k g(a\bar{b}) + \mu^k g(b\bar{b})) & x = a \\ (1 - p^{|\bar{a}\bar{b}|+k+1})(\mu^k g(a\bar{b}) + \mu^k g(b\bar{b})) & x = b \end{cases} \\ &= \begin{cases} p^{|\bar{b}|+(k+1)+1}\mu^{k+1}g(\bar{b}) & x = a \\ (1 - p^{|\bar{b}|+(k+1)+1})\mu^{k+1}g(\bar{b}) & x = b \end{cases}, \end{aligned}$$

woraus die Zwischenbehauptung folgt.

Wir zeigen nun, dass $\text{edim } g = \infty$ gilt. Dafür bemerken wir, dass wir mit der obigen Zwischenbehauptung einfach auf die Gleichung ($k, i \in \mathbb{N}$)

$$\mu^k g(a^i) = p^{\sum_{t=1}^i (k+t)}$$

schließen können.

Wir setzen $a_{ik} := \mu^{k-1}g(a^i)$ für $k, i \geq 1$ und betrachten dann die Matrizen

$$A_n := (a_{ik})_{1 \leq i, k \leq n} = \begin{pmatrix} p & p^2 & \dots & p^n \\ p^{1+2} & p^{2+3} & \dots & p^{n+n+1} \\ \vdots & \vdots & \ddots & \vdots \\ p^{1+\dots+n-1} & p^{2+\dots+n} & \dots & p^{n+\dots+2n-1} \end{pmatrix} \in \mathbb{R}^{n \times n}.$$

Es gilt

$$\det(A_n) = \prod_{k=1}^n p^{2n-1} \prod_{1 \leq i, j \leq n, i < j} (p^i - p^j) \neq 0,$$

denn induktiv

$$\det(A_n) = \prod_{k=1}^n p^{2n-1} \det \begin{pmatrix} 1 & p & \dots & p^{n-1} \\ 1 & p^2 & \dots & p^{2(n-1)} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & p^n & \dots & p^{n(n-1)} \end{pmatrix}$$

und die Determinante auf der rechten Seite ergibt die Vandermonde'sche Zahl

$$\Delta(p, p^2, \dots, p^n) = \prod_{i < j} (p^i - p^j),$$

womit die Behauptung bewiesen wäre. Daher folgt, dass die je n stochastischen Vektoren $g, \mu g, \dots, \mu^{n-1}g$ linear unabhängig sind und damit $\text{edim } g = \infty$.

Wir definieren nun \bar{g} durch

$$\bar{g}(\bar{a}) = \begin{cases} 1 & \text{für } \bar{a} = b^{|\bar{a}|} \\ 0 & \text{sonst.} \end{cases} \quad (4.32)$$

und wir beweisen ($g_n := \mu^n g$), dass

$$\lim_{n \rightarrow \infty} \|g_n - \bar{g}\|_{TV} = 0,$$

was mit untenstehendem Lemma 4.46 aufzeigt, dass g ein ES-Vektor ist und damit den Rest der Behauptung. Sei also $t \in \mathbb{N}$, dann gilt

$$\sum_{\bar{a} \in \Sigma^t} |g_n(\bar{a}) - \bar{g}(\bar{a})| = 1 - g_n(b^t) + \sum_{\bar{a} \in \Sigma^t \setminus \{b^t\}} g_n(\bar{a}). \quad (4.33)$$

Wir haben demnach zu zeigen, dass $g_n(b^t)$ gegen 1 konvergiert, und das gleichmäßig in t . Dazu zeigen wir, dass

$$\log \frac{1}{g_n(b^t)} \leq \frac{p^{n+1}}{(1-p)^2}, \quad (4.34)$$

denn daraus folgt natürlich

$$1 \geq g_n(b^t) \geq \left(\exp\left(\frac{p^{n+1}}{(1-p)^2}\right) \right)^{-1} \xrightarrow{n \rightarrow \infty} 1$$

und damit die Behauptung. Zum Beweis von Gleichung 4.34 betrachten wir zunächst folgendes

LEMMA 4.45. Sei $1 < r \in \mathbb{R}$. Dann gilt

$$\log(r) - \log(r-1) \leq \frac{1}{r-1}. \quad (4.35)$$

(Mit \log sei der natürliche Logarithmus gemeint.)

Beweis des Lemmas: Mit dem Mittelwertsatz folgt die Existenz eines $\xi \in [r-1, r]$, so dass

$$\log(r) - \log(r-1) = \frac{\log(r) - \log(r-1)}{r - (r-1)} = (\log)'(\xi) = \frac{1}{\xi}.$$

Wegen $\xi \geq r-1$ folgt $\frac{1}{\xi} \leq \frac{1}{r-1}$ und damit die Behauptung. \square

Weiter im Beweis des Satzes: Nun folgt

$$\begin{aligned} \log \frac{1}{g_n(b^t)} &= \log \left(\prod_{l=1}^t \frac{1}{1-p^{l+n}} \right) = \log \left(\prod_{l=1}^t \frac{(1/p)^{l+n}}{(1/p)^{l+n}-1} \right) \\ &= \sum_{l=1}^t \log((1/p)^{l+n}) - \log((1/p)^{l+n}-1) \stackrel{(4.35)}{\leq} \sum_{l=1}^t \frac{1}{(1/p)^{l+n}-1} \\ &= \sum_{l=1}^t \frac{p^{l+n}}{1-p^{l+n}} \leq \sum_{l=1}^t \frac{p^{l+n}}{1-p} \\ &= (1-p) \sum_{l=1}^t p^{l+n} = (1-p)p^{n+1} \sum_{l=1}^t p^{l-1} \\ &\leq \frac{p^{n+1}}{(1-p)^2}, \end{aligned} \quad (4.36)$$

und damit die Behauptung. \square

Im Beweis waren wir das folgende Lemma schuldig geblieben:

LEMMA 4.46. Sei $(g_n)_{n \in \mathbb{N}}$ eine Folge von Vektoren aus l_{TV} , die in Norm gegen einen Vektor g konvergieren, d.h.

$$\lim_{n \rightarrow \infty} \|g_n - g\|_{TV} = 0.$$

Dann konvergiert auch die Folge $(\tilde{g}_n := \frac{1}{n} \sum_{i=0}^{n-1} g_i)_{n \in \mathbb{N}}$ in Norm gegen g .

BEWEIS. Das folgt aus der Abschätzung

$$\begin{aligned} \left\| \frac{1}{n} \sum_{i=0}^{n-1} g_i - g \right\|_{TV} &= \left\| \frac{1}{n} \sum_{i=0}^{n-1} g_i - \frac{1}{n} \sum_{i=0}^{n-1} g \right\|_{TV} \\ &\leq \frac{1}{n} \sum_{i=0}^{n-1} \|g_i - g\|_{TV} \end{aligned} \quad (4.37)$$

mit dem nachstehenden elementaren Satz von Cesaro/Cauchy. \square

LEMMA 4.47 (Satz von Cesaro/Cauchy). *Sei $(a_n)_{n \in \mathbb{N}}$ eine konvergente Folge reeller Zahlen, und sei $a := \lim_{n \in \mathbb{N}} a_n$. Dann gilt*

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

Die Folge der arithmetischen Mittel einer konvergenten Folge konvergiert gegen denselben Grenzwert wie die Folge selbst.

BEWEIS. Siehe [Heu86], S. 177, 27.1. □

4.3.4. Bedingte Wahrscheinlichkeiten. In diesem Abschnitt zeigen wir, dass sich die Eigenschaft der Evolutionssummierbarkeit auf Ereignisvektoren unter gewissen Voraussetzungen überträgt. Wir möchten an dieser Stelle darauf hinweisen, dass man unter Zuhilfenahme des später zitierten Ergodensatzes von Birkhoff 4.59 die unten angeführten Resultate im Sprachrahmen der dynamischen Systeme in “starker Konvergenz” und damit unter Berücksichtigung von Satz 4.41 auch für ES-Vektoren erhält, siehe dazu Bemerkung 4.49 weiter unten. Wir versuchen an dieser Stelle wie auch im Folgenden eine nicht auf starke Sätze angewiesene Mathematik zu betreiben. In diesem Sinne hier ein paar Sätze, die allein mit unserem Modell erzielt wurden.

SATZ 4.48. *Sei g ein stochastischer ES-Vektor und B ein invariantes ($T_\Omega^{-1}B = B$) Ereignis mit $|g|(B) > 0$. Dann ist auch g^B ein ES-Vektor.*

BEWEIS. Sei $\epsilon \in \mathbb{R}^+$. Setze $g_n := 1/n \sum_{i=0}^{n-1} \mu^i g$. Da die g_n nach Voraussetzung eine Cauchy-Folge bilden, findet man ein $N_0 \in \mathbb{N}$, so dass für alle $n, m \geq N_0$:

$$\|g_n - g_m\|_{TV} < \frac{|g|(B)}{2\|g\|_{TV}} \epsilon.$$

Damit folgt für $n, m \geq N_0$ mit Satz 2.63

$$\begin{aligned}
& \left\| \frac{1}{n} \sum_{i=0}^{n-1} \mu^i g^B - \frac{1}{m} \sum_{i=0}^{m-1} \mu^i g^B \right\|_{TV} = 2 \sup_{A \in \mathcal{B}_\Sigma} \left| \frac{1}{n} \sum_{i=0}^{n-1} \mu^i g^B(A) - \frac{1}{m} \sum_{i=0}^{m-1} \mu^i g^B(A) \right| \\
& = 2 \sup_{A \in \mathcal{B}_\Sigma} \frac{\|g\|_{TV}}{|g|(B)} \left| \frac{1}{n} \sum_{i=0}^{n-1} g(T_\Omega^{-i} A \cap B) - \frac{1}{m} \sum_{i=0}^{m-1} g(T_\Omega^{-i} A \cap B) \right| \\
& = 2 \sup_{A \in \mathcal{B}_\Sigma} \frac{\|g\|_{TV}}{|g|(B)} \left| \frac{1}{n} \sum_{i=0}^{n-1} g(T_\Omega^{-i} A \cap T_\Omega^{-i} B) - \frac{1}{m} \sum_{i=0}^{m-1} g(T_\Omega^{-i} A \cap T_\Omega^{-i} B) \right| \\
& = 2 \sup_{A \in \mathcal{B}_\Sigma} \frac{\|g\|_{TV}}{|g|(B)} \left| \frac{1}{n} \sum_{i=0}^{n-1} g(T_\Omega^{-i}(A \cap B)) - \frac{1}{m} \sum_{i=0}^{m-1} g(T_\Omega^{-i}(A \cap B)) \right| \\
& = 2 \sup_{A \in \mathcal{B}_\Sigma} \frac{\|g\|_{TV}}{|g|(B)} \left| \frac{1}{n} \sum_{i=0}^{n-1} \mu^i g(A \cap B) - \frac{1}{m} \sum_{i=0}^{m-1} \mu^i g(A \cap B) \right| \\
& \leq 2 \sup_{A \in \mathcal{B}_\Sigma} \frac{\|g\|_{TV}}{|g|(B)} \left| \frac{1}{n} \sum_{i=0}^{n-1} \mu^i g(A) - \frac{1}{m} \sum_{i=0}^{m-1} \mu^i g(A) \right| \\
& \leq \frac{2\|g\|_{TV}}{|g|(B)} \|g_n - g_m\|_{TV} < \epsilon.
\end{aligned} \tag{4.38}$$

Die $\frac{1}{n} \sum_{i=0}^{n-1} \mu^i g^B$ bilden somit eine Cauchy-Folge und konvergieren deshalb gemäß Satz 2.40. \square

BEMERKUNG 4.49. *Wie schon oben angedeutet erübrigen sich obige Arbeiten, bringt man den Ergodensatz von Birkhoff 4.59 ins Spiel. Eine einfache Folgerung desselben ist nämlich, dass für ein AMS-Wahrscheinlichkeitsmaß m und ein Ereignis B auch m^B ein AMS-Wahrscheinlichkeitsmaß ist (wobei $m^B(A) := m(A|B)$ definiert sein soll). In Hinblick auf Satz 4.41 werden die obigen Arbeiten dann hinfällig. Nach wie vor glaubt der Autor jedoch an einen elementaren Beweis, dass g^B für beliebige Ereignisse B ein ES-Vektor ist, gegeben g ist ein ES-Vektor. Dazu bemerke man, dass es reichte,*

$$\tau_a(l_{ES}) \subset l_{ES}$$

für die Prognoseoperatoren τ_a zu zeigen. Die Anwendungen von Lemma 3.12 und des Approximationssatzes 2.71 führten dann zum Ziel.

Für endlich-dimensionale g 's ist es uns zumindest bereits gelungen, die Aussage zu erhalten - eben ohne Ergodensatz von Birkhoff. Im Falle endlich-dimensionaler Vektoren g sind die dazu assoziierten Ereignisvektoren g^B evolutionssummierbar.

SATZ 4.50. *Sei $g \in l_{SM}$ ein endlich-dimensionaler Vektor und $B \in \mathcal{B}_\Sigma$ ein beliebiges Ereignis mit $|g|(B) > 0$. Dann ist g^B ein ES-Vektor und es gilt*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} \mu^i g^B \in \mathcal{V}_g.$$

BEWEIS. Wir bemerken, dass die Behauptung für eine Zylindermenge $Z(B)$ einfach zu zeigen ist. Dazu sei $t := |B|$, dann gilt nach 3.12

$$\mu^t g^B = g_B \in \mathcal{V}_g.$$

Damit ist g^B von endlicher Evolutionsdimension und wegen Satz 4.43 ein ES-Vektor. Offensichtlich gilt noch

$$\bar{\mu} g^B = \bar{\mu} g_B \in \mathcal{V}_g.$$

Sei nun $B \in \mathcal{B}_\Sigma$ beliebig. Gemäß des Approximationssatzes für Vektoren 2.71 wählen wir eine Folge von Zylindermengen mit

$$\lim_{i \rightarrow \infty} g^{B_i} = g^B. \quad (4.39)$$

Da l_{ES} abgeschlossen ist (Satz 4.38), folgt, dass g^B ein ES-Vektor ist. Weiter wegen der Stetigkeit von $\bar{\mu}$ (siehe Proposition 4.35)

$$\lim_{n \rightarrow \infty} \bar{\mu} g^{B_i} = \bar{\mu} g^B$$

und damit wegen $\bar{\mu} g^{B_i} \in \mathcal{V}_g$ für alle i die Behauptung. \square

Der sich anschließende Satz zeigt einen Weg auf, wie der Grenzvektor von g^B praktisch gefunden werden kann, indem man ihn durch Vektoren aus \mathcal{V}_g approximiert.

SATZ 4.51. Sei $g \in l_{SM}$. Sei B mit $|g|(B) > 0$ ein Ereignis, so dass der Limes

$$h := \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} \mu^i g^B$$

(in Norm) existiert. Wählt man eine Folge von Zylindermengen $(B_i)_{i \in \mathbb{N}}$, so dass $|B_i| = i$ und

$$\lim_{i \rightarrow \infty} \|g^{B_i} - g^B\|_{TV} = 0, \quad (4.40)$$

so gilt

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} g_{B_i} = h \in \bar{\mathcal{V}}_g.$$

BEWEIS. Wir setzen

$$a_n := \left\| \frac{1}{n} \sum_{i=0}^{n-1} g_{B_i} - \frac{1}{n} \sum_{i=0}^{n-1} \mu^i g^B \right\|_{TV}$$

und zeigen $\lim_{n \rightarrow \infty} a_n = 0$. Daraus folgt die Behauptung. Dann gilt

$$\begin{aligned} a_n &= \left\| \frac{1}{n} \sum_{i=0}^{n-1} g_{B_i} - \frac{1}{n} \sum_{i=0}^{n-1} \mu^i g^B \right\|_{TV} = \left\| \frac{1}{n} \sum_{i=0}^{n-1} \mu^i g^{B_i} - \frac{1}{n} \sum_{i=0}^{n-1} \mu^i g^B \right\|_{TV} \\ &\leq \frac{1}{n} \sum_{i=0}^{n-1} \|\mu^i g^{B_i} - \mu^i g^B\|_{TV} \leq \frac{1}{n} \sum_{i=0}^{n-1} \|\mu^i\|_{TV} \|g^{B_i} - g^B\|_{TV} \\ &\stackrel{\text{Satz 3.6}}{\leq} \frac{1}{n} \sum_{i=0}^{n-1} \|g^{B_i} - g^B\|_{TV}, \end{aligned} \quad (4.41)$$

woraus wegen (4.40) mit der zusätzlichen Hilfe von Lemma 4.46 die Behauptung folgt. \square

KOROLLAR 4.52. *Sei g ein endlich-dimensionaler stochastischer Vektor und (B_k) eine Folge von Zylindermengen, so dass es eine Menge $B \in \mathcal{B}_\Sigma$ gibt mit $g(B) \neq 0$ und $\lim_{k \rightarrow \infty} g(B_k \triangle B) = 0$. Dann konvergiert die Folge der Vektoren $\frac{1}{n} \sum_{k=0}^{n-1} g_{B_k}$.*

BEWEIS. Das ist eine offensichtliche Folgerung aus den Sätzen 4.43 und 4.48. \square

4.4. Ergodizität

Dieser Abschnitt führt den Begriff des “ergodischen Vektors” ein. Dabei sind ergodische Vektoren so konzipiert, dass stochastische ergodische Vektoren gerade mit den ergodischen stochastischen Prozessen korrespondieren. In den meisten für uns wichtigen Fällen lässt sich Ergodizität dann gewissermaßen als Eindeutigkeit des stationären Grenzmaßes für alle Elemente des Prognoseraums gedeutet werden. Vor allem im Falle endlich-dimensionaler Vektoren lässt sich dies besonders klar formulieren. Diese Formulierung wiederum erlaubt die Einsicht der sich im folgenden Kapitel anschließenden “ergodischen Zerlegung” endlich-dimensionaler Vektoren.

4.4.1. Einführung und Motivation. Als Einleitung erklären wir den aus der Welt der dynamischen Systeme stammenden Begriff der Ergodizität. Dieser zentrale, namensstiftende Begriff kann als Begriff der Ergodentheorie schlechthin gewertet werden. Ergodische dynamische Systeme haben besonders schöne Eigenschaften, was die Konvergenz von Auszählstatistiken “entlang” von Sequenzen betrifft.

Dazu wollen wir als Erstes den ebenfalls wichtigen Begriff der ergodischen Eigenschaften eines dynamischen Systems bezüglich einer messbaren Funktion $f : \Omega \rightarrow \mathbb{R}$ vorstellen.

DEFINITION 4.53. *Sei $(\Omega, \mathcal{B}, m, T)$ ein dynamisches System und f eine messbare Funktion auf Ω . Man sagt, dass das dynamische System die **ergodische Eigenschaft bezüglich f** hat, falls*

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

fast überall existiert, mit anderen Worten

$$m(\{\omega \mid \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i \omega) \text{ ex.}\}) = m(\Omega),$$

was wiederum bedeutet, dass die Menge der Sequenzen, für die eine solche Konvergenz nicht gegeben ist, eine Nullmenge bezüglich m ist.

Die Eigenschaft eines dynamischen Systems $(\Omega, \mathcal{B}, m, T)$, ergodisch zu sein, bedeutet nun anschaulich, dass für messbare Funktionen $f : \Omega \rightarrow \mathbb{R}$, bezüglich derer das System die ergodische Eigenschaft hat, die Limes

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i \omega) = \int_{\Omega} f dm$$

fast überall denselben Grenzwert haben, wobei dieser Grenzwert mit dem Erwartungswert für diese messbare Funktion übereinstimmt. In der Stochastik bedeutet das, dass für fast jede Sequenz, die von einem ergodischen stochastischen Prozess herrührt, die gleichen Statistiken erhoben werden. Es reicht demnach, eine einzige Sequenz des Prozesses zu betrachten, um Erwartungswerte des Prozesses ermitteln zu können.

Die eigentliche Definition des Begriffs sieht jedoch so aus:

DEFINITION 4.54. *Ein dynamisches System $(\Omega, \mathcal{B}, m, T)$ heißt **ergodisch**, falls (sei $\mathcal{I} \subset \mathcal{B}$ die Unter- σ -Algebra der bezüglich des Schiebeoperators invarianten Ereignisse, d.h. $T_{\Sigma}^{-1}I = I$ für $I \in \mathcal{I}$)*

$$\forall I \in \mathcal{I} : m(I) \in \{0, m(\Omega)\}.$$

Mit anderen Worten, $(\Omega, \mathcal{B}, m, T)$ ist ergodisch, falls alle invarianten Ereignisse entweder mit 0 oder mit der "Gesamtmasse" $m(\Omega)$ belegt werden.

Das veranlasst uns zur folgenden Definition:

DEFINITION 4.55. *Wir nennen einen Vektor $g \in l_{SM}$ **ergodisch**, falls (sei $\mathcal{I} \subset \mathcal{B}$ die Unter- σ -Algebra der invarianten Ereignisse wie in der Definition darüber)*

$$\forall I \in \mathcal{I} : |m_g|(I) = m_{|g|}(I) \in \{0, \|g\|_{TV}\}, \quad (4.42)$$

wobei $|g|$ der totale Variationsvektor gemäß Definition 2.54 sei.

Mit anderen Worten ist ein endlicher, signierter Maßvektor genau dann ergodisch, falls alle invarianten Ereignisse "betraglich" mit voller "Wahrscheinlichkeit" oder mit 0 belegt sind. Es ist nun offensichtlich, dass diese Definition im Falle von Maßvektoren mit der für dynamische Systeme gegebenen Definition übereinstimmt:

PROPOSITION 4.56. *Ist $g \in l_{SM}$ ein Maßvektor, so ist das dazu assoziierte dynamische System $(\Omega_{\Sigma}, \mathcal{B}_{\Sigma}, m_g, T_{\Omega})$ genau dann ergodisch, wenn g es ist.*

BEWEIS. Das ist nun tatsächlich offensichtlich nach den Definitionen. \square

Wir erinnern hier und im Folgenden an die Definition 2.54 der Variationsvektoren g^+, g^- und $|g| = g^+ + g^-$ und bemerken, dass wir im Folgenden hemmungslosen Gebrauch der Schreibweise

$$g(B) := m_g(B)$$

für $B \in \mathcal{B}_\Sigma$ machen, wir also Vektoren $g \in l_{SM}$ mit ihren assoziierten endlichen, signierten Maßen identifizieren. Wir erhalten dann:

PROPOSITION 4.57. *Ein Vektor $g \in l_{SM}$ ist genau dann ergodisch, falls g^+ und g^- ergodisch sind und, falls $g^+ \neq 0$ und $g^- \neq 0$, zusätzlich*

$$\forall I \in \mathcal{I} : \quad g^+(I) = 0 \iff g^-(I) = 0 \quad (4.43)$$

gilt.

BEWEIS. Ist g ergodisch, so folgt wegen $|g| = g^+ + g^-$ für $I \in \mathcal{I}$ direkt $g^+(I) = g^-(I) = 0$ für $|g|(I) = 0$. Gilt $|g|(I) = \|g\|_{TV}$, so

$$\|g\|_{TV} = |g|(I) = g^+(I) + g^-(I) \leq \|g^+\|_{TV} + \|g^-\|_{TV} = \|g\|_{TV},$$

woraus sowohl $|g^+|(I) = g^+(I) = \|g^+\|_{TV}$ und ebenso $|g^-|(I) = \|g^-\|_{TV}$ als auch Aussage (4.43) folgen. Umgekehrt ist die Aussage trivial. \square

4.4.2. Einschub: Ergodensätze. Zur besseren Einordnung der folgenden Ergebnisse betrachten wir zwei prominente Ergodensätze. Ergodensätze machen einen großen Teil der Ergodentheorie aus und füllen für sich allein ganze Bücher ([Kre85]). Im Falle allgemeiner dynamischer Systeme $(\Omega, \mathcal{B}, m, T)$ stellen sie eine Beziehung zwischen Eigenschaften des Maßes m selbst und dem Verhalten von Funktionen $f : \Omega \rightarrow \mathbb{R}$ entlang von Orbits $(T^k \omega)_{k \in \mathbb{N}}$ her.

Wir betrachten die zwei bekanntesten Exemplare, von denen wir später dann auch (in geringem Ausmaß) Gebrauch machen werden. Der erste Ergodensatz ist von von (sic!) Neumann, erzielt eine Aussage für stationäre Systeme und ist leichter zu beweisen als der sich anschließende Ergodensatz von Birkhoff. Wir führen ihn der Vollständigkeit halber auf, weil wir unsere Ergebnisse allein durch Berufung auf ihn erzielen können. Für die dafür benötigten Begriffe aus der Integrationstheorie siehe z.B. [Bau90]. Wer Integrationstheorie vermeiden möchte, überspringe jedoch diesen Satz und wende sich dem Birkhoffschen zu. Auch er hilft uns weiter und seine Aussage ist einfacher zu verstehen. Der Beweis jedoch ist um einiges komplizierter.

SATZ 4.58 (Mittlerer Ergodensatz von von Neumann). *Sei $(\Omega, \mathcal{B}, m, T)$ ein stationäres dynamisches System und $f, g \in L^2(\Omega, \mathcal{B}, m)$ quadrat-integrierbar (d.h. f^2 und g^2 sind jeweils integrierbare Funktionen). Dann existieren die Limites*

$$\lim_{n \rightarrow \infty} \sum_{i=0}^{n-1} \int f(T^i \omega) g(\omega) dm(\omega)$$

fast überall. Ist $(\Omega, \mathcal{B}, m, T)$ zusätzlich ergodisch, so fast überall

$$\lim_{n \rightarrow \infty} \sum_{i=0}^{n-1} \int f(T^i \omega) g(\omega) dm(\omega) = \int f dm(\omega) \int g dm(\omega).$$

BEWEIS. Siehe [PY98], Seiten 99-101. \square

Der ‘‘mittlere’’ Ergodensatz ist noch nicht wirklich das, was wir angekündigt hatten, nämlich Aussagen über Statistiken entlang einzelner Sequenzen. Solche Aussagen macht jedoch nun der

SATZ 4.59 (Ergodensatz von Birkhoff). *Ein dynamisches System $(\Omega, \mathcal{B}, m, T)$ besitzt genau dann die ergodische Eigenschaft bezüglich aller beschränkten, messbaren Funktionen (siehe Definition 4.53), wenn es AMS (siehe Definition 4.39) ist.*

BEWEIS. Den Beweis dazu gibt es in zwei Varianten: Eine einfachere Variante für ergodische dynamische Systeme und eine allgemeine. Beide finden sich in [PY98], Seiten 102-105, 111-112. \square

BEMERKUNG 4.60. *Die hier aufgeführten Ergodensätze kennen unzählige Beweisvarianten. Für den von Neumann’schen Satz siehe [Bou87], [Zim90]. Für den Birkhoff’schen siehe z.B. den von Katznelsson und Weiss übernommenen ([KW82]) auch in [Gra01] beschriebenen. Dieser kommt ohne den maximalen Ergodensatz (z.B. [Kre85], Seite 8) aus. In [Kre85] ist die nach unserem Verständnis eleganteste Variante zu finden, die Gebrauch vom maximalen Ergodensatz macht, dessen Beweis ebenso kurz wie wundersam daherkommt. Alternative Varianten wären außerdem die in [Jon83] und [Kam82] vorgestellten.*

Diese beiden Sätze kommen z.B. in folgendem Lemma zum Einsatz, das später auf einfache Art und Weise auf unsere Begriffe übertragen werden kann.

LEMMA 4.61. *Sei mit $(\Omega, \mathcal{B}, m, T)$ ein dynamisches AMS System mit stationärem Grenzmaß \bar{m} gegeben. Sei weiter \mathcal{B} von einem Ring \mathcal{F} erzeugt. Dann sind die folgenden Aussagen äquivalent:*

- (i) *m ist ergodisch.*
- (ii) $\forall G, F \in \mathcal{B} \times \mathcal{B} : \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} m(T^{-i}G \cap F) = \frac{\bar{m}(G)m(F)}{m(\Omega)}.$
- (iii) $\forall F \in \mathcal{F} : \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} m(T^{-i}F \cap F) = \frac{\bar{m}(F)m(F)}{m(\Omega)}.$

BEWEIS. Siehe [Gra01], Seiten 147/148/149. \square

BEMERKUNG 4.62. *In dieser Variante benötigt Lemma 4.61 den obenstehenden Satz von Birkhoff 4.59. Ersetzt man AMS durch stationär, so wird nur der von Neumann’sche mittlere Ergodensatz 4.58 benötigt.*

Nach diesem Exkurs fahren wir nun mit unserer Theorie fort und studieren die Zusammenhänge zwischen Ergodizität und den Begriffen der vorangehenden Abschnitte.

4.4.3. Ergodizität und Vektorräume. Ergodische Vektoren kann mit Hilfe von Untervektorräumen klassifizieren. Diesen Mechanismus wollen wir an dieser Stelle ausführen.

Dazu betrachten wir die Mengenfunktionen

$$\{0, 1\}^{\mathcal{I}_\Sigma} = \{Q : \mathcal{I}_\Sigma \rightarrow \{0, 1\}\}$$

über der σ -Algebra der invarianten Mengen \mathcal{I}_Σ , die ausschließlich die Werte 0 oder 1 annehmen. Wir betrachten nun die folgende

DEFINITION 4.63. *Sei $Q \in \{0, 1\}^{\mathcal{I}_\Sigma}$. Dann nennen wir*

$$\mathcal{V}_Q := \{g \in l_{SM} \mid \forall I \in \mathcal{I}_\Sigma : |g|(I) = \|g\|_{TV} Q(I)\}.$$

den **ergodischen Vektorraum** von Q .

Für etliche solche Q ist der dazu assoziierte ergodische Raum äußerst uninteressant:

BEMERKUNG 4.64. *Ist $Q \in \{0, 1\}^{\mathcal{I}_\Sigma}$ kein Wahrscheinlichkeitsmaß auf $(\Omega, \mathcal{I}_\Sigma)$, so offensichtlich $\mathcal{V}_Q = \{0\}$, da $|g|$ für $g \in \mathcal{V}_Q$ als Maß auf $(\Omega, \mathcal{B}_\Sigma)$ ja auch ein Maß auf $(\Omega, \mathcal{I}_\Sigma)$ induziert, das bis auf Skalierung mit Q übereinstimmt.*

Wir erhalten dann

SATZ 4.65. *Ist $Q \in \{0, 1\}^{\mathcal{I}_\Sigma}$, so ist der ergodische Raum \mathcal{V}_Q ein abgeschlossener Untervektorraum von l_{SM} .*

BEWEIS. Wegen Bemerkung 4.64 können wir ohne Einschränkung davon ausgehen, dass es sich bei Q um ein Wahrscheinlichkeitsmaß handelt. Es gilt also

$$Q(I) = 0 \iff Q(\mathbb{C}I) = 1. \quad (4.44)$$

Ist $g \in \mathcal{V}_Q$, so wegen

$$\forall I \in \mathcal{I}_\Sigma : |\alpha g|(I) = |\alpha| |g|(I) = |\alpha| \|g\|_{TV} Q(I) = \|\alpha g\|_{TV} Q(I)$$

auch $\alpha g \in \mathcal{V}_Q$ für alle $\alpha \in \mathbb{R}$. Sind weiter $g_1, g_2 \in \mathcal{V}_Q$ und ist $I \in \mathcal{I}_\Sigma$ mit $Q(I) = 0$, so

$$0 \leq |g_1 + g_2|(I) \leq |g_1|(I) + |g_2|(I) = 0 = \|g_1 + g_2\|_{TV} Q(I). \quad (4.45)$$

Ist hingegen $I \in \mathcal{I}_\Sigma$, so dass $Q(I) = 1$, so

$$\begin{aligned} |g_1 + g_2|(I) &= |g_1 + g_2|(\Omega_\Sigma) - |g_1 + g_2|(\mathbb{C}I) \\ &\stackrel{(4.44), (4.45)}{=} |g_1 + g_2|(\Omega_\Sigma) = \|g_1 + g_2\|_{TV}, \end{aligned}$$

woraus insgesamt folgt, dass \mathcal{V}_Q ein Vektorraum ist. Ist nun $h \in \overline{\mathcal{V}_Q}$, so gibt es eine Folge $h_n \in \mathcal{V}_Q$, die gegen h konvergiert. Nach Satz 2.55 konvergiert dann die Folge der $|h_n|$ gegen $|h|$. Ist nun $I \in \mathcal{I}_\Sigma$, so konvergiert die Folge $|h_n|(I)$ gegen $|h|(I)$ (siehe Korollar 2.65). Daher

$$\begin{aligned} |h|(I) &= \lim_{n \rightarrow \infty} |h_n|(I) = \lim_{n \rightarrow \infty} \|h_n\|_{TV} Q(I) \\ &= Q(I) \lim_{n \rightarrow \infty} \|h_n\|_{TV} = \|h\|_{TV} Q(I), \end{aligned}$$

womit die Behauptung des Satzes vollständig bewiesen wäre. \square

Weiter erhalten wir

SATZ 4.66. *Seien $Q_1, Q_2 \in \{0, 1\}^{\mathcal{I}_\Sigma}$, so dass $Q_1 \neq Q_2$. Dann*

$$\mathcal{V}_{Q_1} \cap \mathcal{V}_{Q_2} = \{0\}.$$

BEWEIS. Wir nehmen die Existenz eines Elements $0 \neq g \in \mathcal{V}_{Q_1} \cap \mathcal{V}_{Q_2}$ an. Da $Q_1 \neq Q_2$, gibt es eine Menge \mathcal{I}_Σ , so dass $Q_1(I) \neq Q_2(I)$, also ohne Einschränkung $Q_1(I) = 1, Q_2(I) = 0$. Es folgt

$$0 = \|g\|_{TV} Q_2(I) = |g|(I) = \|g\|_{TV} Q_1(I) = \|g\|_{TV},$$

ein Widerspruch zu $g \neq 0$. \square

Ist nun $g \neq 0$ ein ergodischer Vektor, so induziert $|g|$ über die Vorschrift

$$Q_g(I) := \frac{1}{\|g\|_{TV}} |g|(I)$$

eine Mengenfunktion $Q_g \in \{0, 1\}^{\mathcal{I}_\Sigma}$ (im Falle $g = 0$ setzen wir $Q_g \equiv 0$). Es gilt dann der folgende

SATZ 4.67. *Sei $g \in l_{SM}$ ein ergodischer Vektor. Dann*

$$\overline{\mathcal{V}_g} \subset \mathcal{V}_{Q_g}.$$

BEWEIS. Wegen der Abgeschlossenheit von \mathcal{V}_{Q_g} reicht es, $\mathcal{V}_g \subset \mathcal{V}_{Q_g}$ zu zeigen. Wir betrachten zunächst zwei Lemmata.

LEMMA 4.68. *Sei $Q \in \{0, 1\}^{\mathcal{I}_\Sigma}$ und $g \in \mathcal{V}_Q$. Dann auch $g^+, g^- \in \mathcal{V}_Q$.*

BEWEIS. Das folgt mit Proposition 4.57 offensichtlich aus den Definitionen der involvierten Begriffe. \square

LEMMA 4.69. *Sei $Q \in \{0, 1\}^{\mathcal{I}_\Sigma}$ und $g \in \mathcal{V}_Q$ ein Maßvektor, also $g \geq 0$. Dann*

$$\forall \bar{a} \in \Sigma^* : \tau_{\bar{a}} g \in \mathcal{V}_Q. \quad (4.46)$$

BEWEIS. Sei $\bar{a} \in \Sigma^*$. Dann folgt die Behauptung aus der Rechnung

$$g(I) = g(T_\Omega^{-|\bar{a}|} I) = \mu^{|\bar{a}|} g(I) = \sum_{\bar{b} \in \Sigma^{|\bar{a}|}} \tau_{\bar{b}} g(I)$$

sofort im Falle $g(I) = 0$ und im Falle $g(I) = \|g\|_{TV} = g(\square)$ dann aus der weiteren Rechnung

$$\tau_{\bar{a}} g(I) = \tau_{\bar{a}} g(\Omega_\Sigma) - \underbrace{\tau_{\bar{a}} g(\mathbb{C}I)}_{=0} = \tau_{\bar{a}} g(\Omega_\Sigma) = \|\tau_{\bar{a}} g\|_{TV}.$$

\square

Weiter im **Beweis von Satz 4.67**: Sei $\bar{a} \in \Sigma^*$ und $I \in \mathcal{I}_\Sigma$ mit $|g|(I) = 0$. Dann gilt

$$\begin{aligned} |\tau_{\bar{a}} g|(I) &= |\tau_{\bar{a}}(g^+ - g^-)|(I) \leq (|\tau_{\bar{a}} g^+| + |\tau_{\bar{a}} g^-|)(I) \\ &= \tau_{\bar{a}} g^+(I) + \tau_{\bar{a}} g^-(I) \stackrel{L.4.68, L.4.69}{=} 0 \end{aligned}$$

und im Falle $|g|(I) = \|g\|_{TV}$ mit einem schon mehrfach zitierten Argument

$$|\tau_{\bar{a}}g|(I) = |\tau_{\bar{a}}g|(\Omega_{\Sigma}) - |\tau_{\bar{a}}g|(\mathbb{C}I) = \|\tau_{\bar{a}}g\|_{TV}.$$

□

KOROLLAR 4.70. *Sei $g \in l_{SM}$ ein ergodischer Vektor und $h \in \overline{\mathcal{V}}_g, h \neq 0$. Ist $I \in \mathcal{I}$ eine invariante Menge, so*

$$\begin{aligned} |g|(I) = 1 &\implies \exists A \subset I : h(A) \neq 0 \\ |g|(I) = 0 &\implies \forall A \subset I : h(A) = 0. \end{aligned}$$

BEWEIS. Im Fall $g = 0$ ist nichts zu zeigen. Wir können daher $\|g\|_{TV} > 0$ annehmen. Wegen Satz 4.67 gilt $h \in \mathcal{V}_{Q_g}$. Ist h ein Maßvektor, so folgt die Behauptung wegen $h = |h|$ im Fall $|g|(I) = \|g\|_{TV}$, wählt man $A = I$, und im Fall $|g|(I) = 0$ für $A \subset I$ aus der Rechnung

$$0 \leq h(A) = h(A \cap I) \leq h(I) = \|h\|_{TV} \frac{1}{\|g\|_{TV}} |g|(I) = 0.$$

Mutatis mutandis erhalten wir die Aussage analog für den Fall, dass $-h$ ein Maßvektor ist. Bleibt $h = h^+ - h^-$ mit linear unabhängigen Variationsvektoren h^+, h^- zu betrachten. Diese befinden sich ebenfalls in \mathcal{V}_{Q_g} (Lemma 4.68). Ist I eine invariante Menge mit $|g|(I) = 0$ und $A \subset I$ eine Teilmenge, so

$$0 \leq h^+(A) \leq h^+(I) = \|h^+\|_{TV} \frac{1}{\|g\|_{TV}} |g|(I) = 0$$

und die Behauptung folgt für diesen Fall mit einer zusätzlichen analogen Rechnung für h^- . Im Fall $|g|(I) = \|g\|_{TV}$ können wir zunächst aufgrund der linearen Unabhängigkeit der h^+, h^- eine Menge B mit

$$h^+(B) \neq h^-(B)$$

wählen. Setzt man $A := B \cap I$, so erhält man wegen $|g|(\mathbb{C}I) = 0$ und folglich $h^+(B \cap \mathbb{C}I) = h^-(B \cap \mathbb{C}I) = 0$ aufgrund des ersten Falls

$$\begin{aligned} h(A) &= h^+(A) - h^-(A) = h^+(B \cap I) - h^-(B \cap I) \\ &= h^+(B) - h(B \cap \mathbb{C}I) - h^-(B) + h(B \cap \mathbb{C}I) \\ &= h^+(B) - h^-(B) \neq 0, \end{aligned}$$

womit die Behauptung vollständig bewiesen wäre. □

Wir werden uns im weiteren Verlauf mehrfach auf diese Erkenntnisse stützen.

4.4.4. Ergodizität und Stationarität. Die Kombination von Ergodizität und Stationarität ist bei Theoretikern der Dynamischen Systeme gerne gesehen, ergänzen sich doch diese Eigenschaften intuitiv hervorragend. Eine Vielzahl von Sätzen benötigt diese Kombination als Voraussetzung und eine große Menge von Anwendungen funktioniert unter der Annahme, dass die Datenquellen solcher Natur sind (was in der Praxis fast nie der Fall ist).

Auch wir wollen einen Beitrag zu diesem Thema leisten. Wir bringen zunächst in Erfahrung, was wir von voneinander verschiedenen ergodischen,

stationären Vektoren erwarten dürfen: sie können in gewisser Weise als disjunkt betrachtet werden.

LEMMA 4.71. *Seien $g_1, \dots, g_n \in l_{SM}$ eine Menge paarweise verschiedener, ergodischer, stationärer, stochastischer Vektoren. Dann gibt es invariante Mengen $I_i \in \mathcal{I}_\Sigma, i = 1, \dots, n$, so dass*

$$g_i(I_j) = \delta_{ij}, i, j = 1, \dots, n.$$

BEWEIS. Wir betrachten zunächst Pärchen von Vektoren $g_i \neq g_j$. Da $g_i - g_j \neq 0$ stationär ist, gibt es aufgrund von Lemma 4.14 ein invariantes Ereignis \tilde{I}_{ij} , so dass $(g_i - g_j)(\tilde{I}_{ij}) \neq 0$ und damit $(g_i - g_j)(\tilde{I}_{ij}) \in \{-1, 1\}$ wegen der Ergodizität von g_i, g_j . Im Falle $(g_i - g_j)(\tilde{I}_{ij}) = 1$ setzen wir $I_{ij} := \tilde{I}_{ij}, I_{ji} = \complement \tilde{I}_{ij}$ und im Fall $(g_i - g_j)(\tilde{I}_{ij}) = -1$ definieren wir $I_{ij} := \complement \tilde{I}_{ij}, I_{ji} = \tilde{I}_{ij}$ und sehen in beiden Fällen ein, invariante Mengen I_{ij}, I_{ji} gefunden zu haben, so dass

$$\begin{aligned} g_i(I_{ij}) = 1 \quad \text{und} \quad g_j(I_{ij}) = 0 \\ g_j(I_{ji}) = 1 \quad \text{und} \quad g_i(I_{ji}) = 0. \end{aligned}$$

Haben wir invariante Mengen I_{ij} für alle Pärchen g_i, g_j bestimmt, so setzen wir

$$I_i := \bigcap_{j \neq i} I_{ij}$$

und erhalten

$$g_i(I_i) = 1, \forall j \neq i : g_j(I_i) = 0$$

und damit das Gewünschte. \square

KOROLLAR 4.72. *Sei g_1, \dots, g_n eine Menge paarweise verschiedener, ergodischer, stationärer, stochastischer Vektoren. Dann sind die g_1, \dots, g_n linear unabhängig.*

BEWEIS. Gilt $\sum_{i=1}^n \alpha_i g_i = 0$, so

$$\forall j = 1, \dots, n : 0 = \sum_{i=1}^n \alpha_i g_i(I_j) = \alpha_j$$

mit den in Lemma 4.71 bestimmten invarianten Mengen I_j und damit die Behauptung. \square

KOROLLAR 4.73. *Seien $g_1, \dots, g_n \in l_{SM}$ eine Menge ergodischer, stationärer, Vektoren, so dass die Menge der totalen Variationsvektoren $|g_i|$ linear unabhängig ist. Dann gibt es invariante Mengen $I_i \in \mathcal{I}_\Sigma, i = 1, \dots, n$, so dass*

$$|g_i|(I_j) = \delta_{ij} \|g_i\|_{TV}, i, j = 1, \dots, n.$$

BEWEIS. Skaliert man die $|g_i|$ zu stochastischen Vektoren, so erfüllen diese die Voraussetzungen von Lemma 4.71. Die dort gefundenen Mengen leisten das Gewünschte. \square

KOROLLAR 4.74. *Seien $f \neq g$ zwei ergodische, stationäre, stochastische Vektoren. Sei $\alpha \in [0, 1]$ und $h := \alpha f + (1 - \alpha)g$ eine Konvexkombination der beiden. Dann gilt:*

$$h \text{ ergodisch} \iff \alpha \in \{0, 1\}. \quad (4.47)$$

Mit anderen Worten: Jede echte Konvexkombination zweier stationärer, ergodischer, stochastischer Vektoren ist nicht mehr ergodisch.

BEWEIS. Die Rückrichtung ist natürlich klar. Für die andere Richtung finden wir mit Lemma 4.71 eine invariante Menge I , so dass $f(I) = 1$ und $g(I) = 0$. Für $\alpha \in]0, 1[$ folgt dann

$$h(I) = \alpha f(I) + (1 - \alpha)g(I) = \alpha f(I) = \alpha \notin \{0, \|h\|_{TV} = 1\}.$$

Das widerspricht der definierenden Eigenschaft von Ergodizität. \square

Wir kommen nun zum Höhepunkt dieses Abschnitts: Innerhalb eines ergodischen Raums \mathcal{V}_Q sind stationäre Vektoren dünn gesät.

SATZ 4.75. *Sei $Q \in \{0, 1\}^{\mathcal{I}_\Sigma}$. Dann gilt*

$$\dim(\mathcal{V}_Q \cap l_\mu) \leq 1.$$

BEWEIS. Wir nehmen das Gegenteil an und finden damit für passendes $Q \in \{0, 1\}^{\mathcal{I}_\Sigma}$ zwei linear unabhängige, stationäre Vektoren $g_1, g_2 \in \mathcal{V}_Q$. Wir können nun ohne Einschränkung die g_1, g_2 als stochastische Vektoren wählen. Im Fall, dass jeweils g_1 oder $-g_1$ und g_2 und $-g_2$ Maßvektoren sind, erhalten wir linear unabhängige stochastische Vektoren durch Skalieren der beiden. Ist z.B. g_1 nicht solcher Natur, so sind g_1^+, g_1^- linear unabhängige Maßvektoren, die aufgrund Lemma 4.68 beide in \mathcal{V}_Q liegen. Durch Skalieren werden sie stochastisch, und aufgrund Lemma 4.11 wissen wir, dass sie stationär sind. Als Elemente von \mathcal{V}_Q sind sie obendrein ergodisch. Mit Lemma 4.71 finden wir dann eine invariante Menge I , so dass $g_1(I) = 1 \neq 0 = g_2(I)$, was es gemäß der Definition von \mathcal{V}_Q nicht geben darf. \square

Wir erhalten noch ein (möglicherweise verblüffendes) Korollar. Stationäre, ergodische Vektoren sind immer skalierte stochastische Vektoren.

KOROLLAR 4.76. *Sei $g \in l_{SM}$ ein ergodischer, stationärer Vektor. Dann ist entweder g oder $-g$ ein Maßvektor.*

BEWEIS. Nehmen wir das Gegenteil an, so $g = g^+ - g^-$, wobei g^+, g^- linear unabhängige, stationäre Maßvektoren sind mit $g^+, g^- \in \mathcal{V}_{Q_g}$ (siehe dafür die Lemmata 4.11, 4.68). Das ist ein Widerspruch zu Satz 4.75. \square

BEMERKUNG 4.77. *Dies könnte als Aufhänger zur Lösung des Positivitätsproblems 3.49 benutzt werden. Ergodizität und Stationarität sind leicht überprüfbar. Für solche signierten Maßvektoren g muss demnach ein Algorithmus her, der überprüft, ob g ein Element von l_{TV} ist. Dieses Problem scheint einfacher in Angriff zu nehmen zu sein.*

All dies konnte ohne den Einsatz von Ergodensätzen aufgrund elementarer Überlegungen erhalten werden. Als Zusatz formulieren wir jedoch noch ein letztes Lemma, das zeigt, wie sich Prognosevektoren unter Anwendung des Evolutionsoperators verhalten: sie streben zurück zum Grundvektor. Dieses Ergebnis erhalten wir alleine durch Einsatz des von Neumannschen mittleren Ergodensatzes 4.58, siehe Bemerkung 4.60.

SATZ 4.78. *Sei $g \in l_{SM}$ ein stationärer Maßvektor. Dann sind die folgenden Aussagen äquivalent:*

- (i) g ist ergodisch.
- (ii) $\forall(\bar{a}, \bar{b}) \in \Sigma^* \times \Sigma^* : \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=0}^{n-1} \mu^t \tau_{\bar{a}} g(\bar{b}) = \frac{g(\bar{a})}{\|g\|_{TV}} \bar{g}(\bar{b})$.
- (iii) $\forall \bar{a} \in \Sigma^* : \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=0}^{n-1} \mu^t \tau_{\bar{a}} g(\bar{a}) = \frac{g(\bar{a})}{\|g\|_{TV}} \bar{g}(\bar{a})$.

BEWEIS. Wir bemerken, dass

$$\begin{aligned} \mu^t \tau_{\bar{a}} g(\bar{b}) &= \sum_{\bar{c} \in \Sigma^t} g(\bar{a} \bar{c} \bar{b}) \\ &= g(\{\bar{a}\} \cap T^{-(|\bar{a}|+t)} \{\bar{b}\}). \end{aligned} \tag{4.48}$$

Im Falle der Existenz der Limites gilt also

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=0}^{n-1} \mu^t \tau_{\bar{a}} g(\bar{b}) &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=0}^{n-1} g(\{\bar{a}\} \cap T^{-(|\bar{a}|+t)} \{\bar{b}\}) \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=0}^{n-1} g(\{\bar{a}\} \cap T^{-t} \{\bar{b}\}) \end{aligned} \tag{4.49}$$

Man erhält die Behauptung nun mit Lemma 4.61, indem man bemerkt, dass die Existenz der letzten Limites äquivalent zur Existenz der Limites

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=0}^{n-1} g(F \cap T^{-t} G)$$

für alle Zylindermengen F, G induziert, welche ja \mathcal{B}_Σ erzeugen. Anwendung von Lemma 4.61 führt daher zur Behauptung, da damit die dortigen Kriterien (i), (ii), (iii) gerade denen unseres Satzes entsprechen. \square

4.4.5. Ergodizität und Evolutionssummierbarkeit. Die Kombination der Begriffe der Ergodizität ergibt nun in unserer Sprache eine prägnante Charakterisierung der Ergodizität. Im Übrigen erzielt man durch die Kombination von Ergodizität und Evolutionssummierbarkeit im Wesentlichen das Verhalten, das man im Hinterkopf hat, wenn man sich Ergodizität vorstellt, mit anderen Worten, dass Statistiken entlang von Sequenzen immer den gleichen Grenzwert anstreben. Wir starten mit zwei kleinen Lemmata.

LEMMA 4.79. *Sei g ein ES-Maßvektor. Dann*

$$g \text{ ergodisch} \iff \bar{\mu}g \text{ ergodisch.}$$

BEWEIS. Das folgt unmittelbar aus Bemerkung 4.37 und der Definition von Ergodizität unter dem Hinweis auf $|g| = g$. \square

Damit können wir ein weiteres Lemma formulieren:

LEMMA 4.80. *Sei $g \in l_{SM}$ ein ergodischer Maßvektor. Dann gilt*

$$\mathcal{V}_{Q_g} = \mathcal{V}_{Q_{\bar{\mu}g}}.$$

BEWEIS. Wegen Bemerkung 4.37 folgt offensichtlich

$$Q_g = Q_{\bar{\mu}g},$$

und damit die Behauptung. \square

LEMMA 4.81. *Seien g_1, g_2 zwei ergodische ES-Maßvektoren, so dass $\bar{g}_1 := \bar{\mu}g_1$ und $\bar{g}_2 := \bar{\mu}g_2$ linear unabhängig sind. Dann*

$$\overline{\mathcal{V}_{g_1}} \cap \overline{\mathcal{V}_{g_2}} = \{0\}.$$

BEWEIS. Durch eventuelles Skalieren können ohne Einschränkung annehmen, g_1, g_2 seien stochastisch. Dann auch $\bar{\mu}g_1, \bar{\mu}g_2$ stochastisch. Wir erhalten zunächst

$$\overline{\mathcal{V}_{g_1}} \stackrel{S.4.67}{\subset} \mathcal{V}_{Q_{g_1}} \stackrel{L.4.80}{=} \mathcal{V}_{Q_{\bar{\mu}g_1}}, \quad (4.50)$$

$$\overline{\mathcal{V}_{g_2}} \stackrel{S.4.67}{\subset} \mathcal{V}_{Q_{g_2}} \stackrel{L.4.80}{=} \mathcal{V}_{Q_{\bar{\mu}g_2}}. \quad (4.51)$$

Wegen der linearen Unabhängigkeit von $\bar{\mu}g_1, \bar{\mu}g_2$ findet man mit Hilfe von Lemma 4.14 und der Ergodizität der $\bar{\mu}g_1, \bar{\mu}g_2$ (Lemma 4.79) eine invariante Menge I mit $\bar{\mu}g_1(I) = 1 \neq 0 = \bar{\mu}g_2(I)$, d.h. genau $Q_{\bar{\mu}g_1} \neq Q_{\bar{\mu}g_2}$. Daher mit Satz 4.66

$$\mathcal{V}_{Q_{\bar{\mu}g_1}} \cap \mathcal{V}_{Q_{\bar{\mu}g_2}} = \{0\}$$

und daher aufgrund von (4.50) die Behauptung. \square

Wir kommen nun zum Höhepunkt dieses Abschnitts, indem wir Ergodizität für ES-Vektoren charakterisieren. Das sieht so aus:

SATZ 4.82. *Sei $g \in l_{SM}$ ein ES-Maßvektor. Dann ist g genau dann ergodisch, falls*

$$\dim(l_\mu \cap \overline{\mathcal{V}_g}) = 1. \quad (4.52)$$

BEWEIS. Wir können ohne Einschränkung von einem stochastischen g ausgehen. Ist g ergodisch, so auch $\bar{\mu}g$ (Lemma 4.79). Wegen $\bar{\mu}g \in \overline{\mathcal{V}_g}$ (Satz 4.67) folgt

$$\dim(l_\mu \cap \overline{\mathcal{V}_g}) \geq 1.$$

Wegen $\overline{\mathcal{V}_g} \subset \mathcal{V}_{Q_g}$ folgt dann mit Satz 4.75 die Hinrichtung. Ist umgekehrt g nicht ergodisch, so auch $\bar{\mu}g$ nicht (wiederum Lemma 4.79). Somit finden wir eine invariante Menge I , so dass $\bar{\mu}g(I) = p \in]0, 1[$. Mit Lemma 4.15 folgt dann, dass

$$0 \neq (\bar{\mu}g)^I, (\bar{\mu}g)^{\mathbb{G}I} \in \overline{\mathcal{V}_{\bar{\mu}g}} \subset \overline{\mathcal{V}_g}.$$

Wegen $(\bar{\mu}g)^I(I) = 1 \neq 0 = (\bar{\mu}g)^{\mathbb{G}I}$ sind $(\bar{\mu}g)^I, (\bar{\mu}g)^{\mathbb{G}I}$ jedoch linear unabhängig und wir haben den Satz bewiesen. \square

KOROLLAR 4.83. *Sei $g \in l_{SM}$ ein endlich-dimensionaler Maßvektor. Dann ist g genau dann ergodisch, falls*

$$\dim(l_\mu \cap \mathcal{V}_g) = 1.$$

BEWEIS. Folgt mit Satz 4.82, da endlich-dimensionale Vektoren ES-Vektoren sind (Satz 4.43) und weil $\overline{\mathcal{V}}_g = \mathcal{V}_g$, da \mathcal{V}_g endlich-dimensional ist. \square

Für ein letztes Korollar, ebenfalls die endlich-dimensionalen Vektoren betreffend, wollen wir an die Schreibweise

$$g_{\bar{a}} = \frac{\|g\|_{TV}}{g(\bar{a})} \tau_{\bar{a}}g = \frac{g(\square)}{g(\bar{a})} \tau_{\bar{a}}g$$

für Maßvektoren $0 \neq g$ erinnern.

KOROLLAR 4.84. *Ist $0 \neq g \in l_{SM}$ ein endlich-dimensionaler, ergodischer Maßvektor. Dann gilt für alle $\bar{a} \in \Sigma^*$:*

$$\bar{\mu}\tau_{\bar{a}}g = \frac{g(\bar{a})}{g(\square)} \bar{\mu}g$$

bzw. mit der obigen Schreibweise

$$\bar{\mu}g_{\bar{a}} = \bar{\mu}g.$$

BEWEIS. Da die $g_{\bar{a}}$ endlich-dimensional sind, sind sie ebenfalls ES-Vektoren (Satz 4.43). Wegen $\bar{\mu}g_{\bar{a}} \in l_\mu \cap \mathcal{V}_g$ folgt mit Korollar 4.83

$$\bar{\mu}\tau_{\bar{a}}g = \alpha \bar{\mu}g$$

für ein $\alpha \in \mathbb{R}^+$ und weiter wegen $\bar{\mu}g(\square) = g(\square)$ aus

$$\bar{\mu}\tau_{\bar{a}}g(\square) = \tau_{\bar{a}}g(\square) = g(\bar{a})$$

die Behauptung. \square

Der Vollständigkeit halber erwähnen wir noch ein Resultat, das mit dem Ergodensatz von Birkhoff erhalten werden kann. In Kombination mit der Einsicht, dass Prognosevektoren von ES-Vektoren wiederum ES-Vektoren sind (siehe dazu Bemerkung 4.49), erhalten wir ein zu Korollar 4.84 verallgemeinertes Ergebnis.

SATZ 4.85. *Sei $g \in l_{SM}$ ein ES-Maßvektor. Dann sind die folgenden Aussagen äquivalent:*

- (i) g ist ergodisch.
- (ii) $\forall(\bar{a}, \bar{b}) \in \Sigma^* \times \Sigma^* : \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=0}^{n-1} \mu^t \tau_{\bar{a}}g(\bar{b}) = \frac{g(\bar{a})}{\|g\|_{TV}} \bar{g}(\bar{b})$.
- (iii) $\forall \bar{a} \in \Sigma^* : \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=0}^{n-1} \mu^t \tau_{\bar{a}}g(\bar{a}) = \frac{g(\bar{a})}{\|g\|_{TV}} \bar{g}(\bar{a})$.

BEWEIS. Das ist dieselbe Aussage wie die in Satz 4.85 getroffene, mit dem einzigen Unterschied, dass stationär durch ES ersetzt worden ist. Der Beweis verläuft analog wie dort unter Einsatz von Lemma 4.61. Einziger Unterschied ist nun, dass die Aussage von Lemma 4.61 nunmehr nur noch unter Zuhilfenahme des Ergodensatzes von Birkhoff 4.59 erhalten wird. \square

Die ergodische Zerlegung

In diesem Kapitel wollen wir zeigen, dass sich jeder endlich-dimensionale stochastische Vektor als Konvexkombination (im Stochastikerjargon auch endliche Mixtur genannt) von ergodischen, stochastischen Vektoren beschreiben lässt. Das in der Ergodentheorie vorzufindende Analogon ist die Tatsache, dass sich jeder stochastische Prozess, der asymptotisch stationär im Mittel ist, als (nicht zwingend endliche) Mixtur von ergodischen AMS-Prozessen auffassen lässt, dort ebenfalls “ergodische Zerlegung” genannt ([Gra01], S.160 ff.). In diesem Sinne verschärfen wir dieses Ergebnis für endlich-dimensionale Prozesse. Es bleibt anzumerken, dass die Umkehrung (d.h. ist g eine endliche Mixtur ergodischer stochastischer Vektoren, so folgt, dass $\dim g < \infty$) mit der Existenz eines unendlich-dimensionalen ergodischen stochastischen Vektors steht und fällt - eine noch völlig offene Frage.

5.1. Stationäre Vektoren

Wir behandeln zunächst den Fall endlich-dimensionaler, stationärer, stochastischer Vektoren. Hier ist das Hauptergebnis der folgende

SATZ 5.1 (Ergodische Zerlegung stationärer Vektoren). *Sei g ein endlich-dimensionaler, stationärer, stochastischer Vektor. Sei $E_g := l_\mu \cap \mathcal{V}_g$ und $n := \dim(E_g)$. Dann lässt sich g schreiben als Konvexkombination n vieler ergodischer, stationärer, stochastischer Vektoren, also*

$$g = \sum_{i=1}^n \alpha_i g_i$$

mit $0 < \alpha_i \leq 1$, $\sum \alpha_i = 1$ und ergodischen, stationären, stochastischen Vektoren g_i . Mehr noch, es gilt

$$\mathcal{V}_g = \bigoplus_{i=1}^n \mathcal{V}_{g_i},$$

also \mathcal{V}_g eine direkte Summe von Prognoseräumen von ergodischen, stationären Vektoren.

Zum Beweis dieses Satzes erarbeiten wir uns mehrere Lemmata, deren Zusammensetzung den Satz ergibt. Das erste davon präsentiert unmittelbar die der Zerlegung zugrundeliegende zentrale Idee: Die Dimension des Eigenraums des Evolutionsoperators zum Eigenwert 1 verringert sich, wenn man an den Rand des von den Prognosevektoren erzeugten Kegels K_g (für die Definition siehe Satz 3.29) gelangt. Dabei nutzt man aus, dass stationäre

Vektoren immer im Inneren ihres Prognosekegels liegen (siehe Satz 4.17). Wir schreiben im weiteren Verlauf K_g für diesen Kegel und $E_g = l_\mu \cap \mathcal{V}_g$ für den Eigenraum des Evolutionsoperators zum Eigenwert 1 eingeschränkt auf den Prognoseraum \mathcal{V}_g von g . Wir nennen dies auch den **stationären Raum** von g . Begriffe der Topologie wie Inneres, Rand etc. beziehen sich meist auf die durch $\|\cdot\|_{TV}$ induzierte Teilraumtopologie des Prognoseraums \mathcal{V}_g , wobei ein eventueller Wechsel des Vektors g beachtet werden muss.

LEMMA 5.2. *Sei g ein endlich-dimensionaler, stationärer, stochastischer Vektor. Sei $\tilde{g} \in \delta K_g \cap E_g$ ein stationärer, stochastischer Vektor, der im Rand von K_g liegt. Dann gilt*

$$\dim(E_{\tilde{g}}) < \dim(E_g).$$

BEWEIS. Wir zeigen, dass

$$g \notin \mathcal{V}_{\tilde{g}},$$

was zum Behaupteten führt, da dann (siehe auch Bemerkung 3.19) $\mathcal{V}_{\tilde{g}} \subsetneq \mathcal{V}_g$ und somit wegen der Stationarität von g , d.h. $g \in l_\mu$ dann auch

$$E_{\tilde{g}} = \mathcal{V}_{\tilde{g}} \cap l_\mu \subsetneq \mathcal{V}_g \cap l_\mu = E_g$$

folgt. Wir nehmen das Gegenteil, also $g \in \mathcal{V}_{\tilde{g}}$ an. Nach Satz 4.17 gilt $\tilde{g} \in \text{Int}(K_{\tilde{g}})$ (beachte, hier Int bzgl. der Topologie von $\mathcal{V}_{\tilde{g}}$). Wir finden damit ein $\epsilon \in \mathbb{R}^+$, so dass (bzgl. der Topologie von $\mathcal{V}_{\tilde{g}}$)

$$U_\epsilon(\tilde{g}) \subset K_{\tilde{g}}.$$

Wegen $g \in \mathcal{V}_{\tilde{g}}$ und der Volldimensionalität von $K_{\tilde{g}}$ gilt für $d := \|\tilde{g} - g\|_{TV}$

$$g^* := \tilde{g} - \frac{\epsilon}{2d}(g - \tilde{g}) \in U_\epsilon(\tilde{g}).$$

Wegen $K_{\tilde{g}} \subset K_g$ (Korollar 3.30) folgt nun, dass $g^* \in K_g$. Setzt man nun in Lemma 4.18 $V = \mathcal{V}_g$, $G = K_g$, $v = g^*$, $u = g$ und $\alpha = \frac{2d}{\epsilon+2d} \in]0, 1[$ ein, so folgt $\tilde{g} \in \text{Int}(K_g)$ (nun natürlich bzgl. der Topologie von \mathcal{V}_g), ein Widerspruch! \square

Für das nächste Lemma erinnern wir an Proposition 4.35 und daran, dass endlich-dimensionale Vektoren immer ES-Vektoren sind (Satz 4.43). Den stationären Grenzvektor eines endlich-dimensionalen Vektors g bezeichnen wir dann gemäß der in Proposition 4.35 benutzten Notation mit $\bar{\mu}g$.

LEMMA 5.3. *Sei g ein endlich-dimensionaler, stochastischer Vektor. Dann gibt es einen ergodischen, stationären, stochastischen Vektor g^* mit*

$$g^* \in \mathcal{V}_g.$$

BEWEIS. Das zeigen wir mit vollständiger Induktion über $n := \dim(E_g)$. Für $n = 1$ ist der Fall klar: Gemäß Korollar 4.83 ist g selbst ergodisch. Der stationäre Grenzvektor $\bar{\mu}g$ ist dann ebenfalls ergodisch und erfüllt die Anforderungen. Sei nun $n > 1$. Wir bemerken zuerst, dass es reicht, die Behauptung für stationäres g zu zeigen, da man wegen Bemerkung 3.19 ohne Einschränkung zum stationären Grenzvektor $\bar{g} \in \mathcal{V}_g$ von g übergehen kann. Wir wählen zunächst einen zweiten linear unabhängigen stationären Vektor $g^* \in \mathcal{V}_g$, der wegen der Volldimensionalität von K_g als Element von K_g

gewählt werden kann. Mit Bemerkung 3.31 ohne Einschränkung $g^*(\square) = 1$ und damit g^* ein stochastischer Vektor nach Bemerkung 3.32. Wir setzen nun

$$g_t := g + t(g^* - g)$$

und weiter

$$T := \sup\{t \in \mathbb{R}^+ \mid g_t \in K_g\}.$$

Bemerke, dass $T < \infty$, denn setzt man ins nachfolgende Sublemma 5.4 nun $V = \mathcal{V}_g$, $K = K_g$, $g = g$ und $f = g^* - g$, so führte die Annahme $T = \infty$ auf $g_2 - g \in K_g$. Wegen $(g^* - g)(\square) = 0$ folgt mit Bemerkung 3.31 aber $g_2 - g \notin K_g$. Weiter wegen $g_1 = g^* \in K_g$ natürlich $T \geq 1$. Wegen $g_T(\square) = 1$ folgt mit Bemerkung 3.32, dass g_T ein stochastischer Vektor ist. g_T ist als Linearkombination stationärer Vektoren natürlich ebenfalls stationär und nach Definition von T ein Element von δK_g , des Rands von K_g . Mit Lemma 5.2 folgt $\dim E_{g_T} < \dim E_g$ und damit die Behauptung aufgrund der Induktionsannahme. \square

Wir holen nun noch das soeben verwendete Sublemma nach:

SUBLEMMA 5.4. *Sei V ein normierter \mathbb{R} -Vektorraum und $K \subset V$ ein abgeschlossener konvexer Kegel. Seien $g \in K$ und $f \in V$ so, dass $\forall t \in \mathbb{R}^+ : g + tf \in K$. Dann auch $f \in K$.*

BEWEIS. Wegen $g + nf \in K, n \in \mathbb{N}$ aufgrund der Kegeleigenschaften auch

$$\forall n \in \mathbb{N} : \quad \frac{1}{n}g + f = \frac{1}{n}(g + nf) \in K.$$

Die Behauptung folgt daher aus der Abgeschlossenheit von K . \square

BEMERKUNG 5.5. *Wir wollen an dieser Stelle bemerken, dass die im soeben aufgeführten Beweis von Lemma 5.3 verwendete explizite Konstruktion eines Vektors g_T im Rand des Prognosekegels nicht unbedingt nötig ist. Mit einfachen topologischen Argumenten fände man ebenso einen stationären, stochastischen Vektor im Rand des Prognosekegels. Wir haben uns jedoch für diese konstruktivere Variante des Beweises entschieden, da so ersichtlich wird, dass durch Subtrahieren von stationären Komponenten die "ergodische Dimension" iterativ verringert werden kann.*

Mit Hilfe der bereitgestellten Lemmata sind wir nun in der Lage, zu beweisen, dass jedes stationäre g eine Konvexkombination von ergodischen, stationären Vektoren ist. Offen bleibt zunächst nur noch, aus wievielen solcher Vektoren die Mixtur besteht.

LEMMA 5.6. *Sei g ein endlich-dimensionaler, stationärer, stochastischer Vektor. Dann lässt sich g darstellen als*

$$g = \sum_{i=1}^n \alpha_i g_i$$

Konvexkombination von ergodischen, stationären, stochastischen Vektoren g_i .

BEWEIS. Wir führen den Beweis wiederum mit vollständiger Induktion über $n := \dim(E_g)$. Für $n = 1$ ist der Fall abermals klar nach Korollar 4.83: g ist selber ergodisch. Sei nun $n > 1$. Mit Lemma 5.3 finden wir einen ergodischen, stationären Vektor $g^* \in \mathcal{V}_g$. Für $t \in \mathbb{R}^+$ setzen wir nun

$$g_t := g - t(g^* - g)$$

und weiter

$$T := \sup\{t \in \mathbb{R}^+ \mid g_t \in K_g\}.$$

Mit derselben Argumentation wie in Lemma 5.3 folgt $T < \infty$. Weiter ist dann $g_T \in \delta K_g$ und wegen Lemma 5.2 gilt $\dim(E_{g_T}) < \dim(E_g)$. Nach Induktionsannahme daher

$$g_T = \sum_{i=1}^m \alpha_i g_i$$

mit $0 \leq \alpha_i \leq 1$, $\sum_i \alpha_i = 1$ und g_i ergodisch. Es folgt

$$g = \frac{T}{T+1}g^* + \frac{1}{T+1}g_T = \frac{T}{T+1}g^* + \sum_{i=1}^m \frac{\alpha_i}{T+1}g_i, \quad (5.1)$$

und damit die Behauptung. \square

Wir kommen, wie versprochen, zur Bestimmung der Anzahl der ergodischen Mixturkomponenten:

LEMMA 5.7. *Sei g ein stationärer, stochastischer Vektor, der sich als Konvexkombination*

$$g = \sum_{i=1}^n \alpha_i g_i$$

ergodischer, stationärer, stochastischer Vektoren g_i schreiben lässt, wobei $0 < \alpha_i \leq 1$. Dann

$$\dim(E_g) = n.$$

BEWEIS. Wir zeigen zunächst $\dim(E_g) \geq n$. Dazu bemerken wir, dass es reicht, $g_i \in \mathcal{V}_g$ zu zeigen. Da die g_i gemäß Korollar 4.72 linear unabhängig sind und als stationäre Vektoren in E_g liegen, folgt dann die Behauptung. Für ein g_i wählen wir nun gemäß Lemma 4.71 eine invariante Menge I_i mit $g_i(I_i) = 1$ und $g_j(I_i) = 0$ für $j \neq i$. Offensichtlich $g(I_i) = \alpha_i$. Gemäß Korollar 4.16 gilt nun

$$g^{I_i} \in \mathcal{V}_g.$$

Aus

$$\begin{aligned} g^{I_i}(\bar{a}) &= \frac{1}{g(I_i)}g(I_i \cap \{\bar{a}\}) \\ &= \frac{1}{\alpha_i} \sum_{j=1}^n \alpha_j g_j(I_i \cap \{\bar{a}\}) \\ &= \frac{1}{\alpha_i} \alpha_i g_i(I_i \cap \{\bar{a}\}) = g_i(\bar{a}) \end{aligned} \quad (5.2)$$

folgt $g_i = g^{I_i}$ und damit die Behauptung.

Für den Teil $\dim(E_g) \leq n$ nehmen wir das Gegenteil, also $\dim(E_g) > n$ an. Dann gibt es einen stationären Vektor $g^* \in E_g$, der linear unabhängig von den g_i ist. Gemäß Lemma 5.6 lässt sich auch $g^* = \sum_j \beta_j g_j$, $0 < \beta_j \leq 1$ schreiben als Konvexkombination ergodischer, stationärer g_j , wobei natürlich mindestens ein g_j von den g_i verschieden sein muss. Sei \tilde{g} ein solches g_j . Mit Hilfe von Lemma 4.71 finden wir wieder eine invariante Menge I , so dass $\tilde{g}(I) = 1$ und $\forall i \in \{1, \dots, n\} : g_i(I) = 0$. Damit auch $g(I) = 0$. Der Widerspruch ergibt sich damit aus Satz 4.19. \square

Beweis von Satz 5.1: Der erste Teil der Aussage ist jetzt einfach. Wir schreiben g als Konvexkombination von ergodischen, stationären, stochastischen Vektoren g_i gemäß Lemma 5.6 und wissen mit Lemma 5.7, dass es an der gewünschten Zahl viele sind. Für den zweiten Teil der Aussage bemerken wir zunächst, dass jeweils

$$\mathcal{V}_{g_i} \cap \mathcal{V}_{g_j} = \{0\}$$

gemäß Lemma 4.81. Wegen $g_i \in \mathcal{V}_g$ (siehe Beweis von Lemma 5.6) folgt deshalb (siehe Bemerkung 3.19)

$$\bigoplus_{i=1}^n \mathcal{V}_{g_i} \subset \mathcal{V}_g.$$

Wegen $g = \sum_{i=1}^n \alpha_i g_i \in \bigoplus_{i=1}^n \mathcal{V}_{g_i}$ folgt gemäß der Definition eines Prognoserraums natürlich auch

$$\mathcal{V}_g \subset \bigoplus_{i=1}^n \mathcal{V}_{g_i},$$

womit der Satz vollständig bewiesen wäre. \square

Wir zitieren noch einige nützliche Korollare für spätere Arbeiten.

KOROLLAR 5.8. *Sei $g = \sum_{i=1}^n \alpha_i g_i$ eine endliche Mischung endlich-dimensionaler, ergodischer, stationärer Vektoren g_i . Dann*

$$\forall i \in \{1, \dots, n\} : g_i \in \mathcal{V}_g.$$

BEWEIS. Das folgt aus dem Beweis zu Lemma 5.7. \square

KOROLLAR 5.9. *Sei g ein endlich-dimensionaler, stationärer, stochastischer Vektor und sei $g = \sum_{i=1}^n \alpha_i g_i$ seine Zerlegung in stationäre, ergodische Mischturkomponenten. Dann gilt*

$$E_g = \text{span}\{g_1, \dots, g_n\},$$

mehr noch, die g_i bilden eine Basis von E_g .

BEWEIS. Folgt direkt aus Lemma 5.7 und der linearen Unabhängigkeit der g_i . \square

Das nun folgende Korollar liefert ein geometrisches Bild des "stationären Prognosekegels". Er ist immer polyedrisch und die ergodischen Komponenten des stationären Grundvektors können mit seinen Extremalstrahlen identifiziert

werden.

KOROLLAR 5.10. *Sei g ein endlich-dimensionaler, stationärer, stochastischer Vektor und $g = \sum_{i=1}^n \alpha_i g_i$ seine Zerlegung in stationäre, ergodische Mirkurkomponenten. Dann*

$$K_g \cap E_g = \text{cone}\{g_1, \dots, g_n\}.$$

Der "stationäre Prognosekegel" ist demnach polyedrisch und wird von n vielen ergodischen, stationären, stochastischen Vektoren erzeugt.

BEWEIS. Sei $g^* \in K_g \cap E_g$. Dann ist g^* nach Satz 5.1 eine Konvexkombination von endlich vielen, ergodischen, stationären, stochastischen Vektoren

$$g^* = \sum_{j=1}^m \beta_j g_j$$

mit $0 < \beta_j \leq 1$. Wegen $E_{g^*} \subset E_g$ folgt mit Korollar 5.8, dass alle g_j in E_g liegen. Da voneinander verschiedene endlich-dimensionale, ergodische, stationäre, stochastische Vektoren linear unabhängig sind (siehe Korollar 4.72) lassen sich die g_j jeweils mit gewissen g_i identifizieren. \square

Schon einmal spielten polyedrische Kegel in diesem Text eine Rolle. Ließen die Prognoseoperatoren einen polyedrischen Kegel invariant, so handelte es sich um ein Hidden Markov Modell. Daher erhalten wir noch

KOROLLAR 5.11. *Ist g ein endlich-dimensionaler, stochastischer Vektor, so dass*

$$\mathcal{V}_g \subset l_\mu,$$

dann ist g ein Hidden Markov Vektor.

BEWEIS. Das folgt mit Satz 3.44, da in diesem Fall $K_g = K_g \cap l_\mu$. K_g ist demnach polyedrisch aufgrund Korollar 5.10 und erfüllt damit die Anforderungen von Satz 3.44. \square

5.2. Der allgemeine Fall

Wir profitieren nun von den obigen Arbeiten, um das Ergebnis auf alle endlich-dimensionalen, stochastischen Vektoren zu verallgemeinern. Für diesen Zweck schreiben wir wie bisher

$$E_g := l_\mu \cap \mathcal{V}_g$$

für den stationären Raum von g und nun zusätzlich

$$N_g := \ker(\bar{\mu}) \cap \mathcal{V}_g$$

für den Untervektorraum der Vektoren aus \mathcal{V}_g , deren Projektion auf den stationären Raum gerade Null ergibt. Für den allgemeinen Fall sieht die ergodische Zerlegung dann so aus:

SATZ 5.12 (Ergodische Zerlegung). *Sei g ein endlich-dimensionaler, stochastischer Vektor. Sei $n := \dim(E_g)$ die Dimension des stationären Raums $E_g = \mathcal{V}_g \cap l_\mu$ von g . Dann gibt es n viele ergodische, endlich-dimensionale, stochastische Vektoren $(g_i)_{i=1, \dots, n}$, so dass*

$$g = \sum_{i=1}^n \alpha_i g_i,$$

wobei $0 < \alpha_i \leq 1$, $\sum_{i=1}^n \alpha_i = 1$ und darüberhinaus

$$\mathcal{V}_g = \mathcal{V}_{\bar{\mu}g} \oplus \mathcal{O}_g = \mathcal{V}_{\bar{\mu}g_1} \oplus \dots \oplus \mathcal{V}_{\bar{\mu}g_n} \oplus \mathcal{O}_g, \quad (5.3)$$

wobei $\mathcal{O}_g \cong N_g / \oplus_{i=1}^n N_{g_i}$.

Zunächst ein paar Vorbereitungen.

LEMMA 5.13. *Sei g ein endlich-dimensionaler, stochastischer Vektor. Sei $E_g := \mathcal{V}_g \cap l_\mu$ der stationäre Raum von g . Dann hat E_g eine Basis von endlich-dimensionalen, ergodischen, stationären, stochastischen Vektoren.*

BEWEIS. Wir betrachten zunächst das folgende

SUBLEMMA 5.14. *Sei V ein endlich-dimensionaler \mathbb{R} -Vektorraum, $F : V \rightarrow V$ ein Endomorphismus und $K \subset V$ ein echter Kegel, so dass*

$$F(K) \subset K.$$

Bezeichnet weiter $\rho(F)$ den Spektralradius von F , $\deg \lambda$ die Größe des größten Jordan-Blocks eines Eigenwerts λ und $V_{exp}(\lambda)$ die Anzahl solcher größten Blocks. Dann enthält K mindestens $V_{exp}(\rho(F))$ viele Eigenvektoren zum Eigenwert $\rho(F)$.

BEWEIS. Siehe [Sch81], S. 265, Korollar 5.3. Dort wird die Aussage für den Fall eines \mathbb{C} -Vektorraums bewiesen. Der Fall eines \mathbb{R} -Vektorraums ist jedoch ein offensichtliches Korollar. \square

Beweis von Lemma 5.13: Wir setzen nun in Sublemma 5.14 $V := \mathcal{V}_g$, $F = \mu$ und $K = K_g$, den Prognosekegel K_g von g ein. Da der Prognosekegel K_g echt ist und μ ihn invariant lässt (siehe Satz 3.29), liegen also mindestens $V_{exp}(\rho(\mu))$ viele Eigenvektoren zum Eigenwert $\rho(\mu)$ in K_g . Diese können durch eventuelles Skalieren als stochastische Vektoren gewählt werden (siehe Bemerkung 3.32). Da μ stabil ist (siehe Satz 3.8) und den Eigenwert 1 besitzt (siehe Korollar 3.22), folgt $\rho(\mu) = 1$ (siehe das Eigenwertkriterium für stabile Operatoren (10.4) aus Satz 10.9) und außerdem $\deg \rho(\mu) = 1$ (siehe hierzu Gleichung (10.5) aus Satz 10.9). Daher $V_{exp}(1) = \dim E_g$ und wir haben an dieser Stelle eine Basis von endlich-dimensionalen, stationären, stochastischen Vektoren (g_i) von E_g erhalten. Schreiben wir nun

$$g_i = \sum_{j=1}^{n_i} \alpha_{ij} g_j$$

als Konvexkombination von ergodischen g_j gemäß Satz 5.1, so sehen wir wegen $g_j \in E_g$ (Korollar 5.8) ein, dass wir unter den g_j ebenfalls eine Basis von

E_g finden und die Behauptung folgt. \square

Darauf aufbauend erhalten wir nun, dass wir den stationären Raum eines endlich-dimensionalen Vektors g mit dem seines stationären Grenzvektors identifizieren werden können.

LEMMA 5.15. *Sei g ein endlich-dimensionaler, stochastischer Vektor und $\bar{g} := \bar{\mu}g$ (siehe Proposition 4.35) der stationäre Grenzvektor. Dann*

$$E_{\bar{g}} = E_g.$$

BEWEIS. $E_{\bar{g}} \subset E_g$ ist klar wegen $\mathcal{V}_{\bar{g}} \subset \mathcal{V}_g$ (Bem. 3.19). Für die umgekehrte Richtung nehmen wir die Existenz eines $0 \neq g^* \in E_g \setminus E_{\bar{g}}$ an. Dieses kann wegen Lemma 5.13 als endlich-dimensionaler, ergodischer, stationärer, stochastischer Vektor gewählt werden [Wir wählen für beide E_g und $E_{\bar{g}}$ eine Basis ergodischer Vektoren und können die Basisvektoren von $E_{\bar{g}}$ mit gewissen der Basisvektoren von E_g identifizieren. Dabei bleibt mindestens ein Basisvektor von E_g übrig]. Wir können daher mit Hilfe von Lemma 4.71 eine invariante Menge I mit

$$g^*(I) = 1 \text{ und } \bar{g}(I) = 0$$

wählen. Wegen $g^* \in \mathcal{V}_g$ gibt es somit ein Wort \bar{a} (wähle eine Basis von Prognosevektoren und stelle g^* darauf dar), so dass

$$\tau_{\bar{a}}g(I) > 0.$$

Deshalb auch

$$\mu^{|\bar{a}|}g(I) = \sum_{\bar{b} \in \Sigma^{|\bar{a}|}} \tau_{\bar{b}}g(I) \geq \tau_{\bar{a}}g(I) > 0.$$

Wegen $\bar{\mu}\mu^{|\bar{a}|}g = \bar{\mu}g = \bar{g}$ und Bemerkung 4.37 (Übereinstimmung von $\mu^{|\bar{a}|}g$ und \bar{g} auf invarianten Mengen) ergibt sich ein Widerspruch. \square

Beweis von Satz 5.12: Sei $\bar{g} := \bar{\mu}g$ mit $\bar{\mu}$ wie in Proposition 4.35, also \bar{g} der stationäre Grenzvektor von g . Gemäß Satz 5.1 ist

$$\bar{g} =: \sum_{i=1}^n \alpha_i \bar{g}_i$$

eine endliche Mixtur ergodischer, stationärer, stochastischer Vektoren \bar{g}_i und wegen Korollar 5.9 und Lemma 5.15

$$E_g = E_{\bar{g}} = \text{span}\{\bar{g}_i \mid i = 1, \dots, n\}.$$

Mit Lemma 4.81 folgt $\mathcal{V}_{Q_{\bar{g}_i}} \cap \mathcal{V}_{Q_{\bar{g}_j}} = \{0\}$ für $i \neq j$. Wir behaupten nun

$$\mathcal{V}_g \subset \bigoplus_{i=1}^n \mathcal{V}_{Q_{\bar{g}_i}}.$$

Existierte nämlich ein $h \in \mathcal{V}_g \setminus \bigoplus_{i=1}^n \mathcal{V}_{Q_{\bar{g}_i}}$, so erhalten wir wegen $\bar{\mu}h \notin \text{span}\{g_i\} = E_g$ einen Widerspruch. Daher $g \in \bigoplus_{i=1}^n \mathcal{V}_{Q_{\bar{g}_i}}$ und wir schreiben nun

$$g = \sum_{i=1}^n \beta_i g_i$$

mit $g_i \in \mathcal{V}_{Q_{\bar{g}_i}}$. Wir sehen wegen $\overline{\mathcal{V}_{g_i}} \subset \mathcal{V}_{Q_{\bar{g}_i}}$ zunächst ein, dass die g_i endlich-dimensional sein müssen. Wegen

$$\sum_{i=1}^n \alpha_i \bar{g}_i = \bar{\mu}g = \sum_{i=1}^n \beta_i \bar{\mu}g_i = \sum_{i=1}^n \beta_i \bar{g}_i$$

und der linearen Unabhängigkeit der \bar{g}_i folgt $\alpha_i = \beta_i$. Damit haben wir den ersten Teil der Behauptung bewiesen.

Für den zweiten Teil der Behauptung beachten wir, dass die zweite Gleichheit in (5.3) direkt aus Satz 5.1 folgt. Für die erste Gleichheit erinnern wir an die Schreibweise schreiben wir

$$\tau_{\bar{a}}g = \bar{\mu}\tau_{\bar{a}}g + (\tau_{\bar{a}}g - \bar{\mu}\tau_{\bar{a}}g).$$

Es gilt dann

$$\begin{aligned} \bar{\mu}\tau_{\bar{a}}g &= \bar{\mu}\tau_{\bar{a}}\left(\sum_{i=1}^n \alpha_i g_i\right) = \bar{\mu}\left(\sum_{i=1}^n \alpha_i \tau_{\bar{a}}g_i\right) \\ &= \sum_{i=1}^n \alpha_i \bar{\mu}\tau_{\bar{a}}g_i \stackrel{K.4.84}{=} \sum_{i=1}^n \alpha_i g_i(\bar{a})\bar{g}_i \in \mathcal{V}_{\bar{\mu}g} \end{aligned}$$

und

$$\bar{\mu}(\tau_{\bar{a}}g - \bar{\mu}\tau_{\bar{a}}g) = 0.$$

Damit folgt wegen $\mathcal{V}_{\bar{\mu}g} \subset \mathcal{V}_g$ und $N_g \subset \mathcal{V}_g$ insgesamt zunächst

$$\mathcal{V}_g = \mathcal{V}_{\bar{\mu}g} + N_g.$$

Hieraus folgt wegen

$$N_g \cap \mathcal{V}_{\bar{\mu}g} = N_{\bar{\mu}g} = \bigoplus_{i=1}^n N_{\bar{g}_i}$$

die Behauptung aus der bekannten Tatsache $W_1 + W_2 \cong W_1 \oplus W_2 / (W_1 \cap W_2)$ für zwei Untervektorräume W_1, W_2 eines Vektorraums V . \square

5.3. Klassifikation von endlich-dimensionalen Prozessen

Wir beschließen dieses Kapitel mit einem kurzen Abschnitt eher provisorischer Natur, der den Ausgangspunkt einer Theorie der Klassifikation endlich-dimensionaler Prozesse bildet. Eine solche Theorie mag bei der Mathematisierung eher intuitiver Begriffe wie "hochgradig instationär" behilflich sein. Wir liefern erste Vorschläge für solche klassifizierenden Definitionen.

Dafür führen wir zunächst jedoch inzwischen offensichtlich gewordene Kennzahlen für endlich-dimensionale Prozesse ein. Wir starten mit einer sich aus den obigen Sätzen aufdrängenden Begriffsbildung.

DEFINITION 5.16. Sei $g \in l_{SM}$ und $E_g := l_{\mu} \cap \overline{\mathcal{V}_g}$. Dann nennen wir

$$\text{ergdim } g := \dim E_g$$

die **ergodische Dimension** von g .

Notwendigerweise haben endlich-dimensionale Vektoren eine endliche ergodische Dimension. Die ergodische Zerlegung führt auf weitere einleuchtende Kennzahlen, die in Analogie zu Begriffen für Markov-Ketten entwickelt werden.

DEFINITION 5.17. Sei $g \in l_{SM}$ ein endlich-dimensionaler, stochastischer Vektor. Dann nennen wir

$$\text{transdim } g := \dim \mathcal{O}_g$$

die **Flüchtigkeitsdimension** von g und

$$\text{statdim } g := \dim g - \text{transdim } g$$

die **stationäre Dimension** von g .

Diese Werte sind nicht aussagekräftig in dem Sinne, dass pro ergodischer Komponente eine geringe stationäre und geringe Flüchtigkeitsdimension vorliegen kann, der gesamte Prozess aufgrund einer hohen Anzahl ergodischer Komponenten jedoch entsprechend große Werte erzielt. Deshalb schlagen wir vor zu mitteln:

DEFINITION 5.18. Sei $g \in l_{SM}$ ein endlich-dimensionaler, stochastischer Vektor. Wir nennen

$$\text{avgtransdim } g := \frac{\text{transdim } g}{\text{ergdim } g}$$

die **mittlere Flüchtigkeitsdimension** und entsprechend

$$\text{avgstatdim } g := \frac{\text{statdim } g}{\text{ergdim } g}$$

die **mittlere stationäre Dimension** von g .

Die Untersuchung von sogenannten Mischraten in der Theorie der Markov-Ketten ist dem Einsatz von "Markov-Chain-Monte-Carlo- (MCMC)" Techniken in Simulationen zu verdanken. Schnell mischende Markov-Ketten konvergieren zügig gegen ihre stationäre Grenzverteilung und die Verwendung solcher Ketten in MCMC-Simulationen erzielt schneller bessere Ergebnisse. Eine Reihe von Arbeiten und Büchern [Alo86], [JS89], [DS91], [Dia96], [Chu97], [MP03], [AF06] (das ist beileibe keine vollständige Liste) beschäftigte sich mit diesem Thema. In Analogie zu den dort entwickelten Begriffen definieren wir:

DEFINITION 5.19. Sei $g \in l_{SM}$ ein endlich-dimensionaler, stochastischer Vektor. Sei $\mu : \mathcal{V}_g \rightarrow \mathcal{V}_g$ der Evolutionsoperator auf dem Prognoseraum und sei

$$\sigma(\mu) := \{\lambda_1 = 1, \lambda_2, \dots, \lambda_m\}$$

das Spektrum von μ , so dass $|\lambda_1| \geq |\lambda_2| \geq \dots \geq |\lambda_m|$. Dann nennen wir

$$\rho(g) := 1 - |\lambda_2| \in [0, 1]$$

die **Spektrallücke** von g .

Es ist offensichtlich, dass die Geschwindigkeit der Konvergenz der Folge der $g_n := \frac{1}{n} \sum_{i=0}^{n-1} \mu^i g$ gegen den stationären Grenzvektor $\bar{\mu}g$ in einem Zusammenhang mit der Spektrallücke $\rho(g)$ steht. Je größer $\rho(g)$, desto schneller konvergiert die Folge der Maße.

Zum Abschluss dieses Kapitels machen wir einen Vorschlag zur Charakterisierung “hochgradig instationärer” endlich-dimensionaler, stochastischer Vektoren. Dabei hatten wir Prozesse im Auge, die bislang gerne durch “Profile HMMs” [DEKM98b] modelliert worden waren. “Profile HMMs” sind zunächst einmal ergodisch. Die meisten versteckten Zustände müssen unmittelbar nach Betreten wieder verlassen werden. Bringt man diese Erkenntnisse in einen Zusammenhang mit den obigen Begriffen, so schlagen wir Folgendes vor.

DEFINITION 5.20. Sei $g \in l_{SM}$ ein ergodischer, stochastischer Vektor und seien $\epsilon, \delta \in \mathbb{R}^+$. Wir nennen g dann **(ϵ, δ) -instationär**, falls

$$\frac{\text{statdim } g}{\text{dim } g} = \frac{1 - \text{transdim } g}{\text{dim } g} \leq \epsilon \iff \text{transdim } g \geq 1 - \epsilon \text{dim } g$$

und

$$\rho(g) \geq 1 - \delta.$$

“Hochgradig” instationäre Vektoren seien dann (ϵ, δ) -instationäre Vektoren mit hinreichend kleinen $\epsilon, \delta \in \mathbb{R}^+$.

Entropieraten

Dieses Kapitel enthält im Wesentlichen einen elementaren Beweis für die Existenz der Entropieraten für durch stochastische ES-Vektoren induzierte stochastische Prozesse. Wie schon vorher bemerkt (siehe Korollar 2.65 und Satz 4.41), induzieren stochastische ES-Vektoren stochastische Prozesse, die asymptotisch stationär im Mittel sind und umgekehrt. Wir erhalten somit einen Beweis für die Existenz der Entropieraten von AMS-Prozessen.

Diese Tatsache ist bereits bekannt und ist eine Konsequenz des berühmten und in vielzähligen aufwändigen Beweisvarianten, die meist alle Register der Maßtheorie ziehen, erhältlichen Satzes von “Shannon-McMillan-Breiman” [Sha48], [McM53],[Bre57],[Bre60],[Bil65], [Jac59], [Jac62],[GK80] (für Erläuterungen siehe Abschnitt 6.3 weiter unten). Für den Fall stationärer Prozesse existiert jedoch ein elementarer Beweis für die Existenz der Entropieraten. Somit können wir dieses Kapitel als Verallgemeinerung dieses Resultats auf AMS-Prozesse auffassen. Es bleibt zu erwähnen, dass unser Beweis sehr elementar über die Bühne geht; es werden ausschließlich Hilfsmittel der Analysis eingesetzt. Große Sätze der Maßtheorie werden nicht benötigt.

6.1. Einführung: Entropie

Wir bedienen uns im weiteren Verlauf oft bei [HK02]. Mit \log sei im Zweifel der natürliche Logarithmus gemeint. Wir definieren somit

DEFINITION 6.1. *Sei Σ eine endliche Menge und P ein Wahrscheinlichkeitsmaß auf dem Messraum $(\Sigma, \mathcal{P}(\Sigma))$. Dann ist*

$$H(P) := \sum_{x \in \Sigma} P(x) \log \frac{1}{P(x)}$$

*die **Entropie** des Wahrscheinlichkeitsraums $(\Sigma, \mathcal{P}(\Sigma), P)$ bzw. der Wahrscheinlichkeitsverteilung P .*

Die Entropie stellt ein Maß für das Wissen um den Ausgang eines Wahrscheinlichkeitsexperiments auf einem gegebenen Wahrscheinlichkeitsraum dar. Je größer die Entropie, desto “gleichverteilter” die Wahrscheinlichkeitsverteilung und damit um so kleiner unser Wissen um den Ausgang eines Experiments. Es gilt nämlich das elementare

LEMMA 6.2. *Sei*

$$\Delta^{n-1} := \{x = (x_1, \dots, x_n) \in \mathbb{R}^n \mid 0 \leq x_i \leq 1, \sum_i x_i = 1\}.$$

und

$$f : \begin{array}{ccc} \Delta^{n-1} & \longrightarrow & \mathbb{R} \\ x = (x_1, \dots, x_n) & \longmapsto & \sum_{i=1}^n x_i \log \frac{1}{x_i}. \end{array}$$

Dann gilt $0 \leq f \leq \log n$ und

$$f(x) = 0 \quad \Leftrightarrow \quad x = e_i = (0, \dots, 0, \frac{1}{i}, 0, \dots, 0) \text{ f\"ur ein } i = 1, \dots, n$$

und weiter

$$f(x) = \log n \quad \Leftrightarrow \quad x = (1/n, \dots, 1/n).$$

BEWEIS. Siehe z.B. [Gra90], [HK02]. Folgt auch aus Lemma 6.8 weiter unten. \square

Wendet man dieses Lemma auf die Berechnung von Entropien an, so sieht man, dass Gleichverteilungen die h"ochsten Entropien erzielen, wohingegen Wahrscheinlichkeitsr"äume von Experimenten, deren Ausgang klar ist, die Entropie 0 erzielen.

F"ur weitere Zusammenh"änge, Wesen und Urspr"unge der Informationstheorie siehe die B"ucher [Gra90], [Mac03], [HK02], [Bil65] oder den grundsteinlegenden Artikel von Shannon selbst [Sha48].

In diesem Sinne setzen wir nun f"ur einen stochastischen Vektor g

$$H^n(g) := \sum_{\bar{a} \in \Sigma^n} g(\bar{a}) \log \frac{1}{g(\bar{a})},$$

wobei wir $0 \log \infty = 0$ setzen, also $H^n(g)$ die Entropie der durch g auf Σ^n induzierten Wahrscheinlichkeitsverteilung des zu g assoziierten stochastischen Prozesses. Mit Hilfe von Lemma 6.2 folgt dann

$$0 \leq H^n(g) \leq \log \text{card}(\Sigma^n) = \log (\text{card}(\Sigma))^n = n \log \text{card}(\Sigma).$$

6.2. Entropieraten station"arer Vektoren

Studiert man stochastische Prozesse, interessiert man sich nun f"ur das asymptotische Verhalten der beschr"ankten Folge

$$0 \leq \left(\frac{1}{n} H^n(g)\right)_{n \in \mathbb{N}} \leq \log \text{card}(\Sigma)$$

und definiert daher

DEFINITION 6.3. *Sei g ein stochastischer Vektor. Dann ist*

$$\bar{H}(g) := \limsup_{n \rightarrow \infty} \frac{1}{n} H^n(g)$$

die **Entropierate** von g bzw. des zu g gegebenen stochastischen Prozesses.

Interessant ist nun die Frage, wann

$$\bar{H}(g) = \lim_{n \rightarrow \infty} \frac{1}{n} H^n(g),$$

für einen stochastischen Vektor g , die Entropierate von g also mit dem Limes der entsprechenden Folge übereinstimmt. Hier haben wir den folgenden

SATZ 6.4. *Sei g ein stationärer, stochastischer Vektor. Dann*

$$\bar{H}(g) = \lim_{n \rightarrow \infty} \frac{1}{n} H^n(g).$$

Wir wollen den Beweis dieses Satzes erarbeiten, um ein Gefühl für die Problematik des Stoffs zu vermitteln und auch, um später verwendete Begriffe vorzustellen. Dafür definieren wir

$$H_k^n(g) := \sum_{\bar{a} \in \Sigma^k} g(\bar{a}) H^n(g_{\bar{a}}) \stackrel{g_{\bar{a}} = 1/g(\bar{a})\tau_{\bar{a}}g}{=} \sum_{\bar{a} \in \Sigma^k} \sum_{\bar{b} \in \Sigma^n} g(\bar{a}\bar{b}) \log \frac{1}{g_{\bar{a}}(\bar{b})},$$

wobei wir wiederum $0 \log \infty = 0$ setzen und betrachten weiter folgendes

LEMMA 6.5. *Sei g ein Maßvektor. Dann für alle $n, k \geq 1$*

$$H^{n+k}(g) = H^k(g) + H_k^n(g) = H^n(g) + H_n^k(g).$$

BEWEIS. Siehe dazu [HK02], Seite 22, Theorem 2.1. Es gilt nämlich

$$\begin{aligned} H_k^n(g) &= \sum_{\bar{a} \in \Sigma^k} \sum_{\bar{b} \in \Sigma^n} g(\bar{a}\bar{b}) \log \frac{1}{g_{\bar{a}}(\bar{b})} \\ &= \sum_{\bar{a} \in \Sigma^k} \sum_{\bar{b} \in \Sigma^n} g(\bar{a}\bar{b}) \log \frac{g(\bar{a})}{g(\bar{a}\bar{b})} \\ &= \sum_{\bar{a} \in \Sigma^k} \sum_{\bar{b} \in \Sigma^n} g(\bar{a}\bar{b}) \log \frac{1}{g(\bar{a}\bar{b})} + \sum_{\bar{a} \in \Sigma^k} \sum_{\bar{b} \in \Sigma^n} g(\bar{a}\bar{b}) \log g(\bar{a}) \\ &= \sum_{\bar{a} \in \Sigma^k} \sum_{\bar{b} \in \Sigma^n} g(\bar{a}\bar{b}) \log \frac{1}{g(\bar{a}\bar{b})} - \sum_{\bar{a} \in \Sigma^k} g(\bar{a}) \log \frac{1}{g(\bar{a})} \\ &= H^{n+k}(g) - H^k(g), \end{aligned}$$

wobei wiederum darauf geachtet werden muss, $0 \log \infty = 0 = 0 \log 0$ zu setzen. \square

Dieses Lemma hat zunächst eine wichtige Konsequenz.

KOROLLAR 6.6. *Ist g ein Maßvektor, so*

$$H^n(g) = \sum_{i=1}^n H_i^1(g).$$

BEWEIS. Folgt mit Induktion aus Lemma 6.5. \square

Wir betrachten ein weiteres

LEMMA 6.7. *Sei g ein Maßvektor und $n, k, m \geq 1$. Dann*

$$H_{k+m}^n(g) \leq H_m^n(\mu^k g).$$

BEWEIS. Es gilt

$$H_{k+m}^n(g) = \sum_{\bar{a} \in \Sigma^k} \sum_{\bar{b} \in \Sigma^m} \sum_{\bar{c} \in \Sigma^n} g(\bar{a}\bar{b}\bar{c}) \log \frac{g(\bar{a}\bar{b})}{g(\bar{a}\bar{b}\bar{c})}$$

und

$$H_m^n(\mu^k g) = \sum_{\bar{b} \in \Sigma^m} \sum_{\bar{c} \in \Sigma^n} \mu^k g(\bar{b}\bar{c}) \log \frac{\mu^k g(\bar{b})}{\mu^k g(\bar{b}\bar{c})}.$$

Für festes $\bar{b}\bar{c} \in \Sigma^{m+n}$ erhält man

$$\left(\sum_{\bar{a} \in \Sigma^k} g(\bar{a}\bar{b}\bar{c}) \right) \log \frac{\sum_{\bar{a} \in \Sigma^k} g(\bar{a}\bar{b}\bar{c})}{\sum_{\bar{a} \in \Sigma^k} g(\bar{a}\bar{b})} \leq \sum_{\bar{a} \in \Sigma^k} g(\bar{a}\bar{b}\bar{c}) \log \frac{g(\bar{a}\bar{b}\bar{c})}{g(\bar{a}\bar{b})},$$

indem man nachstehendes Lemma anwendet. Hieraus folgt die Behauptung wegen $\mu^k g(\bar{b}\bar{c}) = \sum_{\bar{a} \in \Sigma^k} g(\bar{a}\bar{b}\bar{c})$ bzw. $\mu^k g(\bar{b}) = \sum_{\bar{a} \in \Sigma^k} g(\bar{a}\bar{b})$. \square

LEMMA 6.8. *Seien $(a_n)_{n \in \mathbb{N}}$ und $(b_n)_{n \in \mathbb{N}}$ Folgen nicht-negativer Zahlen, so dass $a := \sum_{n \in \mathbb{N}} a_n < \infty$ sowie $b := \sum_{n \in \mathbb{N}} b_n < \infty$. Dann gilt*

$$\sum_{n \in \mathbb{N}} a_n \log \frac{a_n}{b_n} \leq a \log \frac{a}{b},$$

wobei Gleichheit genau dann erzielt wird, falls es ein $C \in \mathbb{R}^+$ gibt, so dass für alle $n \in \mathbb{N}$ gilt, dass $a_n = Cb_n$.

BEWEIS. Siehe [HK02], S.25, Theorem 2.3. \square

Damit nun ein weiteres, wichtiges

KOROLLAR 6.9. *Ist g ein stationärer Maßvektor, so*

$$H_n^1(g) \leq H_{n-1}^1(g).$$

BEWEIS. Es gilt

$$H_n^1(g) = H_{(n-1)+1}^1(g) \stackrel{L.6.7}{\leq} H_{n-1}^1(\mu g) \stackrel{\mu g = g}{=} H_{n-1}^1(g).$$

\square

Nun zum **Beweis von Satz 6.4**. Siehe dazu auch [HK02], S. 79, Theorem 3.2. Wir setzen

$$\alpha_n := H_{n-1}^1(g)$$

und bemerken, dass die Folge der α_n wegen Korollar 6.9 monoton fallend und beschränkt ist. Daher konvergiert diese Folge gegen einen Grenzwert a . Weiter gilt

$$\frac{1}{n} H^n(g) \stackrel{K.6.6}{=} \frac{1}{n} \sum_{i=1}^n H_i^1(g) = \frac{1}{n} \sum_{i=0}^{n-1} \alpha_i.$$

Wegen dem Satz von Cesaro/Cauchy 4.47 folgt daher $\lim_{n \rightarrow \infty} \frac{1}{n} H^n(g) = a$. \square

Für stationäre Vektoren ist die Sache demnach auf elementare Art und Weise klar.

6.3. Entropieraten von ES-Vektoren

Auch für von ES-Vektoren induzierten stochastischen Prozessen kann auf die Existenz der Entropieraten geschlossen werden. Das ist wiederum kein unbekanntes Resultat, sondern eine Folgerung aus dem Satz von Shannon-McMillan-Breiman. Dieser Satz hat eine lange Geschichte, die im fundamentalen Artikel von Shannon ([Sha48]) ihren Anfang nimmt. Er beweist den Satz zunächst nur für stationäre, ergodische Markov-Ketten. Die Folgeartikel von McMillan ([McM53]) und Breiman ([Bre57], [Bre60]) beweisen ihn für den Fall allgemeiner stationärer, ergodischer Prozesse. Billingsley ([Bil65]) schließlich gelang der Beweis für den Fall allgemeiner stationärer Prozesse. An dieser Stelle sei bemerkt, dass der oben angeführte Satz von Kolmogorov-Sinai auf diesem Weg als Korollar erhalten werden kann. Schließlich erweitern die Arbeiten von Jacobs ([Jac59], [Jac62]) und Gray und Kieffer ([GK80]) den Satz auf Prozesse, die asymptotisch stationär im Mittel sind. An dieser Stelle erhält man die Existenz der Entropieraten für AMS-Prozesse als Korollar. Wir wollen an dieser Stelle noch erwähnen, dass etliche weitere Beweisvarianten dieses Satzes existieren, die alle zu Ehre und Würden gelangten, siehe z.B. [AC88], [OW83].

Tatsächlich scheint jedoch kein elementarer Beweis für die Existenz der Entropieraten im Falle von AMS-Prozessen bekannt zu sein. Dies holen wir an dieser Stelle nach.

SATZ 6.10. *Sei g ein stochastischer ES-Vektor und sei \bar{g} der stationäre Grenzvektor. Dann*

$$\bar{H}(g) = \lim_{n \rightarrow \infty} \frac{1}{n} H_g(X^n) = \bar{H}(\bar{g}). \quad (6.1)$$

Insbesondere existiert der Limes und ist gleich der Entropierate des stationären Grenzvektors.

Die verwendeten Hilfsmittel für den Beweis sind nun elementare, analytische Erkenntnisse. Wir führen diese der Reihe nach auf.

LEMMA 6.11. *Sei g ein stochastischer Vektor. Dann für alle $k \geq 0$*

$$\lim_{n \rightarrow \infty} \frac{1}{n} (H^n(g) - H^n(\mu^k g)) = 0. \quad (6.2)$$

BEWEIS. Mit der Schreibweise

$$H_n^k(g|\mu^k g) := \sum_{\bar{a} \in \Sigma^k} \sum_{\bar{b} \in \Sigma^n} g(\bar{a}\bar{b}) \log \frac{\mu^k g(\bar{b})}{g(\bar{a}\bar{b})}$$

erhält man

$$H^{k+n}(g) = H_n^k(g|\mu^k g) + H^n(\mu^k g)$$

und damit weiter

$$\frac{1}{n}H^n(g) + \frac{1}{n}H_n^k(g) \stackrel{L.6.5}{=} \frac{1}{n}H^{n+k}(g) = \frac{1}{n}H_n^k(g|\mu^k g) + \frac{1}{n}H^n(\mu^k g).$$

Wegen

$$0 \leq \frac{1}{n}H_n^k(g) \leq \frac{1}{n}H^k(\mu^n g) \leq \frac{1}{n} \log \text{card}(\Sigma)^k \xrightarrow{n \rightarrow \infty} 0 \quad (6.3)$$

und

$$0 \leq \frac{1}{n}H_n^k(g|\mu^k g) \leq \frac{1}{n}H^k(g) \leq \frac{1}{n} \log \text{card}(\Sigma)^k \xrightarrow{n \rightarrow \infty} 0 \quad (6.4)$$

folgt die Behauptung des Lemmas durch Anwendung des Sandwich-Theorems. \square

Auf dieses Lemma aufbauend erhalten wir ein Korollar, das später tatsächlich zum Einsatz kommen wird.

KOROLLAR 6.12. *Sei g ein stochastischer Vektor. Dann für alle $k \geq 0$*

$$\lim_{n \rightarrow \infty} \frac{1}{n} [H^n(g) - H^n(\frac{1}{k} \sum_{t=0}^{k-1} \mu^t g)] = 0.$$

BEWEIS. Das folgt mit Hilfe der Gleichung

$$\frac{1}{k} \sum_{t=0}^{k-1} H^n(\mu^t g) \leq H^n(\frac{1}{k} \sum_{i=0}^{k-1} \mu^i g) \leq \frac{1}{k} \sum_{t=0}^{k-1} H^n(\mu^t g) + \log k,$$

welche aus [Gra90], Lemma 2.3.4 mit weiterer Hilfe von Induktion über k hergeleitet werden kann. \square

Wir fahren mit rein analytischen Erkenntnissen fort:

LEMMA 6.13. *Sei $(a_n^k)_{n,k \in \mathbb{N}}$ eine Doppelfolge, so dass für alle n und für alle k die Limites*

$$a^k := \lim_{n \rightarrow \infty} a_n^k$$

und

$$a_n := \lim_{k \rightarrow \infty} a_n^k$$

existieren, wobei die a_n^k gleichmäßig in k für $n \rightarrow \infty$ gegen die a^k konvergieren. Darüberhinaus existiere der Limes

$$a := \lim_{k \rightarrow \infty} a^k.$$

Dann gilt

$$\lim_{n \rightarrow \infty} a_n = a.$$

Das ist wiederum nichts Neues, und kann, nach Einbettung der Behauptung in die Theorie der Netze in Analysis-Büchern wiedergefunden werden ([Heu86]). Der direkte Beweis jedoch umfasst zwei Zeilen, und daher führen wir ihn im Folgenden auf - ohne Netztheorie.

BEWEIS. Sei $\epsilon \in \mathbb{R}^+$. Dann gibt es ein $N_0 \in \mathbb{N}$, so dass

$$\forall n \geq N_0, \forall k \in \mathbb{N} : |a_n^k - a^k| < \epsilon/2. \quad (6.5)$$

Daraus folgt

$$\forall n \geq N_0 : |a_n - a| = \lim_{k \rightarrow \infty} |a_n^k - a^k| \leq \epsilon/2 < \epsilon. \quad (6.6)$$

□

Hier und im Folgenden sei wiederum

$$\Delta^{n-1} := \{x = (x_1, \dots, x_n) \in \mathbb{R}^n \mid 0 \leq x_i \leq 1, \sum_i x_i = 1\}$$

gesetzt. Weiter sei $e := \exp 1$. Das jetzt folgende Lemma zeigt das zentrale technische Ergebnis des Beweises.

LEMMA 6.14. *Setze*

$$h_n : \begin{array}{ccc} \Delta^{n-1} & \longrightarrow & \mathbb{R} \\ x = (x_1, \dots, x_n) & \mapsto & \frac{1}{\log n} \sum_{i=1}^n x_i \log \frac{1}{x_i}. \end{array} \quad (6.7)$$

für $n \geq 2$. Dann sind die h_n "gleichgradig stetig" bezüglich der 1-Norm. Mit anderen Worten (bzw. in Zeichen) gilt

$$\forall \epsilon \in \mathbb{R}^+ \exists \delta \in \mathbb{R}^+ \forall n \geq 2 \forall x, y \in \Delta^{n-1} : \|x - y\|_1 < \delta \implies |h_n(x) - h_n(y)| < \epsilon. \quad (6.8)$$

BEMERKUNG 6.15. *Erstaunlicherweise ist die Aussage nicht mehr gültig, wenn man sie bezüglich einer von 1 verschiedenen p -Norm ($2 \leq p \leq \infty$) formuliert, wie Faigle bemerkt hat ([Fai05b]). Das sieht man folgendermaßen ein. Sei $\epsilon = 1/2$ und $\delta \in \mathbb{R}^+$ beliebig. Dann wählen wir ein $m \in \mathbb{N}$, so dass $m > \frac{1}{\delta}$. Dann gibt es ein $N_0 > m$, so dass für alle $N \geq N_0$ in Δ^{n-1}*

$$\|(\frac{1}{N}, \dots, \frac{1}{N})\|_2 = \sqrt{\frac{1}{N}} < \delta.$$

Es folgt dann

$$\begin{aligned} \|\underbrace{(\frac{1}{m}, \dots, \frac{1}{m}, 0, \dots, 0)}_{m\text{-mal}} - (\frac{1}{N}, \dots, \frac{1}{N})\|_p &\leq \|(\frac{1}{N}, \dots, \frac{1}{N})\|_p \\ &= \sqrt[p]{\frac{1}{N^{p-1}}} = N^{-\frac{p-1}{p}} = N^{\frac{1}{p}-1} \leq N^{-\frac{1}{2}} \\ &= \|(\frac{1}{N}, \dots, \frac{1}{N})\|_2 < \delta, \end{aligned}$$

aber

$$\left| h_N(\underbrace{\frac{1}{m}, \dots, \frac{1}{m}}_{m\text{-mal}}, 0, \dots, 0) - h_N\left(\frac{1}{N}, \dots, \frac{1}{N}\right) \right| = \frac{1}{\log N} |\log m - \log N| \xrightarrow{N \rightarrow \infty} 1$$

und somit finden wir ein $N \in \mathbb{N}$ und passende $x, y \in \Delta^{n-1}$, die der Aussage des obigen Lemmas widersprechen.

Für den Beweis dieses Lemmas benötigen wir zwei weitere Lemmata, die wir zuerst aufführen.

LEMMA 6.16. *Sei*

$$\begin{aligned} h : [0, 1] &\longrightarrow \mathbb{R} \\ x &\longmapsto x \log \frac{1}{x}. \end{aligned}$$

Dann gilt für $x, y \in [0, 1]$:

$$|x - y| \leq \frac{1}{e} \implies |h(x) - h(y)| \leq h(|x - y|). \quad (6.9)$$

BEWEIS. Wir bemerken zunächst, dass $h'(x) = \log \frac{1}{x} - 1$ und weiter $h''(x) = -\frac{1}{x}$. h ist konkav und nimmt bei $\frac{1}{e}$ ein globales Maximum an ($h(\frac{1}{e}) = \frac{1}{e}$). Wir bemerken daher, dass

$$x \leq h(x) \iff x \leq \frac{1}{e}. \quad (6.10)$$

Wegen

$$\begin{aligned} |h(x) - h(y)| &= \left| \left| h(x) - h\left(\frac{1}{e}\right) \right| - \left| h\left(\frac{1}{e}\right) - h(y) \right| \right| \\ &\leq \max\left\{ \left| h(x) - h\left(\frac{1}{e}\right) \right|, \left| h\left(\frac{1}{e}\right) - h(y) \right| \right\} \end{aligned} \quad (6.11)$$

und der Tatsache, dass h auf $[0, \frac{1}{e}]$ monoton steigend ist, können wir ohne Einschränkung annehmen, dass beide x und y entweder größer gleich oder kleiner gleich $\frac{1}{e}$ sind.

Da

$$\forall x \in \left[\frac{1}{e}, 1\right] : |h'(x)| \leq 1, \quad (6.12)$$

folgt mit dem Mittelwertsatz

$$\frac{1}{e} \leq x, y \leq 1 \implies |h(x) - h(y)| \leq |x - y|. \quad (6.13)$$

Aufgrund 6.10 folgt damit im Falle $\frac{1}{e} \leq x, y \leq 1$ die Behauptung.

Es bleibt der Fall (ohne Einschränkung $x < y$) $x < y \leq \frac{1}{e}$. Hier gilt $|h(x) - h(y)| = h(y) - h(x)$. Wir bemerken, dass die Funktion $\log \frac{1}{t} - 1$ auf $[0, \frac{1}{e}]$ positiv und monoton fallend ist (*). Die Behauptung folgt nun aus der

Rechnung

$$\begin{aligned}
|h(x) - h(y)| &= \int_x^y \left(\log \frac{1}{t} - 1\right) dt \\
&\stackrel{(*)}{\leq} \int_x^y \left(\log \frac{1}{t-x} - 1\right) dt \stackrel{s=t-x}{=} \int_0^{y-x} \left(\log \frac{1}{s} - 1\right) ds \quad (6.14) \\
&= \left[s \log \frac{1}{s} \right]_0^{y-x} = h(y-x).
\end{aligned}$$

□

LEMMA 6.17. Sei $K > 0$, $2 \leq n \in \mathbb{N}$ und

$$\Delta_{K,n-1} := K \cdot \Delta^{n-1} := \{x \in \mathbb{R}^n \mid (1/K)x \in \Delta^{n-1}\},$$

also

$$x = (x_1, \dots, x_n) \in \Delta_{K,n-1} \iff \sum_{i=1}^n x_i = K, 0 \leq x_i \leq K.$$

Dann nimmt die Funktion

$$\begin{aligned}
h_{K,n} : \quad \Delta_{K,n-1} &\longrightarrow \mathbb{R} \\
x = (x_1, \dots, x_n) &\mapsto \sum_{i=1}^n x_i \log \frac{1}{x_i} \quad (6.15)
\end{aligned}$$

in $\bar{x} := (K/n, \dots, K/n)$ ihr globales Maximum an.

BEWEIS. Sei $x = (x_1, \dots, x_n) \in \Delta_{K,n-1}$ und weiter

$$\tilde{x} := (1/K)x = (x_1/K, \dots, x_n/K) \in \Delta_{1,n-1} = \Delta^{n-1}.$$

Dann gilt

$$\begin{aligned}
h_{K,n}(x) &= \sum_{i=1}^n x_i \log \frac{1}{x_i} \\
&= \sum_{i=1}^n K \tilde{x}_i \log \frac{1}{K \tilde{x}_i} \quad (6.16) \\
&= K \sum_{i=1}^n \tilde{x}_i \left(\log \frac{1}{\tilde{x}_i} + \log \frac{1}{K} \right) \\
&= K \left(h_{1,n}(\tilde{x}) + \log \frac{1}{K} \right).
\end{aligned}$$

Da $h_{1,n}$ in $(1/n, \dots, 1/n)$ ein globales Maximum annimmt (siehe Lemma 6.2), tut $h_{K,n}$ dasselbe bei $K(1/n, \dots, 1/n)$. □

Nun zum **Beweis von Lemma 6.14**. Sei also $\epsilon \in \mathbb{R}^+$ gegeben. Dann existiert wegen der Stetigkeit von $x \log \frac{1}{x}$ in 0 ein $\delta_1 \in \mathbb{R}^+$, so dass

$$\forall x \in [0, \delta_1] : \quad x \log \frac{1}{x} < \frac{\epsilon}{2}. \quad (6.17)$$

Wir setzen nun

$$\delta := \min\left\{\delta_1, \frac{\epsilon}{2}, \frac{1}{e}\right\}.$$

Sei nun $2 \leq n \in \mathbb{N}$ und $x, y \in \Delta^{n-1}$, so dass $\|x - y\|_1 < \delta$. Es gilt dann

$$\forall i : |x_i - y_i| \leq \|x - y\|_1 < \delta \leq \frac{1}{e}. \quad (6.18)$$

Weiter folgt

$$\begin{aligned} |h_n(x) - h_n(y)| &= \frac{1}{\log n} \left| \sum_{i=1}^n \left(x_i \log \frac{1}{x_i} - y_i \log \frac{1}{y_i} \right) \right| \\ &\leq \frac{1}{\log n} \sum_{i=1}^n \left| x_i \log \frac{1}{x_i} - y_i \log \frac{1}{y_i} \right| \\ &\stackrel{(6.18)+L. 6.16}{\leq} \frac{1}{\log n} \sum_{i=1}^n |x_i - y_i| \log \frac{1}{|x_i - y_i|} \\ &\stackrel{L. 6.17}{\leq} \frac{1}{\log n} \sum_{i=1}^n \frac{\|x - y\|_1}{n} \log \frac{1}{\frac{\|x - y\|_1}{n}} \\ &= \frac{1}{\log n} \sum_{i=1}^n \frac{\|x - y\|_1}{n} \log \frac{n}{\|x - y\|_1} \\ &= \frac{1}{\log n} \|x - y\|_1 \log \frac{n}{\|x - y\|_1} \\ &= \|x - y\|_1 + \frac{1}{\log n} \|x - y\|_1 \log \frac{1}{\|x - y\|_1} \\ &\stackrel{(6.17)}{<} \delta + \frac{1}{\log n} \frac{\epsilon}{2} \leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon. \end{aligned} \quad (6.19)$$

□

Das folgende Korollar zeigt, wie wir das Lemma in der Welt der stochastischen Vektoren zum Einsatz bringen wollen.

KOROLLAR 6.18. *Sei $(g_k)_{k \in \mathbb{N}}$ eine Folge stochastischer Vektoren, die in Norm gegen einen stochastischen Vektor g konvergieren. Dann gilt*

$$\lim_{k \rightarrow \infty} \left(\sup_{n \in \mathbb{N}} \frac{1}{n} |H^n(g_k) - H^n(g)| \right) = 0.$$

BEWEIS. Sei $\epsilon \in \mathbb{R}^+$. Dann gibt es aufgrund Lemma 6.14 ein $\delta \in \mathbb{R}^+$, so dass für alle $n \in \mathbb{N}$ und alle $x, y \in \Delta^{n-1}$

$$\|x - y\|_1 < \delta \implies |h_n(x) - h_n(y)| < \frac{\epsilon}{2 \log(|\Sigma|)}. \quad (6.20)$$

Wegen $\|g_k - g\|_{TV} \xrightarrow{k \rightarrow \infty} 0$ gibt es ein K_0 , so dass für alle $k \geq K_0$: $\|g_k - g\|_{TV} < \delta$, mit anderen Worten, für $k \geq K_0$:

$$\forall n \in \mathbb{N} : \sum_{\bar{a} \in \Sigma^n} |g_k(\bar{a}) - g(\bar{a})| < \delta. \quad (6.21)$$

Wir fassen nun $x_k^n := (g_k(\bar{a}))_{\bar{a} \in \Sigma^n}$, $y^n := (g(\bar{a}))_{\bar{a} \in \Sigma^n}$ als Elemente von $\Delta_{|\Sigma^n|-1}$ auf. Wegen (6.21) gilt dann

$$\forall k \geq K_0 : \|x_k^n - y^n\|_1 < \delta. \quad (6.22)$$

Für $k \geq K_0$ folgt dann

$$\begin{aligned}
\sup_{n \in \mathbb{N}} \frac{1}{n} |H^n(g_k) - H^n(g)| &= \sup_{n \in \mathbb{N}} \frac{1}{n} \left| \sum_{\bar{a} \in \Sigma^n} (g_k(\bar{a}) \log \frac{1}{g_k(\bar{a})} - g(\bar{a}) \log \frac{1}{g(\bar{a})}) \right| \\
&= \sup_{n \in \mathbb{N}} \frac{1}{n} |\log(|\Sigma^n|) h_{|\Sigma^n|}(x_k^n) - h_{|\Sigma^n|}(y^n)| \\
&= \sup_{n \in \mathbb{N}} \frac{1}{n} |n \log(|\Sigma|) h_{|\Sigma^n|}(x_k^n) - h_{|\Sigma^n|}(y^n)| \\
&= \sup_{n \in \mathbb{N}} |\log(|\Sigma|) h_{|\Sigma^n|}(x_k^n) - h_{|\Sigma^n|}(y^n)| \\
&\stackrel{(6.22), (6.20)}{\leq} \log(|\Sigma|) \cdot \frac{\epsilon}{2 \log(|\Sigma|)} < \epsilon.
\end{aligned} \tag{6.23}$$

□

Wir kommen nun zum

Beweis von Satz 6.10: Sei $\bar{g}_k := 1/k \sum_{i=0}^{k-1} \mu^i g$. Wir betrachten die Doppelfolge

$$\bar{h}_n^k := \frac{1}{n} H^n(\bar{g}_k) = \frac{1}{n} \sum_{\bar{a} \in \Sigma^n} \bar{g}_k(\bar{a}) \log \frac{1}{\bar{g}_k(\bar{a})}.$$

Für alle k und alle n gilt

$$0 \leq \bar{h}_n^k \leq \frac{1}{n} \log(|\Sigma^n|) = \frac{1}{n} \log(|\Sigma|^n) = \log(|\Sigma|), \tag{6.24}$$

mit anderen Worten, die Doppelfolge ist beschränkt. Weiter gilt für alle n , da g ein ES-Vektor ist,

$$\lim_{k \rightarrow \infty} \bar{h}_n^k = \frac{1}{n} H^n(\bar{g}). \tag{6.25}$$

Nach Definition eines ES-Vektors und Korollar 6.18 findet diese Konvergenz gleichmäßig in n statt. Da \bar{g} stationär ist, folgt mit dem Satz von Kolmogorov-Sinai 6.4

$$\lim_{n \rightarrow \infty} \frac{1}{n} H^n(\bar{g}) = \bar{H}(\bar{g}). \tag{6.26}$$

Wir betrachten nun die Folgen $(\bar{h}_n^k)_{n \in \mathbb{N}}$. Diese sind wegen 6.24 beschränkt. Sei demnach $C \in [0, \log |\Sigma|]$ ein Häufungspunkt der Folge $(\bar{h}_n^0)_{n \in \mathbb{N}}$. Dann gibt es eine Teilfolge $(\bar{h}_{l(n)}^0)_{n \in \mathbb{N}}$, so dass

$$\lim_{n \rightarrow \infty} \bar{h}_{l(n)}^0 = C. \tag{6.27}$$

Mit Lemma 6.11 folgt nun für alle $k \in \mathbb{N}$ ebenso

$$\lim_{n \rightarrow \infty} \bar{h}_{l(n)}^k = C. \tag{6.28}$$

Durch Anwendung von Lemma 6.13 auf die Doppelfolge $\bar{h}_{l(n)}^k$ erhält man nun

$$C = \lim_{k \rightarrow \infty} C = \lim_{k \rightarrow \infty} \left(\lim_{n \rightarrow \infty} \bar{h}_{l(n)}^k \right) \stackrel{L.6.13}{=} \lim_{n \rightarrow \infty} \left(\lim_{k \rightarrow \infty} \bar{h}_{l(n)}^k \right) \stackrel{(6.26)}{=} \bar{H}(\bar{g}). \tag{6.29}$$

□

Prognosegraphen

Der Glaube an die Klasse der endlich-dimensionalen Prozesse rührt von ihrer endlichen Parametrisierbarkeit her. Eine mögliche und die offensichtlichste Form der Parametrisierung sind Koordinatendarstellungen samt der den Prognoseoperatoren entsprechenden Matrizen. Es ist schnell ersichtlich, dass mit Hilfe dieser endlichen Menge von Parametern der gesamte endlich-dimensionale Vektor beschrieben werden kann. Dies ist im Wesentlichen die Arbeit von Jäger ([Jae97a]) und soll an dieser Stelle, um Wiederholungen zu vermeiden, nicht weiter erörtert werden. Andere Techniken der Parametrisierung sollen im Folgenden dargestellt werden. Die erste davon ist die Möglichkeit einer graphischen Darstellung von endlich-dimensionalen Prozessen.

Die Darstellung endlich-dimensionaler Prozesse als Graphen kann man analog zu der graphischen Darstellung diskretwertiger Hidden Markov Modelle betreiben. Bei den Hidden Markov Modellen jedoch haben Knoten, Kanten sowie Kantengewichte eine anschauliche Interpretation im Sinne von versteckten Zuständen, Übergängen und Übergangswahrscheinlichkeiten. Für allgemeine endlich-dimensionale Prozesse jedoch drängt sich eine derartige Repräsentation nicht auf, da sie schwer interpretierbare negative Zahlen beinhaltet. Nichtsdestotrotz ist die Formulierung als Graph möglich.

7.1. Einführung

Wir starten direkt mit der Definition eines Prognosegraphen, den man mit einem endlich-dimensionalen stochastischen Prozess identifizieren können wird:

DEFINITION 7.1. Sei Σ eine endliches Alphabet,

$$G = (V, E \subset V \times V)$$

ein gerichteter Graph,

$$p_0 : V \longrightarrow \mathbb{R}$$

ein reellwertiger Vektor über den Knoten von G ,

$$M : V \times V \longrightarrow \mathbb{R}$$

eine Kantenbewertungsfunktion, wobei $M(v, j) = 0$ für $(v, j) \notin E$ und

$$Y : V \longrightarrow \Sigma$$

eine Σ -wertige Funktion über den Knoten von G . Dann nennen wir das Tupel

$$\mathcal{G} := (G, Y, M, p_0)$$

einen **Prognosegraphen**, falls die folgenden Bedingungen erfüllt sind:

$$(a) \quad \sum_{v \in V} p_0(v) = 1. \quad (7.1)$$

$$(b) \quad \forall w \in V : \sum_{v \in V} M(v, w) = 1. \quad (7.2)$$

(c) Setzen wir für $a \in \Sigma$ zunächst

$$V_a := \{v \in V \mid Y(v) = a\}$$

und für $\bar{a} = a_1 \dots a_t \in \Sigma^t$ weiter

$$V_{\bar{a}} := V_{a_1} \times \dots \times V_{a_t} = \{v_1 \dots v_t \in V^t \mid Y(v_1) \dots Y(v_t) = \bar{a}\},$$

und definieren wir weiter

$$P_{\mathcal{G}} : \Sigma^+ \longrightarrow \mathbb{R} \\ \bar{a} \longmapsto \sum_{w \in V} p_0(w) \sum_{v_1 \dots v_t \in V_{\bar{a}}} M(v_1, w) \cdot \prod_{s=1}^{t-1} M(v_{s+1}, v_s), \quad (7.3)$$

so gilt

$$\forall \bar{a} \in \Sigma^+ : P_{\mathcal{G}}(\bar{a}) \in [0, 1]. \quad (7.4)$$

Das wirkt so natürlich noch reichlich undurchsichtig. Deshalb geben wir den Dingen von oben zunächst suggestive Namen:

DEFINITION 7.2. Sei $\mathcal{G} = (G, Y, M, p_0)$ ein Prognosegraph. Dann nennen wir V auch **Beobachtungsfunktion**, M eine **Übergangs- oder auch Evolutionsfunktion** bzw., wenn wir M als Matrix $M = (M(v, w))_{(v, w) \in V^2}$ auffassen, **Übergangs- oder Evolutionsmatrix** und zuguterletzt heißen wir p_0 den **Startzustand oder Startvektor**.

Per definitionem treten Prognosegraphen zusammen mit einem Startvektor p_0 auf. Das zugrundeliegende Graphengerüst jedoch lässt sich unter Umständen nicht nur durch die durch p_0 induzierte Wahrscheinlichkeitsverteilung interpretieren, sondern man kann einem Prognosegraphen im Allgemeinen noch andere stochastische Prozesse auslesen. Dies geschieht durch Auswahl eines von p_0 verschiedenen Startvektors, falls existent. Deshalb

DEFINITION 7.3. Sei $\mathcal{G} = (G = (V, V \times V), Y, M, p_0)$ ein Prognosegraph mit $n := |V|$. Für $p \in \mathbb{R}^V$ mit $\sum_{v \in V} p(v) = 1$ setzen wir analog zu oben:

$$P_{\mathcal{G}, p} : \Sigma^+ \longrightarrow \mathbb{R} \\ \bar{a} \longmapsto \sum_{w \in V} p(w) \sum_{v_1 \dots v_t \in V_{\bar{a}}} M(v_1, w) \cdot \prod_{s=1}^{t-1} M(v_{s+1}, v_s). \quad (7.5)$$

Wir nennen p einen **zulässigen Startzustand**, falls

$$\forall \bar{a} \in \Sigma^+ : P_{\mathcal{G}, p}(\bar{a}) \in [0, 1]. \quad (7.6)$$

Wir ergänzen das Ganze um eine weitere Definition, die bereits Analogien zu den vorherigen Kapiteln erkennen lässt:

DEFINITION 7.4. Sei $\mathcal{G} = (G, Y, M, p_0)$ ein Prognosegraph und $n := |V|$. Dann definieren wir zunächst für $a \in \Sigma$ Matrizen T_a durch die Vorschrift

$$\forall (v, w) \in V^2 : (T_a)_{vw} := \begin{cases} M(v, w) & v \in V_a \\ 0 & \text{sonst.} \end{cases} \quad (7.7)$$

Wir setzen weiter für alle $\bar{a} = a_1 \dots a_t \in \Sigma^*$

$$T_{\bar{a}} := T_{a_t} \circ \dots \circ T_{a_1}.$$

Wir nennen die $T_{\bar{a}}$ auch **Prognoseoperatoren** zum Wort \bar{a} . Weiter setzen wir

$$p_{\bar{a}} := \begin{cases} \frac{1}{P_{\mathcal{G}}(\bar{a})} T_{\bar{a}} p_0 & P_{\mathcal{G}}(\bar{a}) \neq 0 \\ 0 & \text{sonst} \end{cases} \quad (7.8)$$

und nennen die $p_{\bar{a}}$ **Prognosevektoren** zum Wort \bar{a} .

Mit dieser Definition fällt uns auf:

PROPOSITION 7.5. Ist $\mathcal{G} = (G = (V, E), Y, M, p_0)$ ein Prognosegraph, so gilt für $\bar{a} = a_1 \dots a_t \in \Sigma^t$ mit obigen Bezeichnungen:

$$(T_{\bar{a}})_{vw} = \begin{cases} \sum_{v_1 \dots v_{t-1} \in V_{a_1 \dots a_{t-1}}} M(v, v_{t-1}) \cdot \prod_{s=1}^{t-2} M(v_{s+1}, v_s) \cdot M(v_1, w) & v \in V_{a_t} \\ 0 & v \notin V_{a_k} \end{cases} \quad (7.9)$$

Daher weiter

$$P_{\mathcal{G}}(\bar{a}) = \sum_{w \in V} p_0(w) \sum_{v \in V} (T_{\bar{a}})_{vw}. \quad (7.10)$$

Betrachtet man weiter $T_{\bar{a}} p_0 \in \mathbb{R}^V$, so

$$\sum_{v \in V} T_{\bar{a}} p_0(v) = P_{\mathcal{G}}(\bar{a}). \quad (7.11)$$

BEWEIS. Gleichung (7.9) folgt einfach aus den Gesetzen der Matrixmultiplikation. Gleichung (7.10) ist dann offensichtlich und (7.11) folgt trivialerweise aus (7.10). \square

Nun kommen wir zum Kern dieses Abschnitts: Wir wollen aus einem Prognosegraphen einen endlich-dimensionalen stochastischen Prozess auslesen können. Umgekehrt müssen wir später die Konstruktion eines Prognosegraphen aus einem endlich-dimensionalen Prozess demonstrieren. Dies alles tun die beiden für diesen Abschnitt fundamentalen Ergebnisse.

SATZ 7.6. Sei $\mathcal{G} = ((V, E), Y, M, p_0)$ ein Prognosegraph. Dann wird durch $P_{\mathcal{G}}$ ein endlich-dimensionaler stochastischer Vektor über Σ induziert.

BEWEIS. Wir fassen $g_{\mathcal{G}} := P_{\mathcal{G}}$ als Element aus l_{∞} auf, indem wir $g_{\mathcal{G}}(\square) = 1$ setzen. Um zu zeigen, dass $g_{\mathcal{G}}$ einen stochastischen Prozess induziert, reicht es gemäß Korollar 2.19 zu zeigen, dass $\sigma g_{\mathcal{G}} = g_{\mathcal{G}}$ und $\forall \bar{a} \in \Sigma^* : g_{\mathcal{G}}(\bar{a}) \in [0, 1]$ gelten. Der zweite Punkt ist klar nach Definition. Für den ersten Punkt beachten wir zunächst

$$\begin{aligned} \sigma g_{\mathcal{G}}(\square) &= \sum_{a \in \Sigma} g_{\mathcal{G}}(a) = \sum_{a \in \Sigma} \sum_{w \in V} p_0(w) \sum_{v \in V_a} M(v, w) \\ &= \sum_{v \in V} \sum_{w \in V} p_0(w) M(v, w) \stackrel{(7.1)}{=} 1 = g_{\mathcal{G}}(\square) \end{aligned} \quad (7.12)$$

und weiter für alle $\bar{a} \in \Sigma^+$ ($\bar{a} =: a_1 \dots a_t$)

$$\begin{aligned} \sigma g_{\mathcal{G}}(\bar{a}) &= \sum_{a \in \Sigma} g_{\mathcal{G}}(\bar{a}a) \\ &\stackrel{(7.11)}{=} \sum_{a \in \Sigma} \sum_{v \in V} T_{\bar{a}a} p_0(v) = \sum_{a \in \Sigma} \sum_{v \in V} T_a \circ T_{\bar{a}} p_0(v) \\ &= \sum_{v \in V} M T_{\bar{a}} p_0(v) \stackrel{(7.2)}{=} \sum_{v \in V} T_{\bar{a}} p_0(v) \\ &\stackrel{(7.11)}{=} g_{\mathcal{G}}(\bar{a}), \end{aligned} \quad (7.13)$$

woraus das Behauptete folgt. Wir haben nun

$$\dim g_{\mathcal{G}} < \infty$$

zu zeigen. Dafür bemerken wir zunächst, dass nach Definition der Prognosevektoren und mit Proposition 7.5:

$$\forall (\bar{a}, \bar{b}) \in \Sigma^+ \times \Sigma^+ : (g_{\mathcal{G}})_{\bar{b}}(\bar{a}) = \sum_{v \in V} T_{\bar{a}} p_{\bar{b}}(v) = \sum_{v \in V} \left(\sum_{w \in V} (T_{\bar{a}})_{vw} \right) p_{\bar{b}}(v). \quad (7.14)$$

Wir betrachten nun

$$\mathcal{P}_g := ((g_{\mathcal{G}})_{\bar{b}}(\bar{a}))_{\bar{a}, \bar{b} \in \Sigma^*} \in \mathbb{R}^{\mathbb{N} \times \mathbb{N}}$$

und haben zu zeigen, dass diese Matrix endlichen Rang hat. Dazu betrachten wir weiter die Matrix

$$\Pi := (p_{\bar{a}}(v))_{v \in V, \bar{a} \in \Sigma^*} \in \mathbb{R}^{V \times \mathbb{N}}$$

und stellen fest, dass jede Zeile $f_{\bar{a}}$ von \mathcal{P}_g vermöge Gleichung (7.14) eine Linearkombination von Zeilen Π_v von Π ist (dafür sei $\alpha_v := \sum_{w \in V} (T_{\bar{a}})_{vw}$)

$$f_{\bar{a}} = \sum_{v \in V} \alpha_v \Pi_v, \quad (7.15)$$

damit \mathcal{P}_g also endlichen Rang hat. \square

Um die Äquivalenz der Klasse von Prozessen, die durch Prognosegraphen dargestellt werden, mit der Klasse der endlich-dimensionalen Prozesse einzusehen, braucht man noch ein Verfahren, aus endlich-dimensionalen Prozessen auch Prognosegraphen zu gewinnen. Hier ist es:

SATZ 7.7. Sei Σ ein endliches Alphabet und $g \in l_\infty^\Sigma$ ein endlich-dimensionaler stochastischer Vektor. Sei desweiteren $n := \dim g$ und $(g_i)_{i=1,\dots,n}$ eine Basis von \mathcal{V}_g , so dass $g_i(\square) = 1$ (die g_i können als stochastische Vektoren gewählt werden). Wir setzen

$$V := \{1, \dots, n\} \times \Sigma$$

und

$$Y : \begin{array}{ccc} V & \longrightarrow & \Sigma \\ (i, a) & \mapsto & a. \end{array}$$

Wir identifizieren nun die Prognoseoperatoren $\tau_a, a \in \Sigma$ mit entsprechenden Matrizen bezüglich der Basis $(g_i)_{i=1,\dots,n}$ und setzen

$$M : \begin{array}{ccc} V \times V & \longrightarrow & \mathbb{R} \\ ((i, a), (j, b)) & \mapsto & (\tau_a)_{ij}. \end{array}$$

Ist nun $g = \sum_i \alpha_i g_i$ eine Darstellung von g über der Basis (g_i) , so wählen wir für alle $i = 1, \dots, n$ ein $a_i \in \Sigma$, so dass $g_i(a_i) \neq 0$. Wir setzen nun

$$p_0((i, a)) = \begin{cases} \alpha_i & a = a_i \\ 0 & \text{sonst.} \end{cases} \quad (7.16)$$

Dann erhalten wir einen Prognosegraphen $\mathcal{G}_g := ((V, V \times V), Y, M, p_0)$, so dass $P_{\mathcal{G}_g} = g$.

BEWEIS. Gleichung 7.1 folgt direkt aus den Voraussetzungen, man beachte $g(\square) = 1$ sowie $g_i(\square) = 1$. Gleichung 7.2 folgt aus

$$\begin{aligned} \sum_{(i,a) \in V} M((i,a), (j,b)) &= \sum_{(i,a) \in V} (\tau_a)_{ij} = \sum_{(i,a) \in V} (\tau_a)_{ij} \underbrace{g_i(\square)}_{=1} \\ &= \sum_{a \in \Sigma} \tau_a g_j(\square) = \sum_{a \in \Sigma} g_j(a) = g_j(\square) = 1. \end{aligned}$$

für alle $(j, b) \in V$. Um (7.4) zu bestätigen, zeigen wir

$$P_{\mathcal{G}_g} = g,$$

womit auch die zusätzliche Behauptung gezeigt wäre. Sei also $\bar{a} \in \Sigma^*$. Wir bemerken zunächst, dass für den Prognoseoperator eines Wortes $\bar{a} = a_1 \dots a_t \in \Sigma^t$ unter Beachtung von Gleichung (7.9) aus Proposition 7.5, der Definition von M und der Regeln der Matrixmultiplikation Folgendes gilt:

$$(T_{\bar{a}})_{(i,a),(j,b)} = \begin{cases} (\tau_{\bar{a}})_{ij} & a_t = a \\ 0 & \text{sonst,} \end{cases} \quad (7.17)$$

woraus

$$\sum_{(i,a) \in V} (T_{\bar{a}})_{(i,a),(j,b)} = \sum_{i=1}^n (\tau_{\bar{a}})_{ij} \quad (7.18)$$

folgt. Damit

$$\begin{aligned}
P_{\mathcal{G}_g}(\bar{a}) &\stackrel{(7.11)}{=} \sum_{(i,a) \in V} T_{\bar{a}} p_0(i,a) = \sum_{(i,a) \in V} \sum_{(j,b) \in V} (T_{\bar{a}})_{(i,a),(j,b)} p_0((j,b)) \\
&= \sum_{(j,b) \in V} p_0((j,b)) \sum_{(i,a) \in V} (T_{\bar{a}})_{(i,a),(j,b)} \stackrel{(7.17)}{=} \sum_{(j,b) \in V} p_0((j,b)) \sum_{i=1}^n (\tau_{\bar{a}})_{ij} \\
&= \sum_{j=1}^n \sum_{i=1}^n (\tau_{\bar{a}})_{ij} \sum_{b \in \Sigma} p_0((j,b)) \stackrel{(7.16)}{=} \sum_{j=1}^n \sum_{i=1}^n (\tau_{\bar{a}})_{ij} \alpha_j \\
&= \sum_{j=1}^n \sum_{i=1}^n (\tau_{\bar{a}})_{ij} \alpha_j \underbrace{g_j(\square)}_{=1} = \tau_{\bar{a}} g(\square) = g(\bar{a}).
\end{aligned} \tag{7.19}$$

□

DEFINITION 7.8. Sei g ein endlich-dimensionaler stochastischer Vektor. Dann nennen wir den in Satz 7.7 konstruierten Prognosegraphen $\mathcal{G}_{g,(g_j)}$ auch den zu g und (g_j) assoziierten **natürlichen Prognosegraphen**.

KOROLLAR 7.9. Die durch Prognosegraphen induzierten stochastischen Prozesse entsprechen genau der Klasse der endlich-dimensionalen stochastischen Prozesse.

BEWEIS. Das folgt nun schnell aus den Sätzen 7.6 und 7.7. □

Im Zusammenhang mit zulässigen Startzuständen erhalten wir

LEMMA 7.10. Sei $\mathcal{G}_{g,(g_i)}$ ein natürlicher Prognosegraph, $n := \dim g$ und $p^* \in \mathbb{R}^{n|\Sigma|}$ ein zulässiger Startzustand (Definition 7.3). Dann gilt für den durch p^* induzierten stochastischen Vektor g^*

$$g^* = \sum_{i=1}^n \left(\sum_{a \in \Sigma} p^*((i,a)) \right) g_i \in \mathcal{V}_g.$$

BEWEIS. Das folgt unter Ausnutzung der Definition eines durch einen zulässigen Startzustand induzierten Vektors, der Definition eines natürlichen Prognosegraphen und Anwendung der Rechenregeln aus Proposition 7.5. □

Desweiteren kann man HMM's in ihrer graphischen Darstellung schnell als Unterklasse der obigen Graphen erkennen:

PROPOSITION 7.11. Sei $\mathcal{G} = (G, Y, M, p_0)$ ein Prognosegraph mit den zusätzlichen Eigenschaften, dass

$$\forall v \in V : p_0(v) \geq 0 \tag{7.20}$$

$$\forall (v,w) \in E : M(v,w) \geq 0. \tag{7.21}$$

Dann ist der durch \mathcal{G} induzierte stochastische Vektor $g_{\mathcal{G}}$ ein Hidden Markov Vektor.

BEWEIS. Das folgt, indem man $Y : V \rightarrow \Sigma$ als Alphabetwechsel auffasst, denn die zusätzlichen Bedingungen machen p_0 und M zu stochastischen Matrizen (siehe die Definitionen 3.33, 3.38 und 3.39). \square

Wir beschließen diesen einführenden Abschnitt mit zwei Definitionen, die auf der Hand liegen und bei der Entwicklung des sich später anschließenden Satzes über die Zusammenhangskomponenten notwendig sind. Beide sind motiviert durch die Einsicht, dass eine Menge verschiedener Prognosegraphen dieselben stochastischen Vektoren induzieren können. Zunächst einmal geben wir der Situation, dass zwei Prognosegraphen im Wesentlichen gleich sind, einen Namen:

DEFINITION 7.12. Zwei Prognosegraphen \mathcal{G}_1 und \mathcal{G}_2 werden **äquivalent** genannt, falls

$$g_{\mathcal{G}_1} = g_{\mathcal{G}_2}, \quad (7.22)$$

sie also, mit anderen Worten, denselben stochastischen Prozess induzieren.

BEMERKUNG 7.13. Zwei Prognosegraphen

$$\mathcal{G}_1 = (G_1, Y_1, M_1, p_1), \mathcal{G}_2 = (G_2, Y_2, M_2, p_2),$$

die beide natürliche Prognosegraphen ein und desselben g sind, sind trivialerweise äquivalent. Insbesondere sind sich die Matrizen M_1 und M_2 ähnlich.

Insbesondere die mit Null belegten Kanten spielen keine Rolle bei der Betrachtung eines Prognosegraphen. Deshalb entfernen wir sie.

DEFINITION 7.14. Sei $\mathcal{G} = ((V, V \times V), Y, M, p_0)$ ein Prognosegraph. Wir nennen den durch Entfernen aller Kanten (v, w) mit $M(v, w) = 0$ entstehenden Graphen den **reduzierten** Prognosegraphen von \mathcal{G} . Offensichtlich sind die beiden Graphen äquivalent.

Wir schließen mit einem Lemma, das äquivalente Prognosegraphen erkennen lassen hilft.

LEMMA 7.15. Seien $\mathcal{G}_1 := \mathcal{G}_{g, (g_i)} = (G_1 = (V_1, E_1), Y_1, M_1, p_0)$ und $\mathcal{G}_2 := \mathcal{G}_{g, (h_i)} = (G_2 = (V_2, E_2), Y_2, M_2, q_0)$ zwei natürliche Prognosegraphen ein und desselben stochastischen Vektors g . Seien $p_1 \in \mathbb{R}^{V_1}, q_1 \in \mathbb{R}^{V_2}$ zwei zulässige Startzustände gemäß Definition 7.3 von \mathcal{G}_1 bzw. \mathcal{G}_2 . Dann sind die Prognosegraphen $\mathcal{G}_{p_1} = (G_1, Y_1, M_1, p_1)$ und $\mathcal{G}_{q_1} = (G_2, Y_2, M_2, q_1)$ genau dann äquivalent, falls im Prognoseraum \mathcal{V}_g

$$\sum_{i=1}^n \left(\sum_{a \in \Sigma} p_1((i, a)) \right) g_i = \sum_{i=1}^n \left(\sum_{a \in \Sigma} p_2((j, a)) \right) h_i$$

gilt.

BEWEIS. Das ist eine direkte Folgerung von Lemma 7.10. \square

7.2. Ergodische Zerlegung von Prognosegraphen

Dieser Unterabschnitt hat zum Ziel, ein zur Theorie der Markov-Ketten analoges Resultat zu erzielen. Dort kann man die Menge der Zustände in disjunkte Mengen von irreduziblen, abgeschlossenen Mengen von Zuständen plus einer Menge von Zuständen, die man als "vergänglich" bezeichnet, zerlegen (siehe z.B. [GS01]). Wir haben in Kapitel 5 eine ergodische Zerlegung endlich-dimensionaler stochastischer Vektoren vorgestellt. Überträgt man dieses Ergebnis auf Prognosegraphen, erhält man ein der Theorie der Markov-Ketten ähnliches Ergebnis.

Wir starten diesen Unterabschnitt mit ein paar illustrierenden Formalitäten. Betrachten wir bei gegebenem g samt Basis (g_i) mit $n := \dim g$ den dazu assoziierten natürlichen Prognosegraphen $\mathcal{G}_{g,(g_i)} =: ((V, V \times V), Y, M, p_0)$ und definiert man $e_{(i,a)} \in \mathbb{R}^V$ durch

$$e_{(i,a)}((j,b)) = \begin{cases} 1 & (j,b) = (i,a) \\ 0 & (j,b) \neq (i,a) \end{cases},$$

so kann man die Prognoseoperatoren als Endomorphismen des Vektorraums

$$\mathbb{R}^{n \cdot |\Sigma|} \cong \mathbb{R}^V$$

samt kanonisch gegebener Basis

$$(e_{(i,a)})_{(i,a) \in V}$$

auffassen. Bezüglich dieser Basis hat M dann die folgende Gestalt (sei $\Sigma = \{a_1, \dots, a_m\}$):

$$M = \begin{pmatrix} [\tau_{a_1}] \in \mathbb{R}^{n^2} & \dots & [\tau_{a_1}] \in \mathbb{R}^{n^2} \\ [\tau_{a_2}] \in \mathbb{R}^{n^2} & \dots & [\tau_{a_2}] \in \mathbb{R}^{n^2} \\ \vdots & \ddots & \vdots \\ [\tau_{a_m}] \in \mathbb{R}^{n^2} & \dots & [\tau_{a_m}] \in \mathbb{R}^{n^2} \end{pmatrix} \in \mathbb{R}^{n|\Sigma| \times n|\Sigma|},$$

wobei die τ_a gerade mit Koordinatendarstellungen bezüglich der Basis (g_i) identifiziert werden. Weiter gilt dann nach Definition

$$T_a = \begin{pmatrix} 0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 0 \\ [\tau_a] & \dots & [\tau_a] \\ 0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 0 \end{pmatrix} \in \mathbb{R}^{n|\Sigma| \times n|\Sigma|}.$$

T_a besteht nämlich nach Definition gerade aus den Zeilen von M , die zu den Basisvektoren $e_{(i,a)}$, $i = 1, \dots, n$ gehören und ansonsten aus Nullen. Setzen wir

$$\forall a \in \Sigma : \quad W_a := \text{span}\{e_{(i,a)} \mid i = 1, \dots, \dim g\},$$

so stellen wir also fest, dass

$$\text{Im}(T_a) \subset W_a. \quad (7.23)$$

Wir bemerken weiter, dass die zu zwei $e_{(i_1, a_1)}, e_{(i_2, a_2)}$ gehörenden Spalten von den T_a und damit auch von M identisch sind, falls $i_1 = i_2$. Daher gilt für alle $a, a_1, a_2 \in \Sigma, i = 1, \dots, n$

$$T_a(e_{(j, a_1)}) = T_a(e_{(j, a_2)}) = \sum_{i=1}^n (\tau_a)_{ij} e_{(i, a)}.$$

Mit diesen Erkenntnissen im Hinterkopf treffen wir nun ein paar Vorbereitungen für den späteren Satz über die Zusammenhangskomponenten der natürlichen Prognosegraphen.

SATZ 7.16. *Sei g ein endlich-dimensionaler stochastischer Vektor ($\dim g =: n$) und $(g_j)_{j=1, \dots, n}$ eine Basis von \mathcal{V}_g , so dass die g_j stochastisch sind, und weiter $\mathcal{G}_{g, (g_j)} =: (G, Y, M, p_0)$ der dazu assoziierte Prognosegraph. Dann sind die Matrix M sowie die Prognoseoperatoren T_a , wie oben als Endomorphismen von \mathbb{R}^V aufgefasst, stabil.*

BEWEIS. Wir zeigen zunächst, dass die T_a stabil sind. Zunächst bemerken wir, dass (siehe Gleichung (7.23))

$$\forall b \in \Sigma : \quad T_a(W_b) \subset W_a.$$

Daher reicht es, die Stabilität für Elemente von W_a zu überprüfen. Mit den Bezeichnungen von oben gilt aber nun (τ_a identifizieren wir wiederum mit einer entsprechende Matrixdarstellung bezüglich (g_j))

$$T_a|_{W_a} = \tau_a$$

und die Behauptung folgt aus der Stabilität der τ_a . Gemäß Satz 3.29 lassen die Prognoseoperatoren τ_a den Prognosekegel $K_g \subset \mathcal{V}_g$ invariant. Wir erhalten nun für alle $a \in \Sigma$ Isomorphismen

$$\begin{aligned} \Phi_a : \mathcal{V}_g &\longrightarrow W_a \\ g_j &\longmapsto e_{j, a} \end{aligned}$$

und somit jeweils einen echten Kegel

$$K_a = \Phi_a(K_g) \subset W_a.$$

also gewissermaßen K_a jeweils eine Kopie von K_g in W_a . Es gilt dann für alle $a, b \in \Sigma$

$$T_a(K_b) \subset K_a.$$

Wir betrachten nun

$$\tilde{K} := \bigoplus_{a \in \Sigma} K_a \subset \mathbb{R}^V$$

und stellen fest, dass es sich um einen echten Kegel handelt, der nun von allen T_a invariant gelassen wird. Damit wird \tilde{K} natürlich auch von $M = \sum_{a \in \Sigma} T_a$ invariant gelassen. Da M komponentensummenerhaltend ist, folgt die Aussage mit Satz 12.10, wenn man beachtet, dass die dort geforderte Eigenschaft (12.9) aufgrund der Eigenschaften des Prognosekegels und der Wahl der Basis (g_j) als stochastische Vektoren erfüllt ist. \square

Die nun folgende Proposition schließlich liefert Aufschluss über die Eigenwerte der T_a und M .

PROPOSITION 7.17. *Sei g ein endlich-dimensionaler stochastischer Vektor, $n := \dim g$ und $(g_i)_{i=1, \dots, n}$ eine Basis von \mathcal{V}_g , $\mu : \mathcal{V}_g \rightarrow \mathcal{V}_g$ der Evolutionsoperator und weiter $\mathcal{G}_{g, (g_i)} = (G, Y, M, p_0)$ der dazu assoziierte Prognosegraph. Dann gilt für die charakteristischen Polynome P_M, P_μ der jeweiligen Operatoren*

$$P_M(X) = X^{n \cdot (|\Sigma|-1)} P_\mu(X). \quad (7.24)$$

BEWEIS. Wir betrachten das folgende

LEMMA 7.18. *Sei $A_i \in \mathbb{C}^{n \times n}$, $i = 1, \dots, m$ eine endliche Familie von Matrizen und sei $B := \sum_{i=1}^m A_i$ deren Summe. Sei weiter*

$$C := \begin{pmatrix} A_1 & \dots & A_1 \\ \vdots & \ddots & \vdots \\ A_m & \dots & A_m \end{pmatrix} \in \mathbb{C}^{mn \times mn}$$

Seien $P_B(X)$ und $P_C(X)$ die jeweiligen charakteristischen Polynome. Dann gilt

$$P_C(X) = X^{n(m-1)} P_B(X). \quad (7.25)$$

Beweis des Lemmas: Es gilt:

$$\begin{aligned} P_C(X) &= \det \begin{pmatrix} A_1 - X \cdot Id_n & A_1 & \dots & A_1 \\ A_2 & A_2 - X \cdot Id_n & \dots & A_2 \\ \vdots & \vdots & \ddots & \vdots \\ A_m & A_m & \dots & A_m - X \cdot Id_n \end{pmatrix} \\ &= \det \begin{pmatrix} B - X \cdot Id_n & B - X \cdot Id_n & \dots & B - X \cdot Id_n \\ A_2 & A_2 - X \cdot Id_n & \dots & A_2 \\ \vdots & \vdots & \ddots & \vdots \\ A_m & A_m & \dots & A_m - X \cdot Id_n \end{pmatrix} \quad (7.26) \\ &= \det \begin{pmatrix} B - X \cdot Id_n & 0 & \dots & 0 \\ A_2 & -X \cdot Id_n & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ A_m & 0 & \dots & -X \cdot Id_n \end{pmatrix} \\ &= \det(B - X \cdot Id_n) \cdot (\det(-X \cdot Id_n))^{m-1} = P_B(X) X^{n(m-1)}, \end{aligned}$$

wobei die zweite Gleichung durch Addition des zweiten bis m -ten Zeilenblocks auf den ersten folgt und die dritte Gleichung durch Subtraktion des ersten Spaltenblocks vom zweiten bis m -ten Spaltenblock. \square

Weiter im Beweis der Proposition: Die Behauptung folgt nun durch Anwendung des Lemmas auf die Zerlegung $M = \sum_{a \in \Sigma} T_a$. \square

Nun zum namensstiftenden Ergebnis dieses Unterabschnitts. Zum genauen Verständnis benötigen wir zunächst eine graphentheoretische

DEFINITION 7.19. Sei $G = (V, E)$ ein gerichteter Graph. Eine Teilmenge von Knoten $W \subset V$ heißt **abgeschlossen**, falls

$$\forall w \in W : \nexists w^* \in \mathcal{CV}_1 : (w, w^*) \in E,$$

d.h. also wenn keine Kanten aus V_1 herausführen. Existieren für zwei Knoten $v_1, v_2 \in V$ jeweils gerichtete Pfade von v_1 nach v_2 und andersherum, so sagt man, dass v_1 und v_2 **kommunizieren**. Eine Teilmenge $W \subset V$ von Knoten heißt dann **irreduzibel**, falls alle Paare $w_1, w_2 \in W$ kommunizieren.

Natürliche Prognosegraphen ergodischer stochastischer Vektoren enthalten nun, unabhängig von der Auswahl einer Basis für den Graphen, immer genau eine irreduzible, abgeschlossene Menge von Knoten. Stochastische Vektoren g , deren stationärer Raum $E_g := l_\mu \cap \mathcal{V}_g$ die Dimension n hat, die nicht ergodisch sind, erlauben die Auswahl einer Basis, so dass der resultierende Prognosegraph genau n irreduzible, abgeschlossene Komponenten hat.

SATZ 7.20 (Ergodische Zerlegung von Prognosegraphen). Sei g ein endlich-dimensionaler stochastischer Vektor. Dann sind die folgenden Aussagen äquivalent:

- (i) g ist ergodisch.
- (ii) $\dim \text{Eig}(M; 1) = 1$ für alle natürlichen Prognosegraphen von g .
- (iii) Für alle Basen (g_j) enthält der reduzierte Prognosegraph $\mathcal{G}_{g, (g_j)}$ des dazu assoziierten natürlichen Prognosegraphen genau eine irreduzible, abgeschlossene Teilmenge von Knoten.

Genauer gilt: Ist $n := \text{ergdim } g$ die ergodische Dimension von g , so findet man einen reduzierten, natürlichen Prognosegraphen von g , der genau n irreduzible, abgeschlossene Komponenten hat.

BEWEIS.

“(i) \implies (ii)”: Ist g ergodisch, so hat der Eigenraum des Evolutionsoperators $\mu : \mathcal{V}_g \rightarrow \mathcal{V}_g$ zum Eigenwert 1 gemäß Korollar 4.83 die Dimension 1. Gemäß Proposition 7.17 und der Stabilität von μ (Satz 3.8) erhalten wir damit, dass der Eigenraum von M zum Eigenwert 1 ebenfalls Dimension 1 hat.

“(ii) \implies (iii)”: Wir betrachten für alle $(j, b) \in \{1, \dots, n\} \times \Sigma$ die Folgen

$$(p_{(j,b)})_n := \frac{1}{n} \sum_{k=1}^{n-1} M^k e_{(j,b)}.$$

Diese konvergieren wegen (ii) und weil M stabil (Satz 7.16) und damit gemäß Satz 10.14 summierbar und komponentensummenerhaltend ist, unabhängig von (j, b) alle gegen den Eigenvektor \bar{p} von M zum Eigenwert 1. Schreiben wir

$$\bar{p} = \sum_{i=1}^n \sum_{a \in \Sigma} \alpha_{(i,a)} e_{(i,a)},$$

Sei (i_0, a_0) nun so, dass $\alpha_{(i_0, a_0)} \neq 0$ (solche gibt es wegen $\sum_{i=1}^n \sum_{a \in \Sigma} \alpha_{(i,a)} = 1$). Wir schreiben weiter

$$M^k e_{(j,b)} = \sum_{i=1}^n \sum_{a \in \Sigma} (\beta_k)_{(i,a)} e_{(i,a)}.$$

Wegen

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^{n-1} M^k e_{(j,b)} = \bar{p} = \sum_{i=1}^n \sum_{a \in \Sigma} \alpha_{(i,a)} e_{(i,a)}$$

finden wir demnach für alle (j, b) ein $k_{(j,b)}$, so dass

$$(\beta_{k_{(j,b)}})_{(i_0, a_0)} \neq 0.$$

Mit anderen Worten erreichen wir im reduzierten Graphen vom Knoten (j, b) den Knoten (i_0, a_0) in $k_{(j,b)}$ Schritten. Daraus folgt offensichtlich, dass es keine zwei abgeschlossene Komponenten geben kann. Die Behauptung folgt mit dem sich anschließenden Lemma. \square

LEMMA 7.21. *Sei $G = (V, E)$ ein gerichteter Graph. Dann enthält G eine nicht-leere, irreduzible, abgeschlossene Komponente.*

Den **Beweis** führen wir [Lät06] mit Induktion über die Anzahl $n := |V|$ der Knoten von V . Im Falle $n = 1$ tut es der einzige vorhandene Knoten. Habe V nun $n + 1$ Knoten. Enthält der Graph keine zwei Knoten, die miteinander kommunizieren, so enthält er keine gerichteten Kreise. Deshalb finden wir in diesem Fall einen Knoten, aus dem keine Kanten hinausführen, und erhalten somit die Behauptung, da dieser Knoten eine irreduzible, abgeschlossene Menge bildet. Finden wir zwei Knoten v_1, v_2 , die miteinander kommunizieren, so sind wir fertig, falls $\{v_1, v_2\}$ abgeschlossen ist. Falls nicht, definieren wir einen neuen Graphen $G^* = (V^*, E^*)$, mit nunmehr n Knoten, indem wir v_1 und v_2 zu einem Knoten v zusammenfassen, d.h.

$$V^* = (V \setminus \{v_1, v_2\}) \cup \{v\}$$

und

$$(u, w) \in E^* \iff \begin{cases} (u, w) \in E & u, w \in V \setminus \{v_1, v_2\} \\ (v_1, w) \in E \vee (v_2, w) \in E & u = v, w \in V \setminus \{v_1, v_2\} \\ (u, v_1) \in E \vee (u, v_2) \in E & w = v, u \in V \setminus \{v_1, v_2\} \end{cases}$$

sowie $(v, v) \in E^*$, falls im ursprünglichen Graphen eine Kante zwischen v_1 und v_2 verlief. Nach Induktionsvoraussetzung enthält V^* dann eine irreduzible, abgeschlossene Komponente. Es ist ein Leichtes, einzusehen, dass deren Übertragung auf den ursprünglichen Graphen ebenfalls irreduzibel und abgeschlossen ist. \square

“(iii) \implies (i)”: Sei g nicht ergodisch. Gemäß dem Satz über die ergodische Zerlegung lässt sich g als Mixtur endlich vieler ergodischer Komponentenvektoren schreiben:

$$g = \sum_{i=1}^n \alpha_i g_i,$$

wobei die g_i linear unabhängig sind und weiter gilt

$$\mathcal{V}_g = \mathcal{V}_{\bar{\mu}g_1} \oplus \dots \oplus \mathcal{V}_{\bar{\mu}g_n} \oplus \mathcal{O}_g.$$

Wir setzen $k_i := \dim \mathcal{V}_{\bar{\mu}g_i}$ und $k_0 := \dim \mathcal{O}_g$ und konstruieren eine passende Basis von stochastischen Vektoren

$$\mathfrak{G} := \{g_{11}, \dots, g_{1k_1}, \dots, g_{n1}, \dots, g_{nk_n}, h_1, \dots, h_{k_0}\},$$

von \mathcal{V}_g , so dass jeweils

$$\text{span}\{g_{ij} \mid j = 1, \dots, k_i\} = \mathcal{V}_{\bar{\mu}g_i}$$

und die h_i als beliebige stochastische Vektoren gewählt seien. Wir behaupten, dass der reduzierte Prognosegraph von $\mathcal{G}_{g, \mathfrak{G}}$ mindestens n irreduzible, abgeschlossene Komponenten hat. Dazu bemerken wir, dass

$$\forall a \in \Sigma : \tau_a g_{ij} \in \mathcal{V}_{\bar{\mu}g_i}$$

und damit, dass die τ_a bezüglich einer solchen Basis die Gestalt

$$\tau_a = \begin{pmatrix} [\tau_a | \mathcal{V}_{\bar{\mu}g_1}] & 0 & \dots & 0 \\ 0 & [\tau_a | \mathcal{V}_{\bar{\mu}g_2}] & \dots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \dots & [\tau_a | \mathcal{V}_{\bar{\mu}g_n}] \\ * & * & \dots & * \end{pmatrix}$$

annehmen. Nach Definition der reduzierten Prognosegraphen gibt es daher keine Kanten, falls die beteiligten Knoten von verschiedenen Komponenten $\mathcal{V}_{\bar{\mu}g_i}, \mathcal{V}_{\bar{\mu}g_j}, i \neq j$ herrühren, was zeigt, dass die entsprechenden Komponenten abgeschlossen sind. Bemerkt man weiter, dass die durch die zu einer Komponente gehörenden, induzierten Teilgraphen bei entsprechender Wahl des Startvektors gerade reduzierten, natürlichen Prognosegraphen von $\bar{\mu}g_i$ entsprechen, so folgt die Behauptung nun aus der bereits bewiesenen Richtung “(ii) \implies (iii)”. \square

Quanten-Prognose-Modelle

A quantum system is a useful abstraction which frequently appears in the literature, but does not really exist in nature.

Asher Peres

Die Beschreibung endlich-dimensionaler Prozesse gelingt auch noch mit einem ganz anderen Mechanismus, dessen Motivation der Beschreibung stochastischer Phänomene der Quantenmechanik zuzuordnen ist. Hier agieren lineare Operatoren auf dem Raum der selbst-adjungierten Matrizen einer gewissen Dimension und Wahrscheinlichkeiten lassen sich als die Spur der durch die Operationen resultierenden Matrizen ablesen. Der in diesem Abschnitt entwickelte Kalkül ist im Wesentlichen Faigle ([Fai05a]) zu verdanken.

Sei in diesem Abschnitt immer V ein Vektorraum über den komplexen Zahlen \mathbb{C} einer endlichen Dimension n . Ist $z = x + iy \in \mathbb{C}$, so bezeichne $z^* := x - iy$ den komplex konjugierten Wert von z . Ist $A = (a_{ij})_{i,j=1,\dots,n} \in \mathbb{C}^{n \times n}$ eine komplexwertige Matrix, so sei

$$A^* := (a_{ji}^*)$$

die dazu komplex konjugierte (oder auch adjungierte) Matrix.

8.1. Der Zustandsraum

Wir starten diesen Abschnitt mit der Definition des Raums der Zustände, im Wesentlichen der Raum der selbst-adjungierten Matrizen über V .

DEFINITION 8.1. *Wir setzen*

$$\mathcal{S}^n := \{Q \in \mathbb{C}^{n \times n} \mid Q = Q^*\},$$

*also \mathcal{S}^n der \mathbb{R} -Vektorraum der Matrizen (auch selbst-adjungierten Matrizen genannt) über V (bemerke, dass Addition zweier selbst-adjungierten Matrizen sowie Multiplikation mit einer reellen Zahl wiederum eine selbst-adjungierte Matrix ergibt). Wir nennen dies in Zukunft den zu V assoziierten **Zustandsraum**. selbst-adjungierte Matrizen Q werden dementsprechend als **Zustand** gehandelt.*

Man bemerke, dass eine selbst-adjungierte Matrix Q ausschließlich reelle Eigenwerte $\{\lambda_1, \dots, \lambda_n\}$ samt einer dazu passenden Orthogonalbasis von Eigenvektoren $\{v_1, \dots, v_n\}$ hat. Es gilt dann

$$Q = \sum_{i=1}^n \lambda_i v_i v_i^*,$$

wobei v_i^* der zu v_i transponierte Vektor sei (bezüglich einer wie auch immer gewählten Koordinatendarstellung) und daher weiter

$$\operatorname{tr} Q := \operatorname{trace} Q = \sum_{i=1}^n q_{ii} = \sum_{i=1}^n \lambda_i.$$

DEFINITION 8.2. Wir nennen einen Zustand Q **nicht-negativ**, falls

$$\forall v \in V : v^T Q v \geq 0.$$

Die Menge der nicht-negativen Zustände $v \in S^n$ bezeichnen wir mit S_+^n . Einen nicht-negativen Zustand Q mit $\operatorname{tr} Q = 1$ nennen wir im Einklang mit der in der Quantenmechanik üblichen Sprechweise **Dichtematrix** oder **Quantenzustand**.

BEMERKUNG 8.3. Der Zustandsraum S^n hat, wie oben definiert, als \mathbb{R} -Vektorraum aufgefasst, die Dimension n^2 , wie man auf die folgende Art und Weise einsieht: Für alle $Q \in S^n$ gibt es $A, B \in \mathbb{R}^{n \times n}$ mit den Eigenschaften

$$\begin{aligned} A &= A^T, \\ B &= -B^T, \\ Q &= A + iB, \end{aligned}$$

d.h. also A eine symmetrische Matrix, B eine schiefsymmetrische Matrix, so dass $Q = A + iB$ (und folglich $Q^* = A - iB$). Da die Diagonalelemente einer schiefsymmetrischen Matrix gleich Null sein müssen, folgt, dass die Dimension von S^n über \mathbb{R} gleich n^2 ist.

BEMERKUNG 8.4. Sei $S_+^n \subset S^n$ die Menge der nicht-negativen Zustände. Dann fällt auf, dass S_+^n einen echten konvexen Kegel bildet und es gilt weiter, dass für alle $Q \in S_+^n$

$$Q = 0 \quad \Leftrightarrow \quad \operatorname{tr} Q = 0$$

gilt. Daraus folgt natürlich, dass $\operatorname{tr} Q \geq 0$ für $Q \in S_+^n$.

8.2. Schatten-Normen

Eine für uns wichtiges technisches Hilfsmittel wird eine Norm auf dem Zustandsraum, die der Klasse der Schatten-Normen zuzurechnen ist ([Sch60]). Zum Verständnis dieser Klasse von Normen benötigt es ein wenig Vorlauf. Ist $A \in \mathbb{C}^{n \times n}$ eine diagonalisierbare Matrix, so dass

$$A = S^{-1} \operatorname{diag}(\lambda_1, \dots, \lambda_n) S$$

und $f : \mathbb{C} \rightarrow \mathbb{C}$ eine Funktion, so definieren wir

$$f(A) := S^{-1} \operatorname{diag}(f(\lambda_1), \dots, f(\lambda_n)) S.$$

In diesem Sinne ist

$$|A| := \sqrt{(AA^*)}$$

zu verstehen, wobei wir bemerken, dass AA^* für beliebige Matrizen $A \in \mathbb{C}^{n \times n}$ immer selbst-adjungiert und damit diagonalisierbar ist. Wir erhalten dann die folgende

DEFINITION 8.5. Für alle $p \geq 1$ definieren wir Abbildungen

$$\begin{aligned} \|\cdot\|_p : \mathbb{C}^{n \times n} &\longrightarrow \mathbb{R}^+ \\ A &\longmapsto \sqrt[p]{\operatorname{tr} |A|^p} \end{aligned}$$

und nennen dies die **p-te Schatten-Norm**.

Dies sind tatsächlich Normen, was in [Sch60] verifiziert wird. Für selbstadjungierte Matrizen machen wir eine

BEMERKUNG 8.6. Ist $A \in \mathbb{C}^{n \times n}$ selbstadjungiert, so dass

$$A = S^{-1} \operatorname{diag}(\lambda_1, \dots, \lambda_n) S,$$

dann gilt

$$\|A\|_p = \sqrt[p]{\sum_{i=1}^n |\lambda_i|^p} = \|(\lambda_1, \dots, \lambda_n)\|_p,$$

wobei der letzte Term mit der üblichen p -Norm auf dem \mathbb{R}^n zu identifizieren ist. Insbesondere gilt demnach

$$\|A\|_1 = \sum_{i=1}^n \lambda_i = \operatorname{tr} A \quad (8.1)$$

für eine nicht-negative, selbstadjungierte Matrix A .

Die Schatten-1-Norm kann man für selbst-adjungierte Matrizen darüberhinaus noch in einem anderen Licht betrachten:

BEMERKUNG 8.7. Ist $D = \operatorname{diag}(\lambda_1, \dots, \lambda_n) \in \mathbb{R}^{n \times n}$ eine Diagonalmatrix, so definieren wir zunächst $\lambda^+ := \max\{\lambda, 0\}$, $\lambda^- := \max\{-\lambda, 0\}$ für $\lambda \in \mathbb{R}$ und weiter

$$\begin{aligned} D^+ &:= \operatorname{diag}(\lambda_1^+, \dots, \lambda_n^+), \\ D^- &:= \operatorname{diag}(\lambda_1^-, \dots, \lambda_n^-), \end{aligned}$$

so dass $D = D^+ - D^-$. Ist $Q \in \mathbb{C}^{n \times n}$ eine selbstadjungierte Matrix, so dass $Q = S^{-1} D_Q S$ mit $D_Q := \operatorname{diag}(\lambda_1, \dots, \lambda_n)$, so definieren wir demgemäß

$$\begin{aligned} Q^+ &:= S^{-1} D_Q^+ S, \\ Q^- &:= S^{-1} D_Q^- S, \end{aligned}$$

wobei offensichtlich beide $Q^+, Q^- \in S_+^n$ und darüberhinaus

$$Q = Q^+ - Q^-. \quad (8.2)$$

Für die Schatten-1-Norm gilt dann weiter

$$\|Q\|_1 = \text{tr } Q^+ + \text{tr } Q^-. \quad (8.3)$$

8.3. Quanten-Prognose-Modelle

Wir fahren nun damit fort, eine alternative Parametrisierung endlich-dimensionaler Prozesse anzugeben.

DEFINITION 8.8. Sei Σ ein endliches Alphabet. Dann nennen wir ein Tupel

$$(\mathcal{V}, (T_a)_{a \in \Sigma}, Q_0)$$

bestehend aus einem linearen Unterraum $\mathcal{V} \subset \mathcal{S}^n$ eines Zustandsraums \mathcal{S}^n , (\mathbb{R} -)linearen Operatoren $T_a : \mathcal{V} \rightarrow \mathcal{V}$ und einem ausgezeichneten Zustand $Q_0 \in \mathcal{V}$ ein **Quanten-Prognose-Modell**, falls die folgenden Bedingungen erfüllt sind:

$$\text{tr } Q_0 = 1, \quad (8.4)$$

$$\forall Q \in \mathcal{V}: \quad \text{tr } \Psi(Q) = \text{tr } Q, \quad (8.5)$$

$$\forall \bar{a} \in \Sigma^*: \quad \text{tr } T_{\bar{a}} Q_0 \in [0, 1], \quad (8.6)$$

wobei der lineare Operator $\Psi : \mathcal{V} \rightarrow \mathcal{V}$ durch die Vorschrift

$$\Psi(Q) := \sum_{a \in \Sigma} T_a(Q)$$

definiert sei und weiter für ein Wort $\bar{a} = a_1 \dots a_t$ der lineare Operator

$$T_{\bar{a}} := T_{a_t} \circ \dots \circ T_{a_1}$$

(und formal $T_{\square} = Id$) gemeint sei.

BEMERKUNG 8.9. Mit den zuletzt aufgeführten Definitionen gilt natürlich

$$\forall Q \in \mathcal{V}, t \in \mathbb{N}: \quad \Psi^t(Q) = \sum_{\bar{a} \in \Sigma^t} T_{\bar{a}}(Q).$$

Wir kommen nun zur Hauptaussage des Abschnitts. Quanten-Prognose-Modelle induzieren endlich-dimensionale stochastische Vektoren und damit einen endlich-dimensionalen stochastischen Prozess.

SATZ 8.10. Sei $(\mathcal{V}, (T_a)_{a \in \Sigma}, Q_0)$ ein Quanten-Prognose-Modell über einem endlichen Alphabet Σ . Dann ist

$$g: \Sigma^* \longrightarrow \mathbb{R} \\ \bar{a} \longmapsto \text{tr } T_{\bar{a}} Q_0$$

ein endlich-dimensionaler stochastischer Vektor.

BEWEIS. Es ist $g(\bar{a}) \in [0, 1]$ wegen Eigenschaft (8.6) und

$$g(\square) = \text{tr } T_{\square} Q_0 = \text{tr } Id Q_0 = \text{tr } Q_0 = 1$$

nach Eigenschaft (8.4). Weiter gilt

$$\begin{aligned} \sigma g(\bar{a}) &= \sum_{a \in \Sigma} g(\bar{a}a) = \sum_{a \in \Sigma} \text{tr } T_{\bar{a}a} Q_0 \\ &= \text{tr} \sum_{a \in \Sigma} T_a T_{\bar{a}} Q_0 = \text{tr } \Psi T_{\bar{a}} Q_0 \\ &\stackrel{(8.5)}{=} \text{tr } T_{\bar{a}} Q_0 = g(\bar{a}) \end{aligned}$$

womit gezeigt wäre, dass g stochastisch ist. Zur Vervollständigung der Aussage bleibt zu zeigen, dass die Matrix

$$\mathcal{P}_g := (g(\bar{a}\bar{b}))_{\bar{a}, \bar{b} \in \Sigma^*} \in \mathbb{R}^{\Sigma^* \times \Sigma^*}$$

endlich-dimensional ist. Dazu wählen wir eine Basis (Q_1, \dots, Q_{n^2}) von \mathcal{V} und betrachten nun die Matrix

$$\Pi_g := (\text{tr } T_{\bar{a}} Q_i)_{i \in \{1, \dots, n^2\}, \bar{a} \in \Sigma^*} \in \mathbb{R}^{n^2 \times \Sigma^*},$$

die wegen der endlichen Anzahl von Zeilen endlichen Rang hat. Wir setzen nun für alle $\bar{a} \in \Sigma^*$

$$T_{\bar{a}} Q_0 =: \sum_{i=1}^{n^2} \alpha_{i, \bar{a}} Q_i$$

und sehen, dass

$$g(\bar{a}\bar{b}) = \text{tr } T_{\bar{b}} T_{\bar{a}} Q_0 = \text{tr } T_{\bar{b}} \sum_{i=1}^{n^2} \alpha_{i, \bar{a}} Q_i = \sum_{i=1}^{n^2} \alpha_{i, \bar{a}} \text{tr } T_{\bar{b}} Q_i,$$

also die Zeilen von \mathcal{P}_g gerade Linearkombinationen von Zeilen von Π_g sind, woraus die Endlichkeit des Rangs von \mathcal{P}_g folgt. \square

Es bleibt zu zeigen, dass umgekehrt jeder endlich-dimensionale stochastische Vektor als von einem Quanten-Prognose-Modell induzierter Vektor rekonstruiert werden kann. Das ist tatsächlich möglich:

SATZ 8.11. *Sei $g \in l_{SM}^{\Sigma}$ ein stochastischer Vektor, so dass $\dim g =: n < \infty$. Sei weiter $(g_i)_{i=1, \dots, n}$ eine Basis des Prognoseriums \mathcal{V}_g , wobei die g_i als stochastische Vektoren gewählt werden. Dann sei S^n der zu $V := \mathbb{C}^n$ assoziierte Zustandsraum und $\mathcal{V} \subset S^n$ der lineare Unterraum der Diagonalmatrizen aus S^n . Seien weiter die Koeffizienten α_{aij} über die Beziehung*

$$\tau_a g_i =: \sum_{j=1}^n \alpha_{aij} g_j$$

gegeben. Wir setzen nun

$$D_i := \text{diag}(0, \dots, 0, \frac{1}{i}, 0, \dots, 0),$$

und definieren nun die Operatoren $T_a : \mathcal{V} \rightarrow \mathcal{V}$ für $a \in \Sigma$ auf der Basis $(D_i)_{i=1, \dots, n}$ von \mathcal{V} durch die Vorschrift:

$$T_a(D_i) := \sum_{j=1}^n \alpha_{aij} D_j,$$

was durch lineare Fortsetzung auf die Gleichung

$$T_a(Q) := \text{diag}\left(\sum_{i=1}^n \alpha_{ai1} Q_{ii}, \dots, \sum_{i=1}^n \alpha_{ain} Q_{ii}\right).$$

Sind weiter die Koeffizienten α_{0i} durch die Beziehung

$$g =: \sum_{i=1}^n \alpha_{0i} g_i$$

gegeben, so setzen wir

$$Q_0 := \text{diag}(\alpha_{01}, \dots, \alpha_{0n}).$$

Dann ist das Tupel $\mathcal{Q}_g := (\mathcal{V}, (T_a)_{a \in \Sigma}, Q_0)$ ein Quanten-Prognose-Modell, so dass der durch \mathcal{Q}_g induzierte stochastische Vektor gerade g entspricht.

BEWEIS. Wir bemerken zunächst, dass die T_a offensichtlich linear sind. Aus

$$\text{tr} Q_0 = \sum_{i=1}^n \alpha_{0i} = \sum_{i=1}^n \alpha_{0i} \underbrace{g_i(\square)}_{=1} = g(\square) = 1$$

folgt (8.4). Weiter gilt für $Q \in \mathcal{V}$ (und $\Psi = \sum_a T_a$)

$$\begin{aligned} \text{tr} \Psi(Q) &= \text{tr} \sum_{a \in \Sigma} T_a(Q) = \sum_{a \in \Sigma} \text{tr} T_a(Q) \\ &= \sum_{a \in \Sigma} \sum_{i=1}^n \sum_{j=1}^n \alpha_{aij} Q_{ii} = \sum_{i=1}^n Q_{ii} \sum_{a \in \Sigma} \sum_{j=1}^n \alpha_{aij} \\ &= \sum_{i=1}^n Q_{ii} \sum_{a \in \Sigma} \sum_{j=1}^n \alpha_{aij} g_j(\square) = \sum_{i=1}^n Q_{ii} \sum_{a \in \Sigma} \tau_a g_i(\square) \\ &= \sum_{i=1}^n Q_{ii} \sum_{a \in \Sigma} g_i(a) \stackrel{\sigma g = g}{=} \sum_{i=1}^n Q_{ii} g_i(\square) \\ &= \sum_{i=1}^n Q_{ii} = \text{tr} Q \end{aligned}$$

und damit (8.5). Sei nun $\bar{a} = a_1 \dots a_t \in \Sigma^t$. Wir werden nun zeigen, dass

$$\text{tr} T_{\bar{a}}(Q_0) = \text{tr} T_{a_t} \circ \dots \circ T_{a_1}(Q_0) = g(\bar{a}) \in [0, 1] \quad (8.7)$$

gilt, woraus (8.6) als auch der Rest der Behauptung folgen. Dafür schreiben wir für ein Wort $\bar{a} \in \Sigma$ und einen Vektor $h \in \mathcal{V}_g$

$$\tau_{\bar{a}} h = (\tau_{\bar{a}} h)_i g_i,$$

also die $(\tau_{\bar{a}} h)_i$ die Koeffizienten der Darstellung der $\tau_{\bar{a}} h$ über der Basis (g_i) und zeigen allgemeiner

$$(T_{\bar{a}} Q_0)_{ii} = (\tau_{\bar{a}} g)_i, \quad (8.8)$$

woraus die Behauptung wegen

$$\text{tr } T_{\bar{a}}(Q_0) = \sum_i (\tau_{\bar{a}}g)_i = \sum_i (\tau_{\bar{a}})_i g_i(\square) = \tau_{\bar{a}}g(\square) = g(\bar{a})$$

folgt. (8.8) zeigen wir nun mit vollständiger Induktion über t . Für $t = 1$ und $a \in \Sigma$ gilt (beachte: $\alpha_{aj} = (\tau_a g_j)$)

$$\begin{aligned} (T_a Q_0)_{ii} &= \sum_{j=1}^n \alpha_{aj} \alpha_{0j} = \sum_{j=1}^n \alpha_{0j} (\tau_a g_j)_i \\ &= \sum_{j=1}^n (\tau_a (\alpha_{0j} g_j))_i = \tau_a \left(\sum_{j=1}^n \alpha_{0j} g_j \right)_i = (\tau_a g)_i, \end{aligned}$$

womit der Anfang gemacht wäre. Sei nun $t \geq 1$ und $\bar{a} = a_1 \dots a_t a_{t+1} \in \Sigma^{t+1}$. Dann

$$\begin{aligned} (T_{\bar{a}} Q_0)_{ii} &= (T_{a_{t+1}} (T_{a_1 \dots a_t} Q_0))_{ii} = \sum_{j=1}^n \alpha_{a_{t+1} j} (T_{a_1 \dots a_t} Q_0)_{jj} \\ &\stackrel{(IV)}{=} \sum_{j=1}^n \alpha_{a_{t+1} j} (\tau_{a_1 \dots a_t} g)_j = \sum_{j=1}^n (\tau_{a_{t+1}} (\tau_{a_1 \dots a_t} g)_j g_j)_i \\ &= (\tau_{a_{t+1}} \left(\sum_{j=1}^n (\tau_{a_1 \dots a_t} g)_j g_j \right))_i = (\tau_{\bar{a}} g)_i. \end{aligned}$$

□

KOROLLAR 8.12. *Die endlich-dimensionalen stochastischen Prozesse entsprechen gerade der Klasse der durch Quanten-Prognose-Modelle induzierten Prozesse.*

BEMERKUNG 8.13. *Wir bemerken, dass im Falle der Existenz eines echten, polyedrischen Kegels $K \subset \mathcal{V}_g$, der von den Prognoseoperatoren τ_a invariant gelassen wird, also gemäß Satz 3.44 im Falle eines Hidden-Markov-Vektors g , die oben gewählte Basis (g_j) durch ein Erzeugendensystem (h_k) dieses Kegels ersetzt werden kann. Es gilt dann $T_a(\mathcal{V}_+) \subset \mathcal{V}_+$ für alle a .*

8.4. Quanten-Markov-Ketten

Ein Beispiel für Quanten-Prognose-Modelle sind Quanten-Markov-Ketten, die man, wenn man so will, in Analogie zum Verhältnis der endlich-dimensionalen stochastischen Prozesse zu den Hidden-Markov-Modellen auffassen kann. Wir geben im Folgenden die Definition an und zitieren einen Ergodensatz für diese Klasse stochastischer Prozesse.

DEFINITION 8.14. *Ein Quanten-Prognose-Modell $(\mathcal{V}, (T_a)_{a \in \Sigma}, Q_0)$ mit $Q_0 \in \mathcal{V}_+ := \mathcal{V} \cap \mathcal{S}_+^n$ über einem endlichen Alphabet Σ nennen wir **Quanten-Markov-Modell**, falls zusätzlich zu den definierenden Eigenschaften eines*

Quanten-Prognose-Modells gilt:

$$\forall a \in \Sigma : T_a(\mathcal{V}_+) \subset \mathcal{V}_+. \quad (8.9)$$

BEMERKUNG 8.15. *Tatsächlich ersetzt Eigenschaft (8.9) Eigenschaft (8.4), da für alle nicht-negativen Zustände $Q \in \mathcal{S}_+^n$ gilt, dass $\text{tr } Q \geq 0$ (siehe Bemerkung 8.4), denn mit Eigenschaft (8.5) folgt*

$$1 = \text{tr } Q_0 = \text{tr } (\Psi)^t Q \stackrel{B.8.9}{=} \sum_{\bar{a} \in \Sigma^t} \text{tr } T_{\bar{a}}(Q_0)$$

und damit wegen $T_{\bar{a}}(Q_0) \subset \mathcal{V}_+$ durch iterierte Anwendung von (8.9) und $\text{tr } Q \geq 0$ für $Q \in \mathcal{V}_+$ (Bemerkung 8.4)

$$\text{tr } T_{\bar{a}}(Q) \in [0, 1].$$

Deshalb sind auch alle Tupel $(\mathcal{V}, (T_a)_{a \in \Sigma}, Q)$, so dass $Q \in \mathcal{V}_+$ und $\text{tr } Q = 1$ Quanten-Markov-Modelle.

Für Quanten-Markov-Modelle lässt sich nun eine Art Ergodensatz beweisen, die solchen Modellen starke ergodische Eigenschaften zusichert. Dafür benötigen wir noch die folgende Definition:

DEFINITION 8.16. *Wir nennen ein Quanten-Prognose-Modell*

$$\mathcal{Q} = (\mathcal{V}, (T_a)_{a \in \Sigma}, Q_0)$$

evolutionssummierbar, falls der Limes

$$\bar{Q}_0 := \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=0}^{n-1} \Psi^t(Q_0)$$

*existiert. Den Limes \bar{Q}_0 nennen wir dann auch **Grenzzustand** von \mathcal{Q} .*

Lange Rede, kurzer Sinn:

SATZ 8.17. *Ein Quanten-Markov-Modell $\mathcal{Q} := (\mathcal{V}, (T_a)_{a \in \Sigma}, Q_0)$ ist evolutionssummierbar. Der resultierende Grenzzustand \bar{Q}_0 von \mathcal{Q} ist ein Eigenvektor von Ψ zum Eigenwert $\lambda = 1$.*

BEWEIS. Das sich anschließende Lemma wird zeigen, dass der Evolutionsoperator Ψ stabil ist. Daher folgt der Satz mit Hilfe der Äquivalenz von Stabilität und Summierbarkeit, siehe Satz 10.14. \square

Hier der fehlende Baustein:

LEMMA 8.18. *Sei $\mathcal{Q} := (\mathcal{V}, (T_a)_{a \in \Sigma}, Q_0)$ ein QMM über einem Alphabet Σ . Dann ist der Evolutionsoperator $\Psi = \sum_{a \in \Sigma} T_a$ stabil.*

BEWEIS. Sei $\|\cdot\|_1$ die durch Definition 8.5 gegebene Schatten-1-Norm. Wir werden zeigen, dass bezüglich der durch diese Norm auf \mathcal{V} induzierte Operatornorm (die wir ebenfalls mit $\|\cdot\|_1$ bezeichnen)

$$\|\Psi\|_1 = 1$$

gilt. Sei dazu zunächst $Q \in \mathcal{V}_+$. Dann wegen $\Psi Q \in \mathcal{V}_+$, Bemerkung 8.6 und Eigenschaft (8.5)

$$\|\Psi Q\|_1 = \text{tr } \Psi Q = \text{tr } Q = \|Q\|_1.$$

Ist nun $Q \in \mathcal{V}$, so gemäß der Bemerkungen 8.6, 8.7

$$\begin{aligned} \|\Psi Q\|_1 &\stackrel{(8.2)}{=} \|\Psi(Q^+ - Q^-)\|_1 \leq \|\Psi Q^+\|_1 + \|\Psi Q^-\|_1 \\ &= \|Q^+\|_1 + \|Q^-\|_1 \stackrel{(8.1)}{=} \text{tr } Q^+ + \text{tr } Q^- \stackrel{(8.3)}{=} \|Q\|_1, \end{aligned}$$

woraus insgesamt die Behauptung des Lemmas folgt. \square

KOROLLAR 8.19. *Ist $\mathcal{Q} := (\mathcal{V}, (T_a)_{a \in \Sigma}, Q_0)$ ein QMM, so ist auch $\mathcal{Q}_Q := (\mathcal{V}, (T_a)_{a \in \Sigma}, Q)$ ein QMM für jede nicht-negative Dichtematrix $Q \in \mathcal{V}_+$. Daher existieren alle Limites*

$$\bar{Q} := \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=0}^{n-1} \Psi^t Q.$$

Insbesondere existieren die Limites

$$\overline{T_{\bar{a}} Q_0} = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=0}^{n-1} \Psi^t T_{\bar{a}} Q_0$$

für alle Wörter $\bar{a} \in \Sigma$.

BEMERKUNG 8.20. *Für ein allgemeines Quanten-Prognose-Modell kann eine zu Satz 8.17 analoge Aussage nicht erzielt werden. Für diese Modelle gilt jedoch noch immer, dass die durch sie induzierten endlich-dimensionalen stochastischen Prozesse evolutionssummierbar sind, siehe dazu Satz 4.43.*

8.5. Beispiele von Quanten-Markov-Ketten

Einige prominente stochastische Objekte können nun als Quanten-Markov-Ketten interpretiert werden.

8.5.1. Hidden-Markov-Prozesse. Unter Beachtung von Bemerkung 8.13 zu Satz 8.11 stellen wir fest, dass Hidden-Markov-Prozesse durch Quanten-Markov-Modelle kodiert werden können.

8.5.2. Evolution von Wellenfunktionen. Ein Vektor $\psi \in \mathbb{C}^n$ wird in der Physik im Falle von $(\langle \cdot, \cdot \rangle)$ bezeichne das Standardskalarprodukt des \mathbb{C}^n)

$$\|\psi\|_2 = \sqrt{\langle \psi, \psi \rangle} = 1$$

auch als **Wellenfunktion** bezeichnet. Ist $U \in \mathbb{C}^{n \times n}$ eine unitäre Matrix, so ist auch $\psi_t := U^t \psi$ eine Wellenfunktion für alle $t \in \mathbb{N}$. Wellenfunktionen ψ kann man nun über die Beziehung (fasse $\psi \in \mathbb{C}^{n \times 1}$ als Matrix auf)

$$Q_\psi := \psi \psi^*$$

mit selbst-adjungierten Matrizen identifizieren. Für die Wellenfunktion ψ_t gilt dann

$$Q_{\psi_t} = U^t \psi (U^t \psi)^* = U^t \psi \psi^* (U^t)^* = U^t \psi \psi^* (U^*)^t = U^t Q_\psi (U^*)^t.$$

Dieses Modell lässt sich durch die Definition

$$T_a Q := \Psi Q := U Q U^*$$

für alle $Q \in \mathcal{S}^n$ mit einem Quanten-Markov-Modell über dem Alphabet $\{a\}$ auffassen (beachte, dass offensichtlich $T_a(\mathcal{S}_+^n) \subset \mathcal{S}_+^n$ nach Definition der T_a gilt).

8.5.3. Quanten-Irrfahrten auf Graphen. In [AAKV02] wird das folgende Modell einer Quanten-Irrfahrt auf einem Graphen entworfen.

DEFINITION 8.21. *Ein Tripel (G, U, ψ) bestehend aus einem gerichteten, d -regulären Graphen $G = (V, E)$, einem unitären linearen Operator $U : \mathbb{C}^N \rightarrow \mathbb{C}^N$ mit $N := d \cdot \text{card}(V)$ und einer Wellenfunktion $\psi \in \mathbb{C}^N$ wird **Quanten-Irrfahrt (auf G)** genannt, falls die folgenden Eigenschaften vorliegen:*

- (a) *Es gibt eine Menge A mit $\text{card}(A) = d$ und eine Abbildung $c : E \rightarrow A$, so dass die Mengen*

$$c^{-1}(a) \subset E \subset V \times V$$

für alle $a \in A$ als Permutation der Knoten aufgefasst werden können.

- (b) *Identifiziert man die Vektoren $e_{(v,a)}$ der Standardorthonormalbasis von \mathbb{C}^N mit Paaren von Knoten und Markierungen $(v, a) \in V \times A$, so gilt*

$$U e_{(v,a)} \in \text{span}\{e_{(v',a')} \mid v' \in N(v) \cup \{v\}\},$$

wobei $N(v)$ die Menge der von v aus erreichbaren Knoten von V ist.

Die Evolution der Wellenfunktion $\psi \in \mathbb{C}^N$ sei nun wie oben in Abschnitt 8.5.2 durch U (d.h. $\psi_t = U^t \psi$) beschrieben. Die Wahrscheinlichkeit, sich zum Zeitpunkt $t \in \mathbb{N}$ in einem Knoten v aufzuhalten, wird dann definiert durch

$$p_t(v \mid \psi) := \sum_{a \in A} |\langle e_{(v,a)}, \psi_t \rangle|^2,$$

wobei mit $\langle \cdot, \cdot \rangle$ wiederum das Standardskalarprodukt auf dem \mathbb{C}^N gemeint sei. Aharonov et. al. ([AAKV02]) zeigen dann, dass die Limites

$$\bar{p}(v, \psi) := \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=0}^{n-1} p_t(v | \psi) \quad (8.10)$$

für alle $v \in V$ existieren.

Wir sind nun in der Lage, ein Quanten-Markov-Modell zu basteln, in welchem die Wahrscheinlichkeiten $p_t(v | \psi)$ auf natürliche Art und Weise ausgelesen werden können. Die Existenz der Limites (8.10) ergibt sich dann als einfaches Korollar von Satz 8.17.

$$\begin{aligned} p_t(v | \psi) &= \sum_{a \in A} |\langle e_{(v,a)}, \psi_t \rangle|^2 = \langle P_v \psi_t, P_v \psi_t \rangle \\ &= \text{tr } P_v \psi_t (P_v \psi_t)^* = \text{tr } T_v \Psi^t Q_\psi \end{aligned}$$

PROPOSITION 8.22. *Sei $\mathcal{W} := (G, U, \psi)$ eine Quanten-Irrfahrt mit $N = d \cdot \text{card}(V)$. Sei weiter*

$$P_v : \begin{array}{ccc} \mathbb{C}^N & \longrightarrow & \mathbb{C}^N \\ \sum_{(v',a)} \alpha_{(v',a)} e_{(v',a)} & \longmapsto & \sum_{\substack{(v',a) \\ v'=v}} \alpha_{(v',a)} e_{(v',a)} \end{array}$$

die Projektion auf den linearen Unterraum von \mathbb{C}^N , der zu einem Knoten $v \in V$ "gehört". Darauf aufbauend betrachten wir

$$T_v : \begin{array}{ccc} \mathcal{S}^N & \longrightarrow & \mathcal{S}^N \\ Q & \longmapsto & P_v U Q (P_v U)^*, \end{array}$$

offensichtlich ein linearer Operator auf \mathcal{S}^N . Dann ist

$$\mathcal{Q}_{\mathcal{W}} := (\mathcal{S}^N, (T_v)_{v \in V}, Q_\psi = \psi \psi^*)$$

ein Quanten-Markov-Modell über dem Alphabet $\Sigma = V$ der Knoten, so dass

$$p_t(v | \psi) = \text{tr } T_v \Psi^t Q_\psi \quad (8.11)$$

für die oben eingeführten Wahrscheinlichkeiten gilt.

BEWEIS. Wir stellen zunächst fest, dass gemäß obiger Notationen wegen $\sum_{v \in V} P_v = \text{Id}_{\mathbb{C}^N}$

$$\Psi Q = \sum_{v \in V} T_v Q = \sum_{v \in V} P_v U Q (P_v U)^* = U Q U^*$$

gilt. Zunächst stellen wir fest, dass es sich bei $\mathcal{Q}_{\mathcal{W}}$ tatsächlich um ein Quanten-Markov-Modell handelt. Offensichtlich handelt es sich bei Q_ψ nach Definition einer Wellenfunktion um eine Dichtematrix, also $Q_\psi \in \mathcal{S}_+^N$ sowie $\text{tr } Q_\psi = 1$. Desweiteren gilt $\text{tr } \Psi Q = \text{tr } U Q U^* = \text{tr } Q$ und offensichtlich nach Definition $T_v(\mathcal{S}_+^N) \subset \mathcal{S}_+^N$, womit gemäß Bemerkung 8.15 die definierenden Eigenschaften eines Quanten-Markov-Modells überprüft wären.

Zum Beweis von Gleichung (8.11) rechnen wir (sei $\psi_t =: \sum_{(v,a)} \alpha_{(v,a)} e_{(v,a)}$)

$$\begin{aligned}
 p_t(v | \psi) &= \sum_{a \in A} |\langle e_{(v,a)}, \psi_t \rangle|^2 = \sum_{a \in A} |\alpha_{(v,a)}|^2 \\
 &= \sum_{a \in A} \alpha_{(v,a)}^* \alpha_{(v,a)} = \left\langle \sum_{a \in A} \alpha_{(v,a)} e_{(v,a)}, \sum_{a \in A} \alpha_{(v,a)} e_{(v,a)} \right\rangle \\
 &= \langle P_v \psi_t, P_v \psi_t \rangle = \text{tr } P_v \psi_t (P_v \psi_t)^* \\
 &= \text{tr } T_v \Psi^t Q_\psi,
 \end{aligned}$$

wobei wir bemerken, dass für $z \in \mathbb{C}^N$ gerade $\text{tr } z z^* = \langle z^*, z \rangle$ gilt. \square

Nun ist offenbar, dass Existenz der Limites (8.10) als Korollar von Satz 8.17 erhalten wird.

Zusammenfassung und Ausblick

*Ich danke für die Möglichkeit,
hier haben sprechen zu können.
Ich will nicht recht behalten -
diese Sätze sind für die Diskussion da.*

Kurt Tucholsky

9.1. Zusammenfassung

In dieser Arbeit haben wir ein neues Modell zur Betrachtung von diskretwertigen, diskretzeitigen stochastischen Prozessen vorgestellt. Wir hatten dabei im Auge, die Klasse der endlich-dimensionalen stochastischen Prozesse hinsichtlich einer Reihe von Kriterien zu überprüfen, die meist mit ergodentheoretischen Begriffen in Zusammenhang stehen. Dies taten wir mit dem Fernziel, für die Unterklasse der “hochgradig nicht-stationären” endlich-dimensionalen Prozesse zunächst eine mit der Intuition konsistente Definition liefern zu können und schließlich Lernalgorithmen für diese Unterklasse zu konstruieren.

Wir haben auf dem Weg dahin Etliches mehr erreicht. Tatsächlich ist die Klassifizierung endlich-dimensionaler Prozesse nun deutlich besser möglich. Wir schlagen vor, die entsprechenden Definitionen analog zu denen der Markov-Prozesse vorzunehmen (siehe Abschnitt 5.3). Diese orientieren sich meist an den Eigenschaften der Übergangsmatrizen. In Kapitel 4 haben wir gezeigt, dass mit dem Studium der Eigenschaften des Evolutionsoperators entsprechende Ergebnisse hervorgebracht werden. Essentielles Argument für diesen Ratschlag ist z.B. die ergodische Zerlegung aus Kapitel 5.

Über die Fähigkeit hinaus, endlich-dimensionale Prozesse bezüglich etlicher Begriffe aus der Ergodentheorie nun griffig charakterisieren zu können, haben wir gezeigt, mit unserer Form der Modellierung weitere Fragestellungen in Angriff nehmen zu können. In Kapitel 2 haben wir erkannt, dass endliche, signierte Prämaße über einem diskretwertigen Sequenzraum eindeutige Fortsetzungen genießen und daher das Studium ihrer Vektorräume mit dem Studium der Sequenzen über den abzählbar vielen Wörtern eines endlichen Alphabets äquivalent ist. Wir haben dabei bemerkt, dass eine Formulierung der Norm der totalen Variation für endliche, signierte Prämaße in unserem Modell viel einfacher möglich ist. “Schwache Skorokhod” Konvergenz lässt sich deshalb mit den Begriffen unseres Modells sehr natürlich ausdrücken.

Wir haben durch die verbesserte Formulierung der Prognoseoperatoren in Kapitel 3 bereits vorhandene Versuche, ähnliche Modelle aufzusetzen, deutlich vereinfacht. Dies gelingt, indem wir die betrachteten Vektorräume als die Untervektorräume größerer und durch die Ergebnisse aus Kapitel 2 wohlbelegter Vektorräume aufgefasst haben. Dadurch ist es uns gelungen, das Identifikationsproblem und das ihm verwandte Problem des Erkennens von Hidden-Markov-Modellen, sicherlich der Anlass für alle derartigen Modellierungsversuche, deutlich elementarer niederzuschreiben.

In Kapitel 4 haben wir, neben vielen anderen Zusammenhängen unter den bis dahin eingeführten Definitionen, erkannt, dass die Norm der totalen Variation zur Überprüfung von ergodischen Eigenschaften ein nur vermeintlich zu starkes Kriterium etabliert (siehe Satz 4.41 in Appendix B). Wir konnten desweiteren zeigen, dass die in der Ergodentheorie üblichen Begriffe für endlich-dimensionale Prozesse meist sehr prägnant formuliert werden können. Die Konsequenz der elementaren Formulierung ist die ergodische Zerlegung aus Kapitel 5, die in Analogie zu gewöhnlichen Markov-Prozessen aufgefasst werden kann.

In Kapitel 6 nutzen wir den soeben angesprochenen Satz 4.41, um einen elementaren Beweis für die Existenz der Entropieraten für im Mittel asymptotisch stationäre stochastische Prozesse anzugeben, was als Verallgemeinerung des Satzes von Kolmogorov-Sinai aufgefasst werden kann.

Als Schlusspunkt dieses Teils der Arbeit sind in den Kapiteln 7 und 8 zwei weitere mögliche Formen der Darstellung endlich-dimensionaler Prozesse ins Leben gerufen worden. Die erste von ihnen ist (erfolgreich) um eine Anbindung an Konzepte der Graphentheorie und deren Begriffe bemüht. Tatsächlich ergibt sich, wiederum analog zur Theorie der Markov-Ketten, eine interessante Charakterisierung ergodischer Prozesse in der Sprache der hier vorgeführten Graphen.

Der zweiten Form der Darstellung gelingt es, statistische Konzepte aus der Quantenmechanik zu integrieren und dort bereits vorhandene Aussagen ergodentheoretischer Natur zu verschärfen.

Darüber hinaus soll die große Menge illustrierender Beispiele, die die geschaffenen Begriffe gegeneinander abgrenzt, nicht unerwähnt bleiben. Das Design von Gegenbeispielen ist meist eine überaus komplizierte und aufwändige Arbeit. Nichts desto trotz scheint sie oft unbefriedigend, da sie oft dem eigenen Wunschenken zuwider läuft.

Zu guter Letzt wollen wir die drei Kapitel des zweiten Teils, die der linearen Algebra zuzurechnen sind, hier nicht außer Acht lassen. Sie liefern nicht nur eine komprimierte Zusammenfassung der Stabilitätstheorie, sondern bereichern sie auch um etliche Kleinigkeiten.

9.2. Ausblick und offene Probleme

Natürlich sind die hier vorgestellten Arbeiten keinesfalls an ein besiegelndes Ende gelangt. Vielmehr kann die hier vorgestellte Schrift als ein Schnappschuss des gegenwärtigen Stands unserer Forschungen aufgefasst werden.

9.2.1. Lernalgorithmen. Ungelöste Fragestellungen sind z.B.

- Die konkrete Einbindung der hier gewonnenen Kenntnisse in das Jägersche Rahmenwerk zum Bau von Lernalgorithmen: Wie lassen sich die anhand der Spektralzerlegung des Evolutionsoperators erkennbaren “ergodischen Verhältnisse” in Form von Nebenbedingungen in die Optimierungsalgorithmen einbinden?
- Sind möglicherweise modifizierte Techniken, z.B. alternative Auswahlen von Basen, besser zur Konstruktion für robuste Schätzer geeignet?
- Lässt sich ein Zusammenhang zwischen der vom Autor vorgenommenen, von Oberstein [Obe02] vorgeschlagenen Implementierung von generalisierten Suffix-Trees (nicht in dieser Arbeit besprochen) und den sich durch diese Arbeit aufdrängenden Kriterien erkennen?

9.2.2. Modell allgemein. Fragestellungen rund um den Ausbau des Modells bzw. unaufgelöster Zusammenhänge zwischen den Begrifflichkeiten des Modells sind

- Die Einbeziehung der Modellierung einzelner Sequenzen im Rahmen des Modells: Dies lieferte möglicherweise einfachere Varianten von Beweisen für “fast überall”-Sätze, wie z.B. den berühmten Satz von Birkhoff.
- Ein Ausbau der Zusammenhänge zwischen den Begriffen der Stationarität und der Ergodizität auf der einen Seite und der Dimensionalität auf der anderen Seite. Konkret z.B.: Gibt es einen unendlich-dimensionalen, stationären und ergodischen stochastischen Prozess? Der Autor vermutet ja, die Konstruktion eines entsprechenden Beispiels ist ihm jedoch nicht gelungen.
- Eine Theorie der Approximation von allgemeinen Prozessen durch endlich-dimensionale Prozesse. Unter welchen, möglichst gutmütigen Voraussetzungen lässt sich ein Prozess durch endlich-dimensionale Prozesse approximieren?
- Schließlich die Erweiterung des Formalismus hin zu beliebigen Alphabeten. Unter welchen Bedingungen kann man entsprechende Dimensionsbegriffe für allgemeinere dynamische Systeme erhalten? Lässt sich Satz 4.41 auf beliebige Alphabete verallgemeinern? Hier haben wir bereits Verallgemeinerungen in Arbeit.
- Lässt sich das Positivitätsproblem nun lösen? Siehe die hinweisende Bemerkung 4.77.

9.2.3. Analogien zu Markov-Prozessen. Lassen sich Begriffe aus der Welt der Markov-Ketten analog formulieren und dementsprechende Ergebnisse erhalten?

- Lassen sich Begriffe wie z.B. “schnell mischend”, “rekurrent” oder auch “aperiodisch” für endlich-dimensionale Prozesse vernünftig ins Spiel bringen? Wir denken, dass dies auf sinnvolle Weise möglich ist.
- Lassen sich z.B. die graphischen Darstellungen endlich-dimensionaler Prozesse einer Analyse bezüglich Begriffen der Spektralgraphentheorie unterziehen wie es z.B. für Markov-Ketten mit den “Cheeger Bounds” [Chu97] geschehen ist? Entsprechende Ergebnisse würden

auch Zusammenhänge zwischen der Besetzung der Matrizendarstellungen der Prognoseoperatoren und den “Mischraten” der endlich-dimensionalen Prozesse aufzeigen.

9.2.4. Quantenrechnen und Quanteninformationstheorie. Offene Fragen hierzu betreffen Modellerweiterungen der Quanten-Prognose-Modelle in Analogie zum hier vorgestellten Modell.

- Wie lässt sich der Formalismus der Quanten-Prognose-Modelle ausbauen, um z.B. auf ähnliche Weise das Quantenrechnen zu modellieren?
- Lässt sich das Modell in kompletter Analogie auf entsprechende Quanten-Modelle übertragen, um z.B. die hier vorgestellte Beweisvariante für die Existenz der Entropieraten wiederzuerhalten? Welche Quanten-Systeme betreffe diese Beweisvariante? Der Autor vermutet, dass entsprechende Ergebnisse z.B. für die in den Arbeiten [Bje04],[Szk04] behandelten Quanten-Spin-Gitter-Systeme erhalten werden können.

Uns bleibt zu erwähnen, dass dies beileibe nicht alle ungelösten und interessanten Fragen waren. Spontan ist z.B. der in [Obe02] aufgestellte Katalog von Fragen zu nennen, von denen in dieser Arbeit schon einige gelöst werden konnten.

Teil 2

Lineare Algebra: Stabilitätstheorie, Kegel und Komponentensummen

Stabilitätstheorie

Dieses Kapitel beinhaltet verschiedene Ergebnisse der linearen Algebra, die meist der Stabilitätstheorie zugeordnet werden können. Hauptergebnisse dieses Kapitels sind die Entwicklung einer reellwertigen Stabilitätstheorie parallel der komplexwertigen Theorie von Brayton und Tong sowie die Einsicht, dass Stabilität äquivalent zur Existenz von Cesaro-Limites ist. Darüberhinaus werden die Begriffe des (asymptotischen) Folgenrangs eines linearen Operators und der (starken) Folgenstabilität eingeführt und in einen Zusammenhang gestellt. Hauptergebnis hier ist die Aussage, dass der Folgenrang eines linearen Operators gleich dem Grad seines Minimalpolynoms ist. Im weiteren Verlauf enthält dieses Kapitel Aussagen über Matrizen, die Kegel und / oder Komponentensummen invariant lassen. Im Licht der vorhergehenden Stabilitätstheorie lassen sich Zusammenhänge zwischen den einzelnen Konzepten herstellen. Hauptergebnis ist die Tatsache, dass stabile, komponentensummenerhaltende Matrizen gewisse Klassen von Kegeln invariant lassen und umgekehrt.

Der folgende Text befasst sich mit einer Stabilitätstheorie für reelle Vektorräume, die analog zu der komplexen Stabilitätstheorie von Brayton und Tong [BT79] entwickelt wird. Wir ergänzen darüberhinaus die von Brayton und Tong vorgeschlagenen Charakterisierungen im komplexen wie im reellen Fall um eine weitere äquivalente Eigenschaft.

10.1. Grundlagen und Notationen

Im weiteren Verlauf sei

$$\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}.$$

Ist $F : (V, \|\cdot\|_V) \rightarrow (W, \|\cdot\|_W)$ eine lineare Abbildung zwischen normierten Vektorräumen, so sei

$$\|F\| := \inf\{C > 0 \mid \forall v \in V : \|Fv\|_W \leq C\|v\|_V\}$$

die Operatornorm von F . Wir bemerken, dass F genau dann stetig ist, falls $\|F\| < \infty$ (siehe [MV92], Satz 5.4). Für lineare Operatoren F auf endlichdimensionalen \mathbb{K} -Vektorräumen V, W ist folglich $\|F\| < \infty$. Die Operatornorm induziert eine echte Norm auf dem Vektorraum der linearen, stetigen Abbildungen

$$L(V, W) := \{F : V \rightarrow W \mid F \text{ linear und stetig}\},$$

wobei Addition und Skalarmultiplikation auf die offensichtliche Art und Weise definiert werden. Dazu und für weitere Eigenschaften der Operatornorm siehe z.B. [MV92], S.29 ff.

DEFINITION 10.1. *Seien V und W normierte Vektorräume und $F : V \rightarrow W$ ein linearer Operator. Wir nennen F **stabil**, falls*

$$\exists K \in \mathbb{R}^+ : \quad \forall N \in \mathbb{N} \quad \|F^N\| \leq K. \quad (10.1)$$

Stetigkeit ist offensichtlich notwendige Voraussetzung für Stabilität. Da die Operatornorm im Allgemeinen unhandlich ist, ist dieses Kriterium zur Überprüfung von Stabilität nicht unbedingt geeignet. Im Falle eines Banachraums V (beachte: W muss kein Banachraum sein) jedoch können wir stabile Operatoren etwas einfacher charakterisieren. Das zeigt das folgende

LEMMA 10.2. *Seien V, W normierte Vektorräume, wobei V ein Banachraum ist (also z.B. V ein endlich-dimensionaler \mathbb{K} -Vektorraum). Dann ist ein linearer Operator $F : V \rightarrow W$ genau dann stabil, falls*

$$\forall v \in V \exists c = c(v) \in \mathbb{R}^+ : \quad \forall N \in \mathbb{N} \quad \|F^N v\| \leq c. \quad (10.2)$$

BEWEIS. Dass 10.2 aus der Stabilität folgt, ist trivial (setze $c := K \|v\|_V$) nach Definition der Operatornorm. Die andere Richtung folgt aus dem Prinzip der gleichmäßigen Beschränktheit, siehe dazu [MV92], Seite 58, Satz 8.10. \square

Für endlich-dimensionale \mathbb{K} -Vektorräume ist darüberhinaus die Definition der Stabilität unabhängig von der Auswahl einer Norm.

BEMERKUNG 10.3. *Seien V, W endlich-dimensionale \mathbb{K} -Vektorräume und $F : V \rightarrow W$ ein linearer Operator, der bezüglich einem Paar von Normen $\|\cdot\|_V, \|\cdot\|_W$ auf V, W stabil ist, so ist F bezüglich aller Normen stabil.*

BEWEIS. Das folgt unter Zuhilfenahme der Äquivalenz von Normen ([MV92], S. 34, 5.14) für endlich-dimensionale \mathbb{K} -Vektorräume und Bemerkung 10.2. \square

10.2. Stabile Matrizen

Der Vollständigkeit halber übertragen wir die Theorie noch auf Matrizen. Mit Hilfe der folgenden Bemerkungen sehen wir dann ein, dass Matrizen genau dann stabil sind, wenn es die durch sie induzierten linearen Operatoren sind. Für fundamentale Eigenschaften von Matrixnormen siehe z.B. [GVL96].

DEFINITION 10.4. *Wir bezeichnen analog eine Matrix $A \in \mathbb{K}^{n \times n}$ als **stabil**, falls es eine Matrixnorm $\|\cdot\| : \mathbb{K}^{n \times n} \rightarrow \mathbb{R}^+$ gibt, so dass*

$$\exists K \in \mathbb{R}^+ : \quad \forall N \in \mathbb{N} \quad \|A^N\| \leq K.$$

BEMERKUNG 10.5. *Da alle Normen auf dem \mathbb{K}^{n^2} äquivalent sind, findet man für eine stabile Matrix für alle Matrixnormen ein solches K .*

Dass die Definition der Stabilität für Matrizen konsistent mit der für lineare Operatoren ist, zeigt nun der folgende

SATZ 10.6. *Ist F ein Endomorphismus auf einem normierten, n -dimensionalen \mathbb{K} -Vektorraum V und A eine Matrix von F bezüglich einer Koordinatendarstellung $\Phi : V \rightarrow \mathbb{K}^n$, so ist F genau dann stabil, wenn A es ist.*

BEWEIS. Ist $\|\cdot\|$ die Norm auf V und $x \in \mathbb{K}^n$ so setzen wir $\|x\| := \|\Phi^{-1}x\|$, was eine Norm auch auf dem \mathbb{K}^n induziert. Diese wiederum induziert eine mit ihr verträgliche Matrixnorm auf $\mathbb{K}^{n \times n}$, die mit der Operatornorm für den durch die Matrix induzierten linearen Operator übereinstimmt. Dann bezüglich der induzierten Operator- bzw. Matrixnormen

$$\|A^N\| = \|(\Phi F \Phi^{-1})^N\| = \|\Phi F^N \Phi^{-1}\| \leq \|\Phi\| \cdot \|\Phi^{-1}\| \cdot \|F^N\|$$

und aus der Stabilität von F folgt die von A . Umgekehrt folgt die Stabilität von F aus der von A mit einer analogen Rechnung. \square

KOROLLAR 10.7. *Die Eigenschaft einer Matrix, stabil zu sein, bleibt unter Ähnlichkeitstransformationen erhalten.*

BEWEIS. Zu jeder Matrix gibt es Vektorräume V, W und eine lineare Abbildung $F : V \rightarrow W$, so dass die Matrix einer Koordinatendarstellung von F entspricht. Ähnlichkeitstransformationen entsprechen dann gerade Basiswechseln von V und W . Das ändert jedoch nichts an der Stabilität von F und damit ist auch die ähnliche Matrix stabil nach Satz 10.6. \square

10.3. Reelle Stabilitätstheorie

Nach diesen Vorbereitungen wollen wir nun zwei äquivalente Kriterien von Stabilität für den Fall reellwertiger, endlich-dimensionaler Vektorräume analog zu den Kriterien für komplexwertige Vektorräume gemäß der Theorie von Brayton und Tong ([BT79]) präsentieren. Wir beginnen, um den Leser nicht zu lästigem Nachschlagen zu nötigen, mit ein paar Erinnerungen an für das Verständnis des folgenden Texts nötige Erkenntnisse der linearen Algebra. Nachschlagen kann man diese in jedem guten Buch über lineare Algebra, z.B. in [Fis86], [Mey00] und [Kow75].

DEFINITION 10.8. *Ist $F : V \rightarrow V$ ein Endomorphismus auf einem endlich-dimensionalen \mathbb{K} -Vektorraum, so sei*

$$\sigma(F) := \{\lambda_1, \dots, \lambda_m\} \subset \mathbb{C}$$

die (endliche) Menge der Eigenwerte von F , auch **Spektrum** von F genannt und

$$\rho(F) := \max\{|\lambda| \mid \lambda \in \sigma(F)\}$$

der **Spektralradius** von F . Weiter sei

$$V_a(\lambda) := \dim \text{Hau}(F; \lambda),$$

die **algebraische Vielfachheit** des Eigenwerts λ , wobei

$$Hau(F; \lambda) = \bigcup_{s \geq 1} Ker(F - \lambda \cdot Id)^s$$

der Hauptraum eines Eigenwerts λ sei. Dabei beachte man, dass es ein $s_0 \in \mathbb{N}$ gibt, so dass für alle $s \geq s_0$: $Ker(F - \lambda \cdot Id)^s = Ker(F - \lambda \cdot Id)^{s_0}$. Das kleinste solche s_0 ist

$$index(\lambda) := \min\{s \in \mathbb{N} \mid Ker(F - \lambda \cdot Id)^s = Ker(F - \lambda \cdot Id)^{s+1}\}.$$

Man überprüft leicht, dass dann auch $\forall s \geq index(\lambda)$: $Ker(F - \lambda \cdot Id)^s = Ker(F - \lambda \cdot Id)^{index(\lambda)}$ gilt. Wir nennen dies den **Index** des Eigenwerts λ . Offensichtlich gilt dann

$$V_a(\lambda) = \dim Ker(F - \lambda \cdot Id)^{index(\lambda)}.$$

Ist weiter $p_F(X) = c \cdot \prod_{i=1}^m (X - \lambda_i)^{l_i}$ das über \mathbb{C} in Linearfaktoren zerlegte charakteristische Polynom von F , so gilt $V_a(\lambda_i) = l_i$. Hingegen sei

$$V_g(\lambda) := \dim Eig(F; \lambda)$$

die **geometrische Vielfachheit** eines Eigenwerts λ . Hier ist

$$Eig(F; \lambda) = Ker(F - \lambda \cdot Id)$$

der Eigenraum zum Eigenwert λ . Eigenwerte λ mit $V_a(\lambda) = V_g(\lambda)$ nennt man auch **nicht-defekt**.

Weitere Ergebnisse der linearen Algebra, z.B. die für den Beweis des folgenden Satzes wichtige Jordan-Normal-Form finden sich in Anhang A.

So vorbereitet, präsentieren wir nun die äquivalenten Kriterien stabiler Operatoren auf endlich-dimensionalen reellen Vektorräumen.

SATZ 10.9. Sei $(V, \|\cdot\|)$ ein normierter \mathbb{R} -Vektorraum mit $n := \dim V < \infty$ und $F : V \rightarrow V$ ein Endomorphismus. Dann sind die folgenden Aussagen äquivalent:

- (a) F ist stabil.
- (b) Es gibt eine abgeschlossene und beschränkte Umgebung $B \subset V$ des Ursprungs, so dass

$$FB \subset B. \tag{10.3}$$

- (c)

$$\rho(F) \leq 1, \tag{10.4}$$

und für Eigenwerte λ_i von F mit $|\lambda_i| = 1$ gilt

$$V_a(\lambda_i) = V_g(\lambda_i), \tag{10.5}$$

mit anderen Worten, solche Eigenwerte sind nicht-defekt.

BEWEIS. Wir zeigen $(a) \Rightarrow (b) \Rightarrow (c) \Rightarrow (a)$.

“(a) \Rightarrow (b)”: Sei F stabil und dementsprechend $K \in \mathbb{R}^+$ so gewählt, dass $\forall N \in \mathbb{N} : \|F^N\| \leq K$. Sei $B := \{v \in V \mid \|v\| = 1\}$ die Einheitskugel. Nach Definition der Operatornorm folgt:

$$\forall N \in \mathbb{N}, \forall v \in B : \|F^N v\| \leq K,$$

also mit $K \cdot B := \{v \in V \mid \|v\| = K\}$

$$\forall N \in \mathbb{N} : F^N B \subset K \cdot B.$$

Damit ist

$$B_F := \overline{\bigcup_{N=0}^{\infty} F^N B}$$

(der Abschluss der Vereinigung aller $F^N B$) beschränkt und wegen der Stetigkeit von F invariant unter F . Wegen $B \subset B_F$ ist B_F auch eine Umgebung der Null. Insgesamt haben wir eine abgeschlossene Umgebung der Null mit den gewünschten Eigenschaften gefunden.

“(b) \Rightarrow (c)”: Sei $B \subset V$ die abgeschlossene und beschränkte Umgebung der Null mit

$$FB \subset B.$$

“ad (10.4)”: Wir nehmen an, F hätte einen Eigenwert $\lambda \in \mathbb{C}$ mit $|\lambda| > 1$. Der Fall, dass $\lambda \in \mathbb{R}$, gibt uns einen Eigenvektor $v \in B$ (B ist eine Umgebung der Null) an die Hand und damit wegen

$$\|F^k v\| = \|\lambda^k v\| = |\lambda|^k \|v\| \xrightarrow{k \rightarrow \infty} \infty$$

einen Widerspruch. Bleibt der Fall $\lambda \in \mathbb{C} \setminus \mathbb{R}$ zu untersuchen. Nach dem Satz über die reelle Jordan-Normal-Form C.6 finden wir Vektoren $x, y \in V$, so dass $F|_W$ für $W := \text{span}\{x, y\}$ bezüglich der Basis $\{x, y\}$ durch die Matrix

$$A := \begin{pmatrix} |\lambda| \cos \phi & |\lambda| \sin \phi \\ -|\lambda| \sin \phi & |\lambda| \cos \phi \end{pmatrix}$$

beschrieben wird. Gemäß Korollar C.13 folgt

$$A^k = \begin{pmatrix} |\lambda|^k \cos(k\phi) & |\lambda|^k \sin(k\phi) \\ -|\lambda|^k \sin(k\phi) & |\lambda|^k \cos(k\phi) \end{pmatrix}$$

und F^k wird auf $\text{span}(x, y)$ durch A^k beschrieben. Mit Lemma D.3 finden wir eine Folge $(k_j)_{j \in \mathbb{N}}$, so dass $\lim_{j \rightarrow \infty} \cos(k_j \phi) = 1$ und $\lim_{j \rightarrow \infty} \sin(k_j \phi) = 0$. Damit sieht man, dass $(F^{k_j} v)$ für $v \in \text{span}(y)$ eine unbeschränkte Folge ist. Ohne Einschränkung kann man nun $v \in B$ annehmen (B ist ja eine Umgebung der Null) und erhält damit einen Widerspruch.

“ad (10.5)”: Hierzu nehmen wir die Existenz eines Eigenwerts λ mit $|\lambda| = 1$ und $V_a(\lambda) > V_g(\lambda)$ an. Der Fall $\lambda \in \mathbb{R}$, also $\lambda \in \{-1, 1\}$ führt nach dem Satz über die reelle Jordan-Normal-Form auf die Existenz linear unabhängiger

Vektoren $x, y \in V$, so dass F auf dem Teilraum $W := \text{span}\{x, y\}$ bezüglich der Basis $\{x, y\}$ durch die Matrix

$$A := \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}$$

dargestellt wird. Korollar C.11 führt auf

$$A^k := \begin{pmatrix} \lambda^k & k\lambda^{k-1} \\ 0 & \lambda^k \end{pmatrix}.$$

Wählt man nun ein $0 \neq v = \alpha y \in \text{span}\{y\} \cap B$ (B Umgebung der Null), so erhält man

$$\begin{aligned} \|F^k v\| &= |\alpha| \|F^k y\| = |\alpha| \|\lambda^k y + k\lambda^{k-1} x\| \\ &\geq |\alpha| (|k\lambda^{k-1} x| - \|\lambda^k y\|) \stackrel{|\lambda|=1}{=} |\alpha| (k\|x\| - \|y\|) \xrightarrow[k \rightarrow \infty]{} \infty, \end{aligned}$$

und damit einen Widerspruch! Im Falle $\lambda =: e^{i\phi} \in \mathbb{C} \setminus \mathbb{R}$ führt der Satz über die reelle Jordan-Normal-Form C.6 auf die Existenz linear unabhängiger Vektoren $x_0, y_0, x_1, y_1 \in V$, so dass F auf dem Teilraum $W := \text{span}(x_0, y_0, x_1, y_1)$ bezüglich der Basis $\{x_0, y_0, x_1, y_1\}$ durch die Matrix

$$A := \begin{pmatrix} \cos\phi & \sin\phi & 1 & 0 \\ -\sin\phi & \cos\phi & 0 & 1 \\ 0 & 0 & \cos\phi & \sin\phi \\ 0 & 0 & -\sin\phi & \cos\phi \end{pmatrix}$$

beschrieben wird. Korollar C.13 zeigt, dass

$$A^k = \begin{pmatrix} \cos(k\phi) & \sin(k\phi) & k \cdot \cos[(k-1)\phi] & k \cdot \sin[(k-1)\phi] \\ -\sin(k\phi) & \cos(k\phi) & -k \cdot \sin[(k-1)\phi] & k \cdot \cos[(k-1)\phi] \\ 0 & 0 & \cos(k\phi) & \sin(k\phi) \\ 0 & 0 & -\sin(k\phi) & \cos(k\phi) \end{pmatrix},$$

und F^k wird bezüglich der Basis (x_0, y_0, x_1, y_1) durch A^k beschrieben. Da B eine Umgebung der Null ist, finden wir wiederum ein $0 \neq v = \alpha y_1 \in \text{span}(y_1) \cap B$. Mit Lemma D.3 finden wir weiter eine Folge natürlicher Zahlen $(k_j)_{j \in \mathbb{N}}$, so dass $\lim_{j \rightarrow \infty} \sin(k_j \phi) = 0$ und $\lim_{j \rightarrow \infty} \cos(k_j \phi) = 1$. Dann gilt

$$\begin{aligned} \|F^{k_j+1} v\| &= |\alpha| \|F^{k_j+1} y_1\| \\ &= |\alpha| \|(k_j+1)(\sin(k_j \phi)x_0 + \cos(k_j \phi)y_0) + \sin[(k_j+1)\phi]x_1 + \cos[(k_j+1)\phi]y_1\| \\ &\geq |\alpha| (k_j+1) |\cos(k_j \phi)y_0 + \sin(k_j \phi)x_0| - |\alpha| (\|x_1\| + \|y_1\|) \end{aligned}$$

und der letzte Term verhält sich asymptotisch wie $|\alpha|(k_j+1)\|y_0\|$, eine unbeschränkte Folge, ein Widerspruch! \square

“(c) \implies (a)”: Gilt $\rho(F) < 1$, so folgt

$$\lim_{k \rightarrow \infty} F^k = \mathbf{0},$$

durch Übergang zu einer Matrixdarstellung und Ausnutzung von Satz C.9. Damit folgt die Behauptung wegen $\lim_{k \rightarrow \infty} \|F^k\| = 0$.

Wir können uns daher auf den Fall $\rho(F) = 1$ beschränken. Nach dem Satz über die reelle Jordan-Normal-Form C.6 zerfällt V in eine direkte Summe

$$V = \bigoplus_{j=1}^m W_j,$$

so dass die W_j gerade den reellen Jordan-Kästchen entsprechen (nicht den Haupträumen) und folglich $F_j := F|_{W_j}$ nur die zueinander komplex konjugierten Eigenwerte $\lambda_j, \bar{\lambda}_j$ (beachte: $\lambda_j = \bar{\lambda}_j$ im Falle $\lambda_j \in \mathbb{R}$) hat. Es gilt weiter

$$F(W_j) \subset W_j.$$

Finden wir für alle $j = 1, \dots, m$ ein K_j , so dass $\|(F_j)^N\| \leq K_j$, und fasst man die F_j als Endomorphismen von V auf, indem man $F_j|_{\oplus_{i \neq j} W_i} = \mathbf{0}$ setzt, so gilt $F = \sum_j F_j$ und weiter wegen $F_j \circ F_i = \mathbf{0}$ für $j \neq i$

$$\|F^N\| = \left\| \left(\sum_j F_j \right)^N \right\| = \left\| \sum_j (F_j)^N \right\| \leq \sum_j \|(F_j)^N\| \leq \sum_j K_j.$$

Es reicht daher zu zeigen, dass F_j stabil ist, mit anderen Worten können wir annehmen, F habe nur einen Eigenwert und die reelle Jordan-Normal-Form von F entspreche genau einem ihrer Kästchen.

Wir unterscheiden zwei Fälle. Im ersten Fall hat F nur einen reellen Eigenwert $\lambda \in \{-1, 1\}$ und wegen Eigenschaft (c) gilt dann für alle $v \in V$:

$$\|Fv\| = \|\lambda v\| = \|v\|$$

Damit folgt offensichtlich die Behauptung. Im zweiten Fall rührt das Kästchen von einem zueinander komplex-konjugierten Eigenwertpärchen $\lambda, \bar{\lambda}$ her, wobei $\lambda =: e^{i\phi}$ mit $|\lambda| = |\bar{\lambda}| = 1$ und bezüglich einer Jordan-Normal-Basis (v_1, v_2) (oBdA wählen wir v_1, v_2 mit $\|v_1\| = \|v_2\| = 1$) hat F^k die folgende Matrixgestalt (siehe Korollar C.13):

$$\begin{pmatrix} \cos(k\phi) & \sin(k\phi) \\ -\sin(k\phi) & \cos(k\phi) \end{pmatrix}.$$

Ist $v = \alpha_1 v_1 + \alpha_2 v_2$, so folgt

$$\begin{aligned} \|F^k v\| &= \|\alpha_1 F^k v_1 + \alpha_2 F^k v_2\| \leq |\alpha_1| \|F^k v_1\| + |\alpha_2| \|F^k v_2\| \\ &= |\alpha_1| \|\cos(k\phi)v_1 + \sin(k\phi)v_2\| + \\ &\quad |\alpha_2| \|\sin(k\phi)v_1 + \cos(k\phi)v_2\| \\ &\leq |\alpha_1| (\|\cos(k\phi)v_1\| + \|\sin(k\phi)v_2\|) + \\ &\quad |\alpha_2| (\|\sin(k\phi)v_1\| + \|\cos(k\phi)v_2\|) \\ &\leq 2(|\alpha_1| + |\alpha_2|) \end{aligned}$$

unabhängig von k . Die Behauptung folgt unter Zuhilfenahme von Lemma 10.2. \square

BEMERKUNG 10.10. Die oben aufgeführten Kriterien für Stabilität sind völlig analog auf komplexe Vektorräume übertragbar. Diese Theorie und eine Verallgemeinerung des Stabilitätsbegriffs auf von endlich vielen Matrizen erzeugten

Halbgruppen findet man in der Arbeit [BT79], die jedoch mit Vorsicht zu genießen ist. Ersetze dort z.B. den Begriff "simple" durch "semi-simple" oder auch "non-defective". Wir bemerken darüberhinaus, dass die Äquivalenz der Stabilität (Eigenschaft (a)) und des Eigenwertkriteriums (c) im reellen Fall eine einfache Reduktion auf den in [BT79] behandelten komplexen Fall ist, vorausgesetzt, man korrigiert die oben angesprochenen Defizite dieser Arbeit. Eigenschaft (b) jedoch lässt sich nicht durch eine einfache Reduktion einbinden.

10.4. Cesaro-Mittel und Stabilität

Der in diesem Abschnitt vorgestellte Satz charakterisiert stabile Abbildungen auf eine zusätzliche Art und Weise. In Worten ausgedrückt, besagt er, dass ein linearer Operator F genau dann stabil ist, falls die Folge der Cesaro-Mittel

$$F_n := \frac{1}{n} \sum_{k=0}^{n-1} F^k$$

konvergiert, wobei Konvergenz von linearen Operatoren hier und im Folgenden immer bezüglich der Operatornormen gemeint sei. Aus dem Beweis wird ersichtlich, dass

$$\bar{F} := \lim_{n \rightarrow \infty} F_n$$

ein linearer Operator ist, der wegen

$$\bar{F} \circ F = \bar{F} = F \circ \bar{F}$$

($\Rightarrow \bar{F}^2 = \bar{F}$) als Projektion auf den Eigenraum von F zum Eigenwert 1 aufgefasst werden kann. Zunächst formal die folgende

DEFINITION 10.11. *Sei V und W normierte Vektorräume und $F : V \rightarrow W$ ein linearer Operator. F wird **summierbar** genannt, falls der Limes*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} F^k =: \bar{F}$$

existiert.

Operatornormen sind im Allgemeinen unhandlich. Auf Banachräumen jedoch, und das sind die meisten Räume dieser Arbeit, lässt sich Summierbarkeit auf den Elementen überprüfen:

LEMMA 10.12. *In Banachräumen V, W ist die Summierbarkeit eines stetigen linearen Operators $F : V \rightarrow W$ äquivalent zu*

$$\forall v \in V : \frac{1}{n} \sum_{k=0}^{n-1} F^k v \text{ existiert.} \quad (10.6)$$

Im Allgemeinen folgt aber aus der Existenz dieser Limites nicht die Summierbarkeit des Operators.

BEWEIS. Im Falle der Summierbarkeit gilt natürlich

$$\bar{F}v = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} F^k v.$$

Umgekehrt folgt die Summierbarkeit des Operators im Falle von Banachräumen mit Hilfe des Prinzips der gleichmäßigen Beschränktheit und einem darauf aufbauenden Lemma, siehe [MV92], S. 58/59, 8.10/8.14. \square

Eine offensichtliche Folgerung ist

KOROLLAR 10.13. *Sind V und W Banachräume, $F : V \rightarrow W$ ein linearer Operator und $(v_i)_{i \in I}$ mit $\text{card}(I) = \dim V$ eine Basis von V , so ist F genau dann summierbar, falls*

$$\forall i \in I : \quad \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} F^k v \text{ existiert.}$$

Wir kommen nun zum Hauptergebnis dieses Abschnitts. Im Falle endlich-dimensionaler komplexer sowie reeller Vektorräume sind Summierbarkeit und Stabilität äquivalente Konzepte.

SATZ 10.14. *Sei $(V, \|\cdot\|)$ ein normierter \mathbb{K} -Vektorraum, so dass $n := \dim V < \infty$ und $F : V \rightarrow V$ ein Endomorphismus. Dann sind die folgenden Aussagen äquivalent:*

- (i) F ist stabil.
- (ii) F ist summierbar.

BEMERKUNG 10.15. *Wir werden im Folgenden das Eigenwertkriterium für Stabilität nutzen, um diesen Satz zu beweisen. Die Äquivalenz des Eigenwertkriteriums und der Summierbarkeit (ohne einen Zusammenhang mit der Stabilität) ist tatsächlich im Falle $\mathbb{K} = \mathbb{C}$ schon bekannt. Da dem Autor die Quelle (das zumindest im deutschsprachigen Raum ungebräuchliche Buch [Mey00], alle anderen dem Autor bekannten Bücher über Lineare Algebra verschweigen diesen Zusammenhang) zum Zeitpunkt der Entwicklung dieses Satzes unbekannt war, geben wir hier den Originalbeweis des Autors vollständig an.*

Bevor wir den Satz beweisen, betrachten wir das folgende technische

LEMMA 10.16. *Sei $\lambda \in \mathbb{C}$. Dann ist die Folge*

$$\frac{1}{n} \sum_{k=0}^{n-1} \lambda^k \tag{10.7}$$

genau dann konvergent, falls $|\lambda| \leq 1$. Weiter ist die Folge

$$\frac{1}{n} \sum_{k=1}^{n-1} k \lambda^{k-1}. \quad (10.8)$$

genau dann konvergent, falls $|\lambda| < 1$.

BEWEIS. Der erste Teil des Lemmas ist einfach: Der Fall $\lambda = 1$ ist sofort klar. Die anderen Fälle folgen aus der Rechnung

$$\begin{aligned} \left| \frac{1}{n} \sum_{k=0}^{n-1} \lambda^k \right| &= \left| \frac{1}{n} \frac{1 - \lambda^n}{1 - \lambda} \right| = \frac{1}{n} \left| \frac{1}{1 - \lambda} \right| |1 - \lambda^n| \\ &\begin{cases} \leq \frac{2}{n} \left| \frac{1}{1 - \lambda} \right| \longrightarrow 0 & |\lambda| \leq 1 \\ \geq \frac{1}{n} \left| \frac{1}{1 - \lambda} \right| (|\lambda^n| - 1) = \frac{|\lambda^n|}{n} \left| \frac{1}{1 - \lambda} \right| - \frac{1}{n} \left| \frac{1}{1 - \lambda} \right| \longrightarrow \infty & |\lambda| > 1 \end{cases} \end{aligned} \quad (10.9)$$

Für $\lambda = 1$ folgt der zweite Teil aus der Gleichung

$$\frac{1}{n} \sum_{k=1}^{n-1} k \lambda^{k-1} = \frac{1}{n} \sum_{k=1}^{n-1} k = \frac{1}{n} \frac{n(n-1)}{2} = \frac{n-1}{2}. \quad (10.10)$$

Für alle anderen λ gilt

$$\begin{aligned} \frac{1}{n} \sum_{k=1}^{n-1} k \lambda^{k-1} &= \left(\frac{1}{n} \sum_{k=0}^{n-1} \lambda^k \right)' \\ &= \left(\frac{1}{n} \frac{1 - \lambda^n}{1 - \lambda} \right)' \\ &= \frac{1}{n} \left(\frac{-n \lambda^{n-1}}{1 - \lambda} + \frac{1 - \lambda^n}{(1 - \lambda)^2} \right) \\ &= \frac{1}{n} \frac{1}{(1 - \lambda)^2} - \frac{\lambda^{n-1}}{(1 - \lambda)} \left(\frac{\lambda}{n(1 - \lambda)} + 1 \right). \end{aligned} \quad (10.11)$$

Die Konvergenz im Falle $|\lambda| < 1$ leuchtet direkt ein. Die Divergenz im Falle $|\lambda| > 1$ folgt nun aus

$$\begin{aligned} &\left| \frac{1}{n} \frac{1}{(1 - \lambda)^2} - \frac{\lambda^{n-1}}{(1 - \lambda)} \left(\frac{\lambda}{n(1 - \lambda)} + 1 \right) \right| \\ &\geq \left| \frac{\lambda^{n-1}}{(1 - \lambda)} \left| \frac{\lambda}{n(1 - \lambda)} + 1 \right| - \frac{1}{n} \frac{1}{(1 - \lambda)^2} \right| \\ &\geq \left| \frac{\lambda^{n-1}}{(1 - \lambda)} \left(1 - \left| \frac{\lambda}{n(1 - \lambda)} \right| \right) - \frac{1}{n} \frac{1}{(1 - \lambda)^2} \right| \\ &\longrightarrow \infty. \end{aligned} \quad (10.12)$$

Die Divergenz im Falle $|\lambda| = 1, \lambda \neq 1$ folgt daraus, dass

$$\frac{1}{n} \frac{1}{(1 - \lambda)^2} - \frac{\lambda^{n-1}}{(1 - \lambda)} \left(\frac{\lambda}{n(1 - \lambda)} + 1 \right) = \underbrace{\frac{1}{n} \frac{1}{(1 - \lambda)^2} - \frac{1}{n} \frac{\lambda^n}{(1 - \lambda)^2}}_{\longrightarrow 0} - \frac{\lambda^{n-1}}{1 - \lambda}, \quad (10.13)$$

die Folge sich also asymptotisch wie $\frac{\lambda^{n-1}}{1 - \lambda}$ verhält, eine offensichtlich divergente Folge.

□

Nun zum BEWEIS VON SATZ 10.14: $\mathbb{K} = \mathbb{C}$: Wir nutzen das Eigenwertkriterium 10.4, 10.5 für Stabilität, das für den komplexen Fall völlig analog zum reellen Fall äquivalent zur Stabilität ist (siehe 10.10). Wir betrachten eine Basis, bezüglich derer eine Matrixdarstellung von F Jordan-Normal-Form (siehe C.2) hat. Um Summierbarkeit zu zeigen reicht es nach Korollar 10.13, die Existenz der Limites $1/n \sum_{k=0}^{n-1} F^k v$ für Elemente v der Jordan-Normal-Basis zu zeigen. Deshalb sei im Folgenden v immer ein Element der Jordan-Normal-Basis.

Ist F nun stabil, haben wir nach dem Eigenwertkriterium die Fälle 10.4, 10.5 zu unterscheiden.

Im ersten Fall gehört v zu einem Eigenwert λ mit $|\lambda| < 1$, und

$$0 = \lim_{n \rightarrow \infty} F^n v = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} F^k v$$

nach C.9.

Gehört v zu einem Eigenwert λ mit $|\lambda| = 1$, so ist v wegen 10.5 ein Eigenvektor von λ und daher

$$\frac{1}{n} \sum_{k=0}^{n-1} F^k v = \frac{1}{n} \sum_{k=0}^{n-1} \lambda^k v = \left(\frac{1}{n} \sum_{k=0}^{n-1} \lambda^k \right) v.$$

Die Konvergenz folgt deshalb aus Gleichung (10.7) des obigen Lemmas.

Ist umgekehrt F nicht stabil, so finden wir nach dem Eigenwertkriterium entweder einen Eigenwert λ mit $|\lambda| > 1$ oder einen defekten Eigenwert λ vom Betrag 1.

Im ersten Fall folgt die Divergenz von

$$\frac{1}{n} \sum_{k=0}^{n-1} F^k v = \frac{1}{n} \sum_{k=0}^{n-1} \lambda^k v = \left(\frac{1}{n} \sum_{k=0}^{n-1} \lambda^k \right) v$$

für den entsprechenden Eigenvektor v wiederum aus Gleichung (10.7) des obigen Lemmas.

Im zweiten Fall findet man gemäß dem Satz über die Jordan-Normal-Form C.2 für λ einen Hauptvektor v und einen Eigenvektor w , so dass

$$F^k v = \lambda^k v + k \lambda^{k-1} w.$$

Weiter folgt

$$\begin{aligned} \frac{1}{n} \sum_{k=0}^{n-1} F^k v &= \frac{1}{n} \sum_{k=0}^{n-1} (\lambda^k v + k \lambda^{k-1} w) \\ &= \frac{1}{n} \sum_{k=0}^{n-1} \lambda^k v + \frac{1}{n} \sum_{k=0}^{n-1} k \lambda^{k-1} w \\ &= \left(\frac{1}{n} \sum_{k=0}^{n-1} \lambda^k \right) v + \left(\frac{1}{n} \sum_{k=0}^{n-1} k \lambda^{k-1} \right) w \end{aligned}$$

und die Divergenz des zweiten Summanden und damit des ganzen Terms folgt aus Gleichung (10.8), wiederum aus dem obigen Lemma 10.16.

$\mathbb{K} = \mathbb{R}$: Diesen Fall führen wir auf den Fall $\mathbb{K} = \mathbb{C}$ zurück. Wir betrachten die Menge

$$\tilde{V} := V \times V = \{(v_1, v_2) \mid v_1, v_2 \in V\},$$

die vermöge komponentenweiser Addition und für $a + ib \in \mathbb{C}$ durch die Skalarmultiplikation $(a + ib)(v_1, v_2) := (av_1 - bv_2, av_2 + bv_1)$ zum \mathbb{C} -Vektorraum wird (die Vektorraum-Axiome zu überprüfen, ist eine sehr einfache Rechnung). Diesen Vektorraum versehen wir mit der Norm

$$\|(v_1, v_2)\|_{\tilde{V}} := \|v_1\|_V + \|v_2\|_V$$

F induziert dann vermöge der Abbildungsregel

$$\tilde{F}(v_1, v_2) := (F(v_1), F(v_2))$$

eine \mathbb{C} -lineare Abbildung \tilde{F} auf \tilde{V} (dies zu überprüfen ist wiederum eine einfache Rechnung). Unter Ausnutzung der Bemerkungen 10.2 bzw. 10.12 stellen wir nun fest, dass F genau dann stabil bzw. summierbar ist, wenn \tilde{F} stabil bzw. summierbar ist. Die Behauptung folgt daher aus dem Teil $\mathbb{K} = \mathbb{C}$. \square

BEMERKUNG 10.17. *Wir wollen darauf hinweisen, dass wir die Theorie der stabilen Operatoren mit Blick auf die später betrachtete Ergodentheorie stochastischer Vektoren entwickelt haben. Tatsächlich benötigen wir später die Äquivalenz von Stabilität und Summierbarkeit, wobei sich die Stabilität im Sinne der ursprünglichen Definition bzw. in der Form von Lemma 10.2 äußert und Summierbarkeit im Sinne von Lemma 10.12. Für diesen speziellen Zweck hat Faigle im Falle $\mathbb{K} = \mathbb{C}$ einen kurzen, einfachen Beweis vorge schlagen. Dieser Beweis vermeidet jegliche Querverweise und benötigt nur allgemein bekannte Sätze. In dieser Arbeit jedoch sind wir um Vollständigkeit, auch in Detailfragen, bemüht - wir scheuen keine Sublemmata und keine Querverweise. Die im Beweis von Faigle verwendeten Konzepte lassen sich, detaillierter und oft als eigenständige Lemmata formuliert, ausnahmslos an anderen Stellen dieser Arbeit wiederfinden. Daher, um an dieser Stelle Redundanz zu vermeiden, geben wir hier nur eine Skizze an. Der Beweis reduziert das Problem, analog wie in Satz 10.9, Teil (c) \Rightarrow (a) zunächst mit Hilfe des Satzes über die Jordan-Normal-Form (es reicht auch der Satz über die Zerlegung in Haupträume) auf Operatoren mit nur einem Eigenwert und überprüft*

Stabilität sowie Summierbarkeit anhand der verschiedenen Fälle des Eigenwertkriteriums und kann in seiner Struktur somit durch

$$\text{“ (10.2) } \Leftrightarrow \text{ (10.4), (10.5) } \Leftrightarrow \text{ (10.6) ”}$$

beschrieben werden. Nachlesen kann man den Beweis in [FS05a].

FAZIT 10.18. *Tatsächlich scheinen die hier präsentierten vier Kriterien für Stabilität - Beschränktheit der Operatornorm, Existenz einer invarianten Umgebung der Null, Eigenwertkriterium und Summierbarkeit - noch nie in einem Atemzug und auch sonst nur im Fall $\mathbb{K} = \mathbb{C}$ genannt worden zu sein. Brayton und Tong ([BT79]) bemerken die Äquivalenz der ersten drei Kriterien im Falle $\mathbb{K} = \mathbb{C}$. In dem dem Autor zum Zeitpunkt der Entwicklung dieses Kapitels unbekanntem, weil ungebräuchlichen, Buch [Mey00] wird die Äquivalenz von Summierbarkeit und Eigenwertkriterium angegeben, ebenfalls nur im Fall $\mathbb{K} = \mathbb{C}$. Tatsächlich sind die Äquivalenz der Stabilität selber, des Eigenwertkriteriums und der Summierbarkeit für den Fall $\mathbb{K} = \mathbb{R}$ jedoch einfache Reduktionen auf den komplexen Fall. Nur die Existenz der invarianten Umgebung der Null bereitet wirklich mehr Schwierigkeiten.*

Folgenrang und Folgenstabilität

In diesem Kapitel zeigen wir Zusammenhänge zwischen dem Grad und der Beschaffenheit des Minimalpolynoms eines linearen Operators sowie Eigenschaften der durch iterierte Anwendung des linearen Operators auf einzelne Elemente entstehende Teilmengen des Vektorraums sowie die dadurch erzeugten Untervektorräume.

Diese Ergebnisse sind nichts Neues und können in Büchern, zum Teil unter dem Stichwort Krylov-Theorie, nachgelesen werden [Mey00]. Auch ältere Bücher [Jac53] führen ähnliche Ergebnisse, aus denen die unsrigen einfach abgeleitet werden können. Dem Autor ist es jedoch nicht gelungen, den hier beschrittenen, konstruktiven Weg des Beweises in Büchern oder Artikeln wiederzufinden. Wir geben diese Beweise deshalb in voller Länge an, auch weil sie die zugrundeliegende Idee mit Hilfe der Jordan-Normal-Form, einem Standarddenkmuster der linearen Algebra, entwickelt. Unser Beweischema erlaubt im Übrigen eine Ausdehnung des Begriffs des Folgenrangs auf den des asymptotischen Folgenrangs. Uns ist nicht klar, ob entsprechende Ergebnisse auf einfache Art und Weise über die in [Jac53] aufgeführten Beweiswege erhalten werden können.

11.1. Folgenrang

In diesem Abschnitt definieren wir den (asymptotischen) Folgenrang eines linearen Operators sowie den Begriff der Folgenstabilität. Wir zeigen, dass der (asymptotische) Folgenrang am Minimalpolynom des Operators abgelesen werden kann. Um lästiges Nachschlagen zu vermeiden, zitieren wir zum Einstieg die folgende

DEFINITION 11.1. *Sei V ein endlich-dimensionaler \mathbb{K} -Vektorraum und $F : V \rightarrow V$ ein Endomorphismus. Ein Polynom $m_F \in \mathbb{K}[X]$ minimalen Grades, für das*

$$m_F(F) = \mathbf{0},$$

*wird **Minimalpolynom** von F genannt (und ist bis auf konstante Faktoren eindeutig bestimmt). Ist $\sigma(F) = \{\lambda_1, \dots, \lambda_m\} \subset \mathbb{C}$ das Spektrum von F und*

$$m_F(X) = \prod_{i=1}^m (X - \lambda_i)^{l_i}$$

die Zerlegung des Minimalpolynoms in Linearfaktoren über \mathbb{C} , so gilt

$$l_i = \text{index}(\lambda_i).$$

Weitere Informationen, die den Index eines Eigenwerts und damit das Minimalpolynom in einen Zusammenhang mit der Jordan-Normal-Form stellen, finden sich in Anhang A, siehe vor allem Korollar C.5.

Nun zur namensgebenden Definition dieses Abschnitts, die mit dem Begriff des Minimalpolynoms in Abgleich gebracht werden wird.

DEFINITION 11.2. Sei V ein \mathbb{K} -Vektorraum mit $\dim V = n$, $F : V \rightarrow V$ ein linearer Operator und $v \in V$. Dann setzen wir

$$\begin{aligned} fdim_F : V &\longrightarrow \{0, \dots, n\} \\ v &\longmapsto \dim \operatorname{span}\{F^k v \mid k \in \mathbb{N}\} \end{aligned} \quad (11.1)$$

und nennen diese Abbildung die **Folgendimension bezüglich F**. Wir nennen weiter

$$\operatorname{frang}(F) := \max_{v \in V} fdim_F(v) \quad (11.2)$$

den **Folgenrang** von F .

Offensichtlich ist der Folgenrang kleiner gleich dem Rang eines linearen Operators. Sogleich stellt sich die Frage, welche Klassen von linearen Operatoren maximalen Folgenrang haben. Wir lösen diese Frage umgehend, indem wir das Hauptresultat dieses Abschnitts vorstellen. In einfachen Worten besagt es, dass der Folgenrang eines linearen Operators durch den Grad seines Minimalpolynoms gegeben ist. Diesem Satz sehr nahestehende Aussagen findet man in [Jac53], S. 66-69 oder auch in [HT73].

SATZ 11.3. Sei V ein \mathbb{K} -Vektorraum mit $n := \dim V < \infty$ und $F : V \rightarrow V$ ein Endomorphismus. Sei weiter m_F das Minimalpolynom von F . Dann gilt

$$\operatorname{frang}(F) = \deg m_F. \quad (11.3)$$

BEWEIS. Wir behandeln zunächst den Fall $\mathbb{K} = \mathbb{C}$. Wir können ohne Einschränkung annehmen (beide Größen sind unabhängig von der Auswahl einer Basis), dass F durch eine Matrix $A \in \mathbb{C}^{n \times n}$ in Jordan-Normal-Form beschrieben wird. Zunächst folgt gemäß Definition des Minimalpolynoms, dass

$$\forall v \in V : m_F(F)(v) = 0,$$

und damit für alle $v \in V$, dass $fdim_F(v) \leq \deg m_F$, also auch

$$\operatorname{frang}(F) \leq \deg m_F. \quad (11.4)$$

Im Folgenden konstruieren wir nun ein Element $v_0 \in V$, für das $fdim_F(v_0) = \deg m_F$ gilt. Zerlegt man das Minimalpolynom in Linearfaktoren

$$m_F = \prod_{i=1}^m (X - \lambda_i)^{l_i},$$

so findet man mit Hilfe des Satzes über die Jordan-Normal-Form (siehe hierzu Korollar C.5) einen linearen Unterraum von W von V , so dass $F(W) \subset W$

und bezüglich dessen A die folgende Gestalt hat:

$$A|_W = \begin{pmatrix} \begin{bmatrix} \lambda_1 & 1 & & \\ & \lambda_1 & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_1 \end{bmatrix} & & & \\ & \begin{bmatrix} \lambda_2 & 1 & & \\ & \lambda_2 & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_2 \end{bmatrix} & & & \\ & & \ddots & & & & \\ & & & \begin{bmatrix} \lambda_m & 1 & & \\ & \lambda_m & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_m \end{bmatrix} & & & \end{pmatrix},$$

wobei die λ_i paarweise verschieden sind und die Größe des Kästchens von λ_i gerade l_i ist und demnach $A|_W \in \mathbb{C}^{\deg m_F \times \deg m_F}$. Im Folgenden sei immer

$$d := \deg m_F$$

gesetzt. Wir setzen

$$w_0 = \sum_{j=1}^d \alpha_j e_j \text{ mit } \alpha_j = \begin{cases} 1 & \exists i^* \in \{1, \dots, m\} : j = \sum_{i=1}^{i^*} l_i = l_1 + \dots + l_{i^*} \\ 0 & \text{sonst} \end{cases} \quad (11.5)$$

(also $w_0 = (0, \dots, 0, \underbrace{1}_{l_1}, 0, \dots, 0, \underbrace{1}_{l_1+l_2}, 0, \dots, 0, \underbrace{1}_d, 0, \dots, 0)$), und zeigen, dass die

Vektoren

$$w_0, Aw_0, A^2w_0, \dots, A^{d-1}w_0$$

linear unabhängig sind. Mit Lemma C.12 gilt für $j = 1, \dots, d-1$ und $l = l_1, l_1+l_2, \dots, l_1+\dots+l_{m-1}, d$ (hier und im Folgenden sei $\binom{a}{b} := 0$ für $b > a$)

$$A^j e_{(l_1+\dots+l_i)} = \sum_{p=1}^{s_i} \binom{j}{l_i-p} \lambda_i^{j-(l_i-p)} e_{l_1+\dots+l_{i-1}+p} \quad (11.6)$$

und folglich

$$A^j w_0 = \sum_{i=1}^m \sum_{p=1}^{l_i} \binom{j}{l_i-p} \lambda_i^{j-(l_i-p)} e_{l_1+\dots+l_{i-1}+p}. \quad (11.7)$$

Schreibt man die Vektoren $A^j w_0, j = 0, \dots, d-1$ als Spaltenvektoren in eine Matrix B , so erhält man (wir lassen die Nullen, die zu W^\perp gehören, weg)

$$B = \begin{pmatrix} 0 & 0 & 0 & 0 & \cdots & 1 & \cdots & \binom{d-1}{s_1-1} \lambda_1^{d-(s_1-1)} \\ \vdots & \vdots & \vdots & \vdots & & \vdots & & \vdots \\ 0 & 0 & 0 & 1 & \cdots & \binom{s_1-1}{3} \lambda_1^{s_1-4} & \cdots & \binom{d-1}{3} \lambda_1^{d-4} \\ 0 & 0 & 1 & \binom{3}{2} \lambda_1 & \cdots & \binom{s_1-1}{2} \lambda_1^{s_1-3} & \cdots & \binom{d-1}{2} \lambda_1^{d-3} \\ 0 & 1 & \binom{2}{1} \lambda_1 & \binom{3}{1} \lambda_1^2 & \cdots & \binom{s_1-1}{1} \lambda_1^{s_1-2} & \cdots & \binom{d-1}{1} \lambda_1^{d-2} \\ 1 & \lambda_1 & \lambda_1^2 & \lambda_1^3 & \cdots & \lambda_1^{s_1-1} & \cdots & \lambda_1^{d-1} \\ \vdots & \vdots & \vdots & \vdots & & \ddots & & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 & \cdots & \binom{d-1}{s_m-1} \lambda_m^{d-(s_m-1)} \\ \vdots & \vdots & \vdots & \vdots & & \vdots & & \vdots \\ 0 & 0 & 0 & 1 & \cdots & \binom{s_m-1}{3} \lambda_m^{s_m-4} & \cdots & \binom{d-1}{3} \lambda_m^{d-4} \\ 0 & 0 & 1 & \binom{3}{2} \lambda_m & \cdots & \binom{s_m-1}{2} \lambda_m^{s_m-3} & \cdots & \binom{d-1}{2} \lambda_m^{d-3} \\ 0 & 1 & \binom{2}{1} \lambda_m & \binom{3}{1} \lambda_m^2 & \cdots & \binom{s_m-1}{1} \lambda_m^{s_m-2} & \cdots & \binom{d-1}{1} \lambda_m^{d-2} \\ 1 & \lambda_m & \lambda_m^2 & \lambda_m^3 & \cdots & \lambda_m^{s_m-1} & \cdots & \lambda_m^{d-1} \end{pmatrix} \in \mathbb{C}^{d \times d},$$

und es ist zu zeigen, dass B regulär ist. Sei also $\alpha \in \mathbb{C}^n$ mit $B\alpha = 0$. Es ist zu zeigen, dass $\alpha = 0$. Wir betrachten nun das Polynom

$$g(X) := \sum_{i=0}^{d-1} \alpha_i X^i \in \mathbb{C}[X].$$

Bezeichnen wir mit $g^{(p)}$ die p -te formale Ableitung von g , so gilt

$$\begin{aligned} g^{(p)}(X) &= \sum_{i=0}^{d-1} \alpha_i (i \cdot \dots \cdot (i - (p-1))) X^{i-p} \\ &= p! \cdot \sum_{i=0}^{d-1} \alpha_i \binom{i}{p} X^{i-p} \end{aligned} \quad (11.8)$$

und damit

$$g^{(p)}(\lambda_i) = 0 \text{ für } 1 \leq i \leq m, 0 \leq p \leq s_i - 1, \quad (11.9)$$

wobei je eine der Gleichungen einer Zeile von $B\alpha = 0$ entspricht. Daraus folgt nun, dass g vom Minimalpolynom

$$m_F(X) := \sum_{i=1}^m (X - \lambda_i)^{s_i}$$

geteilt wird. m_F ist ein Polynom vom Grad d . g ist ein Polynom vom Grad $d-1$. Daraus folgt, dass g das Nullpolynom ist. Bezeichnen wir mit v_0 den

Vektor zur Koordinatendarstellung $w_0 \in \mathbb{C}^n$, so haben wir ein Element v_0 von $W \subset V$ gefunden, für das

$$fdim_F(v_0) = d.$$

Zusammen mit 11.4 folgt damit das Behauptete.

Der Fall $\mathbb{K} = \mathbb{R}$ lässt sich mit Hilfe des komplexen Falls erledigen. Dass $frang(F) \leq deg m_F$ folgt wiederum unmittelbar aus der Definition des Minimalpolynoms. Durch Übergang zu einer Koordinatendarstellung sei ohne Einschränkung $V = \mathbb{R}^n$ und $A \in \mathbb{R}^{n \times n}$ eine Matrix, die F entspricht. Indem man A als Abbildung von \mathbb{C}^n nach \mathbb{C}^n auffasst, findet man analog zu oben ein $v_0 \in \mathbb{C}^n$, so dass die Vektoren $v_0, Av_0, \dots, A^{d-1}v_0$ linear unabhängig sind. Wenn man die obige Konstruktion verfolgt, bemerkt man, dass tatsächlich $v_0 \in \mathbb{R}^n$. Schreiben wir nämlich

$$v_0 = \sum_{i=1}^m v_i,$$

wobei die v_i wie oben zu je einem Jordan-Kästchen gehören, so stellt sich zunächst heraus, dass für reelle λ_i die v_i reelle Hauptvektoren sind, wohingegen echt komplexe Eigenwerte jeweils in Pärchen $\lambda_i, \bar{\lambda}_i$ auftreten (siehe hierzu den Satz über die reelle Jordan-Normal-Form C.6). Man bemerkt nun (siehe z.B. den Beweis von der reellen Jordan-Normal-Form), dass dazu assoziierten Pärchen von v_i, \bar{v}_i jeweils zueinander komplex konjugierte Vektoren sind, so dass $v_i + \bar{v}_i$ einen reellen Vektor ergibt. Damit erweist sich v_0 als reell. \square

Wir erweitern die oben eingeführten Begriffe um eine weitere

DEFINITION 11.4. Sei V ein \mathbb{K} -Vektorraum mit $dim V =: n$, $F : V \rightarrow V$ ein linearer Operator und $v \in V$. Dann setzen wir

$$\begin{aligned} afdim_F : V &\longrightarrow \{0, \dots, n\} \\ v &\longmapsto \min_{S \subset \mathbb{N}, card(S) < \infty} dim \operatorname{span}\{F^k v \mid k \in \mathbb{N} \setminus S\} \end{aligned} \quad (11.10)$$

und nennen diese Abbildung die **asymptotische Folgendimension bezüglich F** . Wir nennen weiter

$$afrang(F) := \max_{v \in V} afdim_F(v) \quad (11.11)$$

den **asymptotischen Folgenrang** von F .

Offensichtlich ist der asymptotische Folgenrang kleiner gleich dem herkömmlichen Folgenrang. Der Unterschied besteht in der Tatsache, dass zur Bestimmung der Dimension endlich viele Vektoren der Form $F^k v$ weggelassen werden können. Die Frage, die sich jetzt unmittelbar stellt, ist die nach einer ebenso so klaren, geschlossenen Formel auch für den asymptotischen Folgenrang. Sie wird beantwortet durch den folgenden

SATZ 11.5. Sei V ein \mathbb{K} -Vektorraum mit $n := \dim V < \infty$ und $F : V \rightarrow V$ ein Endomorphismus. Sei weiter m_F das Minimalpolynom von F . Schreiben wir

$$m_F(X) = X^s f(X),$$

wobei $s \geq 0$ maximal gewählt wird, so gilt

$$afrang(F) = \deg f = \deg m_F - s. \quad (11.12)$$

Mit anderen Worten, die asymptotische Folgenrang ist der Grad des Minimalpolynoms minus des Exponenten, der zum Eigenwert Null gehört.

BEWEIS. Der Einfachheit halber setzen wir wieder $d := \deg m_F$. Durch Einsetzen von F ins Minimalpolynom erhält man gemäß Definition des Minimalpolynoms eine Gleichung der Form

$$F^s f(F) = \mathbf{0},$$

also sind für alle $v \in V$ die $F^s v, F^{s+1} v, \dots, F^{s+\deg f} v$ linear abhängig, woraus unmittelbar $afrang(F) \leq \deg f$ folgt. Um die Behauptung des Satzes zu komplettieren zeigen wir nun, dass für ein geeignetes v_0 jeweils $d - s$ aufeinanderfolgende Vektoren

$$F^{k+1} v_0, F^{k+2} v_0, \dots, F^{k+d-s} v_0$$

linear unabhängig sind. Dafür betrachten wir zunächst den Fall $\mathbb{K} = \mathbb{C}$. Unter Zuhilfenahme des Satzes über die Jordan-Normal-Form erhält man durch Zerlegung von f

$$f(X) = \prod_{i=1}^m (X - \lambda_i)^{s_i}$$

in Linearfaktoren (die λ_i seien paarweise verschieden) einen Unterraum W , bezüglich dessen bei geeigneter Wahl der Basis und Übergang zur entsprechenden Koordinatendarstellung eine Matrix A von F die folgende Form hat (siehe Korollar C.5, zerlege den Unterraum W aus dem Korollar hier weiter zu $W = W^* \oplus W_0$, wobei W_0 zum Eigenwert 0 gehört und arbeite hier mit W^*):

$$A|_W = \begin{pmatrix} \begin{bmatrix} \lambda_1 & 1 & & \\ & \lambda_1 & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_1 \end{bmatrix} & & & \\ & \begin{bmatrix} \lambda_2 & 1 & & \\ & \lambda_2 & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_2 \end{bmatrix} & & & \\ & & \ddots & & & & \\ & & & \begin{bmatrix} \lambda_m & 1 & & \\ & \lambda_m & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_m \end{bmatrix} & & & \end{pmatrix}.$$

Völlig analog zum Beweis von Satz 11.3 wählt man nun w_0 , und erhält, wenn man die Vektoren $A^{k+1} w_0, \dots, A^{k+d-s} w_0$ ebenfalls analog wie oben als

Spaltenvektoren in eine Matrix B schreibt, das folgende Bild:

$$B = \begin{pmatrix} \binom{k+1}{s_1-1} \lambda_1^{k+1-(s_1-1)} & \dots & \binom{k+(d-s)}{s_1-1} \lambda_1^{k+(d-s)-(s_1-1)} \\ \binom{k+1}{s_1-2} \lambda_1^{k+1-(s_1-2)} & \dots & \binom{k+(d-s)}{s_1-2} \lambda_1^{k+(d-s)-(s_1-2)} \\ \vdots & & \vdots \\ \binom{k+1}{1} \lambda_1^k & \dots & \binom{k+(d-s)}{1} \lambda_1^{k+(d-s)-1} \\ \lambda_1^{k+1} & \dots & \lambda_1^{k+(d-s)} \\ \binom{k+1}{s_2-1} \lambda_2^{k+1-(s_2-1)} & \dots & \binom{k+(d-s)}{s_2-1} \lambda_2^{k+(d-s)-(s_2-1)} \\ \vdots & & \vdots \\ \lambda_2^{k+1} & \dots & \lambda_2^{k+(d-s)} \\ \vdots & & \vdots \\ \binom{k+1}{s_m-1} \lambda_m^{k+1-(s_m-1)} & \dots & \binom{k+(d-s)}{s_m-1} \lambda_m^{k+(d-s)-(s_m-1)} \\ \vdots & & \vdots \\ \lambda_m^{k+1} & \dots & \lambda_m^{k+(d-s)} \end{pmatrix} \in \mathbb{C}^{d-s \times d-s}.$$

Es ist zu zeigen, dass B regulär ist. Sei also $\alpha \in \mathbb{C}^n$ mit $B\alpha = 0$. Dann haben wir zu zeigen, dass $\alpha = 0$. Betrachte nun das Polynom

$$g(X) := X^{k+1} \sum_{i=0}^{n-1} \alpha_i X^i \in \mathbb{C}[X].$$

Bezeichnen wir mit $g^{(p)}$ die p -te formale Ableitung von g , so gilt

$$\begin{aligned} g^{(p)}(X) &= X^{k+1} \sum_{i=0}^{(d-s)-1} \alpha_i (k+1+i) \cdot \dots \cdot (k+1+i-(p-1)) X^{i-p} \\ &= (p!) X^{k+1} \sum_{i=0}^{(d-s)-1} \alpha_i \binom{k+i+1}{p} X^{i-p} \end{aligned} \tag{11.13}$$

und damit

$$g^{(p)}(\lambda_i) = 0 \text{ für } 1 \leq i \leq m, \quad 0 \leq p \leq s_i - 1, \tag{11.14}$$

wobei je eine der Gleichungen einer Zeile von $B\alpha = 0$ entspricht. Daraus folgt nun, dass g vom Polynom

$$f(X) := \sum_{i=1}^m (X - \lambda_i)^{s_i}$$

geteilt wird. f aber ist ein Polynom vom Grad $d-s$. g ist ein Polynom vom Grad $k+(d-s)$ mit der $(k+1)$ -fachen Nullstelle 0. Da alle $\lambda_i \neq 0$, wird g von $X^{k+1}f(X)$ geteilt, ein Polynom vom Grad $k+(d-s)+1$. Daraus folgt, dass g das Nullpolynom ist. Wir finden somit ein Element $v_0 \in V$ mit den oben erklärten Eigenschaften. Damit ist der Fall $\mathbb{K} = \mathbb{C}$ erledigt. Der

Fall $\mathbb{K} = \mathbb{R}$ folgt mit einer völlig analogen Argumentation wie im Beweis von Satz 11.3. \square

11.2. Folgenstabilität

Wir erweitern dieses Kapitel um eine letzte Definition, den der (starken) Folgenstabilität. Dieser Begriff schlägt eine Brücke zwischen den oben zitierten Ergebnissen und denen des vorhergehenden Kapitels. Die hier vorgestellten Ergebnisse können als Zusammenführungen derselben aufgefasst werden.

DEFINITION 11.6. *Sei V ein \mathbb{K} -Vektorraum mit $n := \dim V < \infty$ und $F : V \rightarrow V$ ein linearer Operator. Wir nennen F **folgenstabil**, falls es ein $0 \neq v_0 \in V$ gibt, so dass die folgenden beiden Eigenschaft erfüllt sind:*

- (B) $\{z \in V \mid \exists k \in \mathbb{N} : z = F^k v_0\}$ ist beschränkt und
- (F) $\text{span}\{F^k v_0 \mid k \in \mathbb{N}\} = V$.

Ersetzen wir Eigenschaft (F) durch

- (AF) $\forall S \subset \mathbb{N}$ mit $\text{card}(S) < \infty : \text{span}\{z \in V \mid \exists k \in \mathbb{N} \setminus S : z = F^k v_0\} = V$,

so nennen wir F **stark folgenstabil**.

BEMERKUNG 11.7. *Wir stellen schleunigst fest, dass Eigenschaft (F) gerade der Tatsache*

$$f\dim_F(v_0) = n$$

bzw. Eigenschaft (AF)

$$a\dim_F(v_0) = n$$

entspricht, aus denen natürlich maximaler Folgenrang bzw. maximaler asymptotischer Folgenrang folgt.

Es gilt dann folgender Hauptsatz:

SATZ 11.8. *Sei V ein \mathbb{K} -Vektorraum mit $n := \dim V < \infty$ und $F : V \rightarrow V$ ein Endomorphismus. F ist genau dann folgenstabil, falls die folgenden beiden Kriterien erfüllt sind:*

- (i) F ist stabil sowie
- (ii) $\forall \lambda \in \text{Spec}(A) : \dim(\text{Eig}_\lambda(A)) = 1$.

Weiter ist F genau dann stark folgenstabil, falls zusätzlich zu (i) und (ii)

- (iii) F ist ein Isomorphismus

gegeben ist.

BEWEIS. Wir starten mit der *Vorbemerkung*, dass Eigenschaft (ii) äquivalent zur Tatsache ist, dass das charakteristische Polynom das Minimalpolynom teilt, also beide Polynome miteinander identifiziert werden können.

“ \implies ”: Sei F folgenstabil und $v_0 \neq 0$ ein Vektor, für den die geforderten Eigenschaften erfüllt sind.

“ad (i)”: Wir setzen

$$B := \{z \in \mathbb{K}^n \mid \exists \lambda \in \mathbb{K}, |\lambda| \leq 1, k \in \mathbb{N} : z = \lambda F^k v_0\},$$

und weiter sei \bar{B} der Abschluss der konvexen Hülle von B . \bar{B} ist nach Konstruktion eine beschränkte Umgebung der Null. [Mit $d := \min_{z \in \delta B} \|z\|_2$ gilt $U_d(0) \subset \bar{B}$.] Weiter gilt wegen der Stetigkeit von F :

$$F\bar{B} \subset \bar{B}. \quad (11.15)$$

Nach Satz 10.9 ist F somit stabil, also (i) erfüllt.

“ad (ii)”: Mit Hilfe von Bemerkung 11.7 und Satz 11.3 stellen wir fest, dass das Minimalpolynom maximalen Grad hat, was mit der Vorbemerkung zum Ziel führt.

“ad (iii)”: Sei nun F darüberhinaus stark folgenstabil. Da sein Minimalpolynom maximalen Grad hat, folgt mit Satz 11.5, dass 0 kein Eigenwert von F sein kann. F ist folglich ein Isomorphismus.

“ \Leftarrow ”: Sei F zunächst ein Endomorphismus, der die Eigenschaften (i) und (ii) erfüllt, also F ein stabiler Endomorphismus mit Minimalpolynom von maximalem Grad. Aufgrund von Eigenschaft (ii) finden wir mit Satz 11.3 ein $v_0 \in V$ mit $\text{fdim}_F(v_0) = n$, mit anderen Worten

$$\text{span}\{F^k v_0 \mid k \in \mathbb{N}\} = V.$$

Können wir zeigen, dass die Menge $\{F^k v_0 \mid k \in \mathbb{N}\}$ beschränkt ist, so folgt die Behauptung. Da F stabil ist, gibt es eine beschränkte Umgebung der Null $B \subset V$ mit $FB \subset B$. Schreiben wir für $\delta \in \mathbb{K}$

$$\delta V := \{z \in V \mid \exists z_0 \in B, \epsilon \in \mathbb{K}, |\epsilon| \in [0, |\delta|] : z = \epsilon z_0\},$$

so gilt natürlich auch

$$\forall \delta \in \mathbb{K} : F(\delta B) \subset \delta B.$$

Nun gibt es ein $\delta_1 \in \mathbb{K}$, so dass $v_0 \in \delta_1 B$. Es folgt weiter, dass

$$\{z \in V \mid \exists k \in \mathbb{N} : z = F^k v_0\} \subset \delta_1 B,$$

womit die Beschränktheit der zu untersuchenden Menge gezeigt ist.

Sei nun darüberhinaus F ein Isomorphismus. Da nun 0 kein Eigenwert von F ist, finden wir mit Hilfe von Satz 11.5 ein $v_0 \in V$ mit $\text{afdim}_F(v_0) = n$, mit anderen Worten:

$$\forall S \subset \mathbb{N}, \#S < \infty : \text{span}\{z \in V \mid \exists k \in \mathbb{N} \setminus S : z = F^k v_0\} = V.$$

Man argumentiert wie oben, um Eigenschaft (B) bestätigen zu können. \square

Komponentensummen und Kegel

12.1. Komponentensummenerhaltende Matrizen

Dieser Abschnitt handelt von Matrizen, die invariant bezüglich des Komponentensummenfunctionals, also bezüglich der linearen Abbildung in den Grundkörper, deren Bild die Summe der Komponenteneinträge ist. Diese Matrizen können, in der entwickelten Theorie der sich anschließenden Kapitel unter gewissen Zusatzvoraussetzungen als eine Art verallgemeinerte Markoff-Operatoren verstanden werden. Deshalb sind die folgenden Ergebnisse, im Wesentlichen Aussagen über die geometrische Konstellation der Eigen- und Hauptvektoren solcher Matrizen, von Interesse.

12.1.1. Einführung. In diesem Abschnitt sei wiederum immer

$$\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}.$$

$z \in \mathbb{K}^m$ wird bei Bedarf grundsätzlich als Spaltenvektor, also $z \in \mathbb{K}^{m \times 1}$ aufgefasst. Ist $A \in \mathbb{K}^{m \times n}$ eine Matrix, so sei $A^T \in \mathbb{K}^{n \times m}$ die Transponierte dazu. Weiter sei $\langle \cdot, \cdot \rangle$ das Standardskalarprodukt über \mathbb{K} . Es gilt also

$$\forall x, y \in \mathbb{R}^m : \quad \langle x, y \rangle = x^T y$$

im Falle $\mathbb{K} = \mathbb{R}$ und

$$\forall x, y \in \mathbb{C}^m : \quad \langle x, y \rangle = \bar{x}^T y,$$

wobei $\bar{x} := v_1 - iv_2$ für $x = v_1 + iv_2 \in \mathbb{C}^m$. Wichtig ist noch die Bezeichnung

$$\mathbf{1}_m := (1, \dots, 1),$$

also $\mathbf{1}_m$ der Vektor, dessen Einträge alle Eins sind.

DEFINITION 12.1. Sei $A \in \mathbb{K}^{m \times m}$ eine Matrix. Wir nennen A **komponentensummenerhaltend**, falls

$$\forall z \in \mathbb{K}^m : \quad \langle \mathbf{1}_m, z \rangle = \langle \mathbf{1}_m, Az \rangle. \quad (12.1)$$

Wir beginnen mit der folgenden offensichtlichen Feststellung:

PROPOSITION 12.2. Eine Matrix $A \in \mathbb{K}^{m \times m}$ ist genau dann komponentensummenerhaltend, falls für alle Spaltenvektoren $A^i, i = 1, \dots, m$

$$\langle \mathbf{1}_m, A^i \rangle = 1 \quad (12.2)$$

gilt.

BEWEIS. “ \implies ”: Sei $e_i = (0, \dots, 0, \underset{i}{1}, 0, \dots, 0) \in \mathbb{K}^m$ der i -te kanonische Basisvektor. Dann gilt

$$1 = \langle \mathbf{1}_m, e_i \rangle = \langle \mathbf{1}_m, Ae_i \rangle = \langle \mathbf{1}_m, A^i \rangle, \quad (12.3)$$

was zu zeigen war.

“ \impliedby ”: Sei $z \in \mathbb{K}^m$. Schreibe $z = \sum_i \alpha_i e_i$, $\alpha_i \in \mathbb{K}$ und folgere

$$\begin{aligned} \langle \mathbf{1}_m, Az \rangle &= \langle \mathbf{1}_m, A(\sum_i \alpha_i e_i) \rangle = \sum_i \alpha_i \langle \mathbf{1}_m, Ae_i \rangle \\ &= \sum_i \alpha_i \langle \mathbf{1}_m, A^i \rangle = \sum_i \alpha_i \\ &= \sum_i \alpha_i \langle \mathbf{1}_m, e_i \rangle = \langle \mathbf{1}_m, \sum_i \alpha_i e_i \rangle \\ &= \langle \mathbf{1}_m, z \rangle \end{aligned} \quad (12.4)$$

□

PROPOSITION 12.3. *Ist A ein komponentensummenerhaltender Isomorphismus, so ist auch A^{-1} komponentensummenerhaltend.*

BEWEIS. Offensichtlich: Sei $y \in \mathbb{K}^m$, dann gibt es ein $x \in \mathbb{K}^m$ mit $Ax = y \iff A^{-1}y = x$. Es folgt

$$\begin{aligned} \langle \mathbf{1}_m, A^{-1}y \rangle &= \langle \mathbf{1}_m, A^{-1}Ax \rangle = \langle \mathbf{1}_m, x \rangle \\ &= \langle \mathbf{1}_m, Ax \rangle = \langle \mathbf{1}_m, y \rangle \end{aligned} \quad (12.5)$$

□

PROPOSITION 12.4. *Sind $S, A \in \mathbb{K}^{m \times m}$ komponentensummenerhaltende Matrizen und S zusätzlich regulär, so ist auch $B := S^{-1}AS$ komponentensummenerhaltend.*

BEWEIS. Mit Proposition 12.3 offensichtlich:

$$\begin{aligned} \langle \mathbf{1}_m, Bz \rangle &= \langle \mathbf{1}_m, S^{-1}(ASz) \rangle = \langle \mathbf{1}_m, A(Sz) \rangle \\ &= \langle \mathbf{1}_m, Sz \rangle = \langle \mathbf{1}_m, z \rangle. \end{aligned} \quad (12.6)$$

□

12.1.2. Eigenwerte. Wir beschäftigen uns nun mit der Geometrie von Basen von Eigenvektoren komponentensummenerhaltender Matrizen. Zunächst einmal ist das Folgende zu bemerken:

SATZ 12.5. *Eine komponentensummenerhaltende Matrix $A \in \mathbb{K}^{m \times m}$ hat den Eigenwert 1.*

BEWEIS. Mit Proposition 12.2 fällt auf:

$$\begin{aligned} \mathbf{1}_m^T A &= \mathbf{1}_m^T \\ \iff A^T \mathbf{1}_m &= \mathbf{1}_m \end{aligned}$$

Die Transponierte von A hat demnach $\mathbf{1}_m$ als Eigenvektor zum Eigenwert 1. Da die charakteristischen Polynome von A^T und A übereinstimmen, folgt die Behauptung. \square

Wie angekündigt, beschäftigen wir uns nun mit der geometrischen Konstellation der Eigen- und Hauptvektoren komponentensummenerhaltender Matrizen. Eigen- und Hauptvektoren, die nicht zum Eigenwert 1 assoziiert sind, haben Komponentensumme 0. Korollar des Ganzen ist übrigens wiederum, dass es einen Eigenvektor zum Eigenwert 1 geben muss.

SATZ 12.6. *Sei $A \in \mathbb{K}^{m \times m}$ eine komponentensummenerhaltende Matrix und $v \in \mathbb{K}^m$ ein zur (im Falle $\mathbb{K} = \mathbb{R}$ reellen, siehe Satz C.6) Jordan-Normal-Form assoziierter Haupt- oder Eigenvektor zum Eigenwert $\lambda \in \mathbb{C}$. Dann gilt:*

$$\lambda \neq 1 \implies \langle \mathbf{1}_m, v \rangle = 0. \quad (12.7)$$

BEWEIS. Wir bemerken zunächst, dass der Fall $\mathbb{K} = \mathbb{R}$ einfach aus dem Fall $\mathbb{K} = \mathbb{C}$ gefolgert werden kann, indem man A als Element aus $\mathbb{C}^{m \times m}$ auffasst und bemerkt, dass die zu einem von Eins verschiedenen, echt komplexen Eigenwert assoziierten Basisvektoren $x \in \mathbb{R}^m$ der reellen Jordan-Normal-Form nach Konstruktion der reellen Jordan-Normal-Form sich jeweils als die Summe zweier nicht zu Eins gehörenden Basisvektoren z, z^* der Jordan-Normal-Form schreiben lassen:

$$x = \frac{1}{2}(z + z^*).$$

Deshalb

$$\langle \mathbf{1}_m, x \rangle = \frac{1}{2}(\langle \mathbf{1}_m, z \rangle + \langle \mathbf{1}_m, z^* \rangle) = 0$$

nach Fall $\mathbb{K} = \mathbb{C}$. Wir betrachten daher nur noch den Fall $\mathbb{K} = \mathbb{C}$.

Sei $v \in \ker(A - \lambda \cdot id)^l$ ein Eigen- oder Hauptvektor zu einem Eigenwert $\lambda \in \mathbb{C}$ mit $\lambda \neq 1$. Wir verfahren mit vollständiger Induktion über l .

$l = 1$: Dies ist der Fall, dass $Av = \lambda v$, also v ein Eigenvektor zum Eigenwert λ . Die Annahme $\langle \mathbf{1}_m, v \rangle = K \neq 0$ führt auf

$$\begin{aligned} K &= \langle \mathbf{1}_m, v \rangle = \langle \mathbf{1}_m, Av \rangle \\ &= \langle \mathbf{1}_m, \lambda v \rangle = \lambda \langle \mathbf{1}_m, v \rangle = \lambda K, \end{aligned}$$

ein Widerspruch!

$l \rightarrow l + 1$: Sei $v \in \ker(A - \lambda \cdot id)^{l+1}$. Dann gibt es einen Hauptvektor $v^* \in \ker(A - \lambda \cdot id)^l$, so dass

$$Av = \lambda v + v^*.$$

Die Annahme $\langle \mathbf{1}_m, v \rangle = K \neq 0$ führt auf

$$\begin{aligned}
K &= \langle \mathbf{1}_m, v \rangle = \langle \mathbf{1}_m, Av \rangle \\
&= \langle \mathbf{1}_m, \lambda v + v^* \rangle \\
&= \lambda \langle \mathbf{1}_m, v \rangle + \langle \mathbf{1}_m, v^* \rangle \\
&\stackrel{IV}{=} \lambda \langle \mathbf{1}_m, v \rangle = \lambda \cdot K,
\end{aligned}$$

ein Widerspruch! □

Es folgt ein Korollar, das einen geometrischen Aspekt einer stabilen, komponentensummenerhaltenden Matrix aufzeigt.

KOROLLAR 12.7. *Sei $A \in \mathbb{R}^{m \times m}$ eine stabile, komponentensummenerhaltende Matrix. Dann hat A einen Eigenvektor v zum Eigenwert 1 mit der Eigenschaft*

$$\langle \mathbf{1}_m, v \rangle \neq 0. \quad (12.8)$$

BEWEIS. Aufgrund des obigen Satzes 12.6 hat A einen Haupt- oder Eigenvektor v zum Eigenwert 1 mit $\langle \mathbf{1}_m, v \rangle \neq 0$. Weil A stabil ist, gibt es keine Hauptvektoren zum Eigenwert 1 (vergleiche Satz 10.9, 10.5). □

BEMERKUNG 12.8. *Komponentensummeninvarianz alleine reicht nicht, um die Aussage von Korollar 12.7 erzielen zu können. Das zeigt das Beispiel der Matrix*

$$A := \begin{pmatrix} \frac{3}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{pmatrix}.$$

Diese hat Spaltensummen 1 und ist daher nach Proposition 12.2 komponentensummenerhaltend. Jedoch gilt

$$A \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

sowie

$$A \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

$(1, -1)^T$ ist also Eigenvektor zum Eigenwert 1 und $(1, 1)$ Hauptvektor zum Eigenwert 1. Demnach kann es keinen weiteren Eigenvektor zum Eigenwert 1 geben. Es gilt jedoch $\langle \mathbf{1}_2, (1, -1)^T \rangle = 0$.

BEMERKUNG 12.9. *Ein Spezialfall stabiler, komponentensummenerhaltender Matrizen sind Markov-Operatoren (Matrizen mit ausschließlich positiven Einträgen und Spaltensumme 1). In diesem Fall beweist man die Aussage von Korollar 12.7 einfacher mit dem Lemma von Farkas [Fai05b].*

12.2. Invarianz von Kegeln und Komponentensummen, Stabilität

Dieser Abschnitt stellt die Begriffe der Kegelinvarianz, der Stabilität (also der Invarianz bezüglich Umgebungen der Null) und der Komponentensummeninvarianz in einen Zusammenhang. Es stellt sich heraus, dass für komponentensummeninvariante Matrizen Stabilität und Invarianz bezüglich

Kegel einer bestimmten geometrischen Ausprägung äquivalente Konzepte sind. Diese Einsichten mögen sich als hilfreich für die Theorie der “konkreten OOMs” [Jae00], eine Form der Darstellung endlich-dimensionaler stochastischer Prozesse, erweisen

Wir übernehmen für diesen Abschnitt die Bezeichnungen des vorigen Abschnitts. In diesem Abschnitt gilt allerdings immer $\mathbb{K} = \mathbb{R}$, wir betrachten demnach nur reellwertige Vektorräume. Zur Erinnerung: $\mathbf{1}_m = (1, \dots, 1)^T$ ist der Vektor, dessen Komponenten alle Eins sind und $\langle \cdot, \cdot \rangle$ ist das Standardskalarprodukt auf \mathbb{R} . In diesem Abschnitt bezeichnet

$$\text{conv}(B) := \left\{ x \in \mathbb{R}^m \mid x = \sum_{i=1}^m \alpha_i x_i, x_i \in B, \alpha_i \geq 0, \sum_{i=1}^m \alpha_i = 1 \right\}$$

zusätzlich den konvexen Abschluss einer Menge $B \subset V$ eines \mathbb{R} -Vektorraums. Eine Menge B bezeichnen wir als konvex, falls $\text{conv}(B) = B$. Darüberhinaus übernehmen wir alle Bezeichnungen und Begriffe aus Kapitel 3, Unterabschnitt 3.2.2.

Das zentrale Ergebnis dieses Abschnitts ist nun der folgende Satz. Er stellt einen Zusammenhang zwischen invarianten Komponentensummen, Kegeln und Umgebungen der Null her. Im weiteren Verlauf gebrauchen wir, zusätzlich zu den üblichen Begriffen noch die Bezeichnungen

$$\begin{aligned} H_r &:= \{x \in \mathbb{R}^m \mid \langle \mathbf{1}_m, x \rangle = r\}, \\ H_{\{<, \leq, >, \geq\}r} &:= \{x \in \mathbb{R}^m \mid \langle \mathbf{1}_m, x \rangle \{<, \leq, >, \geq\} r\}, \\ (\mathbb{R}^m)^+ &:= \{x \in \mathbb{R}^m \mid x = (x_1, \dots, x_m), \forall i : x_i \geq 0\} \end{aligned}$$

und weiter für beliebige Mengen $B \in \mathbb{R}^m$

$$B^+ := (\mathbb{R}^m)^+ \cap B.$$

SATZ 12.10. *Eine komponentensummenerhaltende Matrix $A \in \mathbb{R}^{m \times m}$ ist genau dann stabil, wenn sie einen echten Kegel K invariant lässt, für den gilt*

$$K \setminus \{0\} \subset H_{>0}. \quad (12.9)$$

BEWEIS. “ \implies ”: Sei A stabil. Nach Satz 10.6 und Satz 10.9 gibt eine abgeschlossene Umgebung der Null B , die von A invariant gelassen wird. Desweiteren kann B ohne Einschränkung konvex gewählt werden (gehe bei Bedarf zu $\text{conv}(B)$ über, offensichtlich wird auch $\text{conv}(B)$ invariant gelassen) und mit der Eigenschaft

$$(H_1)^+ \subset B. \quad (12.10)$$

[Da B eine Umgebung der Null, gibt es ein $\epsilon \in \mathbb{R}^+$, so dass $U_\epsilon(0) \subset B$. Wir setzen $\tilde{B} = \frac{2}{\epsilon}B$ und bemerken, dass $U_2(0) \subset \tilde{B}$ und natürlich auch $A\tilde{B} \subset \tilde{B}$. Für $x \in \mathbb{R}_+^m \cap H_1$ gilt weiter $\|x\|_2 \leq \|x\|_1 = 1$, also $x \in U_2(0) \subset \tilde{B}$.]

Da A komponentensummenerhaltend ist, folgt

$$A(B \cap H_1) \subset B \cap H_1. \quad (12.11)$$

Da $(H_1)_+ \subset B \cap H_1$, ist $B \cap H_1$ nicht-leer, und weiter wie B ebenfalls konvex (da die Komponentensumme bei Konvexkombinationen erhalten bleibt). Wir setzen

$$K = \bigcup_{r \in \mathbb{R}^+} r \cdot (B \cap H_1)$$

(wobei $r \cdot (B \cap H_1) := \{x \in \mathbb{R}^m \mid \exists v \in (B \cap H_1) : x = rv\}$) und stellen fest, dass K offensichtlich von A wegen 12.11 invariant gelassen wird und ein echter Kegel ist:

- *Kegeleigenschaften*: Aus $x \in K$ folgt offensichtlich $\alpha x \in K$ für $\alpha \in \mathbb{R}^+$ nach Konstruktion von K . Seien nun $x_1 = r_1 v_1, x_2 = r_2 v_2 \in K$, $r_1, r_2 \in \mathbb{R}^+, v_1, v_2 \in B \cap H_1$. Dann gilt

$$\frac{1}{r_1 + r_2}(x_1 + x_2) = \frac{r_1}{r_1 + r_2}v_1 + \frac{r_2}{r_1 + r_2}v_2 \in B \cap H_1$$

wegen der Konvexität von $B \cap H_1$. Also ist $x + y \in (r_1 + r_2) \cdot (B \cap H_1) \subset K$.

- *Abgeschlossenheit*: Ist klar, da offensichtlich $\delta K \subset K$.
- *Solidität*: Die Volldimensionalität folgt aus $(H_1)^+ \subset B \cap H_1$.
- *Punktiertheit*: Die Punktiertheit folgt aus der Beschränktheit von $B \cap H_1$.

“ \Leftarrow ”: Wir bezeichnen den echten Kegel, der von A invariant gelassen wird, mit K . Weiter bezeichne

$$K_{\leq 1} := K \cap H_{\leq 1} \cap H_{\geq 0}.$$

Dies ist wegen 12.9 eine nicht-leere, volldimensionale Menge. Wir setzen

$$B := \text{conv}(K_{\leq 1} \cup -K_{\leq 1})$$

und stellen fest, dass B von A invariant gelassen wird. Weiter ist B offensichtlich beschränkt und, da K ein echter, und damit volldimensionaler und abgeschlossener Kegel ist, auch eine abgeschlossene Umgebung der Null. \square

KOROLLAR 12.11. *Eine komponentensummenerhaltende Matrix A ist genau dann stabil, falls es eine zu A ähnliche, komponentensummenerhaltende Matrix gibt, die einen echten Kegel K invariant lässt, mit der zusätzlichen Eigenschaft*

$$K \subset (\mathbb{R}^m)^+. \quad (12.12)$$

BEWEIS. “ \implies ”: Mit Satz 12.10 ist klar, dass A einen echten Kegel K invariant lässt, für den $K \setminus \{0\} \subset H_+$ gilt. Nach Konstruktion (beachte den Beweis von Satz 12.10) gilt weiter $(H_1)^+ \subset K$. Wir basteln einen komponentensummenerhaltenden Isomorphismus S , so dass $B := S^{-1}AS$ die gewünschten Eigenschaften hat (beachte Proposition 12.4). Dafür setze zunächst

$$e := \left(\frac{1}{m}, \dots, \frac{1}{m}\right) \in (H_1)^+ \subset K$$

und weiter

$$D := \max_{x \in K \cap H_1} \|x - e\|_2 < \infty$$

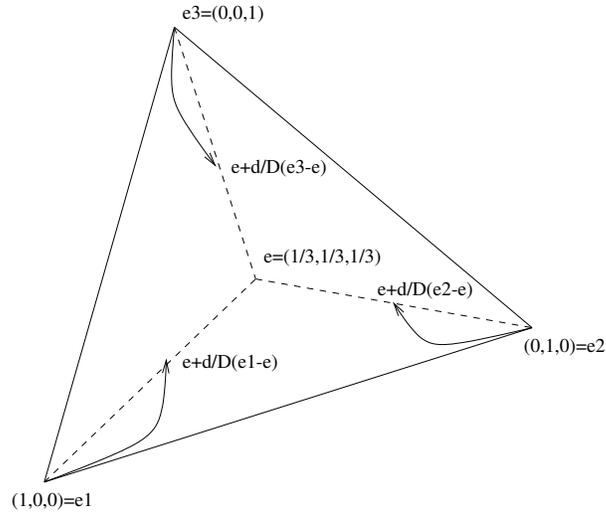


ABBILDUNG 1. Illustration der Abbildung S mit $S(e_i) = e - \frac{d}{D}(e_i - e)$ im Falle $m = 3$.

der größtmögliche Abstand aller in $K \cap H_1$ (offensichtlich kompakt, da K echt) liegenden Vektoren zu $e = (1/m, \dots, 1/m)$. Sei weiter

$$d := \min_{x \in \delta(H_1)^+} \|x - e\|_2 > 0$$

der Radius des größten Kreises um e , der ganz in $(H_1)^+$ liegt. Wegen $\delta(H_1)^+ \in K \cap H_1$ gilt nun

$$0 < \frac{d}{D} < 1.$$

Betrachte die folgende, durch lineare Fortsetzung definierte lineare Abbildung (siehe Abbildung 1):

$$S: \mathbb{R}^m \longrightarrow \mathbb{R}^m \\ e_i \longmapsto e + \frac{d}{D}(e_i - e). \quad (12.13)$$

S ist wegen $d/D > 0$ ein Isomorphismus und wegen

$$\begin{aligned} \langle \mathbf{1}_m, S^i \rangle &= \langle \mathbf{1}_m, S e_i \rangle = \langle \mathbf{1}_m, (1 - \frac{d}{D})e + \frac{d}{D}e_i \rangle \\ &= \langle \mathbf{1}_m, (1 - \frac{d}{D})e \rangle + \langle \mathbf{1}_m, \frac{d}{D}e_i \rangle \\ &= 1 - \frac{d}{D} + \frac{d}{D} = 1 \end{aligned} \quad (12.14)$$

komponentensummenerhaltend. Sei nun $x \in K \cap H_1$. Schreibt man $x = \sum_i \alpha_i e_i$ ($\sum_i \alpha_i = 1$ wegen $x \in H_1$), so gilt

$$\begin{aligned} \|Sx - e\|_2 &= \left\| \sum_i \alpha_i S e_i - e \right\|_2 = \left\| \sum_i \alpha_i (e + \frac{d}{D}(e_i - e)) - e \right\|_2 \\ &\stackrel{\sum_i \alpha_i = 1}{=} \left\| \sum_i \alpha_i \frac{d}{D}(e_i - e) \right\|_2 = \frac{d}{D} \left\| \sum_i \alpha_i e_i - e \right\|_2 \\ &\leq \frac{d}{D} \cdot D = d. \end{aligned} \quad (12.15)$$

Damit ist $Sx \in (H_1)^+$. Daraus folgt, dass $SK \in (\mathbb{R}^m)^+$. SK ist, da alle zu überprüfenden Eigenschaften unter Isomorphismen erhalten bleiben, ein echter Kegel. Setze nun $B := SAS^{-1}$, eine (wg. Proposition 12.4) komponentensummenerhaltende Matrix. Weiter gilt

$$BSK = SAK \subset SK,$$

was zu zeigen war.

“ \Leftarrow ”: Diese Richtung ist einfach. Sei $B := S^{-1}AS$ die zu A ähnliche, ebenfalls komponentensummenerhaltende Matrix und K der von B invariant gelassene Kegel mit den geforderten Eigenschaften. Nach Satz 12.10 ist B stabil. Da Stabilität unter Ähnlichkeitstransformationen erhalten bleibt, ist auch A stabil. \square

FAZIT 12.12. *Die verschiedenen Invarianzen eines linearen Operators, d.h. Umgebungen der Null, Komponentensummen und Kegel, sind zu einem gewissen Grade austauschbar. Wir vermuten, dass bei einer sorgfältig formulierten Verallgemeinerung auf beliebige lineare Funktionale, d.h. bei Invarianz bezüglich beliebiger Hyperebenen, entsprechend allgemeinere Ergebnisse erzielt werden können, bei denen Stabilität und Invarianz bezüglich einer Klasse von entsprechend positionierten Kegeln in Zusammenhang gebracht werden können.*

Teil 3

Appendix

Vier technische Lemmata für (signierte) Maße

An dieser Stelle führen wir vier technische Lemmata auf, die wir in einleitenden Kapiteln benötigt haben. Diese Lemmata sind recht elementar und Ergebnisse, die man in den meisten Büchern als Übungsaufgaben finden kann.

Im Folgenden sei nun (Ω, \mathcal{B}) immer ein Messraum und P sei ein endliches, signiertes Maß darauf (Definitionen siehe Abschnitte 2.1 und 2.2). Mit Ereignissen sind dann Elemente von \mathcal{B} gemeint. Desweiteren sei mit $|P|$ die totale Variation von P gemeint (siehe wiederum Abschnitt 2.3).

LEMMA A.1. *Seien A, B und C drei Ereignisse. Dann gilt*

$$|P|((A \cap C) \Delta (B \cap C)) \leq |P|(A \Delta B).$$

BEWEIS. Zunächst sieht man mit Hilfe der de Morgan'schen Regeln

$$C \cap \complement(C \cap A) = C \cap (\complement C \cup \complement A) = (C \cap \complement C) \cup (C \cap \complement A) = C \cap \complement A$$

ein. Zusammen mit dem analogen Resultat für C und B erhält man

$$\begin{aligned} (C \cap A) \Delta (C \cap B) &= ((C \cap A) \setminus (C \cap B)) \cup ((C \cap B) \setminus (C \cap A)) \\ &= (C \cap B \cap (C \cap A)^c) \cup (C \cap A \cap (C \cap B)^c) = (C \cap B \cap A^c) \cup (C \cap A \cap B^c) \\ &= (C \cap (B \setminus A)) \cup (C \cap (A \setminus B)) \\ &\subset (B \setminus A) \cup (A \setminus B) = A \Delta B. \end{aligned} \tag{A.1}$$

Damit folgt natürlich die Behauptung. \square

LEMMA A.2. *Seien A und B zwei Ereignisse. Dann gilt*

$$|P(A) - P(B)| = |P(A \setminus B) - P(B \setminus A)| \leq |P|(A \Delta B).$$

BEWEIS. Wegen $P(B) = P(B \setminus A) + P(B \cap A)$ und $P(A) = P(A \setminus B) + P(A \cap B)$ folgt

$$\begin{aligned} |P(A) - P(B)| &= |P(A \setminus B) + P(A \cap B) - (P(B \setminus A) + P(B \cap A))| \\ &= |P(A \setminus B) - P(B \setminus A)| \leq |P(A \setminus B)| + |P(B \setminus A)| \leq |P|(A \Delta B). \end{aligned} \tag{A.2}$$

\square

LEMMA A.3. *Seien A, B und C drei Ereignisse. Dann*

$$|P|(A \Delta B) \leq |P|(A \Delta C) + |P|(B \Delta C).$$

BEWEIS. Es gilt

$$A \setminus B \subset (A \setminus C) \cup (C \setminus B),$$

denn sei $x \in A \setminus B$, so dass $x \notin A \setminus C$, dann folgt $x \in (A \cap C) \setminus B \subset C \setminus B$.
Analog folgern wir

$$B \setminus A \subset (B \setminus C) \cup (C \setminus A)$$

und damit

$$\begin{aligned} |P|(A \Delta B) &= |P|(A \setminus B) + |P|(B \setminus A) \\ &\leq |P|((A \setminus C) \cup (C \setminus B)) + |P|((B \setminus C) \cup (C \setminus A)) \\ &\leq |P|(A \setminus C) + |P|(C \setminus B) + |P|(B \setminus C) + |P|(C \setminus A) \quad (\text{A.3}) \\ &= |P|(A \Delta C) + |P|(B \Delta C). \end{aligned}$$

□

LEMMA A.4. Seien A, B, C und D vier Ereignisse. Dann

$$|P|((A \cap B) \Delta (C \cap D)) \leq |P|(A \Delta C) + |P|(B \Delta D).$$

BEWEIS. Das folgt aus

$$\begin{aligned} (A \cap B) \setminus (C \cap D) &= (A \cap B) \cap \mathbb{C}(C \cap D) = (A \cap B) \cap (\mathbb{C}C \cup \mathbb{C}D) \\ &= (A \cap B \cap \mathbb{C}C) \cup (A \cap B \cap \mathbb{C}D) \\ &\subset (A \setminus C) \cup (B \setminus D) \end{aligned}$$

und analog

$$(C \cap D) \setminus (A \cap B) \subset (C \setminus A) \cup (D \setminus B).$$

□

Beweis von Satz 4.41

Wir reichen an dieser Stelle einen Beweis für den folgenden Satz nach:

SATZ B.1. *Ist g ein Maßvektor, so sind die beiden folgenden Aussagen äquivalent:*

- (a) *g ist ein ES-Vektor.*
- (b) *Das zu g assoziierte dynamische System $(\Omega_\Sigma, \mathcal{B}_\Sigma, m_g, T_\Omega)$ ist AMS.*

Das erfordert jedoch den Einsatz von Begriffen, die in diesem Text bislang keine Rolle spielten (und auch keine spielen sollten). Wie bereits in Abschnitt 4.3 angesprochen, ist die Implikation (a) \Rightarrow (b) einfach. Ist \bar{g} der dem stationären Grenzmaß von m_g entsprechende Vektor und schreiben wir ($\mu_n := \frac{1}{n} \sum_{i=0}^{n-1} \mu^i$)

$$m_n := m_{\mu_n g} = \frac{1}{n} \sum_{i=0}^{n-1} m_{\mu^i g} = \frac{1}{n} \sum_{i=0}^{n-1} m \circ T_\Omega^{-i},$$

so haben wir für die umgekehrte Richtung zu zeigen, dass

$$\sup_{B \in \mathcal{B}_\Sigma} |m_n(B) - m_{\bar{g}}(B)| \xrightarrow{n \rightarrow \infty} 0,$$

also dass die Folge der Maße m_n gleichmäßig auf allen Elementen von \mathcal{B}_Σ gegen das stationäre Grenzmaß \bar{m} konvergiert, oder um mit den Worten von Definition 2.59 zu sprechen, die Folge der m_n konvergiert schwach Skorokhod gegen das Maß \bar{m} .

Wir werden im Laufe dieses Kapitels mehr als das zeigen. Zum Verständnis dessen ist jedoch das Verständnis verschiedener Formen der Konvergenz von Zufallsvariablen nötig. Desweiteren beachte man die nachfolgende Bemerkung, die zum weiteren Verständnis des Textes *sehr wichtig* ist:

BEMERKUNG B.2. *Wir werden im weiteren Verlauf ausschließlich dynamische Systeme mit Wahrscheinlichkeitsmaßen m betrachten, da die maßgeblichen Definitionen und die alternativen Formulierungen (als auch die eigentliche Definition, siehe Bemerkung 2.60) der schwachen Skorokhod Konvergenz nur für Folgen von Wahrscheinlichkeitsmaßen angegeben werden. Die gesamte Theorie lässt sich jedoch offensichtlich auf Folgen von Maßen P_n verallgemeinern, so dass es ein $C \in \mathbb{R}^+$ gibt, so dass für alle n*

$$P_n(\Omega_\Sigma) = C$$

gilt, da es sich bei solchen Folgen quasi um Folgen skaliertes Wahrscheinlichkeitsmaße handelt und für solche ist die Übertragung der Theorie eine offensichtlich einfache Angelegenheit. Da es sich bei den hier betrachteten Maßen $m_n := \frac{1}{n} \sum_{i=0}^{n-1} m \circ T_\Omega^{-i}$ um entsprechende Folgen handelt, ist also eine Übertragung der Theorie in unserem Falle zulässig.

So vorbereitet, stürzen wir uns erneut in die Theorie der Konvergenz von Zufallsobjekten.

B.1. Konvergenzformen von Zufallsvariablen

Um den Satz beweisen zu können, wollen wir zunächst verschiedene Formen der Konvergenz von Maßen noch einmal unter die Lupe nehmen. Dabei können die verschiedenen Formen der Konvergenz von Maßen immer unter Zuhilfenahme von Konvergenzbegriffen formuliert werden, die für Folgen von Zufallsvariablen geschaffen worden sind.

DEFINITION B.3. *Ist*

$$(X_n : (\Omega_1, \mathcal{B}_1, P) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R})))_{n \in \mathbb{N}}$$

eine Folge von Zufallsvariablen aus einem Maßraum $(\Omega_1, \mathcal{B}_1, P)$ in den Maßraum der reellen Zahlen versehen mit der von allen Intervallen erzeugten Borel- σ -Algebra $\mathcal{B}(\mathbb{R})$, so sagen wir, dass die X_n **stochastisch** gegen die Zufallsvariable

$$X^* : (\Omega_1, \mathcal{B}_1, P) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$$

konvergieren, falls

$$\forall \epsilon \in \mathbb{R}^+ : P(\{\omega \in \Omega : |X_n(\omega) - X^*(\omega)| > \epsilon\}) \xrightarrow{n \rightarrow \infty} 0,$$

also der Messwert der Menge der $\omega \in \Omega$, auf denen sich die Werte der X_n von denen von X^* um mehr als ϵ unterscheidet, für alle ϵ gegen Null konvergiert.

DEFINITION B.4. *Ist*

$$(X_n : (\Omega_1, \mathcal{B}_1, P) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R})))_{n \in \mathbb{N}}$$

eine Folge von Zufallsvariablen aus einem Maßraum $(\Omega_1, \mathcal{B}_1, P)$ in die reellen Zahlen versehen mit der von allen Intervallen erzeugten Borel- σ -Algebra, so sagen wir, dass die X_n **fast überall** gegen die Zufallsvariable

$$X^* : (\Omega_1, \mathcal{B}_1, P) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$$

konvergieren, falls

$$P(\{\omega \in \Omega : \lim_{n \rightarrow \infty} X_n(\omega) = X^*(\omega)\}) = 1,$$

also die Menge der Elemente von Ω , auf denen die X_n gegen X^* konvergieren, von P mit dem maximalen Messwert 1 belegt wird.

Nun gilt die folgende Implikation:

SATZ B.5. *Konvergiert eine Folge von Zufallsvariablen fast überall, so auch stochastisch.*

BEWEIS. Siehe [GS01], S. 310, Theorem 7.2.3. □

B.2. Absolutstetigkeit und der Satz von Radon-Nikodym

Um den Weg zurück zur Konvergenz von Maßen zu finden, benötigen wir einige weitere Klassiker der Maßtheorie. Zwei davon sind die Definition der Absolutstetigkeit von Maßen bezüglich eines Referenzmaßes und der Satz von Radon-Nikodym, der diesen Begriff auf prägnante und hilfreiche Art und Weise charakterisiert.

DEFINITION B.6. *Seien P, Q zwei Maße auf einem Messraum (Ω, \mathcal{B}) . Wir sagen, dass P **absolutstetig bezüglich Q** ist bzw. Q **dominiert** P , falls*

$$\forall B \in \mathcal{B} : Q(B) = 0 \implies P(B) = 0.$$

Ist dies der Fall, so schreiben wir auch

$$P \ll Q.$$

Der Satz von Radon-Nikodym setzt diesen Begriff nun mit der Existenz von Dichtefunktionen in Beziehung. Für die Definition eines Integrals \int verweisen wir den Leser auf [Bau90] oder aber überlassen ihn seiner Intuition. Eine ausführliche Besprechung dieser Theorie hat an dieser Stelle keinerlei Nutzen im Sinne eines besseren Verständnisses dieses Texts.

SATZ B.7 (Radon-Nikodym). *Seien P, Q zwei Maße auf einem Messraum (Ω, \mathcal{B}) . Dann sind die folgenden beiden Aussagen äquivalent:*

- (a) *Q dominiert P .*
- (b) *Es gibt eine messbare Funktion $f : (\Omega, \mathcal{B}, Q) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, so dass für alle $B \in \mathcal{B}$:*

$$\int_B f dQ = P(B).$$

*Eine Funktion wie in (b) wird auch als **Dichte** von P bezüglich Q bezeichnet.*

BEMERKUNG B.8. *Dichten sind nicht eindeutig bestimmt. Genauer sind zwei messbare Funktionen $f \neq g$ genau dann eine Dichte von P bezüglich Q , falls f eine Dichte von P bezüglich Q ist, und*

$$Q(\{\omega \in \Omega \mid f(\omega) \neq g(\omega)\}) = 0,$$

f und g sich also nur auf einer Nullmenge von Q unterscheiden.

B.3. Der Raum $L_1(P)$

Wir fahren damit fort, ein äußerst beliebtes, weil mit schönen Eigenschaften ausgezeichnetes Objekt der Funktionalanalysis/Maßtheorie vorzustellen und zeigen eine Querverbindung zu den oben definierten Dichten auf. Für die dabei involvierte Definition eines Integrals verweisen wir auf einschlägige Literatur (siehe z.B. [Bau90]).

DEFINITION B.9. *Ist (Ω, \mathcal{B}, P) ein Maßraum, so setzen wir zunächst*

$$\mathcal{L}_1(\Omega, \mathcal{B}, P) := \{f : (\Omega, \mathcal{B}, P) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R})) \text{ messbar} \mid \int_{\Omega} |f| dP < \infty\},$$

also \mathcal{L}_1 der Vektorraum der messbaren Funktionen mit betraglich endlichem Integral und befinden weiter zwei Funktionen $f, g \in \mathcal{L}_1$ für äquivalent, falls $P(\{\omega \in \Omega \mid f(\omega) \neq g(\omega)\}) = 0$, sie sich also bloß auf einer Nullmenge unterscheiden. Wir schreiben dann $f \sim g$ für zwei solche Funktionen f, g und setzen

$$L_1(\Omega, \mathcal{B}, P) := \mathcal{L}_1(\Omega, \mathcal{B}, P) / \sim$$

also L_1 der Vektorraum der Äquivalenzklassen bezüglich \sim aus \mathcal{L}_1 . So formuliert wird L_1 vermöge der (wohldefinierten, da sich die Integrale der Betragsfunktionen der verschiedenen Repräsentanten einer Äquivalenzklasse nicht unterscheiden) Norm

$$\begin{aligned} \|\cdot\|_1 : L_1 &\longrightarrow \mathbb{R}^+ \\ f &\longmapsto \int_{\Omega} |f| dP \end{aligned}$$

zu einem Banachraum.

Siehe z.B. [MV92] für einen Beweis der Vollständigkeit von L_1 . Wir schreiben weiter der Einfachheit halber

$$L_1(P) := L_1(\Omega, \mathcal{B}, P),$$

falls klar ist, um welchen Messraum es sich handelt bzw. im Zweifel

$$L_1(P) := L_1(\Omega_{\Sigma}, \mathcal{B}_{\Sigma}, P)$$

(analog definieren wir $\mathcal{L}_1(P)$).

Wir verfolgen nun eine Argumentation, deren Elemente geläufigen Erkenntnissen entsprechen und z.B. teilweise in [Bau90] nachgelesen werden können. Dafür sei (Ω, \mathcal{B}) ein Messraum. Wir betrachten dann

$$\mathcal{G} := \{P : \mathcal{B} \rightarrow \mathbb{R} \text{ endliches, signiertes Maß}\},$$

die Menge der endlichen, signierten Maße auf (Ω, \mathcal{B}) , die vermöge der elementweisen Addition und Skalarmultiplikation und der Norm

$$\|P\|_* := |P|(\Omega),$$

wobei $|P|$ die totale Variation von P sei (siehe Def 2.48, aus der Minimalitätseigenschaft der Jordan-Zerlegung 2.46 folgt $|P_1 + P_2| \leq |P_1| + |P_2|$, woraus die Normeigenschaft von $\|\cdot\|_*$ ersichtlich wird), zu einem normierten

\mathbb{R} -Vektorraum wird. Sei weiter Q ein Maß (Ω, \mathcal{B}) . Wir betrachten dann die folgende Abbildung

$$\Phi: \mathcal{L}_1(Q) \longrightarrow \mathcal{G} \\ f \longmapsto \left(\begin{array}{ccc} m_f: \mathcal{B} & \rightarrow & \mathbb{R} \\ & B \mapsto & \int_B f dQ \end{array} \right).$$

Das ist eine lineare Abbildung mit

$$\ker(\Phi) = \{f \in \mathcal{L}_1(Q) \mid \int_{\Omega} |f| dQ = 0\} \\ = \{f \in \mathcal{L}_1(Q) \mid Q(\{\omega \mid f(\omega) \neq 0\}) = 0\},$$

woraus offensichtlich

$$L_1(Q) = \mathcal{L}_1(Q) / \ker(\Phi)$$

folgt. Wir studieren nun noch das Bild von Φ . Dazu betrachten wir folgenden Untervektorraum \mathcal{G}_Q von \mathcal{G} :

$$\mathcal{G}_Q := \{P \in \mathcal{G} \mid |P| \ll Q\},$$

also die Menge der endlichen signierten Maße auf (Ω, \mathcal{B}) , deren totale Variation (siehe Definition 2.48) absolutstetig bezüglich Q ist (diese ist ein Vektorraum, da aus $|P_1 + P_2|(B) \leq |P_1|(B) + |P_2|(B)$ für alle $B \in \mathcal{B}$ die Absolutstetigkeit von $|P_1 + P_2|$ aus der von $|P_1|$ und $|P_2|$ folgt). Es gilt nun

$$\Phi(|f|) = |\Phi(f)|,$$

woraus mit dem Satz von Radon-Nikodym $|\Phi(f)| \ll Q$ folgt. Damit erhalten wir

$$\text{Im}(\Phi) \subset \mathcal{G}_Q.$$

Ist umgekehrt ein endliches signiertes Maß P mit $|P| \ll Q$ gegeben, so natürlich auch $P^+, P^- \ll Q$. Diese besitzen nach dem Satz von Radon-Nikodym Dichten f^+, f^- und es gilt dann

$$\Phi(f^+ - f^-) = \Phi(f^+) - \Phi(f^-) = P^+ - P^- = P,$$

und es folgt

$$\text{Im}(\Phi) = \mathcal{G}_Q.$$

Weil allgemein $V/(\ker\phi) \cong \text{Im}\phi$ für eine lineare Abbildung $\phi: V \rightarrow W$ zwischen linearen Vektorräumen gilt, erhalten wir eine Isomorphie von Vektorräumen

$$L_1(Q) \cong \mathcal{G}_Q.$$

Schlussendlich gilt noch für $f \in \mathcal{L}_1(Q)$

$$\|f\|_1 = \int |f| dQ = \Phi(|f|)(\Omega) = |\Phi(f)|(\Omega) = \|\Phi(f)\|_*$$

und wir erhalten insgesamt den folgenden

SATZ B.10. *Mit den obigen Bezeichnungen etabliert*

$$\Phi: L_1(Q) \longrightarrow \mathcal{G}_Q$$

eine Isometrie von \mathbb{R} -Vektorräumen.

Gemäß Korollar 2.52 können wir im Falle des Messraums $(\Omega_\Sigma, \mathcal{B}_\Sigma)$ obigen Vektorraum \mathcal{G} mit dem l_{SM} identifizieren. Dabei entsprechen sich die Normen $\|\cdot\|_*$, $\|\cdot\|_{TV}$ gerade, wie man leicht einsieht (beachte Zerlegungssatz 2.51 und Proposition 2.41). Setzen wir nun für einen Maßvektor $g \in l_{SM}$ nun noch

$$l_g := \{h \in l_{SM} \mid m_{|h|} \ll m_g\},$$

so entspricht l_g gerade obigen \mathcal{G}_{m_g} . Wir erhalten somit

KOROLLAR B.11. *Ist $g \in l_{SM}$ ein Maßvektor, so erhalten wir durch obiges Φ eine Isometrie von Vektorräumen*

$$\Phi : L_1(m_g) \longrightarrow l_g.$$

BEWEIS. Isometrie meint, dass $\|f\|_1 = \|\Phi(f)\|_{TV}$, was aus obiger Argumentation ersichtlich wird. \square

B.4. Formen der Konvergenz von Maßen

So ausgerüstet betrachten wir erneut die schwache Skorokhod Konvergenz als auch eine weitere Form der Konvergenz von Maßen.

Zunächst eine alternative Charakterisierung der schwachen Skorokhod Konvergenz:

PROPOSITION B.12. *Sei $(P_n)_{n \in \mathbb{N}}$ eine Folge von Wahrscheinlichkeitsmaßen auf einem Messraum (Ω, \mathcal{B}) . Dann konvergieren die P_n genau dann **schwach Skorokhod** gegen das Maß P^* , falls es ein Referenzmaß Q gibt, das die P_n für alle $n \in \mathbb{N}$ sowie P^* dominiert und weiter gilt, dass die Folge der Dichten (f_n) der P_n bezüglich Q stochastisch gegen die Dichte f^* von P^* bezüglich Q konvergiert, d.h.*

$$\forall \epsilon \in \mathbb{R}^+ : \quad Q(\{|f_n - f^*| > \epsilon\}) \xrightarrow{n \rightarrow \infty} 0.$$

BEWEIS. Siehe [JR97], Theorem 3.1.3. \square

Nun die (neben starker Konvergenz und schwacher Skorokhod Konvergenz) dritte Konvergenzform von Maßen:

DEFINITION B.13. *Sei $(P_n)_{n \in \mathbb{N}}$ eine Folge von Wahrscheinlichkeitsmaßen auf einem Messraum (Ω, \mathcal{B}) . Wir sagen, dass die P_n **strikt** gegen das Maß P^* **konvergieren**, falls es ein Referenzmaß Q gibt, das die P_n für alle $n \in \mathbb{N}$ sowie P^* dominiert und weiter gilt, dass die Folge der Dichten (f_n) der P_n bezüglich Q fast sicher gegen die Dichte f^* von P^* bezüglich Q konvergiert, d.h.*

$$Q(\{\lim_{n \rightarrow \infty} f_n(\omega) = f^*(\omega)\}) = 1.$$

BEMERKUNG B.14. *Aus Satz B.5 folgt dann, dass eine strikt konvergente Folge von Wahrscheinlichkeitsmaßen auch schwach Skorokhod konvergiert.*

B.5. Positive Kontraktionen

Wir betrachten nun eine spezielle Klasse von Operatoren auf normierten Vektorräumen, die sogenannten Kontraktionen.

DEFINITION B.15. *Ein linearer Operator zwischen partiell geordneten, normierten Vektorräumen*

$$T : (V, \|\cdot\|_V, \leq_V) \longrightarrow (W, \|\cdot\|_W, \leq_W)$$

wird **positive Kontraktion** genannt, falls

$$\forall v \in V : \|Tv\|_W \leq \|v\|_V$$

und

$$v \geq_V 0 \implies Tv \geq_W 0.$$

Positive Kontraktionen auf dem L_1 (wobei $f \leq g \Leftrightarrow g - f \geq 0$) sind ausführlich studiert worden, siehe z.B. [Kre85], Kapitel 3. Einer der aus diesen Studien resultierenden, berühmtesten Sätze ist der von Chacon-Ornstein:

SATZ B.16 (Chacon-Ornstein). *Sei (Ω, \mathcal{B}) ein Messraum, Q ein Wahrscheinlichkeitsmaß darauf und*

$$T : L_1(Q) \longrightarrow L_1(Q)$$

eine positive Kontraktion. Seien weiter $f, g \in L_1(Q)$, so dass $g \geq 0$. Dann konvergieren die Folgen

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

fast überall auf $\{\sum_{i \geq 0} T^i g > 0\}$ gegen ein $f^ \in L_1(Q)$.*

BEWEIS. Siehe z.B. [Kre85], S.119 ff. □

Sehr oft findet der Satz in der folgenden speziellen Form Anwendung:

KOROLLAR B.17. *Sei (Ω, \mathcal{B}) ein Messraum, Q ein Wahrscheinlichkeitsmaß darauf und*

$$T : L_1(Q) \longrightarrow L_1(Q)$$

eine positive Kontraktion. Ist $f \in L_1(Q)$, so konvergieren die Folgen

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

fast überall gegen ein $f^ \in L_1(Q)$, so dass $Tf^* = f^*$.*

BEWEIS. Setze $g \equiv 1$ im Satz von Chacon-Ornstein, der Zusatz $Tf^* = f^*$ folgt offensichtlich. □

B.6. Beweis des Satzes

Der Fahrplan des Beweises ist nun der folgende: Wir suchen zunächst ein Referenzmaß Q für alle m_n und \bar{m} und weisen dann mit Hilfe des Satzes von Chacon-Ornstein die fast sichere Konvergenz der dazu assoziierten Dichten nach. Wir zeigen daher:

SATZ B.18. *Sei $(\Omega_\Sigma, \mathcal{B}_\Sigma, m, T_\Omega)$ ein dynamisches AMS System mit stationärem Grenzmaß \bar{m} . Dann konvergiert die Folge der Maße $m_n := \frac{1}{n} \sum_{i=0}^{n-1} m \circ T^{-i}$ strikt gegen \bar{m} .*

BEWEIS. Wir identifizieren im Verlauf dieses Beweises Vektoren $g \in l_{SM}$ mit ihren assoziierten Maßen und vermeiden daher die Schreibweise m_g . Wir definieren ein Maß Q auf $(\Omega_\Sigma, \mathcal{B}_\Sigma)$ durch die Vorschrift

$$Q(B) := \frac{1}{2}(\bar{m}(B) + \sum_{n \geq 0} 2^{-n-1} m_n(B)),$$

das sich wegen (beachte: $m_n(\Omega_\Sigma) = \bar{m}(\Omega_\Sigma) = m(\Omega_\Sigma)$ für alle $n \in \mathbb{N}$)

$$Q(\Omega_\Sigma) = \frac{1}{2}(\bar{m}(\Omega_\Sigma) + \sum_{n \geq 0} 2^{-n-1} m_n(\Omega_\Sigma)) \stackrel{m_n(\Omega_\Sigma) = \bar{m}(\Omega_\Sigma)}{=} \bar{m}(\Omega_\Sigma)$$

als endlich erweist. Es ist offensichtlich nach Definition, dass Q alle m_n sowie \bar{m} dominiert. Wir betrachten nun weiter

$$l_Q := \text{span}\{g \in l_{SM} \mid |g| \ll g_Q\},$$

den wir mit Hilfe von Korollar B.11 als den $L_1(Q)$ identifizieren. Wir zeigen nun zuerst, dass

$$\mu(l_{g_Q}) \subset l_{g_Q}. \quad (\text{B.1})$$

Zunächst folgt aus

$$0 < \mu g_Q(B) = \frac{1}{2}(\bar{m}(T_\Omega^{-1}B) + \sum_{n \geq 0} 2^{-n-1} m_n(T_\Omega^{-1}B))$$

nämlich $\bar{m} \circ T_\Omega^{-1}(B) = \bar{m}(B) > 0$ oder aber die Existenz eines $i \in \mathbb{N}$, so dass $m(T_\Omega^{-i}(B)) > 0$. Aus beiden Fällen folgt nach Definition von Q wiederum $Q(B) > 0$ und damit die Absolutstetigkeit von $\mu g_Q = Q \circ T_\Omega^{-1}$ bezüglich Q . Ist deshalb $h \in l_{SM}$ ein Maßvektor mit $h \ll g_Q$, so

$$\mu h(B) > 0 \Rightarrow h(T_\Omega^{-1}B) > 0 \Rightarrow Q(T_\Omega^{-1}B) > 0 \Rightarrow Q(B) > 0,$$

also auch $\mu h \ll Q$. (B.1) folgt daher aus $|\mu h| \leq \mu|h|$ für alle $h \in l_{SM}$. Wegen $\|\mu g\|_{TV} \leq \|g\|_{TV}$ induziert μ über Korollar B.11 eine positive Kontraktion T_μ auf dem $L_1(Q)$ (positiv, weil durch μ Maße auf Maße abgebildet werden). Ist f_0 nun die Dichte von m bezüglich Q , so gilt dann offensichtlich, dass $T_\mu^n f_0$ gerade die Dichte von $m \circ T_\Omega^{-n}$ ist. Nach Korollar B.17 des Satzes von Chacon-Ornstein konvergiert nun die Folge $\frac{1}{n} \sum_{i=0}^{n-1} T_\mu^i f_0$ fast überall gegen ein $f^* \in L_1(Q)$. Dies muss wegen der starken Konvergenz der zu den $\frac{1}{n} \sum_{i=0}^{n-1} T_\mu^i f_0$ assoziierten Maße gerade der Dichte von \bar{m} entsprechen. \square

Jordan-Normal-Formen

Dieser Teil des Anhangs liefert Resultate der linearen Algebra, die eher technischer Natur sind und zum Teil in Standardwerken nicht oder nur schwer zu finden sind. Der Satz von der reellen Jordan-Normal-Form ist ein offensichtliches Korollar der komplexen Jordan-Normal-Form, das allgemein bekannt sein dürfte, jedoch in den meisten Lehrbüchern nicht erwähnt wird (eines davon ist [Kow75]). Wir führen deshalb die Aussage samt Beweis an dieser Stelle auf.

C.1. Grundlagen

Um den Einstieg in die Materie deutlich zu erleichtern, beginnen wir diesen Abschnitt mit dem Satz über die Jordan-Normal-Form für komplexwertige Matrizen. Weiter formulieren wir ein reelles Analogon der komplexen Jordan-Normal-Form. Die reelle Jordan-Normal-Form dokumentiert, dass reelle lineare Operatoren im Wesentlichen nur Eines beherrschen: Drehstreckungen. Der Satz ist elementar, und ist ein einfaches Korollar des komplexen Falles. Da er in Standardlehrbüchern jedoch selten zitiert wird ([Kow75] ist, wie bereits erwähnt, eine Ausnahme), führen wir ihn desweiteren an dieser Stelle samt Beweis auf. Wir geben darüber hinaus die Definition des Minimalpolynoms an und illustrieren den Zusammenhang mit der Jordan-Normal-Form. Zunächst erinnern wir an wichtige und hier benötigte Grundbegriffe der linearen Algebra.

DEFINITION C.1. *Ist $A \in \mathbb{K}^{n \times n}$ eine Matrix, die wir als lineare Abbildung $A : \mathbb{K}^n \rightarrow \mathbb{K}^n$ auffassen, so sei*

$$\sigma(A) := \{\lambda_1, \dots, \lambda_m\} \subset \mathbb{C}$$

*die (endliche) Menge der Eigenwerte von A , auch **Spektrum** von A genannt und*

$$\rho(A) := \max\{|\lambda| \mid \lambda \in \sigma(A)\}$$

*der **Spektralradius** von A . Weiter sei*

$$V_a(\lambda) := \dim \text{Hau}(A; \lambda),$$

*die **algebraische Vielfachheit** des Eigenwerts λ , wobei*

$$\text{Hau}(A; \lambda) = \bigcup_{s \geq 1} \text{Ker}(A - \lambda \cdot \text{Id})^s$$

*der **Hauptraum** eines Eigenwerts λ sei. Dabei beachte man, dass es ein $s_0 \in \mathbb{N}$ gibt, so dass für alle $s \geq s_0$: $\text{Ker}(A - \lambda \cdot \text{Id})^s = \text{Ker}(A - \lambda \cdot \text{Id})^{s_0}$. Wir*

wobei die $(J_A)_{k,j} := (J_A)_{\lambda_{k,j}}$ **Jordan-Kästchen** genannt werden und die folgende Gestalt haben:

$$(J_A)_{\lambda_{k,j}} = \begin{pmatrix} \lambda_k & 1 & & \\ & \lambda_k & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_k \end{pmatrix}.$$

Es gelten weiter die folgenden Tatsachen:

- (1) Für die Anzahl j_k der Jordan-Kästchen eines Eigenwerts λ_k gilt:

$$j_k = V_g(\lambda_k).$$

- (2) Bezeichnen wir die Größe eines Jordan-Kästchens $(J_A)_{k,j}$ mit $l_{k,j}$ und setzen weiter $l_k := \max_j l_{k,j}$, so

$$l_k = \text{index}(\lambda_k).$$

- (3) Anzahl, Größe und Beschaffenheit der Jordan-Kästchen sind bis auf die Reihenfolge eindeutig bestimmt und demnach Invarianten unter Ähnlichkeitstransformationen.

BEWEIS. Siehe z.B. [Mey00], Seite 590. \square

KOROLLAR/DEFINITION C.3. Ist V ein n -dimensionaler \mathbb{C} -Vektorraum und $F : V \rightarrow V$ ein Endomorphismus, so gibt es eine Basis von V , bezüglich derer die Matrixdarstellung $A_F \in \mathbb{C}^{n \times n}$ von F Jordan-Normal-Form hat. Eine entsprechende Basis nennen wir auch **Jordan-Normal-Basis**.

BEWEIS. Das ist offensichtlich, da Ähnlichkeitstransformationen Basiswechseln entsprechen. \square

KOROLLAR C.4. Ist V ein endlich-dimensionaler \mathbb{C} -Vektorraum und $F : V \rightarrow V$ ein Endomorphismus und ist $\sigma(F) = \{\lambda_1, \dots, \lambda_m\}$ das Spektrum von F , so

$$V = \bigoplus_{k=1}^m \left(\bigoplus_{j=1}^{j_k} W_{k,j} \right),$$

wobei die Jordan-Normal-Formen von Matrixdarstellungen von $F_{k,j} := F|_{W_{k,j}}$ gerade einzelnen Jordan-Kästchen entsprechen und es gilt $F(W_{k,j}) \subset W_{k,j}$.

BEWEIS. Mit Korollar C.3 finden wir eine Basis von V , bezüglich derer die Matrixdarstellung von F in Jordan-Normal-Form ist. Gruppiert man die Elemente der Jordan-Normal-Basis, so dass je eine Gruppe $((w_{k,j})_1, \dots, (w_{k,j})_{l_{k,j}})$ einem Jordan-Kästchen entspricht, so erhält man die Behauptung, indem man $W_{k,j} := \text{span}\{(w_{k,j})_1, \dots, (w_{k,j})_{l_{k,j}}\}$ setzt. \square

KOROLLAR C.5. Ist V ein endlich-dimensionaler \mathbb{C} -Vektorraum und $F : V \rightarrow V$ ein Endomorphismus, ist $\sigma(F) = \{\lambda_1, \dots, \lambda_m\}$ das Spektrum von F und

$$m_F(X) = \prod_{i=1}^m (X - \lambda_i)^{l_i}$$

das in Linearfaktoren zerlegte Minimalpolynom, so gibt es einen Unterraum W von V , so dass die Matrixdarstellung $(J_A)_{F|_W}$ von $F|_W$ bezüglich ihrer Jordan-Normal-Basis die folgende Gestalt hat:

$$(J_A)_{(F|_W)} = \begin{pmatrix} [(J_A)_{\lambda_1}] & & & \\ & [(J_A)_{\lambda_2}] & & \\ & & \ddots & \\ & & & [(J_A)_{\lambda_m}] \end{pmatrix} \in \mathbb{C}^{\deg m_F \times \deg m_F},$$

wobei die $(J_A)_{\lambda_i}$ gerade Jordan-Kästchen

$$(J_A)_{\lambda_i} = \begin{pmatrix} \lambda_i & 1 & & \\ & \lambda_i & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_i \end{pmatrix} \in \mathbb{C}^{l_i \times l_i}$$

entsprechen.

BEWEIS. Da die Exponenten der Linearfaktoren l_i des Minimalpolynoms mit dem Index der entsprechenden Eigenwerte übereinstimmen, findet man für jeden Eigenwert λ_i in der Jordan-Normal-Darstellung von F ein Jordan-Kästchen der Größe l_i . Wählt man die entsprechenden Basismitglieder der Jordan-Normal-Basis, so erhält man die Behauptung. \square

SATZ C.6 (Reelle Jordan-Normal-Form). Sei $A \in \mathbb{R}^{n \times n}$ eine reellwertige Matrix. Sei $\sigma(A) \subset \mathbb{C}$ das Spektrum der (möglicherweise echt komplexen) Eigenwerte von A . Für jeden echt komplexen Eigenwert λ ist auch der komplex konjugierte Wert $\bar{\lambda}$ ein Eigenwert und wir können das Spektrum schreiben als

$$\sigma(A) = \sigma_{\mathbb{R}}(A) \dot{\cup} \sigma_{\mathbb{C}}(A),$$

wobei

$$\sigma_{\mathbb{R}}(A) = \{\lambda_1, \dots, \lambda_r\} \subset \mathbb{R}$$

und

$$\sigma_{\mathbb{C}}(A) = \{\lambda_{r+1}, \overline{\lambda_{r+1}}, \dots, \lambda_m, \overline{\lambda_m}\} \subset \mathbb{C} \setminus \mathbb{R}.$$

Dann gibt es eine Basis $\{v_1, \dots, v_n\}$, so dass für den Isomorphismus S mit $S(e_l) = v_l, l = 1, \dots, n$ und die Bezeichnung $J_A = S^{-1}AS$ gilt:

$$J_A = \begin{pmatrix} [(J_A)_{11}] & & & & \\ & \ddots & & & \\ & & [(J_A)_{1j_1}] & & \\ & & & \ddots & \\ & & & & [(J_A)_{m1}] \\ & & & & & \ddots \\ & & & & & & [(J_A)_{mj_m}] \end{pmatrix}, \quad (\text{C.1})$$

der Größe l_{kj} , so auch ein Jordan-Kästchen ($\bar{\lambda}_k$ bezeichne den zu λ_k komplex konjugierten Wert)

$$\begin{pmatrix} \bar{\lambda}_k & 1 & & \\ & \bar{\lambda}_k & \ddots & \\ & & \ddots & 1 \\ & & & \bar{\lambda}_k \end{pmatrix}$$

derselben Größe, denn seien $\{z_{(kj)_1}, \dots, z_{(kj)_{l_{kj}}}\}$ die zum Jordan-Kästchen gehörenden Vektoren, so gilt

$$Az_{(kj)_l} = \begin{cases} \lambda z_{(kj)_l} & \text{für } l = 1 \\ z_{(kj)_{l-1}} + \lambda z_{(kj)_l} & \text{für } l = 2, \dots, l_{kj} \end{cases} \quad (\text{C.3})$$

und damit, einer einfachen Rechnung folgend, da A reell ist, (wir setzen $\bar{z} = x - iy$ für $z = x + iy$ mit $x, y \in \mathbb{R}^n$)

$$A\bar{z}_{(kj)_l} = \begin{cases} \bar{\lambda}_k \bar{z}_{(kj)_l} & \text{für } l = 1 \\ \bar{z}_{(kj)_{l-1}} + \bar{\lambda}_k \bar{z}_{(kj)_l} & \text{für } l = 2, \dots, l_{kj} \end{cases}. \quad (\text{C.4})$$

Hieraus folgt das soeben Behauptete, da auch die $\bar{z}_{(kj)_1}, \dots, \bar{z}_{(kj)_{l_{kj}}}$ linear unabhängig sind. Wir setzen nun für alle Pärchen von Jordan-Kästchen $(J_A)_{\lambda_{kj}}, (J_A)_{\bar{\lambda}_{kj}}$ von echt komplexen Eigenwerten λ_k und $l = l_1, \dots, l_{kj}$

$$x_{(kj)_l} = \frac{1}{2}(z_{(kj)_l} + \bar{z}_{(kj)_l}) \text{ und} \quad (\text{C.5})$$

$$y_{(kj)_l} = \frac{1}{2i}(z_{(kj)_l} - \bar{z}_{(kj)_l}). \quad (\text{C.6})$$

Offensichtlich sind alle diese Vektoren linear unabhängig. Für die so definierten Vektoren für ein Pärchen von Jordan-Kästchen gilt nun (der Einfachheit halber bezeichne $(kj)_l$ mit l) für $l = 1$

$$\begin{aligned} Ax_l &= \frac{1}{2}A(z_l + \bar{z}_l) \\ &= |\lambda_k| \frac{1}{2}(e^{i\phi_k} z_l + e^{-i\phi_k} \bar{z}_l) \\ &= \frac{1}{2} \cdot |\lambda_k| \cdot [(\cos\phi_k + i \sin\phi_k)(x_l + iy_l) + (\cos\phi_k - i \sin\phi_k)(x_l - iy_l)] \\ &= |\lambda_k| \cdot (\cos\phi_k x_l - \sin\phi_k y_l) \end{aligned} \quad (\text{C.7})$$

und für $l = 2, \dots, l_{kj}$

$$\begin{aligned} Ax_l &= \frac{1}{2}A(z_l + \bar{z}_l) \\ &= \frac{1}{2}(z_{l-1} + \bar{z}_{l-1}) + |\lambda_k| \frac{1}{2}(e^{i\phi_k} z_l + e^{-i\phi_k} \bar{z}_l) \\ &= x_{l-1} + |\lambda_k| \cdot (\cos\phi_k x_l - \sin\phi_k y_l). \end{aligned} \quad (\text{C.8})$$

Analog schliesst man

$$Ay_l = \begin{cases} |\lambda_k| \cdot (\sin\phi_k x_l + \cos\phi_k y_l) & \text{für } l = 1 \\ y_{l-1} + |\lambda_k| \cdot (\sin\phi_k x_l + \cos\phi_k y_l) & \text{für } l = 2, \dots, l_{kj} \end{cases} \quad (\text{C.9})$$

Bemerkt man die lineare Unabhängigkeit aller x_l, y_l und ergänzt sie mit den reellen Eigen- und Hauptvektoren von A zu einer Basis, erhält man die Behauptungen des Satzes, wobei die Eigenschaften (a), (b) und (c) aus den entsprechenden Eigenschaften der komplexen Jordan-Normal-Form folgen. \square

Wir erhalten zu den obigen Korollaren analoge Resultate:

KOROLLAR C.7. *Ist V ein n -dimensionaler \mathbb{R} -Vektorraum und $F : V \rightarrow V$ ein Endomorphismus, so gibt es eine Basis von V , bezüglich derer die Matrixdarstellung $A_F \in \mathbb{R}^{n \times n}$ von F reelle Jordan-Normal-Form hat. Eine entsprechende Basis nennen wir auch (reelle) **Jordan-Normal-Basis**.*

BEWEIS. Das ist wiederum offensichtlich, da Ähnlichkeitstransformationen Basiswechseln entsprechen. \square

KOROLLAR C.8. *Ist V ein endlich-dimensionaler \mathbb{R} -Vektorraum und $F : V \rightarrow V$ ein Endomorphismus, und ist $\sigma(F)$ das Spektrum von F , so*

$$\sigma(F) = \sigma_{\mathbb{R}}(F) \dot{\cup} \sigma_{\mathbb{C}}(F),$$

wobei

$$\sigma_{\mathbb{R}}(F) = \{\lambda_1, \dots, \lambda_r\} \subset \mathbb{R}$$

und

$$\sigma_{\mathbb{C}}(F) = \{\lambda_{r+1}, \overline{\lambda_{r+1}}, \dots, \lambda_m, \overline{\lambda_m}\} \subset \mathbb{C} \setminus \mathbb{R}.$$

Echt komplexe Eigenwerte treten also als zueinander komplex konjugierte Pärchen auf. Weiter

$$V = \bigoplus_{k=1}^m \left(\bigoplus_{j=1}^{j_k} W_{kj} \right),$$

wobei die Jordan-Normal-Formen von Matrixdarstellungen von $F_{kj} := F|_{W_{kj}}$ gerade einzelnen reellen Jordan-Kästchen entsprechen und es gilt $F(W_{kj}) \subset W_{kj}$.

BEWEIS. Mit Korollar C.7 finden wir eine Basis von V , bezüglich derer die Matrixdarstellung von F in reeller Jordan-Normal-Form ist. Gruppiert man die Elemente der reellen Jordan-Normal-Basis, so dass je eine Gruppe $((w_{kj})_1, \dots, (w_{kj})_{l_{kj}})$ einem reellen Jordan-Kästchen entspricht, so erhält man die Behauptung, indem man $W_{kj} := \text{span}\{(w_{kj})_1, \dots, (w_{kj})_{l_{kj}}\}$ setzt. \square

Mit Hilfe der Jordan-Normal-Form lässt sich der folgende Satz beweisen, der bei uns in Kapitel 2 zur Anwendung kommt.

SATZ C.9 (Neumannsche Reihen). *Sei $A \in \mathbb{C}^{n \times n}$. Dann sind die folgenden Aussagen äquivalent.*

(a) *$(Id - A)$ ist invertierbar und*

$$\lim_{k \rightarrow \infty} \sum_{i=0}^k A^i = (Id - A)^{-1}.$$

(b)

$$\lim_{k \rightarrow \infty} A^k = \mathbf{0}.$$

(c)

$$\rho(A) < 1.$$

BEWEIS. Siehe [Mey00], Seite 618. Der Beweis nutzt im Die Äquivalenz von (b) und (c) folgt im Übrigen leicht aus Lemma C.10. \square

C.3. Zwei technische Lemmata

Für Jordan-Kästchen gilt:

LEMMA C.10. Sei $M \in \mathbb{C}^{n \times n}$ von der Gestalt

$$M = \begin{pmatrix} \lambda & 1 & & \\ & \lambda & \ddots & \\ & & \ddots & 1 \\ & & & \lambda \end{pmatrix},$$

M entspreche also einem einzelnen Jordan-Kästchen. Dann gilt für $j = 1, \dots, n$ und $e_j = (0, \dots, 0, \underset{j}{1}, 0, \dots, 0)^T$

$$M^k e_j = \sum_{i=1}^n \binom{k}{j-i} \lambda^{k-(j-i)} e_i, \quad (\text{C.10})$$

wobei $\binom{j}{l} := 0$ für $l > j, l < 0$ und $\binom{0}{0} := 1$.

BEWEIS. Vollständige Induktion über k :
 $k = 0$:

$$M^0 e_j = e_j = \binom{0}{0} \lambda^{j-j} e_j = \sum_{i=1}^n \binom{0}{j-i} \lambda^{0-(j-i)} e_i.$$

$k \rightarrow k + 1$: Gelte das Behauptete für $k \geq 0$. Zunächst gilt offensichtlich

$$M e_i = \begin{cases} \lambda e_i + e_{i-1} & \text{für } i = 2, \dots, n, \\ \lambda e_i & \text{für } i = 1. \end{cases} \quad (\text{C.11})$$

Damit

$$\begin{aligned}
M^{k+1}e_j &= M(M^k e_j) \stackrel{IV}{=} M\left(\sum_{i=1}^n \binom{k}{j-i} \lambda^{k-(j-i)} e_i\right) \\
&= \sum_{i=1}^n \binom{k}{j-i} \lambda^{k-(j-i)} M(e_i) \\
&\stackrel{C.11}{=} \binom{k}{j-1} \lambda^{k+1-(j-1)} e_1 + \sum_{i=2}^n \binom{k}{j-i} \lambda^{k-(j-i)} (\lambda e_i + e_{i-1}) \\
&\stackrel{\binom{k}{j-i}=0, i>j}{=} \sum_{i=1}^j \binom{k}{j-i} \lambda^{k+1-(j-i)} e_i + \sum_{i=2}^j \binom{k}{j-i} \lambda^{k-(j-i)} e_{i-1} \\
&= \sum_{i=1}^j \binom{k}{j-i} \lambda^{k+1-(j-i)} e_i + \sum_{i=1}^{j-1} \binom{k}{j-(i+1)} \lambda^{k-(j-i-1)} e_i \\
&= \sum_{i=1}^{j-1} \left[\binom{k}{j-(i+1)} + \binom{k}{j-i} \right] \lambda^{k+1-(j-i)} e_i \\
&\quad + \binom{k}{j-j} \lambda^{k+1-(j-j)} e_j \\
&= \sum_{i=1}^{j-1} \binom{k+1}{j-i} \lambda^{k+1-(j-i)} e_i + \binom{k+1}{j-j} \lambda^{k+1-(j-j)} e_j \\
&\stackrel{\binom{k+1}{j-i}=0, i>j}{=} \sum_{i=1}^n \binom{k+1}{j-i} \lambda^{k+1-(j-i)} e_i
\end{aligned} \tag{C.12}$$

□

Als Korollar erhalten wir

KOROLLAR C.11. Ist $M \in \mathbb{C}^{n \times n}$ von der Gestalt

$$M = \begin{pmatrix} \lambda & 1 & & \\ & \lambda & \ddots & \\ & & \ddots & 1 \\ & & & \lambda \end{pmatrix},$$

so für $i, j = 1, \dots, n$

$$(M^k)_{ij} = \binom{k}{j-i} \lambda^{k-(j-i)},$$

wobei $\binom{k}{l} = 0$ für $l > k$ oder $l < 0$ gesetzt sei.

BEWEIS. Das folgt aus obigem Lemma, bemerkt man, dass $(M^k)_{ij} = (M^k e_j)_i$. □

Analog gilt für reelle Jordan-Kästchen:

Zunächst gilt offensichtlich:

$$\begin{aligned} Me_i &= \begin{cases} x \cos \phi e_i - x \sin \phi f_i + e_{i-1} & \text{für } i = 2, \dots, n, \\ x \cos \phi e_i - x \sin \phi f_i & \text{für } i = 1, \end{cases} \\ Mf_i &= \begin{cases} x \sin \phi e_i + x \cos \phi f_i + f_{i-1} & \text{für } i = 2, \dots, n, \\ x \sin \phi e_i + x \cos \phi f_i & \text{für } i = 1. \end{cases} \end{aligned} \quad (\text{C.18})$$

Wir setzen $a_{k,i} := \cos(k-j+i)\phi$ und $b_{k,i} := \sin(k-j+i)\phi$. Zunächst gilt offensichtlich

$$a_{k,i+1} = a_{k+1,i} \text{ und } b_{k,i+1} = b_{k+1,i}. \quad (\text{C.19})$$

Mit den Additionstheoremen folgt weiter

$$\begin{aligned} a_{k+1,i} &= a_{k,i} \cos \phi - b_{k,i} \sin \phi \\ b_{k+1,i} &= b_{k,i} \cos \phi + a_{k,i} \sin \phi. \end{aligned} \quad (\text{C.20})$$

Mit diesen Bezeichnungen folgt unter Beachtung der Induktionsvoraussetzung, Gleichung C.18 und weiter wegen $\binom{k}{l} = 0, l < 0$ mit Indexverschiebung zunächst:

$$\begin{aligned} M^{k+1}e_j &= M(M^k e_j) \stackrel{IV}{=} M\left(\sum_{i=1}^n \binom{k}{j-i} x^{k-(j-i)} (a_{k,i} e_i - b_{k,i} f_i)\right) \\ &= \sum_{i=1}^n \binom{k}{j-i} x^{k-(j-i)} (a_{k,i} M(e_i) - b_{k,i} M(f_i)) \\ &\stackrel{\text{C.18}}{=} \binom{k}{j-1} x^{k-(j-1)} (a_{k,1} (x \cos \phi \cdot e_1 - x \sin \phi \cdot f_1) \\ &\quad - b_{k,1} (x \sin \phi \cdot e_1 + x \cos \phi \cdot f_1)) \\ &\quad + \sum_{i=2}^n \binom{k}{j-i} x^{k-(j-i)} (a_{k,i} (x \cos \phi \cdot e_i - x \sin \phi \cdot f_i + e_{i-1}) \\ &\quad - b_{k,i} (x \sin \phi \cdot e_i + x \cos \phi \cdot f_i + f_{i-1})). \\ &= \sum_{i=1}^j \binom{k}{j-i} x^{k+1-(j-i)} (a_{k,i} (\cos \phi \cdot e_i - \sin \phi \cdot f_i) \\ &\quad - b_{k,i} (\sin \phi \cdot e_i + \cos \phi \cdot f_i)) \\ &\quad + \sum_{i=1}^{j-1} \binom{k}{j-(i+1)} x^{k-(j-(i+1))} (a_{k,i+1} e_i - b_{k,i+1} f_i) \end{aligned}$$

Korollare des Satzes von Jacobi

An dieser Stelle führen wir zwei technische Lemmata auf, die als Korollare des Satzes von Jacobi interpretiert werden können. Diese Lemmata sind im ersten Teil dieser Arbeit für die Stabilitätstheorie hilfreich.

SATZ D.1 (Satz von Jacobi). *Sei $a \in \mathbb{R}^+ \setminus \mathbb{Q}$ eine positive, irrationale Zahl und sei $b \in \mathbb{Q}^+$ eine positive rationale Zahl. Dann gilt:*

$$\forall \epsilon \in \mathbb{R}^+ \quad \exists (k, l) \in (\mathbb{N}^+ \times \mathbb{N}^+) : \quad |k \cdot a - l \cdot b| < \epsilon. \quad (\text{D.1})$$

BEWEIS. Siehe [Bil65], Seite 10. □

LEMMA D.2. *Sei $\phi \in]0, 2\pi[$. Dann gelten*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \cos(k\phi) = 0 \text{ und} \quad (\text{D.2})$$

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \sin(k\phi) = 0. \quad (\text{D.3})$$

BEWEIS. Betrachte die komplexe Zahl $e^{i\phi} = \cos\phi + i\sin\phi$. Weiter gilt $e^{i(k\phi)} = \cos(k\phi) + i\sin(k\phi)$. Die endliche geometrische Reihe liefert weiter

$$\begin{aligned} \forall n \in \mathbb{N} : \left| \sum_{k=0}^{n-1} e^{i(k\phi)} \right| &= \left| \frac{1 - e^{i(n\phi)}}{1 - e^{i\phi}} \right| \\ &\leq (1 + |e^{i(n\phi)}|) \cdot \left| \frac{1}{1 - e^{i\phi}} \right| = 2 \cdot \left| \frac{1}{1 - e^{i\phi}} \right|. \end{aligned} \quad (\text{D.4})$$

Damit folgt

$$\begin{aligned} 0 &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} e^{i(k\phi)} \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \cos(k\phi) + i \sin(k\phi) \\ &= \lim_{n \rightarrow \infty} \left(\frac{1}{n} \sum_{k=0}^{n-1} \cos(k\phi) + i \frac{1}{n} \sum_{k=0}^{n-1} \sin(k\phi) \right). \end{aligned} \quad (\text{D.5})$$

Es folgt die Behauptung, da eine Folge komplexer Zahlen genau dann konvergiert, wenn es die Folgen der Real- und Imaginärteile tun. □

LEMMA D.3. Sei $\phi \in [0, 2\pi[$. Dann gibt es eine Folge monoton wachsender, natürlicher Zahlen $(k_j)_{j \in \mathbb{N}}$ mit

$$\lim_{j \rightarrow \infty} \cos(k_j \phi) = 1 \text{ und} \quad (\text{D.6})$$

$$\lim_{j \rightarrow \infty} \sin(k_j \phi) = 0. \quad (\text{D.7})$$

BEWEIS. Wir haben zwei Fälle zu unterscheiden:

$$(i) \quad \phi = q \cdot \pi \text{ mit } q \in \mathbb{Q} \text{ und}$$

$$(ii) \quad \phi = s \cdot \pi \text{ mit } s \in \mathbb{R} \setminus \mathbb{Q}.$$

In Fall (i) schreiben wir $q = \frac{n_1}{n_2}$ mit $n_1, n_2 \in \mathbb{N}$. Für $n_j := 2jn_2, j \in \mathbb{N}$ gilt

$$\begin{aligned} \cos(n_j \phi) &= \cos(2jn_2 \phi) = \cos(jn_1(2\pi)) \\ &= \cos(2\pi) = 1 \end{aligned} \quad (\text{D.8})$$

Damit haben wir eine Teilfolge $(a_{n_j} := \cos(n_j \phi))_{j \in \mathbb{N}}$, die gegen 1 konvergiert.

Fall (ii) ist etwas komplizierter. Mit Satz D.1 findet man nun für alle $n \in \mathbb{N}$ ein $k_n, l_n \in \mathbb{N}$, so dass

$$|k_n \cdot s - l_n \cdot 2| < \frac{1}{n\pi}. \quad (\text{D.9})$$

Man bemerkt

$$\forall n \in \mathbb{N} \exists N > n : k_N > k_n,$$

da ansonsten die Menge der k_n und damit auch die der l_n endlich wäre und die Rationalität von s folgen würde. Mit dieser Erkenntnis konstruieren wir Teilfolgen $(k_{n_j}, l_{n_j})_{j \in \mathbb{N}}$, so dass

$$\lim_{j \rightarrow \infty} (|k_{n_j} s \pi - l_{n_j} 2\pi|) = 0.$$

Wegen $\forall x, y \in \mathbb{R} : |\cos(x) - \cos(y)| \leq |x - y|$ folgt weiter

$$\forall j \in \mathbb{N} : 0 \leq |\cos(k_{n_j} s \pi) - \cos(l_{n_j} 2\pi)| \leq |k_{n_j} s \pi - l_{n_j} 2\pi|$$

und damit (Sandwich-Theorem)

$$\lim_{j \rightarrow \infty} (\cos(k_{n_j} s \pi) - \cos(l_{n_j} 2\pi)) = 0.$$

Wegen $\cos(l_{n_j} 2\pi) = 1$ für alle $j \in \mathbb{N}$ folgt

$$\lim_{j \rightarrow \infty} \cos(k_{n_j} s \pi) = 1.$$

Aus $\cos^2(k_{n_j} s \pi) + \sin^2(k_{n_j} s \pi) = 1$ folgt auch

$$\lim_{j \rightarrow \infty} \sin(k_{n_j} s \pi) = 0$$

und damit die Behauptung. \square

Literaturverzeichnis

- [AAKV02] D. Aharonov, A. Ambainis, J. Kempe, and U. Vazirani. Quantum walks on graphs. Technical report, arXiv, Cornell University Library, May 2002.
- [AC88] P.H. Algoet and T.M. Cover. A sandwich proof of the shannon-mcmillan-breiman theorem. *The Annals of Probability*, 16(2):899–909, 1988.
- [AF06] D. Aldous and J. Fill. *Reversible Markov Chains and Random Walks on Graphs*. Forthcoming book, www.stat.berkeley.edu/aldous/RWG/book.html, 2006.
- [Alo86] N. Alon. Eigenvalues and expanders. *Combinatorica*, 6:83–96, 1986.
- [Bal93] V. Balasubramanian. Equivalence and reduction of hidden markov models. Technical Report 1370, MIT AI Lab, 1993.
- [Bau90] H. Bauer. *Maß- und Integrationstheorie*. de Gruyter, 1990.
- [Bau91] H. Bauer. *Wahrscheinlichkeitstheorie*. de Gruyter, 1991.
- [Ber70] Abraham Berman. *Cones, Matrices and Mathematical Programming*. Number 79 in Lecture Notes in Economics and Mathematical Systems. Springer-Verlag, 1970.
- [Bil65] P. Billingsley. *Ergodic Theory and Information*. Wiley, 1965.
- [Bil95] P. Billingsley. *Probability and measure*. Wiley, 1995.
- [Bil98] J.A. Bilmes. A gentle tutorial of the em algorithm and its application to parameter estimation for gaussian mixture and hidden markov models. Technical Report TR-97-021, International Computer Science Institute, Berkeley, 1998.
- [Bil99] P. Billingsley. *Convergence of probability measures*. Wiley, 1999.
- [Bje04] Igor Bjelakovic. *Limit Theorems for Quantum Entropies*. PhD thesis, Technische Universität Berlin, 2004.
- [BK57] D. Blackwell and L. Koopmans. On the identifiability problem for functions of finite markov chains. *Annals of Mathematical Statistics*, 28:1011–1015, 1957.
- [BK97] C. Burge and S. Karlin. Prediction of complete gene structures in human genomic dna. *Journal of Molecular Biology*, 268:78–94, 1997.
- [Bou87] J. Bourgain. *An approach to pointwise ergodic theorems*. Lecture Notes in Mathematics. Springer, 1987. GAFA-Seminar, 1987.
- [BP66] L.E. Baum and T. Petrie. Statistical inference for probabilistic functions of finite markov chains. *Annals of Mathematical Statistics*, 37:1554–1563, 1966.
- [BPSW70] L.E. Baum, T. Petrie, G. Soules, and N. Weiss. A maximization technique occurring in the statistical analysis of probabilistic functions. *Annals of Mathematical Statistics*, 41:164–171, 1970.
- [Bre57] L. Breiman. The individual ergodic theorem of information theory. *Annals of Mathematical Statistics*, 28:809–811, 1957.
- [Bre60] L. Breiman. A correction to 'the individual ergodic theorem of information theory'. *Annals of Mathematical Statistics*, 31:809–810, 1960.
- [BS98] C.H. Bennett and P.W. Shor. Quantum information theory. *IEEE Transactions on Information Theory*, 1998.
- [BT79] R.K. Brayton and C.H. Tong. Stability of dynamical systems: A constructive approach. *IEEE Transactions on Circuits and Systems*, CAS-26(4):224–234, 1979.
- [Chu97] F.R.K. Chung. *Spectral Graph Theory*. American Mathematical Society, 1997.
- [DEKM98a] R. Durbin, S. Eddy, A. Krogh, and G. Mitchison. *Biological Sequence Analysis*. Cambridge University Press, 1998.

- [DEKM98b] R. Durbin, S.R. Eddy, A. Krogh, and G. Mitchison. *Biological sequence analysis*. Cambridge University Press, 1998.
- [Dha63a] S.W. Dharmadhikari. Functions of finite markov chains. *Annals of Mathematical Statistics*, 34:1022–1032, 1963.
- [Dha63b] S.W. Dharmadhikari. Sufficient conditions for a stationary process to be a function of a finite markov chain. *Annals of Mathematical Statistics*, 34:1033–1041, 1963.
- [Dha65] S.W. Dharmadhikari. A characterization of a class of functions of finite markov chains. *Annals of Mathematical Statistics*, 36:524–528, 1965.
- [Dia96] P. Diaconis. The cutoff phenomenon in finite markov chains. *Proceedings of the National Academy of Sciences, USA*, 93:1659–1664, 1996.
- [Doo53] J.L. Doob. *Stochastic Processes*. Wiley, 1953.
- [DS91] P. Diaconis and D. Stroock. Geometric bounds for eigenvalues of markov chains. *The Annals of Applied Probability*, 1(1):36–61, 1991.
- [EAM95] R.J. Elliott, L. Aggoun, and J.B. Moore. *Hidden Markov Models - Estimation and Control*. Applications of Mathematics 29. Springer-Verlag, 1995.
- [EM02] Y. Ephraim and N. Merhav. Hidden markov processes. *IEEE Transactions on Information Theory*, 48(6):1518–1569, 2002.
- [Eri70] R.V. Erickson. Functions of markov chains. *Annals of Mathematical Statistics*, 41(3):843–850, 1970.
- [Fai05a] U. Faigle. Discrete quantum markov processes. preprint, 2005.
- [Fai05b] U. Faigle. Persönliche unterhaltung, 2005.
- [Fis86] G. Fischer. *Lineare Algebra*. vieweg, 1986.
- [FR68] M. Fox and H. Rubin. Functions of processes with markovian states i. *Annals of Mathematical Statistics*, 39(3):938–946, 1968.
- [FR69] M. Fox and H. Rubin. Functions of processes with markovian states ii. *Annals of Mathematical Statistics*, 40(3):865–869, 1969.
- [FR70] M. Fox and H. Rubin. Functions of processes with markovian states iii. *Annals of Mathematical Statistics*, 41(2):472–479, 1970.
- [FS05a] U. Faigle and A. Schoenhuth. Note on evolution and asymptotic stationarity of discrete random sources. preprint (submitted to IEEE Transactions on Information Theory), 2005.
- [FS05b] U. Faigle and A. Schoenhuth. Note on negative probabilities and observable processes. In *Algorithms for Optimization with Incomplete Information*, Dagstuhl Seminar Proceedings 05031, pages 1–14, 2005.
- [Gil59] E.J. Gilbert. On the identifiability problem for functions of finite markov chains. *Annals of Mathematical Statistics*, 30:688–697, 1959.
- [GK80] R.M. Gray and J.C. Kieffer. Asymptotically mean stationary measures. *Annals of Probability*, 8:962–973, 1980.
- [Gra90] Robert M. Gray. *Entropy and Information Theory*. Springer Verlag, 1990.
- [Gra01] Robert M. Gray. *Probability, Random Processes and Ergodic Properties*. Springer Verlag, 2001.
- [GS01] G. Grimmett and D. Stirzaker. *Probability and Random Processes*. Oxford University Press, 2001.
- [GVL96] G.H. Golub and C.F. Van Loan. *Matrix Computations*. Johns Hopkins University Press, 1996.
- [Hel65] A. Heller. On stochastic processes derived from markov chains. *Annals of Mathematical Statistics*, 36:1286–1291, 1965.
- [Heu86] H. Heuser. *Lehrbuch der Analysis, Teil 1*. B.G.Teubner, Stuttgart, 1986.
- [HJ99] B.M. Hochwald and P.R. Jelenkovic. State learning and mixing in entropy of hidden markov processes and the gilbert-elliott channel. *IEEE Transactions on Information Theory*, 45:128–138, 1999.
- [HK02] T.S. Han and K. Kobayashi. *Mathematics of Information and Coding*. American Mathematical Society, 2002.
- [Hol01] A.S. Holevo. *Statistical Structure of Quantum Theory*. Springer Verlag, 2001.
- [HT73] M. Heymann and J.A. Thorpe. A note on minimal polynomials. *Linear Algebra and its Applications*, 7:279–280, 1973.

- [IAK92] H. Ito, S.-I. Amari, and K. Kobayashi. Identifiability of hidden markov information sources and their minimum degrees of freedom. *IEEE transactions on information theory*, 38(2):324–333, 1992.
- [IfMS] SUNY at Stony Brook Institute for Mathematical Sciences. Dynamical systems homepage. <http://www.math.sunysb.edu/dynamics>.
- [Jac53] N. Jacobson. *Lectures in Abstract Algebra*, volume II: Linear Algebra. Springer-Verlag, graduate texts in mathematics edition, 1953.
- [Jac59] K. Jacobs. Die Übertragung diskreter informationen durch periodische und fastperiodische kanäle. *Mathematische Annalen*, 137:125–135, 1959.
- [Jac62] K. Jacobs. Über die struktur der mittleren entropie. *Mathematisches Zentralblatt*, 78:33–43, 1962.
- [Jae97a] Herbert Jaeger. Observable operator models and conditioned continuation representations. Technical report, GMD - Forschungszentrum Informationstechnik GmbH, January 1997.
- [Jae97b] Herbert Jaeger. Observable operator models ii: Interpretable models and model induction. Technical report, GMD - Forschungszentrum Informationstechnik GmbH, June 1997.
- [Jae00] H. Jaeger. Observable operator models for discrete stochastic time series. *Neural Computation*, 12(6):1371–1398, 2000.
- [Jay06] E.T. Jaynes. *Probability Theory: The Logic of Science*. Omega Server, University at Albany, omega.albany.edu:8008/JaynesBook.html, 2006.
- [Jel97] F. Jelinek. *Statistical Models for Speech Recognition*. MIT Press, Cambridge, Massachusetts, 1997.
- [Jon83] R.L. Jones. New proofs for the maximal ergodic theorem and the hardy-littlewood maximal theorem. *Proceedings of the American Mathematical Society*, 87(4):681–684, 1983.
- [JR97] S.D. Jacka and G.O. Roberts. On strong forms of weak convergence. *Stochastic Processes and Applications*, 67:41–53, 1997.
- [JS89] M.R. Jerrum and A.J. Sinclair. Approximating the permanent. *SIAM Journal on Computing*, 18:1149–1178, 1989.
- [JS04] M. James and S. Singh. Learning and discovery of predictive state representations in dynamical systems with reset. In *Proceedings 21st Int. Conf. Machine Learning*, pages 417–424, 2004.
- [JZK⁺05] H. Jäger, M. Zhao, K. Kretschmar, T. Oberstein, D. Popovici, and A. Kolling. *Learning observable operator models via the ES algorithm*, chapter to appear. MIT Press, 2005.
- [Kam82] T. Kamae. A simple proof of the ergodic theorem using nonstandard analysis. *Israel Journal of Mathematics*, 42:284–290, 1982.
- [Kna00] B. Knab. *Erweiterungen von Hidden-Markov-Modellen zur Analyse ökonomischer Zeitreihen*. PhD thesis, Universität zu Köln, 2000.
- [Kow75] H.-J. Kowalski. *Lineare Algebra*. deGruyter, 1975.
- [KR81] J.C. Kieffer and M. Rahe. Markov channels are asymptotically mean stationary. *SIAM J. Math. Anal.*, 12(3):293–305, 1981.
- [Kre85] U. Krengel. *Ergodic Theorems*. de Gruyter, 1985.
- [Kre03] Klaus Kretschmar. Learning symbol sequences with observable operator models. Technical report, GMD - Forschungszentrum Informationstechnik GmbH, March 2003.
- [KSSW02] B. Knab, A. Schliep, B. Steckemetz, and B. Wichern. Model-based clustering with hidden markov models and its application to financial time series data. In *Proceedings of the GfKL 2002*, 2002.
- [KW82] Y. Katznelson and B. Weiss. A simple proof of some ergodic theorems. *Israel Journal of Mathematics*, 42:291–296, 1982.
- [LSS01] M.L. Littman, R.S. Sutton, and S. Singh. Predictive representation of state. In *Advances in Neural Information Processing Systems 14 - Proceedings NIPS 2001*, pages 1555–1561, 2001.
- [Lät06] M. Lättsch, 2006. Freundliche Unterstützung.

- [Mac03] D.J.C. MacKay. *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, 2003.
- [McM53] B. McMillan. The basic theorems of information theory. *Annals of Mathematical Statistics*, 24:196–219, 1953.
- [Mey00] C.D. Meyer. *Matrix Analysis and Applied Linear Algebra*. Society for Industrial and Applied Mathematics, 2000.
- [MP03] B. Morris and Y. Peres. Evolving sets and mixing. In *Proceedings of the thirty-fifth annual ACM symposium on theory of computing*, pages 279–286, 2003.
- [MV92] R. Meise and D. Vogt. *Einführung in die Funktionalanalysis*. vieweg, 1992.
- [MZ97] I.L. MacDonald and W. Zucchini. *Hidden Markov and Other Models for Discrete-valued Time Series*. Monographs on Statistics and Applied Probability 70. Chapman and Hall, 1997.
- [Obe02] T.G. Oberstein. Efficient training of observable operator models using context graphs. Diplomarbeit, Universität zu Köln, November 2002.
- [OW83] D. Ornstein and B. Weiss. The shannon-mcmillan-breiman theorem for a class of amenable groups. *Israel Journal of Mathematics*, 44:53–60, 1983.
- [PSS⁺02] P. Pipenbacher, A. Schliep, S. Schneckener, A. Schönhuth, D. Schomburg, and R. Schrader. Proclust: Improved clustering of protein sequences with an extended graph-based approach. *Bioinformatics*, 18(Suppl. 2):182–191, 2002.
- [PY98] M. Pollicott and M. Yuri. *Dynamical Systems and Ergodic Theory*. Cambridge University Press, 1998.
- [Rab89] L. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989.
- [RKTH00] M. Reese, D. Kulp, H. Tammana, and D. Haussler. Genie - gene finding in drosophila melanogaster. *Genome Research*, 10:529–538, 2000.
- [Ros71] M. Rosenblatt. *Markov Processes, Structure and Asymptotic Behavior*. Springer-Verlag, 1971.
- [Sch60] R. Schatten. *Norm ideals of completely continuous operators*. Springer Verlag, 1960.
- [Sch81] H. Schneider. Geometric conditions for the existence of positive eigenvalues of matrices. *Linear Algebra and its Applications*, 38:253–271, 1981.
- [Sch02] A. Schliep. *A Bayesian Approach to Learning Hidden Markov Topology with Applications to Biological Sequence Analysis*. PhD thesis, Universität zu Köln, 2002.
- [SCSS05] A. Schliep, I. Costa, C. Steinhoff, and A. Schönhuth. Analyzing gene expression time courses. *IEEE/ACM Transactions on Computational Biology and Bioinformatics, Special Issue on Machine Learning for Bioinformatics*, 2(3):179–193, 2005.
- [Sha48] C. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 1948.
- [Shi96] P.C. Shields. *The Ergodic Theory of Discrete Sample Paths*, volume 13 of *Graduate Studies in Mathematics*. American Mathematical Society, 1996.
- [Sin94] Y. G. Sinai. *Topics in Ergodic Theory*. Princeton University Press, 1994.
- [SO93] A. Stolcke and S. Omohundro. *Hidden Markov Model Induction by Bayesian Model Merging*, chapter unknown. Number 5 in *Advances in Neural Information Processing Systems*. San Mateo, CA, Morgan Kaufman, 1993.
- [SSS03] A. Schliep, A. Schönhuth, and C. Steinhoff. Using hidden markov models to analyze gene expression time course data. *Bioinformatics*, 19(Suppl. 1):255–263, 2003.
- [SSS04] A. Schliep, C. Steinhoff, and A. Schönhuth. Robust inference of groups in gene expression time-courses using mixtures of hmms. *Bioinformatics*, 20(Suppl. 1):283–289, 2004.
- [Szk04] Arleta Szkola. *The Shannon-McMillan Theorem and Related Results for Ergodic Quantum Spin Lattice Systems and Applications in Quantum Information Theory*. PhD thesis, Technische Universität Berlin, 2004.

- [Upp97] D. R. Upper. *Theory and Algorithms for Hidden Markov Models and Generalized Hidden Markov Models*. PhD thesis, University of California at Berkeley, 1997.
- [Wic01] B. Wichern. *Hidden-Markov-Modelle zur Analyse und Simulation von Finanzzeitreihen*. PhD thesis, Universität zu Köln, 2001.
- [Zim90] R. Zimmer. *Essential Results of Functional Analysis*. Chicago University Press, 1990.

Index

- (Ω, \mathcal{B}) , 17
 (Ω, \mathcal{B}, P) , 20
 (ϵ, δ) -instationär, 139
 2^A , 14
 \ll , 95, 223
 $A \triangle B$, 14
 C_g , 64
 E_g , 129
 $H_n^k(g, \mu^k g)$, 145
 $H^n(g)$, 142
 $H_k^n(g)$, 143
 K_g , 64
 $L_1(\Omega, \mathcal{B}, P)$, 224
 N_g , 134
 P_g , 24
 $P_{\mathcal{G}}$, 154
 Q_g , 120
 T_{Ω} , 26
 $V_a(\lambda)$, 229
 $V_g(\lambda)$, 230
 V_{exp} , 135
 $Z(A)$, 16
 \mathcal{B}_{Σ} , 18
 \square , 14
 Δ^n , 142
 \mathcal{F}_{Σ} , 16
 Ω_{Σ} , 15
 $\Omega_{\bar{a}}$, 15
 Ψ , 170
 Σ , 14
 Σ^* , 14
 Σ^t , 14
 $\Sigma_{\bar{a}}^*$, 15
 $\Sigma_{\bar{a}}^s$, 15
 $\bar{H}(g)$, 142
 $\bar{\mu}$, 105
 \bar{a} , 14
 $\bar{a} \leq \bar{b}$, 15
 $\bar{a} \leq \omega$, 15
 $\mathbb{C}A$, 14
 l_{σ}^{Σ} , 19
 \mathcal{E}_g , 59
 \mathcal{G} , 154
 \mathcal{I}_{Σ} , 91
 $\mathcal{L}_1(\Omega, \mathcal{B}, P)$, 224
 \mathcal{O}_g , 135
 \mathcal{P} , 60
 $\mathcal{P}(A)$, 14
 \mathcal{V}_Q , 119
 \mathcal{V}_g , 59
 μ , 54
 ω , 15
 $\rho(A)$, 229
 $\rho(g)$, 138
 σ -Algebra, 16
 erzeugte, 18
 $\sigma(\mathcal{F})$, 18
 τ_A , 54
 $\tau_{\bar{a}}$, 54
 $a^t, a \in \Sigma$, 15
avgstatdim g , 138
avgtransdim g , 138
card(A), 14
conv(B), 62
*d*_{TV}, 42
dim g , 59
edim g , 59
ergdim g , 137
 $g(A)$, 20
 $g(Z(A))$, 20
 g^B , 46
 $g^{\bar{b}}$, 46
 g_A , 57
 g_m , 23
 $g_{\bar{a}}$, 57
index(λ), 230
 l_{ES} , 104
 l_{SM} , 41
 l_{TV} , 29
 l_{∞} , 18
 l_{∞}^{Σ} , 18
 l_{μ} , 89
 l_{σ} , 19
 m_g , 23
statdim g , 138

trace, 168
transdim g , 138
Übergangsfunktion, 154
Übergangsmatrix, 154
äquivalent
 Prognosegraph, 159
 stochastischer Prozess, 25

- Abbildung
 - messbare, 24
- abgeschlossen
 - Teilmenge von Knoten, 163
- absolutstetig, 223
- adjungiert
 - Matrix, 167
- Algebra, 16
 - σ -, 16
- algebraisch
 - Vielfachheit, 188, 229
- Algorithmus
 - Lern-, 2, 180
- Alphabet, 14
- Alphabetwechsel, 69
- AMS, 103
 - dynamisches System, 107, 221
- Approximationssatz
 - für Maße, 23
 - für Vektoren, 49
- asymptotisch stationär im Mittel
 - dynamisches System, 107
 - stochastischer Prozess, 103
- Basis
 - charakteristische, 77
 - quasi-charakteristische, 77
- Beobachtungsfunktion, 154
- Buchstabe, 14
- Cesaro-Mittel, 103, 192
- charakteristisch
 - Basis, 77
 - Ereignis, 77
 - Polynom, 61, 188
 - System, 77
- Definierende Grundmenge, 16
- Dichte
 - eines Maßes, 223
- Dichtematrix, 168
- Dimension
 - asymptotische Folgen-, 203
 - eines stochastischen Prozesses, 60
 - eines stochastischen Prozesses, 1
 - ergodische, 137
 - Flüchtigkeits-, 138
 - Folgen-, 200
 - mittlere Flüchtigkeits-, 138
 - mittlere stationäre, 138
 - stationäre, 138
 - stochastische, 59
 - stochastische Evolutions-, 59
- dominieren, 223
- Dynamisches System, 6, 86
 - AMS-, 107, 221
 - ergodisches, 115, 116
 - im Mittel asymptotisch stationäres, 107
 - kanonisch induziertes, 87
 - stationäres, 87
- echt
 - Kegel, 63
- Eigenraum, 188
- Eigenwert, 188
 - Index eines, 188, 230
 - nicht-defekter, 188, 230
- Eindeutigkeitssatz, 22
- endlich
 - Prämaß, 20
- Entropie, 141
 - rate, 142
- Ereignis
 - charakteristisches, 77
 - invariantes, 91
- Ereignisvektor, 46
- Ergodensatz
 - von Birkhoff, 103, 113, 118
 - von von Neumann, 117
- Ergodentheorie, 3, 6
- ergodisch
 - Dimension, 137
 - dynamisches System, 115, 116
 - Eigenschaft, 115
 - Vektor, 116
 - Vektorraum, 119
- Ergodische Zerlegung, 129, 134
 - für Prognosegraphen, 163
- Ergodizität, 115
- ES
 - Vektor, 104, 221
- Evolutiondimension
 - eines stochastischen Prozesses, 60
 - stochastische, 59
- Evolutionseffekt, 154
- evolutionskonvergent
 - Vektor, 104
 - wortweise, 104
- Evolutionmatrix, 154
- Evolutionoperator, 54
- Evolutionsraum, 59
- evolutionssummierbar
 - punktweise, 103
 - Quanten-Prognose-Modell, 174
 - Vektor, 104
 - wortweise, 103
- Flüchtigkeitsdimension, 138
 - mittlere, 138
- Folgendimension, 200
 - asymptotische, 203
- Folgenrang, 200
 - asymptotischer, 203
- folgenstabil, 61
 - Operator, 206
- Fortsetzungssatz, 22

- Funktion
 - Übergangs-, 154
 - Beobachtungs-, 154
 - Evolutions-, 154
 - Wellen-, 176
- geometrisch
 - Vielfachheit, 188, 230
- GHMMs, 4
- Graph
 - natürlicher Prognose-, 158
 - Prognose-, 154
- Grenzmaß
 - stationäres, 107
- Grenzvektor, 104
 - stationärer, 104
- Grenzzustand, 174
- Hahn-Zerlegung, 34
- Hauptraum, 188
- Hermite'sch
 - Matrix, 167
- Hidden-Markov-Modell, 4
- Hidden-Markov-Modelle, 5
- Hidden-Markov-Prozesse, 2, 5, 175
- Hidden-Markov-Vektor, 69
- HMMs, 4
- Identifikationsproblem, 3, 74
- Identifikationssatz, 74
- Index
 - eines Eigenwertes, 230
 - eines Eigenwerts, 188
- Inneres einer Menge, 63
- instationär
 - (ϵ, δ) -, 139
 - hochgradig, 139
- invariant
 - Ereignis, 91
 - Menge, 91
- irreduzibel
 - Teilmenge von Knoten, 163
- Jordan
 - Kästchen, 231
 - Normal-Basis, 231
 - Normal-Form, 230
 - Reelle Jordan-Normal-Form, 232
- Jordan-Zerlegung, 35
- Kegel
 - echter, 63
 - konvexer, 62
 - polyedrischer, 67
 - Prognose-, 64, 93
 - reproduzierender, 63
 - solider, 63
 - spitzer, 63
- kommunizieren, 163
- komplex konjugiert
 - Matrix, 167
 - Zahl, 167
- komponentensummenerhaltend
 - Matrix, 209
- Kontraktion
 - positive, 227
- Konvergenz
 - fast überall (von Zufallsvariablen), 222
 - in Norm, 31
 - punktweise, 28
 - schwache, 42
 - schwache Skorokhod-, 41, 226
 - starke, 41
 - stochastische Konvergenz von Zufallsvariablen, 222
 - strikte, 226
 - von Maßen, 41, 226
 - von Zufallsvariablen, 222
 - wortweise, 28
- konvex
 - Kegel, 62
 - Menge, 62, 213
- LDPs, 4
- Lernalgorithmus, 2, 180
- Maß, 20
 - endliches Prä-, 20
 - endliches, signiertes, 224
 - endliches, signiertes Prä-, 20
 - Prä-, 20
 - signiertes, 219
 - signiertes Prä-, 20
 - stationäres Grenz-, 107
 - Wahrscheinlichkeits-, 20
- Maßraum, 20
- Maßvektor, 19
 - signierter, 19
- Markov
 - Operator, 212
 - Prozess, 181
 - Vektor, 65, 66
 - Quanten-Markov-Modell, 173
- Markov-Chain-Monte-Carlo, 138
- Matrix
 - Übergangs-, 154
 - adjungierte, 167
 - Dichte-, 168
 - Evolutions-, 154
 - Hermite'sch, 167
 - komplex konjugierte, 167
 - komponentensummenerhaltende, 209
 - selbstadjungierte, 167
 - stabile, 186
- MCMC, 138
- Menge
 - definierende Grund-, 16

- Inneres einer, 63
- invariante, 91
- konvexe, 62, 213
- messbare, 17
- Zylinder-, 16
- messbar
 - Abbildung, 24
 - Menge, 17
- Messraum, 17
- Metrik
 - der totalen Variation, 42
- Minimalpolynom, 61, 199, 230
- Modell
 - Hidden-Markov-, 4, 5
 - Quanten-Markov-, 173
 - Quanten-Prognose-, 170
 - verallgemeinertes Hidden-Markov-, 4
- negativ
 - Variation, 35
 - Variationsvektor, 39
- nicht-defekt
 - Eigenwert, 188, 230
- nicht-negativ
 - Zustand, 168
- Norm
 - der totalen Variation, 29
 - Operator-, 185
 - Schatten-, 169
- Operator
 - Norm, 185
 - Evolutions-, 54
 - folgenstabiler, 61, 206
 - Markov-, 212
 - Prognose-, 54
 - Prognoseoperator eines Prognosegraphen, 155
 - Schiebe-, 26
 - stabiler, 56, 186
 - stark folgenstabiler, 206
 - stochastischer, 19
 - summierbarer, 192
 - Wechsel-, 69
- PES
 - Vektor, 103
- polyedrisch
 - Kegel, 67
- Polynom
 - charakteristisches, 61, 188
 - Minimal-, 61, 199, 230
- positiv
 - Kontraktion, 227
 - Variation, 35
 - Variationsvektor, 39
- Positivitätsproblem, 76, 123
- Prämaß, 20
 - endliches, 20
 - endliches, signiertes, 20
 - signiertes, 20
- Problem
 - Identifikations-, 3, 74
 - Positivitäts-, 76, 123
- Prognosegraph, 154
 - äquivalenter, 159
 - natürlicher, 158
 - reduzierter, 159
- Prognosekegel, 64, 93
- Prognoseoperator, 54
 - eines Prognosegraphen, 155
- Prognoseraum, 59
- Prognosevektor, 57
 - eines Prognosegraphen, 155
- Prozess
 - AMS-, 103
 - endlich-dimensionaler, 1, 3
 - Hidden-Markov-, 2, 5, 175
 - im Mittel asymptotisch stationärer, 103
 - Markov-, 181
 - stationärer, stochastischer, 88
 - stochastischer, 25
- Punktweise Konvergenz, 28
- Quanten
 - Informationstheorie, 182
 - Irrfahrt, 176
 - Markov-Modell, 173
 - Prognose-Modell, 170
 - Rechnen, 182
 - Spin-Gitter-Systeme, 182
 - Zustand, 168
- quasi-charakteristisch
 - Basis, 77
 - System, 77
- Raum
 - Eigen-, 188
 - Evolutions-, 59
 - Haupt-, 188
 - Maß-, 20
 - Mess-, 17
 - Prognose-, 59
 - Vektor- (der totalen Variation, l_{TV}), 29
 - Vektorraum der endlichen, signierten Maßvektoren (l_{SM}), 41
 - Wahrscheinlichkeits-, 20
 - Zustands-, 167
- reduziert
 - Prognosegraph, 159
- Reelle Jordan-Normal-Form, 232
- reproduzierend
 - Kegel, 63
- Ring, 16

- Satz
 - Approximations- (für Maße), 23
 - Approximations- (für Vektoren), 49
 - Eindeutigkeits-, 22
 - Ergodensatz von Birkhoff, 103, 113, 118
 - Ergodensatz von von Neumann, 117
 - Fortsetzungs-, 22
 - Identifikations-, 74
 - von Cesaro/Cauchy, 111
 - von Chacon-Ornstein, 227
 - von den Neumannschen Reihen, 235
 - von der ergodischen Zerlegung, 134
 - von der Ergodischen Zerlegung von Prognosegraphen, 163
 - von der Jordan-Normal-Form, 230
 - von der reellen Jordan-Normal-Form, 232
 - von Jacobi, 241
 - von Radon-Nikodym, 223
 - von Shannon-McMillan-Breiman, 141
 - Zerlegungs-, 36
- Schatten-Norm, 169
- Schiebeoperator, 26
- schwach
 - Konvergenz, 42
 - Skorokhod Konvergenz, 226
 - Skorokhod-Konvergenz, 41
- selbstadjungiert
 - Matrix, 167
- Sequenz (über Σ), 15
- Shannon-McMillan-Breiman
 - Satz von, 141
- signiert
 - endliches, signiertes Maß, 224
 - Maßvektor, 19
 - Prämaß, 20
- signiertes Maß, 219
- solide
 - Kegel, 63
- Spektral
 - Radius, 187, 229
 - Zerlegung, 181
 - lücke, 139
- Spektrum, 187, 229
- spitz
 - Kegel, 63
- stabil
 - Matrix, 186
 - Operator, 56, 186
- stark
 - folgenstabiler Operator, 206
 - Konvergenz, 41
- Startvektor, 154
- Startzustand, 154
 - zulässiger, 154
- stationär
 - Dimension, 138
 - dynamisches System, 87
 - Grenzmaß, 107
 - Grenzvektor, 104
 - mittlere stationäre Dimension, 138
 - stochastischer Prozess, 88
 - Vektor, 87
- stochastisch
 - Dimension, 59
 - Evolutionsdimension, 59
 - Operator, 19
 - Prozess, 25
 - Vektor, 19
 - Vektorraum, 19
- Stochastischer Prozess, 25
 - AMS-, 103
 - endlich-dimensionaler, 3
 - im Mittel asymptotisch stationärer, 103
 - linear abhängiger, 4
 - stationärer, 88
- strikt
 - Konvergenz, 226
- Suffix Trees, 181
- summierbar
 - Operator, 192
- Symbol, 14
- System
 - charakteristisches, 77
 - dynamisches, 6, 86
 - quasi-charakteristisches, 77
- total
 - Variation, 35
 - Variationsvektor, 39
- Variation
 - Metrik der totalen, 42
 - negative, 35
 - Norm der totalen, 29
 - positive, 35
 - totale, 35
- Variationsvektor
 - negativer, 39
 - positiver, 39
 - totaler, 39
- Vektor
 - Ereignis-, 46
 - ergodischer, 116
 - ES-, 104, 221
 - evolutionskonvergenter, 104
 - evolutionssummierbarer, 104
 - Grenz-, 104
 - Hidden-Markov-, 69
 - Maß-, 19
 - Markov-, 65, 66
 - negativer Variations-, 39
 - PES-, 103
 - positiver Variations-, 39
 - Prognose-, 57

- Prognosevektor eines Prognosegraphen, 155
- punktweise evolutionssummierbarer, 103
- signierter Maß-, 19
- Start-, 154
- stationärer, 87
- stationärer Grenz-, 104
- stochastischer, 19
- totaler Variations-, 39
- Variations-, 39
- wortweise evolutionskonvergenter, 104
- wortweise evolutionssummierbarer, 103
- Zylinder-, 46
- Vektorraum
 - der endlichen, signierten Maßvektoren (l_{SM}), 41
 - der ES-Vektoren (l_{ES}), 104
 - der stationären, endlichen, signierten Maßvektoren (l_{μ}), 89
 - der totalen Variation (l_{TV}), 29
 - ergodischer, 119
 - stochastischer (l_{σ}), 19
- Verallgemeinertes Hidden-Markov-Modell, 4
- Vielfachheit
 - algebraische, 188, 229
 - geometrische, 188, 230
- Wahrscheinlichkeitsmaß, 20
- Wahrscheinlichkeitsraum, 20
- Wechseloperator, 69
- Wellenfunktion, 176
 - Evolution, 175
- Wort, 14
 - Länge eines, 14
 - leeres, 14
- wortweise
 - evolutionskonvergenter Vektor, 104
- Wortweise Konvergenz, 28
- Zerlegung
 - ergodische, 129, 134
 - ergodische Zerlegung für Prognosegraphen, 163
 - Hahn-, 34
 - Jordan-, 35
- Zerlegungssatz, 36
- Zufallsvariable, 24
 - Konvergenz, 222
- zulässig
 - Startzustand, 154
- Zustand, 167
 - Grenz-, 174
 - nicht-negativer, 168
 - Quanten-, 168
 - Start-, 154
- Zustandsraum, 167
- Zylindermenge, 16
- Zylindervektor, 46

Nachwort und Danksagung

*Die Summe unserer Existenz,
durch Vernunft dividiert,
geht niemals rein auf,
sondern immer bleibt ein wunderlicher Rest.*

Goethe

Im September 1999, vor sechseinhalb Jahren, habe ich meine Stelle am Zentrum für Angewandte Informatik der Universität zu Köln angetreten. Für mein Empfinden ist seitdem eine kleine Ewigkeit vergangen. Mit dieser Dissertation wird mir allerdings immer auch einer der schwierigsten und unsichersten Abschnitte meines Lebens in Erinnerung bleiben. Ich erwähne dies aus zwei Gründen. Zum einen will ich den an ihr beteiligten Personen entsprechend danken. Und zum anderen, einem an meinem Lebenslauf interessierten Leser die Umstände der Entstehung dieser Arbeit näher zu bringen.

Schon wenige Wochen nach Antritt der Stelle kristallisierte sich heraus, dass meine Doktorarbeit Teil der Bioinformatik werden würde - ein Thema, auf das ich mich mit großem Interesse stürzte. Bereits gegen Ende des Jahres 1999 begann ich, an einem Dissertationsprojekt mit dem Thema 'Computergestützte Analyse von 2D-Gel-Elektrophorese-Experimenten' in Kooperation mit Wissenschaftlern der University Edinburgh zu forschen. Im Juli 2000 jedoch diagnostizierten Ärzte bei mir ein 'Non-Hodgkin-Lymphom'. Es folgte eine sich über ein dreiviertel Jahr erstreckende Periode von Therapien und eine sich anschließende Phase von einem weiteren Jahr der Rehabilitation. Danach hatte sich mein Dissertationsprojekt in dieser schnelllebigen "Branche" erübrigt.

Obwohl in der Phase meiner Rehabilitation regelmäßige Arbeit nicht möglich war, las ich viel mathematische Fachliteratur und auf diesem Wege wurde mir bewusst, dass, über mein Gefallen an Denksport (ich bin ein begeisterter Schachspieler) hinaus, die Mathematik für mich an Bedeutung gewonnen hatte. Die Arbeit, Aussagen zu bestätigen oder zu widerlegen und der dazu erforderliche Mut zum Schritt in unbekannte Welten, schien nun auch ein wenig Abbild meines persönlichen Schicksals, schwierige und gänzlich unsichere Situationen meistern zu müssen, zu sein. Hierzu fällt mir ein, dass Andrew Wiles seine zehnjährigen(!) Arbeiten am Beweis des Fermatschen Satzes mit dem Herumirren in einem großen, finsternen Raum voller

Möbel mit scharfen Kanten verglichen hat. Der Moment, in dem er den Lichtschalter fand, hätte ihn aber für Jahre schmerzlicher Unsicherheit reichlich entlohnt.

Zu Beginn des Jahres 2002 habe ich schließlich die Arbeiten wieder voll aufgenommen, nun im Rahmen des in Köln neu aufgesetzten "Cologne University Bioinformatics Centre (CUBIC)". Im Herbst 2002 dann begann ich mit den Arbeiten an meinem neuen Dissertationsprojekt. Dabei mögen oben genannte Gründe dazu beigetragen haben, dass aus dieser Arbeit ein Stück echte Mathematik geworden ist.

Diesen für eine Dissertation ungewöhnlichen Weg haben viele Menschen begleitet, denen ich danken möchte und ich beginne mit denen, die unmittelbar an dieser Arbeit beteiligt gewesen sind.

Herrn Prof. Dr. Faigle, meinem Erstgutachter, möchte ich zuvorderst und ganz besonders danken. Zuerst einmal für sein geduldiges Verständnis während meiner Krankheit. Aber vor allem dafür, diese Arbeit in ganz besonders hohem Maße gefördert zu haben. Vieles in der Arbeit ist aufgrund seiner Anregungen entstanden und etliche Sätze sind Antworten auf seine Fragen oder sind Ausarbeitungen seiner Entwürfe. Ebenso danke ich ihm für die Vermittlung einer mir neuen Sichtweise der Mathematik, an der ich Gefallen gefunden habe. Mathematik ist, vor allem Anderen, eben auch Ansichtssache. Die gemeinsamen Diskussionen haben mir großen Spaß bereitet.

An zweiter Stelle möchte ich Herrn Prof. Dr. Jäger, dem Vater dieser Arbeit, danken. Er hat mir dieses auch ihm so wichtige Thema vermittelt und mir, vor allem in seiner Bonner Zeit, mit einer großen Zahl von Tipps und Ratschlägen zur Seite gestanden. Ebenso möchte ich mich bei ihm dafür bedanken, mir einen Aufenthalt in Bremen ermöglicht zu haben, der mir in schöner Erinnerung bleiben wird.

Herrn Prof. Dr. Schrader, der mich anno 1999 eingestellt hat, möchte ich ebenso ganz ausdrücklich danken. Er hat durch dauerhafte Unterstützung, Geduld und Großzügigkeit die Grundlage für diese Dissertation geschaffen.

Herzlich danke ich auch Herrn Priv.-Doz. Dr. Randerath für die Protokollführung.

Meinen Kollegen vom ZAIK möchte ich für die angenehme Arbeitsatmosphäre danken. Ich möchte besonders denen danken, die diese Arbeit korrigieren geholfen haben. Und hierbei vor allem Dominique Andres und Susanne Motameny für die gründlichsten und tiefsten Korrekturen. Aber auch bei Jutta Gebert, Lars Kaderali, Ralf Müller und Nicole Radde will ich mich bedanken. Und bei meiner Mutter und meiner Schwester, die sich der Textpassagen in vielerlei Hinsicht angenommen haben.

Vier weitere Personen waren sehr wichtig für mich und diese Arbeit, und das sind, in alphabetischer Reihenfolge, Prof. Dr. Winfried Hochstättler, Lars Kaderali, Martin Olschewski und Dr. Alexander Schliep.

Winfried Hochstättler will ich zunächst danken für unzählige Hinweise, die mich aus gedanklichen Sackgassen befreit haben. Die Aussagen z.B., ich bräuchte nicht zu verzweifeln, “es ist eben schwer, zu beweisen, dass ein unendlich-dimensionaler Vektorraum endlich-dimensional ist” und “Mathematiker sind wie Katzen, die große Mäuse jagen und dabei viele kleine fressen” haben dieser Arbeit sehr geholfen. Ich will ihm auch danken für eine Vielzahl nützlicher wissenschaftlicher Tipps.

Lars Kaderali danke ich dafür, ein über Jahre hinweg äußerst hilfsbereiter, freundschaftlicher und freundlicher Büro- und Zockerkollege gewesen zu sein. Leider kann ich trotz in Aussicht gestellter fürstlicher Entlohnung seinem letzten Wunsch an diese Arbeit nicht entsprechen.

Martin Olschewski will ich dafür danken, mir unter Schachspielerkollegen die Welt des Internet-Schachs eröffnet zu haben, sowie dafür, durch ihn, mehr als durch alle Anderen, denen ich begegnet bin, Praxis im Umgang mit Rechnern erlangt zu haben.

Alexander Schliep will ich einfach dafür danken, ein Freund in dieser Welt der Wissenschaftler zu sein und nicht zuletzt dafür, dass mir unsere gemeinsame Projektarbeit in vielerlei Hinsicht sehr geholfen hat und in schöner Erinnerung bleiben wird.

Bei Frau Teuner möchte ich mich dafür bedanken, ein ums andere Mal bürokratische Dschungel zu lichten.

Ich bedanke mich bei den anderen Zockern, Niels Lange, Lars Kaderali und Stefan Neuhaus dafür, mir ein ums andere Mal den Nachhauseweg erspart zu haben.

Noch einmal möchte ich mich bei den übrigen Kollegen der CUBIC-Arbeitsgruppe, das sind Jutta Gebert und Nicole Radde für viele gemeinsame Diskussionen bedanken. Ebenso bei allen CUBIC-Kollegen der anderen Institute und hier besonders Mark Lohmann und Beate Marx für die Unterstützung bei der Organisation meiner Vorlesungen.

Ich möchte mich für die gemeinsame Projektarbeit bei den mir zugeteilten Studenten bedanken. Im Einzelnen waren dies Guoqing Xia, Xue Li, Olof Persson, Marcelo Reis, Dr. Hiwot Chemu, Mateus Patricio und Rileen Sinha. Danke auch an die letztgenannten Vier für das Verständnis gegenüber einem Supervisor, der selber eine Arbeit zu schreiben hat.

Nicht zuletzt will ich Herrn Prof. Dr. Schomburg als dem Leiter von CUBIC für die Bereitstellung der Stelle danken. Und dem BMBF für die nötige Finanzierung.

Ich danke außerdem allen, denen ich bei Gastaufenthalten und auf Konferenzen begegnet bin und die mir gute Tipps gegeben haben. Besonders möchte ich hierbei den Mitarbeitern der Arbeitsgruppe Vingron des Max-Planck-Instituts für Molekulare Genetik in Berlin danken und hier vor allem Dr. Christine Steinhoff und Ivan Gesteira Costa da Filho. Ich danke dem Institut auch für die Finanzierung meiner zahlreichen Gastaufenthalte.

Ebenso bedanke ich mich bei den ehemaligen und aktuellen Mitarbeitern von Prof. Jäger, das sind Klaus Kretzschmar, Tobias Oberstein und Mingjie

Zhao für nützliche Hinweise und Diskussionen.

Viele Menschen, die nicht unmittelbar mit dieser Arbeit zu tun hatten, waren dennoch sehr wichtig dafür. Das sind zuallererst meine Eltern, die es geschafft haben, ein meist grübelndes Monster vorbildlich zu unterstützen. Das ist meine Schwester Friederike, deren Ratschläge immer aufbauend für mich waren. Und das ist meine Großtante Johanna Wagner, die mir dauerhaft unter die Arme gegriffen hat, vielleicht auch, damit die Mathematik in unserer Familie mit ihr nicht an ein Ende gelangt wäre.

Nicht alle haben eine Ersatzfamilie, ich jedoch schon, nämlich Sinéad Traynor und Roland Popp. Auch wenn ich finde, dass regelmäßige warme Mahlzeiten in einer Familie dazugehören, so will ich mich "anstandshalber" dafür bedanken. Natürlich vielen Dank für alles Sonstige.

Vielen Dank an meinen Freund Till Hanisch, einen meiner treuesten und verständnisvollsten Begleiter während meiner Krankheit. Ich wünsche ihm alles Gute an dieser Stelle.

Vielen Dank an Fiammetta Ciuffi für bedingungslose Unterstützung.

Vielen Dank an Daniel einfach so, aber vor allem auch für die Beschaffung einer alternativen Therapie in seiner Funktion als Dr. med. Schiefer, die 2000 noch kein Standard war (es jetzt aber ist).

Vielen Dank an "Mr. Paralleluniversum" Urs Roth, nicht nur für kompetente Ratschläge bezüglich der Theorie der Paralleluniversen (im Volksmund Quantenmechanik), sondern auch für praktische Umsetzungen eigener Universen, wie z.B. der (T/F)ango-Welt.

Vielen Dank an Andreas Schröder dafür, dass Mittagessen nicht immer gleich Mensa war und für einen endlosen Sommer in dem Moment, in dem ich einen brauchte.

Vielen Dank an Lorenz Jüngling für große Zeiten und jede Menge guter Ratschläge. Von mir aus kann die Reform des deutschen Bildungssystems jetzt losgehen.

Vielen Dank an Patrick Kiefer für den Aufbau wissenschaftlicher Netzwerke in Südfrankreich.

Vielen Dank an Matthias Futschik für die Nach-ISMB-Urlaube, vor allem den auf Neuseeland. Ich wünsche ihm weiterhin eine kühle Stirn im Umgang mit Ko-Autoren und Acknowledgments.

Zu guter Letzt will ich mich bei höheren Mächten dafür bedanken, meine Freundin Katrin kennengelernt zu haben, ein sehr erfreulicher Schicksalschlag. Ich will mich bei ihr für ihre Unterstützung in schwierigen Zeiten bedanken, die kaum zu vermeiden sind, schreibt man an einer Dissertation. Vielleicht gibt es ja wirklich eine dieser schönen, friedlichen, grünen Inseln, die wir uns dann vorstellten.

Erklärung

Ich versichere, daß ich die von mir vorgelegte Dissertation selbständig angefertigt, die benutzten Quellen und Hilfsmittel vollständig angegeben und die Stellen der Arbeit – einschließlich Tabellen, Karten und Abbildungen –, die anderen Werken im Wortlaut oder dem Sinn nach entnommen sind, in jedem Einzelfall als Entlehnung kenntlich gemacht habe, daß diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat, daß sie – abgesehen von unten angegebenen Teilpublikationen – noch nicht veröffentlicht worden ist sowie, daß ich eine solche Veröffentlichung vor Abschluß des Promotionsverfahrens nicht vornehmen werde. Die Bestimmungen der Promotionsordnung sind mir bekannt. Die von mir vorgelegte Dissertation ist von Prof. Dr. U. Faigle betreut worden.

Teilpublikationen: [FS05b],[FS05a]