

Vorlesungsskript Symbolische Dynamik und Kodierung

PD Dr. Christoph Kawan
Fakultät für Informatik und Mathematik
Universität Passau
94032 Passau

e-Mail: christoph.kawan@uni-passau.de

Datum: 23.03.2016

Inhaltsverzeichnis

1	Motivation und Einführung	4
1.1	Historische Motivation	4
1.2	Anwendung in der Speicherung von Daten	6
2	Shifträume	7
2.1	Volle Shifts	7
2.2	Shifträume	8
2.3	Sprachen	10
2.4	Höhere Blockshifts und höhere Potenzshifts	11
2.5	Gleitende Blockcodes	13
3	Shifts von endlichem Typ	19
3.1	Einschränkungen von endlichem Typ	19
3.2	Graphen und ihre Shifts	22
3.3	Graphendarstellungen von Shifts von endlichem Typ	27
3.4	Zustandsaufspaltung	33
4	Sofische Shifts	37
4.1	Darstellung sofischer Shifts	38
4.2	Charakterisierungen sofischer Shifts	43
4.3	Minimale rechtsauflösende Darstellungen	46
4.4	Konstruktionen und Algorithmen	54
5	Entropie	65
5.1	Definitionen und elementare Eigenschaften	65
5.2	Perron-Frobenius-Theorie	70
5.3	Berechnung der Entropie	75
5.4	Irreduzible Komponenten	78
5.5	Zyklische Struktur	84
6	Finite-State-Codes	92
6.1	Straßenfärbungen und rechtsabschließende Beschriftungen	92
6.2	Finite-State-Codes	99
6.3	Approximative Eigenvektoren	102

6.4	Konstruktion von Codes	106
6.5	Gleitende Block-Dekodierer	111
7	Übungsaufgaben	114

Dies ist das Skript zu meiner Vorlesung *Symbolische Dynamik und Kodierung*, die ich im Wintersemester 2015/16 an der Universität Passau gehalten habe. Der Inhalt umfasst den größten Teil der Kapitel 1 bis 5 des Buchs *An Introduction to Symbolic Dynamics and Coding* von Lind und Marcus [1]. Alle Definitionen und Resultate sind aus diesem Buch entnommen. Die einzige Abweichung stellt der Satz von Perron-Frobenius (Satz 5.18) dar, der hier in voller Allgemeinheit bewiesen wird und in [1] nur für die Dimension 2. Ich bedanke mich bei meinen Zuhörern für ihre Geduld und ihr Feedback. Für Hinweise auf Fehler oder Ungenauigkeiten bin ich weiterhin dankbar.

1 Motivation und Einführung

Die folgende kurze Einführung in die Theorie der symbolischen dynamischen Systeme setzt Kenntnisse gewisser Begriffe der Analysis voraus. In der weiteren Vorlesung werden diese jedoch nicht mehr benötigt und es reichen für weite Teile Kenntnisse der zentralen Begriffe der Linearen Algebra.

1.1 Historische Motivation

Die Theorie der symbolischen dynamischen Systeme ist ursprünglich motiviert durch das Studium von Systemen, die von gewöhnlichen Differenzialgleichungen

$$\dot{x}(t) = f(x(t)), \quad f : M \rightarrow \mathbb{R}^n,$$

erzeugt werden. Dabei kann M eine offene Teilmenge von \mathbb{R}^n oder eine Untermannigfaltigkeit kleinerer Dimension sein (z.B. eine Kugeloberfläche im \mathbb{R}^3). Unter Standardannahmen hat obige Gleichung zu jedem Anfangswertproblem $x(0) = x_0$ eine eindeutige lokale Lösung, d.h. eine differenzierbare Kurve $t \mapsto \varphi(t, x_0)$, definiert auf einem offenen Intervall $I(x_0)$ mit $0 \in I(x_0)$, so dass

$$\varphi(0, x_0) = x_0 \quad \text{und} \quad \frac{\partial}{\partial t} \varphi(t, x_0) = f(\varphi(t, x_0)) \quad \text{für alle } t \in I(x_0).$$

Ist $I(x_0) = \mathbb{R}$ für alle $x_0 \in M$, so nennt man die Abbildung

$$\varphi : \mathbb{R} \times M \rightarrow M, \quad (t, x) \mapsto \varphi(t, x),$$

den Fluss der Differenzialgleichung.¹ Dieser hat die Eigenschaften

$$\varphi(0, x) = x \quad \text{und} \quad \varphi(t + s, x) = \varphi(t, \varphi(s, x)). \quad (1)$$

In the Theorie der dynamischen Systeme interessiert man sich für das Verhalten der Lösungen (auch Trajektorien genannt) $t \mapsto \varphi(t, x)$, wenn t gegen $\pm\infty$ geht. Im Allgemeinen kann es sehr schwierig sein, darüber sinnvolle und interessante Aussagen zu treffen.

¹Einfaches Beispiel: $\dot{x}(t) = ax(t)$ mit $a \in \mathbb{R}$ und $M = \mathbb{R}$. Dann ist $\varphi(t, x_0) = x_0 e^{at}$.

Die symbolische Dynamik liefert eine Methode, den Fluss einer Differenzialgleichung zu verstehen, indem man eine Diskretisierung sowohl in der Zeit als auch im Raum vornimmt. Eine Diskretisierung in der Zeit erhält man ganz einfach, indem man die Abbildung $\varphi : \mathbb{R} \times M \rightarrow M$ in der ersten Komponente auf eine Menge der Form $\tau \cdot \mathbb{Z} = \{\tau \cdot k : k \in \mathbb{Z}\}$ einschränkt, wobei τ eine positive reelle Zahl ist. Definiert man $\phi := \varphi(\tau, \cdot) : M \rightarrow M$, so erhält man aus der Flusseigenschaft (1), dass

$$\varphi(\tau k, x) = \phi^k(x),$$

wobei die Abbildung $\phi^k : M \rightarrow M$ definiert ist durch

$$\phi^k(x) := \begin{cases} \phi \circ \dots \circ \phi \circ \phi(x) & (k \text{ mal}) & \text{falls } k > 0, \\ x & & \text{falls } k = 0, \\ \phi^{-1} \circ \dots \circ \phi^{-1} \circ \phi^{-1}(x) & (-k \text{ mal}) & \text{falls } k < 0. \end{cases}$$

Man erhält so ein diskretes dynamisches System, das durch die Iteration der Abbildung ϕ erzeugt wird.

Für die Diskretisierung im Raum betrachtet man eine (in der Regel geschickt gewählte) endliche Menge von offenen disjunkten Teilmengen P_1, \dots, P_r , deren Abschlüsse $\overline{P}_1, \dots, \overline{P}_r$ eine Überdeckung von M bilden, d.h. $M = \bigcup_{i=1}^r \overline{P}_i$. Jedem $x \in M$ können wir dann mindestens eine beidseitig unendliche Folge

$$a(x) = \dots a_{-2} a_{-1} a_0 a_1 a_2 \dots = (a_k)_{k \in \mathbb{Z}} \quad \text{mit } a_k \in \{1, \dots, r\}$$

zuordnen, so dass $\phi^k(x) \in \overline{P}_{a_k}$ für alle $k \in \mathbb{Z}$. Wenden wir die Abbildung ϕ auf x an, d.h. gehen wir einen Zeitschritt weiter, so erhalten wir die Folge

$$a(\phi(x)) = (a_{k+1})_{k \in \mathbb{Z}}.$$

Dies bedeutet, dass eine Anwendung der Abbildung ϕ einer Verschiebung der Folgenglieder a_k um eine Stelle nach links entspricht. Wir definieren

$$\Sigma_\phi := \{(a_k)_{k \in \mathbb{Z}} : \exists x_0 \in M \text{ mit } \phi^k(x_0) \in \overline{P}_{a_k} \text{ f.a. } k \in \mathbb{Z}\}.$$

Die Zuordnung $x \mapsto a(x)$ ist nicht eindeutig, da die Mengen $\overline{P}_1, \dots, \overline{P}_r$ sich an den Rändern überschneiden können, so dass ein Punkt in M mehreren dieser Mengen angehören kann. Setzen wir allerdings voraus, dass die Menge²

$$\bigcap_{n=0}^{\infty} \overline{\bigcap_{k=-n}^n \phi^{-k}(P_{a_k})}$$

für jede Folge $(a_k)_{k \in \mathbb{Z}}$ in Σ_ϕ aus genau einem Punkt besteht, so können wir umgekehrt eine eindeutige Abbildung

$$\pi : \Sigma_\phi \rightarrow M$$

²Man beachte, dass $\overline{\bigcap_{k=-n}^n \phi^{-k}(P_{a_k})} \subseteq \bigcap_{k=-n}^n \phi^{-k}(\overline{P}_{a_k})$.

definieren, die jede Folge $(a_k)_{k \in \mathbb{Z}}$ in Σ_ϕ auf den entsprechenden Punkt abbildet. Dies liefert ein kommutatives Diagramm

$$\begin{array}{ccc} \Sigma_\phi & \xrightarrow{\sigma} & \Sigma_\phi \\ \pi \downarrow & & \downarrow \pi \\ M & \xrightarrow[\phi]{} & M \end{array}$$

Dabei ist σ die „Schiebeabbildung“ (im Folgenden *Shift-Abbildung* genannt), die eine Folge $(a_k)_{k \in \mathbb{Z}}$ auf $(a_{k+1})_{k \in \mathbb{Z}}$ abbildet. Man kann das diskrete dynamische System (Σ_ϕ, σ) als Digitalisierung von (M, ϕ) auffassen. Man nennt (Σ_ϕ, σ) ein symbolisches dynamisches System. Die Idee ist, die Dynamik von ϕ (und damit die Dynamik der Differenzialgleichung) zu verstehen durch Analyse des einfacheren Systems (Σ_ϕ, σ) . Diese Idee taucht zum ersten Mal in einer Arbeit von Jacques Hadamard aus dem Jahr 1898 auf.

1.2 Anwendung in der Speicherung von Daten

Eine andere Motivation für die Untersuchung symbolischer dynamischer Systeme findet man in der Speicherung und Übertragung von Daten. Aus technischen Gründen werden binäre Daten auf magnetischen (und optischen) Speichermedien nicht direkt abgespeichert. Ein Grund dafür ist, dass beim Lesen der Daten zwei Einsen, die jeweils einen elektrischen Impuls im Lesekopf induzieren, nicht zu nahe beieinander liegen sollten, weil es sonst zu Überlagerungseffekten kommt. Deshalb werden zusätzliche Nullen eingefügt, die dies verhindern. Aus anderen technischen Gründen sollten aber auch nicht zu viele Nullen hintereinander vorkommen. Dies führt zu Codes, die bestimmten Regeln gehorchen müssen, welche die Menge der erlaubten Bitfolgen einschränken. Es stellt sich die Frage, wie man auf effiziente Weise beliebige Bitfolgen in Bitfolgen mit solchen Einschränkungen transformieren kann.

Oft ist es möglich, dies mathematisch durch einen *Shift von endlichem Typ* zu beschreiben. Sei dazu $\mathcal{A} = \{0, 1, \dots, N-1\}$ eine endliche Menge, die wir *Alphabet* nennen. Die Elemente von \mathcal{A} werden *Symbole* (oder *Buchstaben*) genannt. Wir betrachten die Menge

$$\mathcal{A}^{\mathbb{Z}} = \{(x_k)_{k \in \mathbb{Z}} : x_k \in \mathcal{A} \text{ für alle } k \in \mathbb{Z}\}$$

aller Folgen mit Werten in \mathcal{A} . Auf $\mathcal{A}^{\mathbb{Z}}$ können wir, wie im vorherigen Abschnitt, die Shift-Abbildung

$$\sigma : \mathcal{A}^{\mathbb{Z}} \rightarrow \mathcal{A}^{\mathbb{Z}}, \quad (x_k)_{k \in \mathbb{Z}} \mapsto (x_{k+1})_{k \in \mathbb{Z}},$$

betrachten. Wollen wir gewisse Folgen von Symbolen ausschließen, so können wir dies z.B. tun, indem wir für jedes Paar $(i, j) \in \mathcal{A} \times \mathcal{A}$ festlegen, ob in einer Folge $(x_k)_{k \in \mathbb{Z}}$ das Symbol j auf das Symbol i folgen darf. Diese Einschränkungen

lassen sich elegant in einer $N \times N$ Matrix $A = (a_{ij})_{0 \leq i, j \leq N-1}$ zusammenfassen, die nur Nullen und Einsen als Einträge hat. Die Menge der erlaubten Folgen ist dann gegeben durch

$$\Sigma_A := \{(x_k)_{k \in \mathbb{Z}} \in \mathcal{A}^{\mathbb{Z}} : a_{x_k x_{k+1}} = 1 \text{ für alle } k \in \mathbb{Z}\}.$$

Diese Menge ist invariant unter der Shift-Abbildung σ , d.h. aus $(x_k)_{k \in \mathbb{Z}} \in \Sigma_A$ folgt $(x_{k+1})_{k \in \mathbb{Z}} \in \Sigma_A$. Wir erhalten also ein diskretes dynamisches System durch Iteration der Abbildung

$$\sigma|_{\Sigma_A} : \Sigma_A \rightarrow \Sigma_A.$$

Daran sieht man zwar noch nicht, wie genau sich die oben beschriebenen Einschränkungen in den Bitfolgen beschreiben lassen, aber dies wird im Lauf der Vorlesung klar werden. Zunächst müssen wir einiges an Terminologie und elementaren Konstruktionen einführen.

2 Shifträume

2.1 Volle Shifts

Information wird meist durch Folgen von diskreten Symbolen dargestellt, die einem endlichen Alphabet entnommen sind (z.B. geschriebener Text oder binäre Daten). Um dies mathematisch allgemein zu beschreiben, sei \mathcal{A} eine endliche Menge, die wir *Alphabet* nennen. Die Elemente von \mathcal{A} heißen *Symbole*. Typischerweise werden wir die Symbole mit a, b, c, \dots oder $0, 1, 2, \dots$ bezeichnen. Wir werden unendliche Folgen mit Werten in \mathcal{A} betrachten, gleichwohl im wahren Leben natürlich keine unendlichen Folgen auftauchen. Dies ist eine Idealisierung und führt zu einer eleganteren mathematischen Beschreibung (wie auch die reellen Zahlen und andere Konzepte der Analysis). Wir bezeichnen eine solche Folge mit

$$x = (x_i)_{i \in \mathbb{Z}} \quad \text{oder} \quad x = \dots x_{-2} x_{-1} x_0 x_1 x_2 \dots,$$

wobei jedes x_i ein Symbol in \mathcal{A} ist. Um die Position der Stelle x_0 zu kennzeichnen, ist es gebräuchlich einen „Dezimalpunkt“ zu verwenden, der die x_i mit $i < 0$ von denen mit $i \geq 0$ trennt, z.B.

$$x = \dots 010.1101 \dots$$

2.1 Definition: *Ist \mathcal{A} ein endliches Alphabet, so ist der volle \mathcal{A} -Shift die Menge aller bi-unendlichen Folgen mit Werten in \mathcal{A} . Als r -Shift bezeichnen wir den vollen \mathcal{A} -Shift über dem Alphabet $\mathcal{A} = \{0, 1, \dots, r-1\}$, wobei $r \in \mathbb{N} = \{1, 2, 3, \dots\}$. Bezeichnung:³*

$$\mathcal{A}^{\mathbb{Z}} = \{(x_i)_{i \in \mathbb{Z}} : x_i \in \mathcal{A} \text{ für alle } i \in \mathbb{Z}\}.$$

Die Elemente von $\mathcal{A}^{\mathbb{Z}}$ bezeichnen wir auch als Punkte.

³Für beliebige Mengen X und Y ist Y^X die Standardnotation für die Menge aller Abbildungen von X nach Y .

Ein *Block* (oder *Wort*) über \mathcal{A} ist eine endliche Sequenz von Symbolen aus \mathcal{A} . Ist z.B. $\mathcal{A} = \{a, b\}$, so ist $aababb$ ein Block über \mathcal{A} . Für eine mathematisch elegantere Beschreibung führen wir den *leeren Block* ein, der aus Null Symbolen besteht und bezeichnen diesen mit ε . Die *Länge* eines Blocks u ist die Anzahl seiner Symbole und wird mit $|u|$ bezeichnet, z.B. $|aababb| = 6$ oder $|\varepsilon| = 0$. Als *k-Block* bezeichnen wir einen Block der Länge k . Wir schreiben \mathcal{A}^k für die Menge aller k -Blöcke. Ein Teilblock von $u = a_1a_2 \dots a_k$ ist ein Block der Form $a_i a_{i+1} \dots a_j$ mit $1 \leq i \leq j \leq k$. Wir betrachten zudem ε als Teilblock von jedem Block.

Ist $x \in \mathcal{A}^{\mathbb{Z}}$ und $i \leq j$, so schreiben wir

$$x_{[i,j]} := x_i x_{i+1} \dots x_j.$$

Ist $i > j$, so definieren wir $x_{[i,j]} := \varepsilon$. Außerdem schreiben wir

$$x_{[i,j]} := x_i x_{i+1} \dots x_{j-1}.$$

Analog führen wir $x_{[i,\infty)} = x_i x_{i+1} x_{i+2} \dots$ und $x_{(-\infty, i]} = \dots x_{i-2} x_{i-1} x_i$ ein (was allerdings keine Blöcke sind).

Zwei Blöcke u und v können *verkettet* werden, indem v an u „angehängt“ wird. Der so entstehende Block uv hat die Länge $|uv| = |u| + |v|$. Ist u ein Block, so schreiben wir u^n für die Verkettung von n Kopien von u , und wir definieren $u^0 := \varepsilon$. Dann gilt

$$u^m u^n = u^{m+n} \quad \text{für alle } m, n \geq 0.$$

Wir schreiben u^∞ für die Folge $\dots uuu.uuu \dots$.

2.2 Definition: Die Shift-Abbildung $\sigma : \mathcal{A}^{\mathbb{Z}} \rightarrow \mathcal{A}^{\mathbb{Z}}$ bildet x auf $y = \sigma(x)$ ab, wobei $y_i = x_{i+1}$ für alle $i \in \mathbb{Z}$.

Es ist leicht zu sehen, dass σ bijektiv ist. Die Umkehrabbildung σ^{-1} verschiebt jedes Folgenglied um eine Stelle nach rechts.

2.3 Definition: Ein Punkt $x \in \mathcal{A}^{\mathbb{Z}}$ heißt *periodisch* unter σ , falls $\sigma^n(x) = x$ für ein $n \in \mathbb{N}$ und wir sagen, dass x die *Periode* n hat bzw. dass x ein *n-periodischer Punkt* ist. Die *kleinste Zahl* n mit dieser Eigenschaft heißt *Primperiode*. Gilt $\sigma(x) = x$, so heißt x *Fixpunkt*.

2.2 Shifträume

Die Symbolfolgen, die wir studieren werden, unterliegen oft gewissen Einschränkungen. In diesem Abschnitt führen wir den fundamentalen Begriff des *Shiftraums* ein. Damit wird eine Teilmenge eines vollen \mathcal{A} -Shifts bezeichnet, bestehend aus allen Folgen, die gewissen fest gewählten Einschränkungen unterliegen.

Ist $x \in \mathcal{A}^{\mathbb{Z}}$ und w ein Block über \mathcal{A} , so sagen wir, dass w in x *vorkommt*, falls $w = x_{[i,j]}$ für gewisse i, j . Sei \mathcal{F} eine Menge von Blöcken über \mathcal{A} , die wir uns als verbotene Blöcke denken. Für jedes solche \mathcal{F} bezeichnen wir mit $X_{\mathcal{F}}$ diejenige Teilmenge von $\mathcal{A}^{\mathbb{Z}}$, die aus allen Folgen besteht, in denen kein Block aus \mathcal{F} vorkommt.

2.4 Definition: Ein Shiftraum (oder Shift) ist eine Teilmenge X eines vollen \mathcal{A} -Shifts von der Form $X = X_{\mathcal{F}}$ für eine Menge \mathcal{F} verbotener Blöcke über \mathcal{A} .

Die Menge \mathcal{F} kann endlich oder unendlich sein. In jedem Falle ist sie höchstens abzählbar unendlich, da ihre Elemente aufgelistet werden können. Dazu schreibt man zuerst alle Elemente der Länge 1, dann alle der Länge 2 usw. auf. Da das Alphabet endlich ist, gibt es davon jeweils nur endlich viele. Zu einem gegebenen Shiftraum kann es viele verschiedene Mengen \mathcal{F} geben, die ihn beschreiben. Insbesondere ist die leere Menge \emptyset ein Shiftraum, wenn wir $\mathcal{F} = \mathcal{A}$ wählen. Ist ein Shiftraum X in einem Shiftraum Y enthalten, so bezeichnen wir X als *Teilshift* von Y .

2.5 Beispiel:

- (1) Der volle Shift $X = \mathcal{A}^{\mathbb{Z}}$ kann beschrieben werden als $X = X_{\mathcal{F}}$ mit $\mathcal{F} = \emptyset$.
- (2) Sei X die Menge aller Folgen in $\{0, 1\}^{\mathbb{Z}}$, in denen keine zwei Einsen aufeinander folgen. Dann ist $X = X_{\mathcal{F}}$ für $\mathcal{F} = \{11\}$.
- (3) Sei X die Menge aller Folgen in $\{0, 1\}^{\mathbb{Z}}$, so dass die Anzahl der Nullen zwischen je zwei Einsen gerade ist. Dann ist $X = X_{\mathcal{F}}$ für

$$\mathcal{F} = \{10^{2n+1}1 : n \geq 0\}.$$

- (4) Sei $\mathcal{A} = \{e, f, g\}$ und X die Teilmenge aller Folgen in $\mathcal{A}^{\mathbb{Z}}$, in denen auf e nur e oder f folgt, auf f nur g und auf g nur e oder f . Dies ist ein Beispiel für einen Shift von endlichem Typ. Wir können ihn durch die Menge

$$\mathcal{F} = \{eg, fe, ff, gg\}.$$

beschreiben.

Jeder Shift ist invariant unter der Shift-Abbildung und unter ihrer Inversen, d.h. $\sigma(X_{\mathcal{F}}) \subseteq X_{\mathcal{F}}$ und $\sigma^{-1}(X_{\mathcal{F}}) \subseteq X_{\mathcal{F}}$. Daraus folgt

$$\sigma(X_{\mathcal{F}}) = X_{\mathcal{F}}.$$

Die Einschränkung der Shift-Abbildung auf $X = X_{\mathcal{F}}$ bezeichnen wir mit σ_X . Nicht jede Teilmenge von $\mathcal{A}^{\mathbb{Z}}$ ist invariant unter σ und nicht jede invariante Teilmenge ist ein Shift.

2.6 Beispiel: Sei $X \subseteq \{0, 1\}^{\mathbb{Z}}$ die Menge aller Punkte, die genau eine Eins enthalten und sonst nur Nullen. Offensichtlich ist X shift-invariant. Wäre X ein Shiftraum, so dürften keine Blöcke verboten sein, die nur aus Nullen bestehen. Dann wäre aber auch $0^{\infty} = \dots 000.000 \dots$ ein Element von X .

2.3 Sprachen

Manchmal ist es einfacher einen Shiftraum durch Angabe der Blöcke zu beschreiben, die erlaubt sind. Dies führt zum Begriff der *Sprache* eines Shifts.

2.7 Definition: Sei X eine Teilmenge eines vollen Shifts, und sei $\mathcal{B}_n(X)$ für jedes $n \in \mathbb{N}_0 = \{0, 1, 2, \dots\}$ die Menge aller n -Blöcke, die in Punkten von X vorkommen. Die Sprache von X ist die Menge

$$\mathcal{B}(X) := \bigcup_{n=0}^{\infty} \mathcal{B}_n(X).$$

2.8 Beispiel: Der volle 2-Shift hat die Sprache

$$\{\varepsilon, 0, 1, 00, 01, 10, 11, 000, 001, 010, 011, 100, \dots\}.$$

Nicht jede Menge von Blöcken ist die Sprache eines Shiftraums. Der folgende Satz charakterisiert diejenigen, die es sind.

2.9 Satz: Es gelten folgende Aussagen:

- (1) Sei X ein Shiftraum und $\mathcal{L} = \mathcal{B}(X)$ seine Sprache. Ist $w \in \mathcal{L}$, dann gilt:
 - (a) jeder Teilblock von w gehört zu \mathcal{L} und
 - (b) es gibt nichtleere Blöcke u und v in \mathcal{L} , so dass $uvw \in \mathcal{L}$.
- (2) Die Sprachen von Shifträumen sind durch (1) charakterisiert. Das heißt, ist \mathcal{L} eine Menge von Blöcken, dann ist $\mathcal{L} = \mathcal{B}(X)$ für einen Shift X genau dann, wenn \mathcal{L} (1) erfüllt.
- (3) Die Sprache eines Shiftraums bestimmt den Shiftraum vollständig. Genauer gilt $X = X_{\mathcal{B}(X)^c}$.⁴ Zwei Shifträume sind also genau dann identisch, wenn sie dieselbe Sprache haben.

Beweis: (1) Das ist offensichtlich.

(2) Sei \mathcal{L} eine Menge von Blöcken, die (1) erfüllt und $X = X_{\mathcal{L}^c}$. Wir beweisen, dass $\mathcal{L} = \mathcal{B}(X)$. Gilt $w \in \mathcal{B}(X)$, so kommt w in einem $x \in X_{\mathcal{L}^c}$ vor, so dass $w \notin \mathcal{L}^c$ bzw. $w \in \mathcal{L}$. Daher gilt $\mathcal{B}(X) \subseteq \mathcal{L}$. Sei nun umgekehrt $w = x_0x_1 \dots x_m \in \mathcal{L}$. Durch iterierte Anwendung von (1b) finden wir Symbole x_j mit $j > m$ und x_i mit $i < 0$, so dass nach (1a) jeder Teilblock von $x = (x_i)_{i \in \mathbb{Z}}$ in \mathcal{L} liegt. Dann gilt $x \in X_{\mathcal{L}^c}$. Da w in x vorkommt, folgt $w \in \mathcal{B}(X)$, also $\mathcal{L} \subseteq \mathcal{B}(X)$.

(3) Ist $x \in X$, dann ist kein Block, der in x vorkommt, enthalten in $\mathcal{B}(X)^c$, also gilt $X \subseteq X_{\mathcal{B}(X)^c}$. Umgekehrt gilt: Da X ein Shiftraum ist, ist $X = X_{\mathcal{F}}$ für eine Menge \mathcal{F} von Blöcken. Ist $x \in X_{\mathcal{B}(X)^c}$, dann liegt jeder Block in x in $\mathcal{B}(X) =$

⁴Wir bezeichnen mit A^c das Komplement einer Teilmenge A in einer gegebenen Menge X , d.h. $A^c = X \setminus A$.

$\mathcal{B}(X_{\mathcal{F}})$, kann also nicht in \mathcal{F} liegen. Daher gilt $x \in X_{\mathcal{F}}$, also $X_{\mathcal{B}(X)^c} \subseteq X_{\mathcal{F}} = X$. \square

Der obige Satz zeigt, dass es stets eine größte Menge verbotener Blöcke gibt, die einen Shiftraum beschreibt, nämlich das Komplement der Sprache des Shifts. Wir erhalten folgendes Korollar.

2.10 Korollar: *Eine Teilmenge X eines vollen Shifts ist genau dann ein Shift-
raum, wenn aus $x \in \mathcal{A}^{\mathbb{Z}}$ und $x_{[i,j]} \in \mathcal{B}(X)$ für alle i, j folgt, dass $x \in X$.*

Beweis: Sei X ein Shiftraum und $x \in \mathcal{A}^{\mathbb{Z}}$ mit $x_{[i,j]} \in \mathcal{B}(X)$ für alle i, j . Dann gilt $w \notin \mathcal{B}(X)^c$ für alle Blöcke w , die in x vorkommen, also $x \in X_{\mathcal{B}(X)^c} = X$ (nach (3)). Umgekehrt sei $X \subseteq \mathcal{A}^{\mathbb{Z}}$ eine Teilmenge mit der Eigenschaft, dass aus $x \in \mathcal{A}^{\mathbb{Z}}$ mit $x_{[i,j]} \in \mathcal{B}(X)$ für alle i, j folgt, dass $x \in X$. Wir zeigen, dass $\mathcal{B}(X)$ die Eigenschaften in (1) erfüllt. Ist $w \in \mathcal{B}(X)$, so kommt w in einem $x \in X$ vor. Also ist $w = x_{[i,j]}$ für gewisse i, j . Dann gilt aber $x_{[i',j']} \in \mathcal{B}(X)$ für alle i', j' . Folglich sind auch alle Teilblöcke von w in $\mathcal{B}(X)$ enthalten und es gilt (1a). Um (1b) zu zeigen, sei wieder $w = x_{[i,j]}$ für ein $x \in X$. Dann kommt $x_{[i-1,j+1]} = x_{i-1}wx_{j+1}$ in x vor, also gilt $x_{[i-1,j+1]} \in \mathcal{B}(X)$. Damit ist (1b) gezeigt. \square

Der erste Teil von Satz 2.9 zeigt, dass jeder Block nach beiden Seiten hin fortgesetzt werden kann. Umgekehrt ist es aber nicht stets so, dass zu je zwei Blöcken u, v ein Block w existiert, so dass $uwv \in \mathcal{B}(X)$. Shifts mit dieser Eigenschaft spielen eine wichtige und besondere Rolle.

2.11 Definition: *Ein Shiftraum X heißt irreduzibel, falls zu je zwei $u, v \in \mathcal{B}(X)$ ein $w \in \mathcal{B}(X)$ existiert mit $uwv \in \mathcal{B}(X)$.*

Shifts, die nicht irreduzibel sind, können manchmal in irreduzible Teile zerlegt werden. Darauf werden wir in Unterabschnitt 5.4 genauer eingehen.

2.4 Höhere Blockshifts und höhere Potenzshifts

Eine der grundlegenden Konstruktionen in der Theorie der symbolischen dynamischen Systeme besteht im Übergang von einem einzelnen Symbol zu einem Block von Symbolen und der Interpretation solcher Blöcke als Symbole eines neuen Alphabets.

Sei X ein Shiftraum über dem Alphabet \mathcal{A} und $\mathcal{A}_X^{[N]} = \mathcal{B}_N(X)$ die Menge aller N -Blöcke in X für ein fest gewähltes N . Wir betrachten $\mathcal{A}_X^{[N]}$ als neues Alphabet und können den vollen Shift darüber bilden. Der N -te höhere Blockcode ist definiert als

$$\beta_N : X \rightarrow (\mathcal{A}_X^{[N]})^{\mathbb{Z}}, \quad (\beta_N(x))_{[i]} := x_{[i, i+N-1]}.$$

Folglich ersetzt β_N die i -te Koordinate von x durch den Block der Länge N , der bei x_i beginnt. Dies lässt sich besser veranschaulichen, wenn wir uns die

Symbole in $\mathcal{A}_X^{[N]}$ als vertikal geschrieben vorstellen:

$$\beta_4(x) = \dots \begin{bmatrix} x_0 \\ x_{-1} \\ x_{-2} \\ x_{-3} \end{bmatrix} \begin{bmatrix} x_1 \\ x_0 \\ x_{-1} \\ x_{-2} \end{bmatrix} \begin{bmatrix} x_2 \\ x_1 \\ x_0 \\ x_{-1} \end{bmatrix} \cdot \begin{bmatrix} x_3 \\ x_2 \\ x_1 \\ x_0 \end{bmatrix} \begin{bmatrix} x_4 \\ x_3 \\ x_2 \\ x_1 \end{bmatrix} \begin{bmatrix} x_5 \\ x_4 \\ x_3 \\ x_2 \end{bmatrix} \dots \in (\mathcal{A}_X^{[4]})^{\mathbb{Z}}. \quad (2)$$

2.12 Definition: Sei X ein Shiftraum. Der N -te höhere Blockshift $X^{[N]}$ ist das Bild $X^{[N]} := \beta_N(X)$ im vollen Shift über $\mathcal{A}_X^{[N]}$.

Wir beobachten, dass in (2) aufeinanderfolgende Symbole aus $\mathcal{A}_X^{[N]}$ sich überschneiden. Falls $u = u_1u_2\dots u_N$ und $v = v_1v_2\dots v_N$ N -Blöcke sind, so sagen wir, dass u und v sich *progressiv überschneiden*, falls $u_2u_3\dots u_N = v_1v_2\dots v_{N-1}$.

2.13 Satz: Der N -te höhere Blockshift $X^{[N]}$ ist ein Shiftraum.

Beweis: Sei $X = X_{\mathcal{F}}$ für eine Menge \mathcal{F} von Blöcken über \mathcal{A} . Wir erzeugen eine neue Menge $\tilde{\mathcal{F}}$ von Blöcken, indem wir jeden Block $u \in \mathcal{F}$ mit $|u| < N$ durch alle N -Blöcke über \mathcal{A} ersetzen, die u enthalten. Dann gilt offensichtlich $X = X_{\tilde{\mathcal{F}}}$ und jeder Block in $\tilde{\mathcal{F}}$ hat mindestens die Länge N .

Für jedes $w = a_1a_2\dots a_m \in \tilde{\mathcal{F}}$ sei

$$w^{[N]} := (a_1a_2\dots a_N)(a_2a_3\dots a_{N+1})\dots(a_{m-N+1}a_{m-N+2}\dots a_m)$$

der entsprechende $(m-N+1)$ -Block über \mathcal{A}^N . Sei \mathcal{F}_1 die Menge aller Blöcke über \mathcal{A}^N der Form $w^{[N]}$ für ein $w \in \tilde{\mathcal{F}}$. Dies stellt eine Menge von Einschränkungen in $X^{[N]}$ dar, nämlich derjenigen, die von den Einschränkungen des ursprünglichen Shifts herkommen. Es folgt $X^{[N]} \subseteq X_{\mathcal{F}_1}$.

Punkte in $X^{[N]}$ erfüllen außerdem die Überschneidungsbedingung, die wir in (2) sehen. Folglich definieren wir

$$\mathcal{F}_2 := \{uv : u \in \mathcal{A}^N, v \in \mathcal{A}^N, u \text{ und } v \text{ überschneiden sich nicht progressiv}\}.$$

Dann folgt $X^{[N]} \subseteq X_{\mathcal{F}_2}$, so dass nach Übungsaufgabe (3d) auf Blatt 1 gilt

$$X^{[N]} \subseteq X_{\mathcal{F}_1} \cap X_{\mathcal{F}_2} = X_{\mathcal{F}_1 \cup \mathcal{F}_2}.$$

Nun nehmen wir umgekehrt an, dass $y \in X_{\mathcal{F}_1 \cup \mathcal{F}_2}$, und definieren x als diejenige Folge in $\mathcal{A}^{\mathbb{Z}}$, die aus den untersten Symbolen in der Darstellung (2) besteht. Dann gilt $x \in X = X_{\mathcal{F}}$, da y die Einschränkungen aus \mathcal{F}_1 erfüllt und $y = \beta_N(x)$ wegen der Überschneidungsbedingungen aus \mathcal{F}_2 . Also gilt $X^{[N]} \supseteq X_{\mathcal{F}_1 \cup \mathcal{F}_2}$ und $X^{[N]} = X_{\mathcal{F}_1 \cup \mathcal{F}_2}$ ist ein Shiftraum. \square

2.14 Beispiel: Sei X der Shift über $\{0, 1\}$ aus Beispiel 2.5(2), also $X = X_{\mathcal{F}}$, $\mathcal{F} = \{11\}$. Dann gilt

$$\mathcal{A}_X^{[2]} = \{a = 00, b = 01, c = 10\}.$$

Der 2-te höhere Blockshift $X^{[2]}$ kann beschrieben werden durch die Einschränkungen

$$\{ac, ba, bb, cc\}.$$

Alle diese Blöcke sind verboten, weil sie sich nicht progressiv überschneiden. Der Block 11 ist auch verboten, da er in dem ursprünglichen Shiftraum nicht vorkommt. Dies zeigt sich darin, dass er in $\mathcal{A}_X^{[2]}$ nicht auftaucht.

In der Konstruktion des N -ten höheren Blockshifts von X verwenden wir sich überschneidende Blöcke. Dieselbe Art von Konstruktion lässt sich auch mit sich nicht überschneidenden Blöcken durchführen. Dies führt zum Begriff des N -ten höheren Potenzshifts von X .

Unter Verwendung derselben Notation wie zu Beginn dieses Abschnitts definieren wir den N -ten höheren Potenzcode

$$\gamma_N : X \rightarrow (\mathcal{A}_X^{[N]})^{\mathbb{Z}}, \quad (\gamma_N(x))_{[i]} := x_{[iN, iN+N-1]}.$$

Hier zerschneidet γ_N die Koordinaten von x in aufeinanderfolgende N -Blöcke und fügt die Teile wieder zusammen zu einem Punkt in $(\mathcal{A}_X^{[N]})^{\mathbb{Z}}$. Das Bild von $x = (x_i)_{i \in \mathbb{Z}}$ hat die Form

$$\gamma_4(x) = \dots \begin{bmatrix} x_{-9} \\ x_{-10} \\ x_{-11} \\ x_{-12} \end{bmatrix} \begin{bmatrix} x_{-5} \\ x_{-6} \\ x_{-7} \\ x_{-8} \end{bmatrix} \begin{bmatrix} x_{-1} \\ x_{-2} \\ x_{-3} \\ x_{-4} \end{bmatrix} \cdot \begin{bmatrix} x_3 \\ x_2 \\ x_1 \\ x_0 \end{bmatrix} \begin{bmatrix} x_7 \\ x_6 \\ x_5 \\ x_4 \end{bmatrix} \begin{bmatrix} x_{11} \\ x_{10} \\ x_9 \\ x_8 \end{bmatrix} \dots \in (\mathcal{A}_X^{[4]})^{\mathbb{Z}}. \quad (3)$$

2.15 Definition: Sei X ein Shiftraum. Der N -te höhere Potenzshift X^N ist das Bild $X^N = \gamma_N(X)$ von X im vollen Shift über $\mathcal{A}_X^{[N]}$.

2.16 Satz: Die höheren Potenzshifts sind Shifträume.

Beweis: Übungsaufgabe. □

2.5 Gleitende Blockcodes

Sei $x = \dots x_{-1}x_0x_1\dots$ eine Folge von Symbolen in einem Shiftraum X über \mathcal{A} . Wir können x in eine neue Folge $y = \dots y_{-1}y_0y_1\dots$ über einem anderen Alphabet \mathcal{A}' wie folgt transformieren. Seien $m, n \in \mathbb{Z}$ mit $-m \leq n$. Um die i -te Koordinate y_i der transformierten Folge zu berechnen, verwenden wir eine Funktion Φ , die von dem „Fenster“ der Koordinaten von x zwischen $i - m$ und $i + n$ abhängt. Die Funktion $\Phi : \mathcal{B}_{m+n+1}(X) \rightarrow \mathcal{A}'$ ist eine fest gewählte Blockabbildung, genannt eine $(m + n + 1)$ -Blockabbildung von den erlaubten $(m + n + 1)$ -Blöcken in X in die Symbole von \mathcal{A}' , so dass

$$y_i = \Phi(x_{i-m}x_{i-m+1}\dots x_{i+n}) = \Phi(x_{[i-m, i+n]}). \quad (4)$$

2.17 Definition: Sei X ein Shiftraum über \mathcal{A} und $\Phi : \mathcal{B}_{m+n+1}(X) \rightarrow \mathcal{A}'$ eine Blockabbildung. Dann nennen wir die Abbildung $\phi : X \rightarrow (\mathcal{A}')^{\mathbb{Z}}$, definiert durch $y = \phi(x)$ mit y_i wie in (4) den gleitenden Blockcode mit Gedächtnis m und Antizipation n induziert durch Φ . Wir schreiben $\phi = \Phi_{\infty}^{[-m,n]}$, oder einfach $\phi = \Phi_{\infty}$, falls m und n aus dem Kontext heraus klar sind. Wenn über das Gedächtnis nichts gesagt wird, nehmen wir stets $m = 0$ an. Ist Y ein Shiftraum in $(\mathcal{A}')^{\mathbb{Z}}$ und $\phi(X) \subseteq Y$, so schreiben wir $\phi : X \rightarrow Y$.

2.18 Bemerkung: Unter einem *Code* verstehen wir in der Theorie der symbolischen dynamischen Systeme also eine Abbildung, die Symbolfolgen in Symbolfolgen überführt. Eine präzisere Bezeichnung wäre vielleicht *Kodierungsabbildung*.

Die Wirkungsweise eines gleitenden Blockcodes lässt sich folgendermaßen veranschaulichen:

$$\begin{array}{c} \dots x_{i-m-1} \boxed{x_{i-m} x_{i-m+1} \dots x_{i+n-1} x_{i+n}} x_{i+n+1} \dots \\ \downarrow \Phi \\ \dots y_{i-1} \boxed{y_i} y_{i+1} \dots \end{array}$$

Die einfachsten gleitenden Blockcodes sind diejenigen mit $m = n = 0$. In diesem Fall hängt y_i nur von x_i ab und wir sprechen von *1-Blockcodes*.

2.19 Beispiel: Sei X ein Shiftraum über einem Alphabet \mathcal{A} , $\mathcal{A}' = \mathcal{A}$, $m = 0$, $n = 1$ und $\Phi(a_0 a_1) = a_1$. Dann ist $\phi = \Phi_{\infty}^{[0,1]} = \Phi_{\infty}$ die Shiftabbildung σ_X . Ist $\Phi(a_0 a_1) = a_0$, so ist Φ_{∞} die Identität.

Jetzt sei $m = 1$, $n = 0$ and $\Psi(a_{-1} a_0) = a_{-1}$. Dann ist $\psi = \Psi_{\infty}^{[-1,0]} = \Psi_{\infty}$ die Umkehrabbildung der Shiftabbildung, also σ_X^{-1} .

Definieren wir $\Theta(a) = a$ für $a \in \mathcal{A}$, so gilt $\psi = \Theta_{\infty}^{[1,1]}$ und $\phi = \Theta_{\infty}^{[-1,-1]}$. Folglich kann es mehrere Möglichkeiten geben, eine Abbildung zwischen Shifträumen als gleitenden Blockcode darzustellen.

2.20 Beispiel: Sei $\mathcal{A} = \{0, 1\} = \mathcal{A}'$ und $\Phi(a_0 a_1) = a_0 + a_1 \pmod{2}$. Dies ist der Code aus Aufgabe 2 auf dem Präsenzübungsblatt.

2.21 Beispiel: Sei X ein Shiftraum über \mathcal{A} und $\mathcal{A}' = \mathcal{A}_X^{[N]}$, $m = 0$, $n = N - 1$ und $Y = X^{[N]}$. Ferner sei

$$\Phi(a_0 a_1 \dots a_{N-1}) = a_0 a_1 \dots a_{N-1} \in \mathcal{A}_X^{[N]}.$$

Dann ist Φ_{∞} der N -te höhere Blockcode β_N aus dem vorherigen Abschnitt.

Sei $\Phi : \mathcal{B}_{m+n+1}(X) \rightarrow \mathcal{A}'$ eine Blockabbildung, die einen gleitenden Blockcode mit Gedächtnis m und Antizipation n erzeugt. Dann macht es manchmal Sinn anzunehmen, dass Φ ein größeres „Fenster“ hat und die zusätzlichen

Koordinaten zu ignorieren. Ist also $M \geq m$ und $N \geq n$, so definieren wir $\hat{\Phi} : \mathcal{B}_{M+N+1}(X) \rightarrow \mathcal{A}'$ als

$$\hat{\Phi}(x_{[-M,N]}) := \Phi(x_{[-m,n]}).$$

Offensichtlich gilt $\hat{\Phi}_\infty = \Phi_\infty$.

Wir können Φ erweitern, so dass $(m+n+k)$ -Blöcke auf k -Blöcke abgebildet werden, indem wir das Fenster von Φ wie folgt vergrößern. Für $x_{[-m,n+k-1]} \in \mathcal{B}_{m+n+k}(X)$ definieren wir

$$\Phi(x_{[-m,n+k-1]}) := \Phi(x_{[-m,n]})\Phi(x_{[-m+1,n+1]}) \cdots \Phi(x_{[-m+k-1,n+k-1]}). \quad (5)$$

2.22 Beispiel: Sei $\mathcal{A} = \mathcal{A}' = \{0, 1\}$ und $X = X_{\mathcal{F}}$, $\mathcal{F} = \{11\}$ (siehe Beispiel 2.5(2)). Sei ferner Y der in Beispiel 2.5(3) beschriebene Shift. Sei Φ die 2-Blockabbildung, definiert durch $\Phi(00) := 1$, $\Phi(01) := 0$ und $\Phi(10) := 0$. Wir zeigen, dass der induzierte gleitende Blockcode $\phi = \Phi_\infty : X \rightarrow Y$ wohldefiniert und surjektiv ist. (Hier ist $m = 0$ und $n = 1$.)

Falls der Block 10^k1 in $\phi(x)$ vorkommt, muss er das Bild unter Φ des Blocks $0(01)^r00$ sein, so dass $k = 2r$ gerade ist. Daher gilt $\phi(X) \subseteq Y$. Da in jedem $y \in Y$ die Einsen durch eine gerade Anzahl von Nullen getrennt sind, zeigt dieselbe Überlegung, wie man ein $x \in X$ konstruiert, so dass $\phi(x) = y$. Also ist ϕ surjektiv.

2.23 Satz: Seien X und Y Shifträume. Ist $\phi : X \rightarrow Y$ ein gleitender Blockcode, so gilt $\phi \circ \sigma_X = \sigma_Y \circ \phi$, d.h. folgendes Diagramm kommutiert:

$$\begin{array}{ccc} X & \xrightarrow{\sigma_X} & X \\ \phi \downarrow & & \downarrow \phi \\ Y & \xrightarrow{\sigma_Y} & Y \end{array}$$

Beweis: Sei ϕ induziert von der Blockabbildung $\Phi : \mathcal{B}_{m+n+1}(X) \rightarrow \mathcal{A}'$. Dann gilt für alle $i \in \mathbb{Z}$, dass

$$\phi(\sigma_X(x))_{[i]} = \Phi(x_{[i+1-m, i+1+n]})$$

und

$$\sigma_Y(\phi(x))_{[i]} = \sigma_Y((\Phi(x_{[i-m, i+n]})_{i \in \mathbb{Z}})_{[i]}) = \Phi(x_{[i+1-m, i+1+n]}).$$

Damit ist der Satz bewiesen. \square

Jedoch ist nicht jede Abbildung, die mit dem Shift kommutiert, ein gleitender Blockcode (siehe Übungsblatt 2).

2.24 Satz: Seien X und Y Shifträume (über Alphabeten \mathcal{A} und \mathcal{A}'). Eine Abbildung $\phi : X \rightarrow Y$ ist ein gleitender Blockcode genau dann, wenn $\phi \circ \sigma_X = \sigma_Y \circ \phi$ und ein $N \geq 0$ existiert, so dass $\phi(x)_{[0]}$ eine Funktion von $x_{[-N, N]}$ ist.

Beweis: Die Beweisrichtung „ \Rightarrow “ ist klar nach Definition von gleitenden Blockcodes und Satz 2.23. Umgekehrt sei nun $\phi : X \rightarrow Y$ eine Abbildung mit $\phi \circ \sigma_X = \sigma_Y \circ \phi$, so dass $\phi(x)_{[0]}$ eine Funktion von $x_{[-N, N]}$ ist. Dann definieren wir eine $(2N + 1)$ -Blockabbildung

$$\Phi : \mathcal{B}_{2N+1}(X) \rightarrow \mathcal{A}', \quad \Phi(w) := \phi(x)_{[0]},$$

wobei $x \in X$ so gewählt ist, dass $x_{[-N, N]} = w$. Das ist möglich, da w in einem $x \in X$ vorkommt und X shift-invariant ist. Dann ist $\phi = \Phi_{\infty}^{[-N, N]}$, denn für jedes $i \in \mathbb{Z}$ gilt

$$\begin{aligned} \Phi_{\infty}^{[-N, N]}(x)_{[i]} &\stackrel{\text{def}}{=} \Phi(x_{[i-N, i+N]}) \\ &\stackrel{\text{def}}{=} \phi(\sigma_X^i(x))_{[0]} = \sigma_Y^i(\phi(x))_{[0]} = \phi(x)_{[i]}. \end{aligned}$$

Damit ist der Beweis abgeschlossen. \square

2.25 Definition: Einen surjektiven gleitenden Blockcode $\phi : X \rightarrow Y$ nennen wir auch Faktorcode. Existiert ein Faktorcode von X nach Y , so heißt Y Faktor von X . Ist $\phi : X \rightarrow Y$ ein injektiver gleitender Blockcode, so heißt ϕ eine Einbettung. Ist $\phi : X \rightarrow Y$ ein gleitender Blockcode, so dass ein gleitender Blockcode $\psi : Y \rightarrow X$ existiert mit $\phi(\psi(y)) = y$ für alle $y \in Y$ und $\psi(\phi(x)) = x$ für alle $x \in X$, so heißt ϕ invertierbar. In diesem Fall nennen wir ϕ auch eine Konjugation. Falls eine Konjugation $\phi : X \rightarrow Y$ existiert, so sagen wir, X und Y sind konjugiert, in Zeichen $X \cong Y$.

Sind zwei Shifträume konjugiert, so können wir uns den einen als eine lediglich „rekodierte“ Version des anderen vorstellen. Im Wesentlichen sind die Shifträume dann gleich.

2.26 Beispiel: Sei X ein Shiftraum über \mathcal{A} und $X^{[N]}$ sein N -ter höherer Blockshift. Nach Beispiel 2.21 ist $\beta_N : X \rightarrow X^{[N]}$ ein gleitender Blockcode. Definieren wir die 1-Blockabbildung $\Psi : \mathcal{A}_X^{[N]} \rightarrow \mathcal{A}$ durch $\Psi(a_0 a_1 \dots a_{N-1}) = a_0$, so sehen wir leicht, dass $\psi = \Psi_{\infty} : X^{[N]} \rightarrow X$ die Inverse von β_N ist. Also ist β_N eine Konjugation.

Der nächste Satz beschreibt das Verhalten von periodischen Punkten unter gleitenden Blockcodes.

2.27 Satz: Sei $\phi : X \rightarrow Y$ ein gleitender Blockcode. Ist $x \in X$ ein periodischer Punkt von σ_X mit Periode n , so ist $y = \phi(x) \in Y$ ein periodischer Punkt von σ_Y mit Periode n , und die Primperiode von y ist ein Teiler der Primperiode von x . Ist ϕ eine Konjugation, so stimmen die Primperioden überein.

Beweis: Gilt $\sigma_X^n(x) = x$, so folgt $\sigma_Y^n(\phi(x)) = \phi(\sigma_X^n(x)) = \phi(x)$. Also hat $\phi(x)$ auch Periode n . Ist hier n als Primperiode von x gewählt, so muss n

ein ganzzahliges Vielfaches der Primperiode von y sein. Um dies zu zeigen, bezeichnen wir mit m die Primperiode von y . Dann gilt $1 \leq m \leq n$ und wir können $n = km + r$ mit $k \in \mathbb{N}_0$ und $r \in \{0, 1, \dots, m-1\}$ schreiben. Es folgt

$$y = \sigma_Y^n(y) = \sigma_Y^r((\sigma_Y^m)^k(y)) = \sigma_Y^r(y).$$

Ist $r > 0$, so muss r damit auch eine Periode von y sein. Das kann nicht sein, da $r < m$. Also gilt $n = km$. Ist ϕ invertierbar, so sind die Primperioden jeweils ganzzahlige Vielfache voneinander. Also sind sie identisch. \square

Der obige Satz liefert zahlreiche numerische Konjugations-Invarianten, die uns dabei helfen können zu erkennen, dass zwei Shifträume *nicht* konjugiert zueinander sind. Diese Invarianten sind zum einen die Anzahlen der n -periodischen Punkte für jedes n , aber auch die Anzahlen der Punkte mit Primperiode n .

Der folgende Satz erlaubt es, einen beliebigen gleitenden Blockcode zu einem 1-Blockcode zu „rekodieren“. Das Alphabet wird dadurch allerdings komplizierter.

2.28 Satz: Sei $\phi : X \rightarrow Y$ ein gleitender Blockcode. Dann existiert ein höherer Blockshift \tilde{X} von X , eine Konjugation $\psi : X \rightarrow \tilde{X}$ und ein 1-Blockcode $\tilde{\phi} : \tilde{X} \rightarrow Y$, so dass $\tilde{\phi} \circ \psi = \phi$, d.h. folgendes Diagramm kommutiert:

$$\begin{array}{ccc} X & \xrightarrow{\psi} & \tilde{X} \\ \phi \downarrow & \cong & \swarrow \tilde{\phi} \\ Y & & \end{array}$$

Beweis: Sei ϕ induziert durch eine Blockabbildung Φ mit Gedächtnis m und Antizipation n . Sei $\mathcal{A}' := \mathcal{B}_{m+n+1}(X)$. Definiere $\psi : X \rightarrow (\mathcal{A}')^{\mathbb{Z}}$ durch

$$\psi(x)_{[i]} := x_{[i-m, i+n]} \quad \text{für alle } i \in \mathbb{Z}.$$

Dann gilt $\psi = \sigma^{-m} \circ \beta_{m+n+1}$. Daraus folgt, dass $\tilde{X} := \psi(X) = X^{[m+n+1]}$ ein Shiftraum ist. Da σ und β_{m+n+1} Konjugationen sind, ist es auch ψ . Setze nun $\tilde{\phi} := \phi \circ \psi^{-1}$. Die Abbildung $\tilde{\phi}$ ist ein 1-Blockcode. Ist nämlich $\tilde{x} \in \tilde{X}$, so ist $\tilde{x} = \beta_{m+n+1}(x)$ für ein $x \in X$. Dann folgt

$$\begin{aligned} \phi(\psi^{-1}(\tilde{x}))_{[i]} &= \phi(\beta_{m+n+1}^{-1}(\sigma^m(\beta_{m+n+1}(x))))_{[i]} \\ &= \phi(\beta_{m+n+1}^{-1}(\beta_{m+n+1}(\sigma^m(x))))_{[i]} \\ &= \phi(\sigma^m(x))_{[i]} \\ &= \phi((x_{m+j})_{j \in \mathbb{Z}})_{[i]} \\ &= \Phi(x_{[i, i+m+n]}) = \Phi(\tilde{x}_{[i]}). \end{aligned}$$

Damit ist der Beweis abgeschlossen. \square

Wir zeigen als nächstes, dass das Bild eines gleitenden Blockcodes stets ein Shiftraum ist.

2.29 Satz: *Das Bild eines gleitenden Blockcodes ist ein Shiftraum.*

Beweis: Seien X und Y Shifträume und $\phi : X \rightarrow Y$ ein gleitender Blockcode. Nach Satz 2.28 dürfen wir annehmen, dass ϕ ein 1-Blockcode ist. Sei Φ eine 1-Blockabbildung, die ϕ induziert. Wir betrachten $\mathcal{L} := \{\Phi(w) : w \in \mathcal{B}(X)\}$ und zeigen, dass $\phi(X) = X_{\mathcal{L}^c}$, woraus folgt, dass $\phi(X)$ ein Shiftraum ist.

Die Inklusion $\phi(X) \subseteq X_{\mathcal{L}^c}$ ist leicht zu zeigen. Sei dazu $x \in X$. Dann ist jeder Block, der in $\phi(x)$ vorkommt, in \mathcal{L} enthalten. Also gilt $\phi(x) \in X_{\mathcal{L}^c}$.

Nun sei umgekehrt $y \in X_{\mathcal{L}^c}$. Für jedes $n \geq 0$ ist der zentrale $(2n+1)$ -Block von y das Bild unter Φ des zentralen $(2n+1)$ -Blocks eines Punktes $x^{(n)} \in X$, d.h.⁵

$$\Phi(x_{[-n,n]}^{(n)}) = \phi(x^{(n)})_{[-n,n]} = y_{[-n,n]}.$$

Wir werden $x^{(n)}$ verwenden, um einen Punkt x zu konstruieren mit $y = \phi(x)$.

Zunächst stellen wir fest, dass es aufgrund der Endlichkeit des Alphabets eine unendliche Menge $S_0 \subseteq \mathbb{N}_0$, so dass $x_{[0]}^{(n)}$ identisch für alle $n \in S_0$ ist. Dann gibt es eine unendliche Menge $S_1 \subseteq S_0$, so dass für alle $n \in S_1$ die 3-Blöcke $x_{[-1,1]}^{(n)}$ identisch sind. Führen wir diese Konstruktion fort, so erhalten wir Mengen $S_k \subseteq S_{k-1}$, so dass für alle $n \in S_k$ die Blöcke $x_{[-k,k]}^{(n)}$ identisch sind.

Nun definieren wir x als diejenige Folge mit $x_{[-k,k]} = x_{[-k,k]}^{(n)}$ für alle $n \in S_k$. Da $S_{k-1} \subseteq S_k$, ist dieses x wohldefiniert. Jeder Block in x kommt in einem der Blöcke $x_{[-k,k]}^{(n)}$ vor, der nach Konstruktion in $\mathcal{B}(X)$ liegt. Also ist $x \in X$. Nun ist leicht zu sehen, dass $\phi(x) = y$. \square

Der nächste Satz vereinfacht den Nachweis, dass eine gegebene Abbildung eine Konjugation ist.

2.30 Satz: *Sei $\phi : X \rightarrow Y$ ein bijektiver gleitender Blockcode. Dann ist die zu ϕ inverse Abbildung ein gleitender Blockcode und ϕ damit eine Konjugation.*

Beweis: Durch eine Rekodierung wie in Satz 2.28 dürfen wir annehmen, dass ϕ ein 1-Blockcode ist. Sei Φ eine 1-Blockabbildung, die ϕ induziert und sei ψ die zu ϕ inverse Abbildung, von der wir zeigen müssen, dass sie ein gleitender Blockcode ist.

Ist $y = \phi(x)$, so stellen wir fest, dass

$$\sigma_X(\psi(y)) = \sigma_X(x) = \psi(\phi(\sigma_X(x))) = \psi(\sigma_Y(\phi(x))) = \psi(\sigma_Y(y)).$$

Also gilt $\psi \circ \sigma_Y = \sigma_X \circ \psi$. Nach Satz 2.24 reicht es noch zu zeigen, dass es ein $n \geq 0$ gibt, so dass $\psi(y)_{[0]}$ nur von $y_{[-n,n]}$ abhängt. Wäre das nicht so, so gäbe es für jedes n zwei Punkte $y^{(n)}$ und $\tilde{y}^{(n)}$ mit $y_{[-n,n]}^{(n)} = \tilde{y}_{[-n,n]}^{(n)}$ und $\psi(y^{(n)})_{[0]} \neq \psi(\tilde{y}^{(n)})_{[0]}$. Sei $x^{(n)} := \psi(y^{(n)})$ und $\tilde{x}^{(n)} := \psi(\tilde{y}^{(n)})$.

⁵Hier verwenden wir die in (5) definierte Erweiterung von Φ .

Da das Alphabet endlich ist, gibt es eine unendliche Teilmenge $S_0 \subseteq \mathbb{N}_0$ und $a, b \in \mathcal{A}$, so dass $a = x_{[0]}^{(n)} \neq \tilde{x}_{[0]}^{(n)} = b$ für alle $n \in S_0$. Da auch die Anzahl der 3-Blöcke endlich ist, gibt es eine unendliche Teilmenge $S_1 \subseteq S_0$, so dass die Blöcke $x_{[-1,1]}^{(n)}$ für alle $n \in S_1$ gleich sind so wie auch die Blöcke $\tilde{x}_{[-1,1]}^{(n)}$. Ähnlich wie im Beweis des vorherigen Satzes erhalten wir, wenn wir diese Konstruktion fortführen, am Ende Punkte x, \tilde{x} mit $x_{[0]} \neq \tilde{x}_{[0]}$, wobei $x_{[-k,k]} = x_{[-k,k]}^{(n)}$ und $\tilde{x}_{[-k,k]} = \tilde{x}_{[-k,k]}^{(n)}$ für alle $n \in S_k$ und $n \geq k$, so folgt

$$\begin{aligned} \Phi(x_{[-k,k]}) &= \Phi(x_{[-k,k]}^{(n)}) = \phi(x^{(n)})_{[-k,k]} = y_{[-k,k]}^{(n)} \\ &= \tilde{y}_{[-k,k]}^{(n)} = \phi(\tilde{x}^{(n)})_{[-k,k]} = \Phi(\tilde{x}_{[-k,k]}^{(n)}) = \Phi(\tilde{x}_{[-k,k]}). \end{aligned}$$

Daraus folgt $\phi(x) = \phi(\tilde{x})$, also $x = \tilde{x}$, da ϕ injektiv ist. Dies steht im Widerspruch zu $x_{[0]} \neq \tilde{x}_{[0]}$. Damit ist der Satz bewiesen. \square

2.31 Bemerkung: Man kann auf Shifträumen eine Metrik einführen, so dass zwei Folgen nahe beieinander liegen, falls sie in hinreichend langen zentralen Blöcken übereinstimmen. In dieser Metrik sind Shifträume genau die kompakten shift-invarianten Teilräume eines vollen Shifts und die gleitenden Blockcodes sind genau die stetigen Abbildungen zwischen Shifträumen, die mit der Shiftabbildung kommutieren. Damit ergeben sich die Aussagen der letzten zwei Sätze sehr leicht mit Hilfe topologischer Argumente. Satz 2.29 erhält man durch die Tatsache, dass stetige Abbildungen kompakte Mengen auf kompakte Mengen abbilden. Satz 2.30 ergibt sich daraus, dass eine stetige bijektive Abbildung zwischen kompakten metrischen Räumen automatisch ein Homöomorphismus ist, d.h. eine stetige Umkehrabbildung hat.

3 Shifts von endlichem Typ

Eine Klasse von Shifts, die sich sehr gut verstehen lassen, bilden die Shifts, die durch eine endliche Menge von verbotenen Blöcken beschrieben werden können.

3.1 Einschränkungen von endlichem Typ

3.1 Definition: Ein Shiftraum X über \mathcal{A} heißt Shift von endlichem Typ, falls es eine endliche Menge \mathcal{F} von Blöcken über \mathcal{A} gibt, so dass $X = X_{\mathcal{F}}$.

Die Definition sagt lediglich, dass es eine endliche Menge \mathcal{F} mit $X = X_{\mathcal{F}}$ geben muss. Das bedeutet nicht, dass jede Menge \mathcal{F} mit dieser Eigenschaft endlich ist. Wir können jedoch ohne Einschränkung der Allgemeinheit annehmen, dass es eine endliche Menge \mathcal{F} gibt, in der alle Blöcke dieselbe Länge haben. Wenn \mathcal{F} diese Eigenschaft nicht hat, setzen wir $N = \max_{u \in \mathcal{F}} |u|$ und ersetzen jedes $u \in \mathcal{F}$ mit $|u| < N$ durch alle N -Blöcke über \mathcal{A} , die u enthalten.

3.2 Beispiel: Der Shift aus Beispiel 2.5(3) ist nicht von endlichem Typ. Wenn es so wäre, gäbe es ein N , so dass $X = X_{\mathcal{F}}$ für eine Menge \mathcal{F} von N -Blöcken. Dann betrachten wir den Punkt $x = 0^\infty 10^{2N+1} 10^\infty$. Keiner der Blöcke in x ist verboten, also müsste $x \in X$ gelten, was aber nicht der Fall ist.

3.3 Definition: Ein Shift von endlichem Typ ist ein M -Schritt-Shift (bzw. hat Gedächtnis M) für ein $M \in \mathbb{N}_0$, falls er sich durch eine Menge von verbotenen $(M+1)$ -Blöcken beschreiben lässt.

In 0-Schritt-Shifts sind lediglich einzelne Symbole verboten, so dass es sich um volle Shifts handelt. In 1-Schritt-Shifts sind gewisse Symbolübergänge verboten. Solche Shifts lassen sich durch 0-1-Matrizen beschreiben.

Der nächste Satz ist offensichtlich.

3.4 Satz: Sei X ein Shift von endlichem Typ. Dann existiert ein $M \geq 0$, so dass X ein M -Schritt-Shift ist.

Der folgende Satz liefert eine Charakterisierung von M -Schritt-Shifts.

3.5 Satz: Ein Shift X ist ein M -Schritt-Shift genau dann, wenn aus $uv, vw \in \mathcal{B}(X)$ mit $|v| \geq M$ folgt, dass $uvw \in \mathcal{B}(X)$.

Beweis: Wir nehmen zunächst an, dass X ein M -Schritt-Shift ist, also $X = X_{\mathcal{F}}$ und \mathcal{F} ist eine endliche Menge von $(M+1)$ -Blöcken. Es gelte $uv, vw \in \mathcal{B}(X)$ mit $|v| = n \geq M$. Dann gibt es Punkte $x, y \in X$ mit $x_{[-k, n]} = uv$ und $y_{[1, l]} = vw$, so dass $x_{[1, n]} = y_{[1, n]} = v$. Wir zeigen, dass der Punkt $z = x_{(-\infty, 0]} v y_{[n+1, \infty)}$ in X liegt. Falls ein Block aus \mathcal{F} in z vorkommt, muss er entweder in $x_{(-\infty, 0]} v = x_{(-\infty, n]}$ oder in $v y_{[n+1, \infty)} = y_{[1, \infty)}$ vorkommen, da $|v| = n \geq M$. Dies widerspricht entweder $x \in X$ oder $y \in X$. Daher gilt

$$uvw = x_{[-k, 0]} v y_{[n+1, l]} = z_{[-k, l]} \in \mathcal{B}(X).$$

Umgekehrt nehmen wir nun an, dass X ein Shiftraum über \mathcal{A} ist, so dass aus $uv, vw \in \mathcal{B}(X)$ mit $|v| \geq M$ folgt, dass $uvw \in \mathcal{B}(X)$. Sei \mathcal{F} die Menge aller $(M+1)$ -Blöcke über \mathcal{A} , die nicht in $\mathcal{B}_{M+1}(X)$ enthalten sind. Wir zeigen, dass $X = X_{\mathcal{F}}$.

Ist $x \in X$, dann kann kein Block aus \mathcal{F} in x vorkommen, also gilt $x \in X_{\mathcal{F}}$, und damit $X \subseteq X_{\mathcal{F}}$. Jetzt sei $x \in X_{\mathcal{F}}$. Dann sind $x_{[0, M]}$ und $x_{[1, M+1]}$ in $\mathcal{B}(X)$ nach Definition von \mathcal{F} . Da sie sich in M Symbolen überschneiden, ist auch $x_{[0, M+1]}$ in $\mathcal{B}(X)$. (Verwende $u = x_{[0]}$, $v = x_{[1, M]}$ und $w = x_{[M+1]}$.) Auch $x_{[2, M+2]}$ ist in $\mathcal{B}(X)$ und überschneidet sich mit $x_{[0, M+1]}$ in M Symbolen. Also gilt $x_{[0, M+2]} \in \mathcal{B}(X)$. Die iterierte Anwendung dieses Arguments in beide Richtungen ergibt $x_{[-k, l]} \in \mathcal{B}(X)$ für alle $k, l \in \mathbb{Z}$. Nach Korollar 2.10 gilt $x \in X$, womit $X_{\mathcal{F}} \subseteq X$ bewiesen ist. \square

Ein Faktor eines Shifts von endlichem Typ ist nicht zwangsläufig auch von endlichem Typ, wie folgendes Beispiel zeigt.

3.6 Beispiel: Sei $\mathcal{A} = \mathcal{A}' = \{0, 1\}$. Sei $X = X_{\mathcal{F}}$ mit $\mathcal{F} = \{11\}$ und $Y = X_{\mathcal{G}}$ mit $\mathcal{G} = \{10^{2n+1} : n \geq 0\}$. Sei $\Phi : \mathcal{B}_2(X) \rightarrow \mathcal{A}'$ die 2-Blockabbildung, definiert durch $\Phi(00) := 1$, $\Phi(01) := 0$ und $\Phi(10) := 0$. Dann ist $\phi = \Phi_{\infty} : X \rightarrow Y$ ein Faktorcode. Kommt der Block $10^k 1$ in $\phi(x)$ vor, so muss er das Bild von $0(01)^r 00$ unter Φ sein, so dass $k = 2r$ gerade ist. Also gilt $\Phi(X) \subseteq Y$. Da in jedem $y \in Y$ die Einsen durch eine gerade Anzahl von Nullen separiert sind, zeigt dieselbe Überlegung, wie man ein Urbild von y konstruiert. Nach Beispiel 3.2 ist Y nicht von endlichem Typ.

Auch das Urbild eines Shifts von endlichem Typ unter einem gleitenden Blockcode muss kein Shift von endlichem Typ sein. Zum Beispiel ist $Y = \{0^{\infty}\}$ ein Faktor jedes Shifts. Deshalb ist der folgende Satz subtiler als man vielleicht auf den ersten Blick denkt.

3.7 Satz: *Ein Shiftraum, der konjugiert zu einem Shift von endlichem Typ ist, ist selbst von endlichem Typ.*

Beweis: Sei X ein Shiftraum, der konjugiert zu einem Shift Y von endlichem Typ ist. Sei Φ eine Blockabbildung, die eine Konjugation von X nach Y induziert, und sei Ψ eine Blockabbildung, die deren Inverse induziert. Die Idee des Beweises ist es, Satz 3.5 anzuwenden unter Verwendung der Beobachtung, dass die Bilder zweier sich hinreichend überschneidender Blöcke in $\mathcal{B}(X)$ unter Φ sich ebenfalls hinreichend überschneiden, so dass man sie zusammensetzen kann, um einen Block in $\mathcal{B}(Y)$ zu bilden. Wenn man Ψ auf diesen Block anwendet, erhält man die ursprünglichen Blöcke zurück, die entlang ihrer Überlappung verklebt, aber an ihren Enden verkürzt wurden. Um diese Verkürzung zu kompensieren, müssen wir zunächst die ursprünglichen Blöcke erweitern.

Aber nun zum eigentlichen Beweis: Gemäß Satz 3.5 ist es unser Ziel, eine Zahl $M \geq 0$ zu finden, so dass aus $v \in \mathcal{B}(X)$ mit $|v| \geq M$ und $uv, vw \in \mathcal{B}(X)$ folgt, dass $uvw \in \mathcal{B}(X)$.

Sei $\phi : X \rightarrow Y$ eine Konjugation, induziert durch eine Blockabbildung Φ und sei $\psi = \phi^{-1} : Y \rightarrow X$ die Inverse von ϕ , induziert durch Ψ . Indem wir bei Bedarf die Fenstergröße anpassen, können wir annehmen, dass Φ und Ψ dasselbe Gedächtnis m und dieselbe Antizipation n haben und dass $m = n = l$ gilt. Da $\psi(\phi(x)) = x$ für alle $x \in X$, wählt die Blockabbildung $\Psi \circ \Phi : \mathcal{B}_{4l+1}(X) \rightarrow \mathcal{B}_1(X)$ einfach nur das zentrale Symbol in einem Block aus. (Hier verwenden wir wieder die erweiterte Blockabbildung, wie in (5) definiert.) Genauer: $\psi(\phi(x))_{[i]} = x_{[i]}$ für alle $i \in \mathbb{Z}$ ist äquivalent zu

$$\begin{aligned} \Psi(\phi(x)_{[i-l, i+l]}) &= x_{[i]} \quad \text{f.a. } i \in \mathbb{Z} \\ \Leftrightarrow \Psi(\phi(x)_{[i-l]} \phi(x)_{[i-l+1]} \cdots \phi(x)_{[i+l]}) &= x_{[i]} \quad \text{f.a. } i \in \mathbb{Z} \\ \Leftrightarrow \Psi(\Phi(x_{[i-2l, i]}) \Phi(x_{[i-2l+1, i+1]}) \cdots \Phi(x_{[i, i+2l]})) &= x_{[i]} \quad \text{f.a. } i \in \mathbb{Z} \\ \Leftrightarrow \Psi(\Phi(x_{[i-2l, i+2l]})) &= x_{[i]} \quad \text{f.a. } i \in \mathbb{Z}. \end{aligned}$$

Da Y von endlichem Typ ist, gibt es nach Satz 3.5 ein $N \geq 0$, so dass zwei Blöcke in Y , die sich in mindestens N Koordinaten überschneiden, entlang ihrer Überlappung zusammengeklebt werden können, um einen Block in Y zu bilden.

Sei $M := N + 4l$. Um zu zeigen, dass mit dieser Wahl die Bedingungen von Satz 3.5 erfüllt sind, und damit, dass X von endlichem Typ ist, seien $uv, vw \in \mathcal{B}(X)$ mit $|v| \geq M$. Nach Satz 2.9 gibt es Blöcke $s, t \in \mathcal{B}_{2l}(X)$ mit $su, vt \in \mathcal{B}(X)$. Da jeder $(4l+1)$ -Block in $suvt$ in $\mathcal{B}(X)$ enthalten ist, gilt $\Psi(\Phi(suvt)) = uvw$. (Hier verwenden wir wieder die Erweiterung (5).)

Weiter gilt $\Phi(su) = u'\Phi(v)$ und $\Phi(vt) = \Phi(v)w'$, wobei $u', w' \in \mathcal{B}(Y)$ und $|\Phi(v)| = |v| - 2l \geq N$. Daher können $u'\Phi(v)$ und $\Phi(v)w'$ zusammengefügt werden und wir erhalten $u'\Phi(v)w' \in \mathcal{B}(Y)$. Dann gilt

$$uvw = \Psi(\Phi(suvt)) = \Psi(u'\Phi(v)w') \in \mathcal{B}(X),$$

womit bewiesen ist, dass X von endlichem Typ ist. \square

3.2 Graphen und ihre Shifts

Eine Methode, um Shifts von endlichem Typ zu erzeugen, ist es mit einem endlichen gerichteten Graphen zu beginnen und den Shift zu erzeugen über die Menge aller bi-unendlichen Kantenfolgen des Graphen. In einem noch zu präzisierenden Sinne kann jeder Shift von endlichem Typ so rekodiert werden, dass er wie der Kantenshift eines Graphen aussieht.

Einem endlichen gerichteten Graphen kann man seine Übergangsmatrix zuordnen. Aussagen über den Kantenshift lassen sich dann treffen durch die Untersuchung dieser Matrix mit Methoden der Linearen Algebra.

3.8 Definition: Ein Graph G besteht aus einer endlichen Menge $\mathcal{V} = \mathcal{V}(G)$ von Knoten (engl. vertices) zusammen mit einer endlichen Menge $\mathcal{E} = \mathcal{E}(G)$ von Kanten (engl. edges). Jede Kante $e \in \mathcal{E}(G)$ beginnt bei einem Knoten $i(e) \in \mathcal{V}(G)$ und endet bei einem Knoten $t(e) \in \mathcal{V}(G)$. Zwischen zwei Knoten kann es mehrere Kanten geben. Eine Kante e mit $i(e) = t(e)$ nennen wir eine Schleife.

Ist I ein Knoten, so bezeichnen wir mit $\mathcal{E}_I(G)$ die Menge der bei I beginnenden Kanten und mit $\mathcal{E}^I(G)$ die Menge der bei I endenden Kanten.

3.9 Definition: Seien G und H Graphen. Ein Graphenhomomorphismus ist ein Paar $(\partial\Phi, \Phi)$ von Abbildungen, wobei $\partial\Phi : \mathcal{V}(G) \rightarrow \mathcal{V}(H)$ und $\Phi : \mathcal{E}(G) \rightarrow \mathcal{E}(H)$ mit $i(\Phi(e)) = \partial\Phi(i(e))$ und $t(\Phi(e)) = \partial\Phi(t(e))$ für alle $e \in \mathcal{E}(G)$. Sind $\partial\Phi$ und Φ beide injektiv, so heißt $(\partial\Phi, \Phi)$ Grapheneinbettung. Sind beide bijektiv, so heißt $(\partial\Phi, \Phi)$ Graphenisomorphismus. Falls zwischen zwei Graphen G und H ein Graphenisomorphismus existiert, so sagen wir, dass G und H isomorphe Graphen sind und schreiben $G \cong H$.

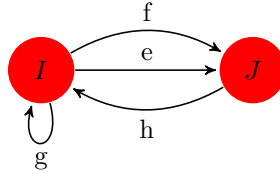


Abbildung 1: Ein einfacher Graph

Als nächstes definieren wir Übergangsmatrizen. Es ist üblich, die Knoten eines Graphen zu ordnen, z.B. $\mathcal{V} = \{I_1, I_2, \dots, I_r\}$. Oft werden wir die Knoten mit verschiedenen ganzen Zahlen benennen, also z.B. $\mathcal{V} = \{1, 2, \dots, r\}$ oder $\mathcal{V} = \{0, 1\}$, oder mit verschiedenen Buchstaben, z.B. $\mathcal{V} = \{I, J, K\}$. In diesen Fällen ist die Ordnung der Knoten die natürliche, entweder die der ganzen Zahlen oder die alphabetische.

3.10 Definition: Sei G ein Graph mit Knotenmenge \mathcal{V} . Für Knoten $I, J \in \mathcal{V}$ sei A_{IJ} die Anzahl der Kanten, die bei I beginnen und bei J enden. Dann ist die Übergangsmatrix von G gegeben durch $A = [A_{IJ}]$ und wir schreiben $A = A(G)$ oder $A = A_G$, um zu verdeutlichen, dass A von dem Graphen G erzeugt wird.

Die Ordnung der Knoten von G sagt uns lediglich, wie wir die Zahlen A_{IJ} zu einer Matrix zusammenstellen müssen. Wenn wir die Knoten anders anordnen, dann ergibt sich eine neue Übergangsmatrix A' , die aber mittels einer Permutationsmatrix P zur alten konjugiert ist, d.h. $A' = PAP^{-1}$.

3.11 Satz: Sei G ein Graph mit Knotenmenge $\mathcal{V} = \{1, \dots, r\}$. Sei $\theta : \mathcal{V} \rightarrow \mathcal{V}$ eine bijektive Abbildung, also eine Permutation der Knoten von G . Sei A die Übergangsmatrix von G bzgl. der Ordnung $1, 2, \dots, r$ der Knoten und A' die Übergangsmatrix bzgl. der Ordnung $\theta(1), \theta(2), \dots, \theta(r)$. Dann existiert eine $r \times r$ -Permutationsmatrix P , so dass

$$A' = PAP^{-1}.$$

Beweis: Wir bezeichnen mit e_1, \dots, e_r die Elemente der Standardbasis von \mathbb{R}^r , also $e_i = (x_{i1}, \dots, x_{ir})$ mit $x_{ij} = 1$, falls $i = j$ und $x_{ij} = 0$ andernfalls. Sei P die Matrix, deren j -te Spalte der Vektor $e_{\theta^{-1}(j)}$ ist. Dies ist offensichtlich eine Permutationsmatrix und es gilt

$$\begin{aligned} PAP^{-1}e_i &= PAe_{\theta(i)} = P \sum_{k=1}^r A_{k\theta(i)}e_k \\ &= \sum_{k=1}^r A_{k\theta(i)}Pe_k = \sum_{k=1}^r A_{k\theta(i)}e_{\theta^{-1}(k)} = \sum_{k=1}^r A_{\theta(k)\theta(i)}e_k. \end{aligned}$$

Außerdem gilt

$$A'e_i = \sum_{k=1}^r A'_{ki} e_k = \sum_{k=1}^r A_{\theta(k)\theta(i)} e_k.$$

Da die Bilder der Vektoren e_1, \dots, e_r die Spalten der jeweiligen Matrix sind, folgt $A' = PAP^{-1}$. Hier haben wir verwendet, dass der Eintrag A'_{ki} nach Definition die Anzahl der Kanten ist, die bei $\theta(k)$ beginnen und bei $\theta(i)$ enden, also gerade $A_{\theta(k)\theta(i)}$. \square

Die Matrix A_G enthält alle wesentlichen Informationen über den Graphen G . Bis auf Graphenisomorphismen können wir den Graphen aus der Matrix A_G rekonstruieren.

3.12 Definition: Sei $A = [A_{IJ}]$ eine $r \times r$ -Matrix mit nichtnegativen ganzzahligen Einträgen. Dann ist der Graph von A der Graph $G = G(A) = G_A$ mit Knotenmenge $\mathcal{V}(G) = \{1, 2, \dots, r\}$ und mit A_{IJ} verschiedenen Kanten, die bei I beginnen und bei J enden.

Es gilt

$$A = A(G(A)) \quad \text{und} \quad G \cong G(A(G)).$$

Der Graphenisomorphismus \cong resultiert aus einer möglichen Umbenennung der Knoten und der Kanten. Diese Korrespondenz zwischen Graphen und ihren Übergangsmatrizen erlaubt es, einen Graphen entweder durch G oder durch A anzugeben. Dies drücken wir in der Notation aus, indem wir z.B. $\mathcal{V}(A)$ für die Menge der A zugrundeliegenden Indizes schreiben.

Jeder Graph erzeugt einen Shift von endlichem Typ.

3.13 Definition: Sei G ein Graph mit Kantenmenge \mathcal{E} und Übergangsmatrix A . Der Kantenshift X_G oder X_A ist der Shift über dem Alphabet \mathcal{E} , definiert durch

$$X_G = X_A := \{(\xi_i)_{i \in \mathbb{Z}} : \xi_i \in \mathcal{E}, t(\xi_i) = i(\xi_{i+1}) \text{ für alle } i \in \mathbb{Z}\}.$$

Die Einschränkung der Shiftabbildung auf X_G heißt Kantenshiftabbildung und wird mit σ_G oder σ_A bezeichnet.

Nach dieser Definition ist eine bi-unendliche Kantenfolge genau dann in X_G , wenn der Knoten, bei dem eine Kante endet, der Anfang der nächsten Kante ist. Die Folgen in X_G beschreiben also beidseitig unendliche Pfade (engl. *walks*) auf dem Graphen G .

Es ist leicht zu sehen, dass die Kantenshifts zweier isomorpher Graphen konjugiert zueinander sind (Übungsaufgabe auf Blatt 3).

3.14 Satz: Ist G ein Graph mit Übergangsmatrix A , dann ist der Kantenshift $X_G = X_A$ ein 1-Schritt-Shift.

Beweis: Sei $\mathcal{A} = \mathcal{E}$ das Alphabet von X_G . Betrachte die endliche Menge

$$\mathcal{F} := \{ef : e, f \in \mathcal{A}, t(e) \neq i(f)\}.$$

Nach Definition des Kantenshifts liegt ein Punkt $\xi \in \mathcal{A}^{\mathbb{Z}}$ genau dann in X_G , wenn kein Block aus \mathcal{F} in ξ vorkommt. Also gilt $X_G = X_{\mathcal{F}}$ und X_G ist ein Shift von endlichem Typ. Da \mathcal{F} nur 2-Blöcke enthält, handelt es sich um einen 1-Schritt-Shift. \square

3.15 Beispiel: Sei $r \geq 1$ und sei A die 1×1 -Matrix $A = [r]$. Dann hat G_A nur einen Knoten und r Schleifen bei diesem Knoten. Nennen wir diese $0, 1, \dots, r-1$, so ist X_G der volle r -Shift, den wir deshalb auch mit $X_{[r]}$ bezeichnen.

Manchmal tauchen gewisse Kanten von G nicht in X_G auf und sind damit unwesentlich für den Kantenshift. Endet z.B. eine Kante in einem Knoten, bei dem keine Kanten beginnen, so handelt es sich um eine „Sackgasse“. Wir nennen einen Knoten $I \in \mathcal{V}$ *gestrandet*, falls keine Kanten bei I beginnen oder keine Kanten bei I enden. In der Übergangsmatrix drückt sich dies dadurch aus, dass die I -te Zeile oder die I -te Spalte von A_G nur Nullen enthält.

3.16 Definition: Ein Graph heißt wesentlich, falls keiner seiner Knoten gestrandet ist.

Um gestrandete Knoten zu entfernen, benötigen wir den Begriff eines *Teilgraphen* von G , d.h. eines Graphen H mit $\mathcal{V}(H) \subseteq \mathcal{V}(G)$ und $\mathcal{E}(H) \subseteq \mathcal{E}(G)$, so dass Kanten in H bei denselben Knoten beginnen und enden wie in G .

3.17 Satz: Ist G ein Graph, dann gibt es einen eindeutigen Teilgraphen H von G , so dass H wesentlich ist und $X_H = X_G$.

Beweis: Sei $\mathcal{E}(H)$ die Mengen aller Kanten in G , die in bi-unendlichen Pfaden durchlaufen werden und sei $\mathcal{V}(H) = \{i(e) : e \in \mathcal{E}(H)\}$ die Menge der Knoten, die auf solchen Pfaden passiert werden. Dann ist H ein Teilgraph von G , und nach Definition ist jeder bi-unendliche Pfad auf G eigentlich ein Pfad auf H . Also ist H wesentlich und $X_H = X_G$.

Um zu zeigen, dass H eindeutig ist, stellen wir zunächst fest, dass nach Definition H der größte wesentliche Teilgraph von G ist. Ist also H' ein wesentlicher Teilgraph von G mit $H' \neq H$, so gibt es eine Kante in H , die nicht in H' vorkommt. Da jede Kante in H in X_H auftaucht, folgt $X_H \neq X_{H'}$ und die Eindeutigkeit von H ist gezeigt. \square

Da der Teilgraph in obigem Satz alle für die symbolische Dynamik relevante Information enthält, werden wir uns im Folgenden auf die Betrachtung wesentlicher Teilgraphen beschränken.

3.18 Definition: Ein Pfad $\pi = e_1 e_2 \dots e_m$ auf einem Graphen G ist eine endliche Folge von Kanten e_i aus G , so dass $t(e_i) = i(e_{i+1})$ für $1 \leq i \leq m-1$. Die

Länge von $\pi = e_1 e_2 \dots e_m$ ist $|\pi| = m$. Der Pfad $\pi = e_1 e_2 \dots e_m$ beginnt bei dem Knoten $i(\pi) := i(e_1)$ und endet bei dem Knoten $t(\pi) := t(e_m)$ und π ist ein Pfad von $i(\pi)$ nach $t(\pi)$. Eine Schleife oder ein geschlossener Pfad ist ein Pfad, der bei demselben Knoten beginnt und endet. Ein einfacher Pfad ist ein Pfad, der sich nicht selbst überkreuzt, d.h. die Knoten $i(e_1), \dots, i(e_m)$ sind paarweise verschieden. Für jeden Knoten I von G gibt es einen leeren Pfad ε_I , der bei I beginnt und endet.

Wir können die Potenzen der Übergangsmatrix A_G verwenden, um die Anzahl von Pfaden einer bestimmten Länge zu zählen.

3.19 Satz: Sei G ein Graph mit Übergangsmatrix A und sei $m \geq 0$. Dann gelten folgende Aussagen:

- (1) Die Anzahl der Pfade der Länge m von I nach J ist $(A^m)_{IJ}$, der (IJ) -te Eintrag der Matrix A^m .
- (2) Die Anzahl der Schleifen der Länge m in G ist $\text{Spur}(A^m)$, die Spur von A^m , und dies ist zugleich die Anzahl der periodischen Punkte in X_G der Periode m .

Beweis: (1) Wir beweisen die Aussage mittels vollständiger Induktion. Für $m = 0$ ist die Aussage trivialerweise wahr, da die einzigen Pfade der Länge 0 die leeren Pfade bei jedem Knoten sind. Für $m = 1$ ist die Aussage wahr, da nach Definition A_{IJ} die Anzahl der Kanten ist, die bei I beginnen und bei J enden. Die Kanten sind aber gerade die Pfade der Länge 1. Jetzt nehmen wir an, dass die Aussage für ein $m \in \mathbb{N}_0$ wahr ist. Es gilt

$$(A^{m+1})_{IJ} = (A^m A)_{IJ} = \sum_K A_{IK}^m A_{KJ}.$$

Für jeden Pfad $e_1 \dots e_{m+1}$ der Länge $m + 1$ von I nach J gibt es einen Knoten K , so dass $e_1 \dots e_m$ bei K endet. Wir erhalten also alle solchen Pfade, indem wir jeden Knoten K und die dort endenden Pfade der Länge m betrachten, die bei I beginnen und die Anzahl dieser Pfade mit der Anzahl der Kanten der Länge 1 von K nach J multiplizieren und über alle K aufsummieren. Das ist aber gerade obige Summe. Damit ist (1) gezeigt.

(2) Die erste Aussage folgt unmittelbar aus (1). Ist π eine Schleife in G der Länge m , so ist π^∞ ein m -periodischer Punkt in X_G . Ist andererseits $x \in X_G$ ein m -periodischer Punkt, dann muss $x_{[0, m-1]}$ eine Schleife in G der Länge m sein. Also besteht eine eins-zu-eins-Korrespondenz zwischen den m -periodischen Punkten in X_G und den Schleifen in G der Länge m . \square

Nach Definition 2.11 heißt ein Shift irreduzibel, falls es zu jedem Paar (u, v) von erlaubten Blöcken einen Block w gibt, so dass $u w v$ erlaubt ist. Wir wollen nun die Graphen klassifizieren, deren Kantenshifts irreduzibel sind.

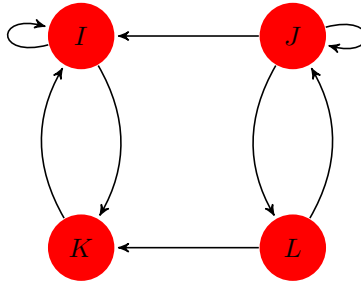


Abbildung 2: Ein reduzibler Graph

3.20 Definition: Ein Graph G heißt irreduzibel, falls zu jedem Paar (I, J) von Knoten in G ein Pfad existiert, bei I beginnt und bei J endet.

3.21 Satz: Ein wesentlicher Graph ist irreduzibel genau dann, wenn sein Kantenshift irreduzibel ist.

Beweis: Sei G ein irreduzibler Graph und seien $\pi, \tau \in \mathcal{B}(X_G)$. Sei I der Knoten, bei dem der Pfad π endet und sei J der Knoten, bei dem der Pfad τ beginnt. Da G irreduzibel ist, existiert ein Pfad $\omega \in \mathcal{B}(X_G)$ von I nach J . Dann ist $\pi\omega\tau$ ein Pfad in G und somit gilt $\pi\omega\tau \in \mathcal{B}(X_G)$.

Jetzt nehmen wir an, dass G wesentlich und X_G ein irreduzibler Shift ist. Seien I, J Knoten in G . Dann gibt es Kanten e, f , so dass e bei I endet und f bei J beginnt. Da X_G irreduzibel ist, gibt es ein $\omega \in \mathcal{B}(X_G)$, so dass $e\omega f \in \mathcal{B}(X_G)$. Dann ist ω ein Pfad von I nach J , also ist G irreduzibel. \square

3.3 Graphendarstellungen von Shifts von endlichem Typ

In diesem Unterabschnitt werden wir zeigen, dass jeder Shift von endlichem Typ so rekodiert werden kann (mittels einer höheren Blockdarstellung), dass er mit dem Kantenshift eines Graphen identifiziert werden kann. Das ist erstaunlich, da die Kantenshifts sehr spezielle Shifts von endlichem Typ zu sein scheinen.

Das folgende Beispiel zeigt, dass nicht jeder 1-Schritt-Shift ein Kantenshift ist.

3.22 Beispiel: Sei $\mathcal{A} = \{0, 1\}$, $\mathcal{F} = \{11\}$ und $X = X_{\mathcal{F}}$. Wäre X der Kantenshift eines Graphen G , so können wir nach Satz 3.17 annehmen, dass G ein wesentlicher Graph ist. Dann hätte G genau zwei Kanten, die 0 und 1 heißen. Für so einen Graphen gibt es zwei Möglichkeiten. Entweder er hat nur einen Knoten, so dass 0 und 1 Schleifen bei diesem Knoten sind, oder er hat zwei Knoten, so dass man zwischen den Knoten mittels der Kanten 0 und 1 hin und her laufen kann. Im ersten Fall ist der Kantenshift der volle 2-Shift und im

zweiten Fall besteht er nur aus den zwei Punkten $(01)^\infty$ und $(10)^\infty$. In beiden Fällen ist er also nicht der Shift X .

Dennoch ist es möglich, jeden M -Schritt-Shift zu einem Kantenshift zu rekodieren.

3.23 Satz: *Sei X ein M -Schritt-Shift. Dann gibt es einen Graphen G , so dass $X^{[M+1]} = X_G$.*

Beweis: Ist $M = 0$, so ist X ein voller Shift und wir können den Graphen G so wählen, dass er nur einen Knoten hat und eine Kante für jedes Symbol des Alphabets (siehe Beispiel 3.15). Wir können also $M \geq 1$ annehmen. Dann sei die Knotenmenge von G gegeben durch $\mathcal{B}_M(X)$, die Menge der M -Blöcke in X . Wir definieren die Kantenmenge wie folgt. Seien $I = a_1 a_2 \dots a_M$ und $J = b_1 b_2 \dots b_M$ Knoten in G . Falls $a_2 a_2 \dots a_M = b_1 b_2 \dots b_{M-1}$ und $a_1 \dots a_M b_M (= a_1 b_1 \dots b_M)$ in $\mathcal{B}(X)$ liegt, zeichne genau eine Kante von I nach J und nenne diese $a_1 a_2 \dots a_M b_M = a_1 b_1 b_2 \dots b_M$. Andernfalls gibt es keine Kante von I nach J . Nach dieser Konstruktion ist klar, dass ein bi-unendlicher Pfad auf G einer Folge von $(M + 1)$ -Blöcken in $\mathcal{B}_{M+1}(X)$ entspricht, die sich progressiv überschneiden. Also ist $X_G = X^{[M+1]}$. \square

3.24 Bemerkung: Für den Shift aus Beispiel 3.22 gilt also $X^{[2]} = X_G$.

Als nächstes führen wir eine Konstruktion für Graphen ein, die der Bildung des N -ten höheren Blockshifts für Shifträume entspricht.

3.25 Definition: *Sei G ein Graph. Für $N \geq 2$ führen wir den N -ten höheren Kantengraphen $G^{[N]}$ ein, dessen Knotenmenge die Menge der Pfade der Länge $N - 1$ in G ist und der genau eine Kante von $e_1 e_2 \dots e_{N-1}$ nach $f_1 f_2 \dots f_{N-1}$ hat, falls $e_2 \dots e_{N-1} = f_1 \dots f_{N-2}$ (oder $t(e_1) = i(f_2)$, falls $N = 2$), und sonst keine weiteren Kanten. Die Kante heißt $e_1 \dots e_{N-1} f_{N-1} = e_1 f_1 \dots f_{N-1}$. Für $N = 1$ sei $G^{[1]} := G$.*

3.27 Satz: *Sei G ein Graph. Dann gilt $(X_G)^{[N]} = X_{G^{[N]}}$.*

Beweis: Die Symbole für $(X_G)^{[N]}$ sind die N -Blöcke aus X_G , die die Pfade der Länge N in G sind. Das sind aber auch die Symbole für $X_{G^{[N]}}$. Eine bi-unendliche Folge dieser Symbole ist in jedem der beiden Shifts enthalten genau dann, wenn die Symbole (d.h. N -Blöcke) sich progressiv überschneiden. \square

Für $N \geq 2$ hat die Übergangsmatrix von $G^{[N]}$ nur Nullen und Einsen als Einträge. Eine solche Matrix heißt 0-1-Matrix. Falls die Übergangsmatrix von G eine 0-1-Matrix ist, dann gibt es zwischen zwei Knoten höchstens eine Kante. Daher kann ein Pfad auf G auch durch die Folge der besuchten Knoten beschrieben werden.

3.26 Beispiel:

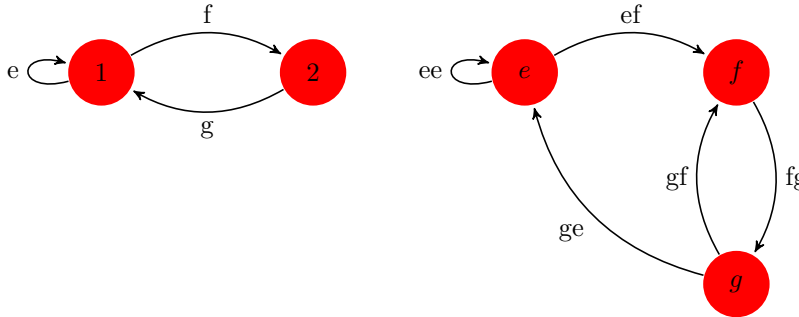


Abbildung 3: $G = G^{[1]}$ und $G = G^{[2]}$

3.28 Definition: Sei B eine $r \times r$ -Matrix mit Nullen und Einsen als Einträgen, oder äquivalent dazu, die Übergangsmatrix eines Graphen G , so dass zwischen zwei beliebigen Knoten höchstens eine Kante verläuft. Der Knotenshift $\widehat{X}_B = \widehat{X}_G$ ist der Shiftraum über dem Alphabet $\mathcal{A} = \{1, 2, \dots, r\}$, definiert durch

$$\widehat{X}_B = \widehat{X}_G := \{x = (x_i)_{i \in \mathbb{Z}} \in \mathcal{A}^{\mathbb{Z}} : B_{x_i x_{i+1}} = 1 \text{ für alle } i \in \mathbb{Z}\}.$$

Die Knotenshiftabbildung ist die Einschränkung der Shiftabbildung auf $\widehat{X}_B = \widehat{X}_G$ und wird mit $\widehat{\sigma}_B = \widehat{\sigma}_G$ bezeichnet.

Nach Satz 3.14 ist jeder Kantenshift ein 1-Schritt-Shift. Das folgende Resultat zeigt den Zusammenhang zwischen Kantenshifts, Knotenshifts und 1-Schritt-Shifts.

3.29 Satz: Es gelten folgende Aussagen:

- (1) Bis auf Umbenennung der Symbole sind die 1-Schritt-Shifts dasselbe wie die Knotenshifts.
- (2) Bis auf Umbenennung der Symbole ist jeder Kantenshift ein Knotenshift (eines anderen Graphen).
- (3) Ist X ein M -Schritt-Shift, dann ist $X^{[M]}$ ein 1-Schritt-Shift oder, äquivalent, ein Knotenshift. In der Tat gibt es einen Graphen G mit $X^{[M]} = \widehat{X}_G$ und $X^{[M+1]} = X_G$.

Beweis: Der Knotenshift \widehat{X}_B ist der Shift, der durch die Menge $\mathcal{F} = \{ij : B_{ij} = 0\}$ von verbotenen 2-Blöcken definiert wird, also ein 1-Schritt-Shift. Umgekehrt ist jeder 1-Schritt-Shift bis auf Umbenennung der Symbole ein Knotenshift. Ist nämlich $X = X_{\mathcal{F}}$, wobei \mathcal{F} aus 2-Blöcken besteht, dann kann X als der Knotenshift \widehat{X}_B betrachtet werden, wobei B die 0-1-Matrix ist, die durch das Alphabet von X indiziert wird mit $B_{ij} = 0$ genau dann, wenn $ij \in \mathcal{F}$. Dies beweist (1).

Aussage (2) folgt aus (1) und Satz 3.14.

Die erste Aussage in (3) folgt unmittelbar aus der Tatsache, dass jeder $(M+1)$ -Block in X als 2-Block in $X^{[M]}$ betrachtet werden kann. Um die zweite Aussage zu beweisen, verwenden wir den im Beweis von Satz 3.23 konstruierten Graphen. (Die Details werden ausgelassen.) \square

Teil (2) des obigen Satzes besagt, dass jeder Kantenshift auch als Knotenshift betrachtet werden kann. Um direkt zu sehen, wie das funktioniert, beginnen wir mit einem Kantenshift X_A mit Übergangsmatrix A und bezeichnen mit \mathcal{E} die Menge der Kanten des Graphen. Es gibt $\sum_{I,J} A_{IJ}$ Elemente in \mathcal{E} . Wir bilden die durch \mathcal{E} indizierte 0-1-Matrix B durch

$$B_{ef} := \begin{cases} 1 & \text{falls } t(e) = i(f), \\ 0 & \text{falls } t(e) \neq i(f). \end{cases}$$

Dann ist \widehat{X}_B eine Version von X_A , in der die Kanten von G_A den Knoten von G_B entsprechen; im Wesentlichen gilt $\widehat{X}_B = X_A$ und $G_B = G_A^{[2]}$.

Da es einfacher erscheint mit 0-1-Matrizen zu arbeiten als mit allgemeinen nicht-negativen ganzzahligen Matrizen, warum verwenden wir dann nicht ausschließlich Knotenshifts statt Kantenshifts? Dafür gibt es zwei Gründe. Der erste ist ein ökonomischer. Drücken wir z.B. den Kantenshift der Übergangsmatrix

$$A = \begin{bmatrix} 3 & 2 \\ 2 & 1 \end{bmatrix}$$

als Knotenshift aus, so erhalten wir die viel größere Matrix

$$B = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix}.$$

Der zweite Grund ist, dass gewisse natürliche Operationen wie z.B. die Potenzbildung die Eigenschaft, 0-1-Matrix zu sein, nicht erhalten. Diese Operation wird z.B. benötigt, wenn man den wie folgt definierten höheren Potenzgraphen verwendet.

3.30 Definition: Sei G ein Graph und $N \geq 1$ eine ganze Zahl. Dann definieren wir den N -ten höheren Potenzgraphen G^N von G mit Knotenmenge $\mathcal{V}(G^N) := \mathcal{V}(G)$ mit genau einer Kante von I nach J für jeden Pfad in G der Länge N von I nach J .

Folglich gilt $G^1 = G$, während G^N typischerweise mehr Kanten als G hat, wenn N groß ist.

3.31 Satz: Sei G ein Graph mit Übergangsmatrix A_G . Dann ist die Übergangsmatrix des N -ten höheren Potenzgraphen G^N die N -te Potenz von A_G , also

$$A_{G^N} = A_G^N.$$

Darüberhinaus gilt $X_{G^N} = (X_G)^N$.

Beweis: Nach Definition haben G^N und G dieselbe Knotenmenge. Nach Satz 3.19 gibt es für Knoten I und J in G genau $(A_G^N)_{IJ}$ Pfade der Länge N von I nach J und dies ist die Anzahl der Kanten in G^N von I nach J . Daher gilt $A_{G^N} = A_G^N$.

Sowohl X_{G^N} als auch $(X_G)^N$ haben ein Alphabet, das aus Pfaden der Länge N in G besteht, und eine bi-unendliche Folge solcher Pfade ist in X_{G^N} (äquivalent, in $(X_G)^N$) genau dann, wenn sie einen bi-unendlichen Pfad auf G beschreibt. Daher gilt $X_{G^N} = (X_G)^N$. \square

Es gibt einen engen Zusammenhang zwischen den Shifts, die wir in diesem Abschnitt betrachten und Markov-Ketten (aus der Wahrscheinlichkeitstheorie).

3.32 Definition: Sei S eine endliche Menge. Eine Markov-Kette μ auf S ist gegeben durch eine Zuordnung von Zustandswahrscheinlichkeiten $\mu(I) \geq 0$ für alle $I \in S$ und bedingten Wahrscheinlichkeiten $\mu(J|I) \geq 0$ für alle $I, J \in S$, so dass

$$\sum_{I \in S} \mu(I) = 1$$

und

$$\sum_{J \in S} \mu(J|I) = 1 \quad \text{für alle } I \in S.$$

Die Idee ist, dass S die Menge der möglichen Ausgänge eines Experiments darstellt, die auch Zustände genannt werden. Die Wahrscheinlichkeit, dass eine Folge $I_0 I_1 \dots I_n$ von Ausgängen beobachtet wird, ist definiert als

$$\mu(I_0 I_1 \dots I_n) := \mu(I_0) \mu(I_1|I_0) \mu(I_2|I_1) \cdots \mu(I_n|I_{n-1}).$$

Gegeben, dass eine gewisse Folge $I_0 I_1 \dots I_{n-1}$ von Ausgängen bereits aufgetaucht ist, ist die Wahrscheinlichkeit, dass der nächste Ausgang eine bestimmtes I_n ist,

$$\frac{\mu(I_0 I_1 \dots I_n)}{\mu(I_0 I_1 \dots I_{n-1})} = \mu(I_n|I_{n-1})$$

und hängt folglich nur von dem vorherigen Ausgang I_{n-1} ab.

Jeder Markov-Kette kann ein Graph $G = (\mathcal{V}, \mathcal{E})$ zugeordnet werden, der wie folgt definiert ist. Die Knoten von G sind die Zustände in S mit positiver Wahrscheinlichkeit, also

$$\mathcal{V} := \{I \in S : \mu(I) > 0\}$$

und die Kanten in G sind die Übergänge von einem Zustand in einen anderen, die positive Wahrscheinlichkeiten haben, also

$$\mathcal{E} := \{(I, J) : \mu(J|I) > 0\}.$$

In G gibt es höchstens eine Kante zwischen jedem Paar von Knoten. Daher sind die Ausgangsfolgen mit positiver Wahrscheinlichkeit genau die Zustandsfolgen der Pfade im Graphen G , mit anderen Worten die Blöcke, die in der Sprache des Knotenshifts \widehat{X}_G vorkommen. Folglich ist ein Knotenshift eine „extremale“ Version einer Markov-Kette in dem Sinne, dass er nur eine Menge von erlaubten Ausgangsfolgen festlegt ohne Auskunft über die Wahrscheinlichkeiten zu geben. Aus diesem Grunde wurden Knotenshifts ursprünglich *intrinsische Markov-Ketten* oder *topologische Markov-Ketten* genannt.

Nun betrachten wir einen beliebigen Graphen $G = (\mathcal{V}, \mathcal{E})$, der mehrere Kanten zwischen jedem Knotenpaar haben darf. Legen wir Zustandswahrscheinlichkeiten für die Knoten in \mathcal{V} und bedingte Wahrscheinlichkeiten für die Kanten in \mathcal{E} fest, dann können wir Wahrscheinlichkeiten für die Pfade in G definieren.

3.33 Definition: Eine Markov-Kette auf einem Graphen $G = (\mathcal{V}, \mathcal{E})$ ist eine Zuordnung von Wahrscheinlichkeiten $\mu(I) \geq 0$ für alle $I \in \mathcal{V}$ und bedingten Wahrscheinlichkeiten $\mu(e|i(e)) \geq 0$ für alle $e \in \mathcal{E}$, so dass

$$\sum_{I \in \mathcal{V}} \mu(I) = 1$$

und

$$\sum_{e \in \mathcal{E}_I} \mu(e|I) = 1 \quad \text{für alle } I \in \mathcal{V}.$$

Für eine Markov-Kette auf einem Graphen G definieren wir die Wahrscheinlichkeit eines Pfads $e_1 \dots e_n$ durch

$$\mu(e_1 \dots e_n) := \mu(i(e_1))\mu(e_1|i(e_1))\mu(e_2|i(e_2)) \cdots \mu(e_n|i(e_n)).$$

Folglich repräsentieren die *Kanten* von G die Ausgänge des Experiments und die Pfade in G repräsentieren Folgen von Ausgängen.

3.34 Bemerkung: Wir fordern nicht, dass die Wahrscheinlichkeiten, die wir den Zuständen und Zustandsübergängen zuordnen, positiv sind. Sind sie aber alle positiv, dann sind die möglichen Folgen genau die Blöcke, die in der Sprache von X_G auftauchen. Andernfalls sind die möglichen Folgen die Blöcke, die in der Sprache des Kantenshifts X_H auftauchen, der durch den Teilgraphen H bestimmt ist, der aus allen Zuständen mit positiver Wahrscheinlichkeit und allen Kanten mit positiven bedingten Wahrscheinlichkeiten besteht.

Für eine Markov-Kette auf einem Graphen ist es üblich, die Zustandswahrscheinlichkeiten $\mu(I)$ in einem Zeilenvektor zusammenzufassen, den wir die *Zustandsverteilung* nennen. Dies ist ein Vektor \mathbf{p} , indiziert durch die Zustände von

G , definiert durch

$$p_I := \mu(I).$$

Es ist auch üblich die bedingten Wahrscheinlichkeiten $\mu(e|I)$ in einer quadratischen Matrix zusammenzufassen, die wir die *bedingte Wahrscheinlichkeitsmatrix* nennen. Dies ist die Matrix P , ebenfalls durch die Zustände von G indiziert, definiert durch

$$P_{I,J} := \sum_{e \in \mathcal{E}_I^J} \mu(e|I), \quad \mathcal{E}_I^J = \mathcal{E}_I \cap \mathcal{E}^J.$$

Der Vektor \mathbf{p} ist ein *Wahrscheinlichkeitsvektor*, d.h. ein Vektor mit nichtnegativen Einträgen, deren Summe 1 ergibt. Die Matrix P ist eine *stochastische Matrix*, d.h. eine Matrix mit nichtnegativen Einträgen, deren Zeilensummen 1 ergeben.

Wir können \mathbf{p} und P verwenden, um gewisse natürliche Größen darzustellen, die mit der Markov-Kette auf G zu tun haben. Zum Beispiel ist $\mathbf{p}P^m$ der Vektor, dessen I -te Komponente die Summe der Wahrscheinlichkeiten aller Pfade der Länge m in G ist, die bei I beginnen.

3.4 Zustandsaufspaltung

Zustandsaufspaltung ist ein Verfahren, um neue Graphen aus einem gegebenen Graphen zu konstruieren. Ausgehend von einer Partition der Kantenmenge wird jeder Knoten aufgespalten in eine bestimmte Anzahl von abgeleiteten Knoten.⁶ Obwohl der resultierende Graph sehr anders aussehen kann als der ursprüngliche Graph, haben die beiden Graphen konjugierte Kantenshifts. Eine Motivation für diese Konstruktion ist die Tatsache, dass sich jede Konjugation zwischen Kantenshifts als Hintereinanderausführung von endlich vielen Zustandsaufspaltungen und Inversen von Zustandsaufspaltungen (sogenannten Amalgamationen) schreiben lässt.

Wir beschreiben zunächst, wie man einen einzelnen Knoten aufspalten kann. Sei G ein Graph mit Knotenmenge \mathcal{V} und Kantenmenge \mathcal{E} . Wir wählen einen Knoten $I \in \mathcal{V}$ und nehmen zunächst an, dass es bei I keine Schleifen gibt. Die Menge \mathcal{E}_I der Kanten, die bei I beginnen, wird in zwei disjunkte Teilmengen \mathcal{E}_I^1 und \mathcal{E}_I^2 zerlegt. Wir konstruieren einen neuen Graphen H , basierend auf dieser Zerlegung, wie folgt.

Die Knoten von H sind die von G mit der Ausnahme, dass I durch zwei Knoten I^1 und I^2 ersetzt wird, d.h.

$$\mathcal{W} = \mathcal{V}(H) = (\mathcal{V} \setminus \{I\}) \cup \{I^1, I^2\}.$$

Für jedes $e \in \mathcal{E}_I^i$ ($i \in \{1, 2\}$) zeichne eine Kante in H von I^i nach $t(e)$, die ebenfalls den Namen e trägt. (Beachte, dass $t(e) \neq I$, da wir annehmen, dass

⁶Die Begriffe *Knoten* und *Zustand* werden hier synonym verwendet. Wir könnten auch von einer *Knotenaufspaltung* sprechen.

es bei I keine Schleifen gibt.) Für jede Kante $f \in \mathcal{E}$, die bei J beginnt und bei I endet, zeichne zwei Kanten, genannt f^1 und f^2 , in H , wobei f^1 von J nach I^1 und f^2 von J nach I^2 verläuft. Alle anderen Kanten in G werden ohne Änderung nach H übertragen.

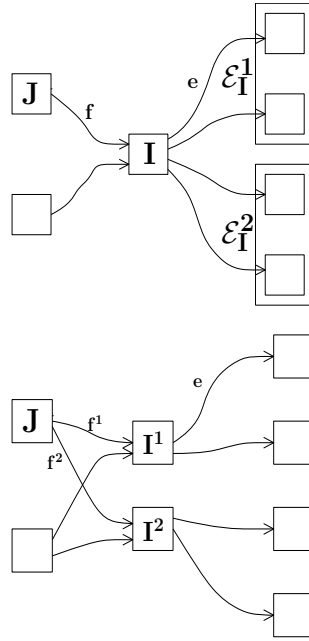


Abbildung 4: Zustandsaufspaltung an einem Knoten

Wir nehmen an, dass H aus G durch Aufspaltung eines Zustands I , wie oben beschrieben, entsteht. Dann konstruieren wir eine Konjugation zwischen den Kantenshifts X_G und X_H . (Zur Erinnerung: \mathcal{E}^I bezeichnet die Menge der bei I endenden Kanten.) Definiere die 1-Blockabbildung $\Psi : \mathcal{B}_1(X_H) \rightarrow \mathcal{B}_1(X_G)$ durch $\Psi(f^i) := f$, falls $f \in \mathcal{E}^I$ und $\Psi(e) := e$, falls $e \notin \mathcal{E}^I$. Mit anderen Worten: Ψ entfernt einfach die Indizes. Man überlegt sich leicht, dass durch die Entfernung von Indizes Pfade in H auf Pfade in G abgebildet werden, so dass Ψ einen 1-Blockcode $\psi = \Psi_\infty : X_H \rightarrow X_G$ induziert (d.h. der von Ψ induzierte Blockcode bildet nach X_G ab).

Umgekehrt definieren wir eine 2-Blockabbildung $\Phi : \mathcal{B}_2(X_G) \rightarrow \mathcal{B}_1(X_H)$ durch

$$\Phi(fe) := \begin{cases} f & \text{falls } f \notin \mathcal{E}^I, \\ f^1 & \text{falls } f \in \mathcal{E}^I \text{ und } e \in \mathcal{E}_I^1, \\ f^2 & \text{falls } f \in \mathcal{E}^I \text{ und } e \in \mathcal{E}_I^2. \end{cases} \quad (6)$$

Die Abbildung Φ „sieht“ also das jeweils nächste Symbol voraus und fügt einen Index hinzu, abhängig davon, was sie sieht. Man kann leicht zeigen, dass Φ

Pfade in G auf Pfade in H abbildet, so dass Φ einen gleitenden Blockcode $\phi = \Phi_\infty : X_G \rightarrow X_H$ mit Gedächtnis 0 und Antizipation 1 induziert.

Da Hinzufügen und anschließendes Entfernen von Indizes keinen Effekt hat, gilt $\psi(\phi(x)) = x$ für alle $x \in X_G$. Umgekehrt sind alle Indizes eindeutig bestimmt durch (6), da \mathcal{E}_I^1 und \mathcal{E}_I^2 die Menge \mathcal{E}^I partitionieren, so dass $\phi(\psi(y)) = y$ für alle $y \in X_H$. Folglich ist ϕ eine Konjugation von X_G nach X_H .

3.35 Beispiel: Die Wirkung von ϕ und ψ auf typischen Punkten $x \in X_G$ und $y \in X_H$ sieht wie folgt aus.

$$\begin{array}{c} x = \dots d a d b d c e . a d c e b d \dots \\ \phi \downarrow \uparrow \psi \\ y = \dots d^1 a d^2 b d^2 c e^1 . a d^2 c e^2 b \dots \end{array}$$

Beachte, dass in y das letzte d ausgelassen wird, weil es von dem Symbol abhängt, das in x nach dem letzten d kommt.

Das allgemeine Zustandsaufspaltungsverfahren verallgemeinert das soweit dargestellte in zwei Richtungen: Die Kanten, die bei einem gegebenen Knoten beginnen, können in beliebig viele Teilmengen partitioniert werden statt nur in zwei, und die Aufspaltung kann simultan bei allen Knoten statt nur bei einem stattfinden. Hierbei sind auch Schleifen erlaubt.

3.36 Definition: Sei G ein Graph mit Knotenmenge \mathcal{V} und Kantenmenge \mathcal{E} . Für jeden Knoten $I \in \mathcal{V}$ sei $\mathcal{P}_I = \{\mathcal{E}_I^1, \mathcal{E}_I^2, \dots, \mathcal{E}_I^{m(I)}\}$ eine Partition von \mathcal{E}_I , wobei $m(I) \geq 1$. Sei $\mathcal{P} = \bigcup_{I \in \mathcal{V}} \mathcal{P}_I$ die resultierende Partition von \mathcal{E} . Der zugehörige Zustandsaufspaltungsgraph $G^{[\mathcal{P}]}$ hat die Knoten $I^1, I^2, \dots, I^{m(I)}$, wobei I alle Knoten in \mathcal{V} durchläuft, und die Kanten e^j , wobei e eine Kante in \mathcal{E} ist und $1 \leq j \leq m(t(e))$. Falls $e \in \mathcal{E}$ von I nach J verläuft, dann ist $e \in \mathcal{E}_I^i$ für ein i , und wir definieren den Anfangs- und Endknoten von e^j in $G^{[\mathcal{P}]}$ durch $i(e^j) := I^i$ und $t(e^j) := J^j$, d.h. e^j verläuft von I^i nach J^j . Eine elementare Zustandsaufspaltung von G bei I liegt vor, wenn $m(I) = 2$ und $m(J) = 1$ für alle $J \neq I$.

Da in der Konstruktion von $G^{[\mathcal{P}]}$ eine Partition der von den Knoten ausgehenden Kanten verwendet wird, wird $G^{[\mathcal{P}]}$ manchmal als *Out-Split-Graph* bezeichnet. Es gibt einen entsprechenden Begriff eines *In-Split-Graphen*, der eine Partition der eingehenden Knoten verwendet.

3.37 Definition: Sei G ein Graph mit Knotenmenge \mathcal{V} und Kantenmenge \mathcal{E} . Für jeden Knoten $J \in \mathcal{V}$ zerlege \mathcal{E}^J in disjunkte Mengen $\mathcal{E}_1^J, \mathcal{E}_2^J, \dots, \mathcal{E}_{m(J)}^J$ mit $m(J) \geq 1$. Sei \mathcal{P} die resultierende Partition von \mathcal{E} . Der In-Split-Graph $G_{[\mathcal{P}]}$ hat die Knoten $J_1, J_2, \dots, J_{m(J)}$, wobei J die Knoten von \mathcal{V} durchläuft, und Kanten e_i , wobei e eine beliebige Kante in \mathcal{E} ist und $1 \leq i \leq m(i(e))$. Falls $e \in \mathcal{E}$ von I nach J verläuft, dann ist $e \in \mathcal{E}_j^J$ für ein j und wir definieren den Anfangs- und Endknoten von e_i in $G_{[\mathcal{P}]}$ durch $i(e_i) := I_i$ und $t(e_i) := J_j$.

3.38 Satz: Wenn H aus G durch eine Zustandsaufspaltung entsteht, dann sind die Kantenshifts X_H und X_G konjugiert.

Beweis: Der Satz ist richtig für In-Split- und für Out-Split-Graphen. Wir führen den Beweis jedoch nur für Out-Split-Graphen. Unter Verwendung der Notation aus Definition 3.36 nehmen wir an, dass $H = G^{[\mathcal{P}]}$ für eine Partition \mathcal{P} der Kanten von G . Wir definieren eine 1-Blockabbildung $\Psi : \mathcal{B}_1(X_H) \rightarrow \mathcal{B}_1(X_G)$ durch $\Psi(e^j) := e$. Ist $e^j f^k \in \mathcal{B}_2(X_H)$, so gilt $ef \in \mathcal{B}_2(X_G)$. Deshalb bildet Ψ Pfade in H auf Pfade in G ab. Der induzierte gleitende 1-Blockcode $\psi = \Psi_\infty$ erfüllt also $\psi(X_H) \subseteq X_G$.

Nun definieren wir eine 2-Blockabbildung $\Phi : \mathcal{B}_2(X_G) \rightarrow \mathcal{B}_2(X_H)$. Gilt $fe \in \mathcal{B}_2(X_G)$, so ist e in genau einer der Mengen \mathcal{E}_j^j enthalten und wir definieren $\Phi(fe) := f^j$. Wie Ψ bildet auch Φ Pfade auf Pfade ab, denn aus $fe, eg \in \mathcal{B}_2(X_G)$ mit $e \in \mathcal{E}_j^j$ und $g \in \mathcal{E}_k^k$ folgt $\Phi(fe) = f^j$, $\Phi(eg) = e^k$ und $f^j e^k \in \mathcal{B}_2(X_H)$, da $t(f^j) = J^j = i(e^k)$. Folglich induziert Φ einen gleitenden Blockcode $\phi = \Phi_\infty^{[0,1]} : X_G \rightarrow X_H$. Ist $x = \dots e_{-1} e_0 e_1 \dots \in X_G$, so hat $\phi(x)$ die Form

$$\phi(x) = \dots e_{-1}^{j_{-1}} e_0^{j_0} e_1^{j_1} \dots,$$

so dass $\psi(\phi(x)) = x$. Ist umgekehrt $y = \dots e_{-1}^{j_{-1}} e_0^{j_0} e_1^{j_1} \dots \in X_H$, so gilt

$$\psi(y) = \dots e_{-1} e_0 e_1 \dots$$

Da $e_i^{j_i} e_{i+1}^{j_{i+1}}$ ein 2-Block in X_H ist, gilt $e_{i+1} \in \mathcal{E}_{i(e_i)}^{j_i}$. Folglich gilt $\Phi(e_i e_{i+1}) = e_i^{j_i}$ für alle $i \in \mathbb{Z}$, woraus $\phi(\psi(y)) = y$ folgt. Also sind X_G und X_H konjugiert. \square

3.39 Bemerkung: Der Beweis zeigt insbesondere Folgendes: Ist H ein Out-Split-Graph von G , so existiert ein 1-Blockcode $X_H \rightarrow X_G$; dieser wird auch als *Out-Amalgamation-Code* bezeichnet. Der zu diesem inverse 2-Blockcode $X_G \rightarrow X_H$ mit Gedächtnis 0 und Antizipation 1 wird auch als *Out-Splitting-Code* bezeichnet. Ist H stattdessen ein In-Splitting von G , so gibt es einen *In-Amalgamation-Code* mit einem 2-Blockcode mit Gedächtnis 1 und Antizipation 0 als Inversem, genannt *In-Splitting-Code*.

Aus obigem Satz folgt: Wenn ein Graph H aus einem Graphen G durch eine Folge von Splitting- und Amalgamationscodes erzeugt werden kann, dann sind X_G und X_H konjugiert. Die zugehörige Konjugation ist eine Hintereinanderschaltung von Splitting- und Amalgamationscodes. Ein wichtiges Resultat in der Theorie der symbolischen dynamischen Systeme besagt, dass umgekehrt jede Konjugation zwischen Shifts von endlichem Typ als eine solche Hintereinanderschaltung geschrieben werden kann (siehe [1, Ch. 7]).

Um zu beschreiben, wie eine Zustandsaufspaltung die Übergangsmatrix eines Graphen transformiert, führen wir folgenden Begriff ein.

3.40 Definition: Sei G ein Graph und $H = G^{[\mathcal{P}]}$ der Zustandsaufspaltungsgraph, der aus G durch die Partition \mathcal{P} entsteht. Seien $\mathcal{V} = \mathcal{V}(G)$ und $\mathcal{W} = \mathcal{V}(H)$. Die Divisionsmatrix D für \mathcal{P} ist die $\mathcal{V} \times \mathcal{W}$ -Matrix, definiert durch

$$D_{IJ^k} := \begin{cases} 1 & \text{falls } I = J, \\ 0 & \text{andernfalls.} \end{cases}$$

Die Kantenmatrix E für \mathcal{P} ist die $\mathcal{W} \times \mathcal{V}$ -Matrix, definiert durch

$$E_{I^k J} := |\mathcal{E}_I^k \cap \mathcal{E}^J|.$$

Die Divisions- und Kantenmatrizen können verwendet werden, um die Übergangsmatrizen der Graphen G und $H = \mathcal{G}^{[\mathcal{P}]}$ zu berechnen.

3.41 Satz: Sei G ein Graph und $H = G^{[\mathcal{P}]}$ der Zustandsaufspaltungsgraph, der aus G durch die Partition \mathcal{P} entsteht. Falls D die Divisions- und E die Kantenmatrix für \mathcal{P} ist, gilt

$$DE = A_G \quad \text{und} \quad ED = A_H.$$

Beweis: Es gilt

$$\begin{aligned} (DE)_{IJ} &= \sum_{i=1}^{m(I)} D_{II^i} E_{I^i J} = \sum_{i=1}^{m(I)} E_{I^i J} \\ &= \sum_{i=1}^{m(I)} |\mathcal{E}_I^i \cap \mathcal{E}^J| = \left| \left(\bigcup_{i=1}^{m(I)} \mathcal{E}_I^i \right) \cap \mathcal{E}^J \right| = |\mathcal{E}_I \cap \mathcal{E}^J| = (A_G)_{IJ}. \end{aligned}$$

Daher gilt $DE = A_G$. Die Gültigkeit der anderen Formel sieht man wie folgt:

$$\begin{aligned} (ED)_{I^i J^j} &= E_{I^i J} D_{J^j I^i} = E_{I^i J} \\ &= |\mathcal{E}_I^i \cap \mathcal{E}^J| = |\mathcal{E}_{I^i} \cap \mathcal{E}^{J^j}| = (A_H)_{I^i J^j}. \end{aligned}$$

Die vorletzte Gleichheit gilt, da bei der Zustandsaufspaltung aus jeder Kante in \mathcal{E}_I^i , die bei J endet, genau eine Kante entsteht, die bei \mathcal{E}_{I^i} beginnt und bei \mathcal{E}^{J^j} endet (für alle $j \in \{1, \dots, m(J)\}$). \square

4 Sofische Shifts

Beschriften wir die Kanten eines Graphen G mit Symbolen aus einem Alphabet \mathcal{A} , wobei mehrere Kanten dieselbe Beschriftung haben dürfen, so liefert jeder bi-unendliche Pfad auf G einen Punkt im vollen Shift $\mathcal{A}^{\mathbb{Z}}$. Die Menge aller dieser Punkte wird als *sofischer Shift* bezeichnet.

Sofische Shifts sind aus verschiedenen Gründen wichtig. Wir werden u.a. beweisen, dass es sich bei den sofischen Shifts genau um die Faktoren der Shifts von

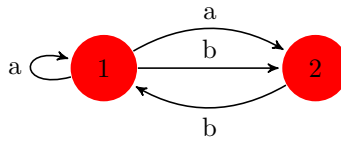


Abbildung 5: Ein beschrifteter Graph

endlichem Typ handelt. Daraus folgt, dass die Klasse der sofischen Shifts die kleinste Menge von Shifts ist, die alle Shifts von endlichem Typ enthält und unter Faktorbildung abgeschlossen ist. Sofische Shifts entsprechen den regulären Sprachen in der Automatentheorie. Sofische Shifts und Shifts von endlichem Typ sind außerdem natürliche Modelle für Informationsspeicherung und -übertragung.

4.1 Darstellung sofischer Shifts

Wir beginnen mit der Definition beschrifteter Graphen.

4.1 Definition: Ein beschrifteter Graph \mathcal{G} ist ein Paar (G, \mathcal{L}) , wobei G ein Graph mit Kantenmenge \mathcal{E} ist und die Beschriftung $\mathcal{L} : \mathcal{E} \rightarrow \mathcal{A}$ jeder Kante e ein Symbol $\mathcal{L}(e)$ aus einem endlichen Alphabet \mathcal{A} zuordnet. Der \mathcal{G} zugrundeliegende Graph ist G . Ein beschrifteter Graph heißt irreduzibel, falls sein zugrundeliegender Graph irreduzibel ist.

Die Beschriftung \mathcal{L} kann eine beliebige Zuordnung von Symbolen (die wir in diesem Zusammenhang auch *Buchstaben* nennen) aus einem Alphabet \mathcal{A} sein. Ein Extrem liegt vor, wenn $\mathcal{A} = \mathcal{E}$ und $\mathcal{L}(e) = e$ für alle $e \in \mathcal{E}$, d.h. die Kanten werden eins-zu-eins durch ihre Namen beschriftet. Das andere Extrem besteht in der Beschriftung aller Kanten durch einen einzigen Buchstaben a eines einelementigen Alphabets, so dass $\mathcal{L}(e) = a$ für alle $e \in \mathcal{E}$. Wir können ohne Einschränkung der Allgemeinheit stets davon ausgehen, dass die Beschriftung \mathcal{L} eine surjektive Abbildung ist, da die Buchstaben, die nicht im Bild von \mathcal{L} liegen, keine Rolle spielen.

So wie ein Graph G durch seine Übergangsmatrix A_G beschrieben wird, hat ein beschrifteter Graph \mathcal{G} eine *symbolische Übergangsmatrix* $A_{\mathcal{G}}$. Der Eintrag an der Stelle (I, J) von $A_{\mathcal{G}}$ enthält die formale „Summe“ der Buchstaben aller Kanten von I nach J , oder das Symbol \emptyset , falls es keine solchen Kanten gibt.

Für den Graphen in Abbildung 5 ergibt sich als symbolische Übergangsmatrix

$$A_{\mathcal{G}} = \begin{bmatrix} a & a+b \\ b & \emptyset \end{bmatrix}.$$

Es gibt einen zu Graphenhomomorphismen analogen Begriff für beschriftete Graphen, der fordert, dass die Namen der Kanten erhalten bleiben.

4.2 Definition: Seien $\mathcal{G} = (G, \mathcal{L}_G)$ und $\mathcal{H} = (H, \mathcal{L}_H)$ beschriftete Graphen, wobei \mathcal{L}_G und \mathcal{L}_H Werte in derselben Alphabet \mathcal{A} haben. Ein Homomorphismus beschrifteter Graphen von \mathcal{G} nach \mathcal{H} ist ein Graphenhomomorphismus $(\partial\Phi, \Phi) : G \rightarrow H$, so dass $\mathcal{L}_H(\Phi(e)) = \mathcal{L}_G(e)$ für alle Kanten $e \in \mathcal{E}(G)$. In diesem Fall schreiben wir $(\partial\Phi, \Phi) : \mathcal{G} \rightarrow \mathcal{H}$. Falls $\partial\Phi$ und Φ beide bijektiv sind, heißt $(\partial\Phi, \Phi)$ Isomorphismus beschrifteter Graphen. Zwei beschriftete Graphen sind isomorph, falls ein Isomorphismus (beschrifteter Graphen) zwischen ihnen existiert.

Ist $\mathcal{G} = (G, \mathcal{L})$ ein beschrifteter Graph, so kann \mathcal{L} dazu verwendet werden, Pfade und bi-unendliche Pfade auf G zu beschriften. Wir definieren die *Beschriftung* eines Pfads $\pi = e_1 e_2 \dots e_n$ als

$$\mathcal{L}(\pi) := \mathcal{L}(e_1)\mathcal{L}(e_2)\dots\mathcal{L}(e_n).$$

Dies ist ein n -Block über \mathcal{A} , den wir manchmal als *Beschriftungsblock* bezeichnen werden. Für jeden leeren Pfad ε_I in G definieren wir $\mathcal{L}(\varepsilon_I) := \varepsilon$ (der leere Block über \mathcal{A}).

Ist $\xi = \dots e_{-1}e_0e_1\dots$ ein bi-unendlicher Pfad auf G , so dass $\xi \in X_G$, so definieren wir die *Beschriftung* von ξ als

$$\mathcal{L}_\infty(\xi) := \dots\mathcal{L}(e_{-1})\mathcal{L}(e_0)\mathcal{L}(e_1)\dots \in \mathcal{A}^{\mathbb{Z}}.$$

Die Menge aller Folgen $\mathcal{L}_\infty(\xi)$ mit $\xi \in X_G$ wird bezeichnet mit

$$X_G := \{x \in \mathcal{A}^{\mathbb{Z}} : x = \mathcal{L}_\infty(\xi) \text{ für ein } \xi \in X_G\} = \mathcal{L}_\infty(X_G).$$

4.3 Definition: Eine Teilmenge X eines vollen Shifts $\mathcal{A}^{\mathbb{Z}}$ heißt *sofischer Shift*, falls ein beschrifteter Graph \mathcal{G} existiert, so dass $X = X_G$. Eine Darstellung eines sofischen Shifts X ist ein beschrifteter Graph \mathcal{G} mit $X = X_G$. Die Einschränkung der Shiftabbildung auf X_G wird mit σ_G bezeichnet.

4.4 Bemerkung: Die Bezeichnung „sofisch“ stammt von dem israelischen Mathematiker B. Weiss und ist von dem hebräischen Wort für „endlich“ abgeleitet.

Es ist wichtig zu verstehen, dass ein und derselbe sofische Shift verschiedene nicht-isomorphe Darstellungen haben kann. In Abbildung 6 sind vier Darstellungen des vollen 2-Shifts abgebildet, von denen keine zwei isomorph zueinander sind.

Ist X ein sofischer Shift, $\mathcal{G} = (G, \mathcal{L})$ eine Darstellung von X und w ein Block in $\mathcal{B}(X)$, so sagen wir, dass ein Pfad π in G eine *Darstellung* von w ist, falls $\mathcal{L}(\pi) = w$. Ein Block kann durch mehrere Pfade dargestellt werden. Zum Beispiel hat der Pfad 010001 im vierten Graphen in Abbildung 6 drei Darstellungen, eine

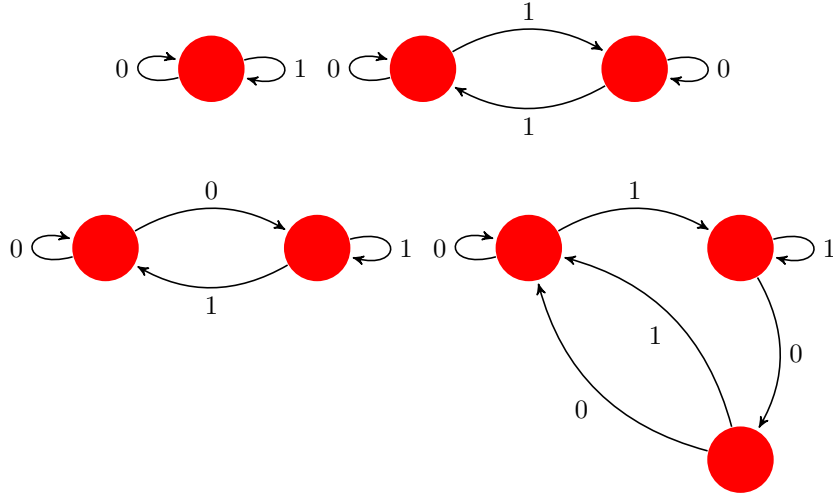


Abbildung 6: Nicht-isomorphe Darstellungen des vollen 2-Shifts

für jeden Knoten als Anfangsknoten des Pfades. Ist $x \in X_G$, so nennen wir einen bi-unendlichen Pfad π auf G eine Darstellung von x , falls $\mathcal{L}_\infty(\pi) = x$. Wie bei den (endlichen) Pfaden können mehrere Darstellungen existieren.

Nun beweisen wir, dass ein sofischer Shift tatsächlich stets ein Shift ist.

4.5 Satz: *Sofische Shifts sind Shifts.*

Beweis: Sei X ein sofischer Shift über \mathcal{A} und $\mathcal{G} = (G, \mathcal{L})$ eine Darstellung von X . Dann induziert $\mathcal{L} : \mathcal{E} \rightarrow \mathcal{A}$ einen 1-Blockcode $\mathcal{L}_\infty : X_G \rightarrow X_G$. Dessen Bild $X = \mathcal{L}_\infty(X_G)$ ist nach Satz 2.29 ein Shift. \square

Kantenshifts sind sofische Shifts, denn wir können einfach die Kantenmenge \mathcal{E} als Alphabet und die Identität $\mathcal{L} : \mathcal{E} \rightarrow \mathcal{E}$ verwenden. Das folgende Resultat zeigt, dass alle Shifts von endlichem Typ sofisch sind.

4.6 Satz: *Jeder Shift von endlichem Typ ist sofisch.*

Beweis: Sei X ein Shift von endlichem Typ. Nach Satz 3.4 ist X ein M -Schritt-Shift für ein $M \geq 0$. Im Beweis von Satz 3.23 wird ein Graph G konstruiert, so dass $X_G = X^{[M+1]}$. Dabei sind die Knoten in G gerade die erlaubten M -Blöcke in X und es gibt eine Kante e von $a_1 a_2 \dots a_M$ nach $b_1 b_2 \dots b_M$ genau dann, wenn $a_2 \dots a_M = b_1 \dots b_{M-1}$. In diesem Fall bezeichnen wir die Kante e mit $a_1 \dots a_M b_M$ und wir beschriften e durch $\mathcal{L}(e) := a_1$. Dies liefert einen beschrifteten Graphen, von dem wir zeigen werden, dass er eine Darstellung von X ist.

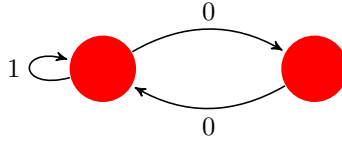


Abbildung 7: Ein beschrifteter Graph mit strikt sofishem Shift

Sei $\beta_{M+1} : X \rightarrow X^{[M+1]} = X_G$ der höhere Blockcode, den wir durch

$$\beta_{M+1}(x)_{[i]} = x_{[i, i+M]}$$

definiert hatten. Da $\mathcal{L}(x_{[i, i+M]}) = x_i$, gilt $\mathcal{L}_\infty(\beta_{M+1}(x)) = x$ für alle $x \in X$, woraus $X \subseteq X_G$ folgt. Umgekehrt hat jeder Punkt $\xi \in X_G = X^{[M+1]}$ die Form $\xi = \beta_{M+1}(x)$ für ein $x \in X$, so dass $\mathcal{L}_\infty(\xi) = \mathcal{L}_\infty(\beta_{M+1}(x)) = x \in X$. Folglich gilt $X_G = \mathcal{L}_\infty(X_G) \subseteq X$, also $X = X_G$. \square

Der 1-Blockcode in obigem Beweis ist also die Inverse der Konjugation β_{M+1} .

4.7 Beispiel: Nicht alle sofishen Shifts sind von endlichem Typ. Dazu betrachten wir den beschrifteten Graphen in Abbildung 7. Dies ist der Shift $X = X_{\mathcal{F}}$ mit $\mathcal{F} = \{10^{2n+1}1 : n \geq 0\}$, von dem wir gezeigt haben, dass er nicht von endlichem Typ ist (siehe Beispiel 3.2).

4.8 Definition: *Ein sofisher Shift, der kein Shift von endlichem Typ ist, heißt strikt sofish.*

Der folgende Satz charakterisiert diejenigen sofishen Shifts, die von endlichem Typ sind.

4.9 Satz: *Ein sofisher Shift X ist ein Shift von endlichem Typ genau dann, wenn es eine Darstellung $\mathcal{G} = (G, \mathcal{L})$ von X gibt, so dass \mathcal{L}_∞ eine Konjugation ist.*

Beweis: Ist X ein sofisher Shift, der durch $\mathcal{G} = (G, \mathcal{L})$ dargestellt wird, so dass \mathcal{L}_∞ eine Konjugation ist, dann ist X durch \mathcal{L}_∞ zum Kantenshift X_G konjugiert, der ein Shift von endlichem Typ ist. Nach Satz 3.7 ist dann X selbst ein Shift von endlichem Typ.

Sei nun umgekehrt X ein Shift von endlichem Typ. Dann betrachten wir den im Beweis von Satz 4.6 konstruierten beschrifteten Graphen, für den \mathcal{L}_∞ eine Konjugation ist. \square

Das folgende Beispiel zeigt, dass nicht jeder Shift sofish ist. Tatsächlich gibt es nur abzählbar unendlich viele sofishen Shifts über einem Alphabet \mathcal{A} (Übungsaufgabe auf Blatt 5).

4.10 Beispiel: Sei $X \subset \{a, b, c\}^{\mathbb{Z}}$ der Shift, der aus denjenigen Punkten $x \in \{a, b, c\}^{\mathbb{Z}}$ besteht, in denen ein Block der Form $ab^m c^n a$ nur dann vorkommt, wenn $m = n$. Wir nehmen an, X wäre sofish. Sei $\mathcal{G} = (G, \mathcal{L})$ eine Darstellung von X . Sei r die Anzahl der Knoten in G . Da $w = ab^{r+1}c^{r+1}a$ erlaubt ist, gibt es einen Pfad π in G , der w darstellt. Sei τ der Teilpfad von π , der b^{r+1} darstellt. Da G nur r Knoten hat, müssen mindestens 2 Knoten in τ identisch sein. Daher können wir $\tau = \tau_1 \tau_2 \tau_3$ schreiben, wobei τ_2 ein geschlossener Pfad ist (falls der erste oder der letzte Knoten in τ einer der wiederholten Knoten ist, dann ist τ_1 oder τ_3 der leere Pfad). Dann wäre auch $\tau' = \tau_1 \tau_2 \tau_2 \tau_3$ ein Pfad. Indem wir τ durch τ' in π ersetzen, erhalten wir einen Pfad π' in G mit $\mathcal{L}(\pi') = ab^{r+1+s}c^{r+1}a$, wobei s die Länge von τ_2 ist. Aber dieser Block ist nicht erlaubt. Darum kann X nicht sofish sein.

Die folgende Grafik zeigt, wie die bisher eingeführten Klassen von Shifts ineinander enthalten sind.

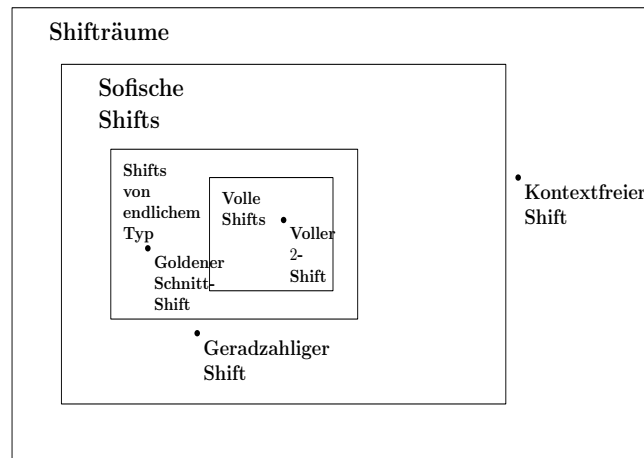


Abbildung 8: Verschiedene Klassen von Shifts

Dabei geben wir den in den Beispielen 2.5(2), 2.5(3) und 4.10 eingeführten Shifts Namen:

- Goldener-Schnitt-Shift: $X = X_{\mathcal{F}}$ über $\mathcal{A} = \{0, 1\}$ mit $\mathcal{F} = \{11\}$.⁷
- Geradzahliges Shift: $X = X_{\mathcal{F}}$ über $\mathcal{A} = \{0, 1\}$ mit $\mathcal{F} = \{10^{2n+1}1 : n \geq 0\}$.
- Kontextfreier Shift: $X = X_{\mathcal{F}}$ über $\mathcal{A} = \{a, b, c\}$ mit $\mathcal{F} = \{ab^n c^m a : n \neq m\}$.

⁷Eine Begründung dieses Namens folgt in Beispiel 5.4.

4.2 Charakterisierungen sofischer Shifts

In diesem Unterabschnitt werden wir zwei verschiedene Charakterisierungen der sofischen Shifts geben. Die erste liefert der folgende Satz.

4.11 Satz: *Ein Shift X ist sofisch genau dann, wenn er ein Faktor eines Shifts von endlichem Typ ist.*

Beweis: Wir nehmen zunächst an, dass X sofisch ist und betrachten eine Darstellung $\mathcal{G} = (G, \mathcal{L})$ von X . Die 1-Blockabbildung \mathcal{L} induziert einen gleitenden Blockcode $\mathcal{L}_\infty : X_G \rightarrow X_G$, der nach Definition surjektiv ist. Daher ist $X = X_G$ ein Faktor eines Kantenshifts, der nach Satz 3.14 von endlichem Typ ist.

Umgekehrt nehmen wir nun an, dass X ein Shift ist, für den es einen Shift Y von endlichem Typ und einen Faktorcode $\phi : Y \rightarrow X$ gibt. ϕ habe Gedächtnis m und Antizipation n . Dann wird ϕ von einer Blockabbildung Φ auf $\mathcal{B}_{n+m+1}(Y)$ induziert. Wenn wir bei Bedarf m durch eine größere Zahl ersetzen, können wir annehmen, dass Y ein $(m+n)$ -Schritt-Shift ist.

Definiere $\psi : Y \rightarrow Y^{[m+n+1]}$ durch

$$\psi(y)_{[i]} := y_{[i-m, i+n]}.$$

(ψ ist fast dasselbe wie die höhere Blockabbildung β_{m+n+1} , außer dass wir die Koordinaten von y vom Index $i-m$ statt i beginnend verwenden.) Da Y ein $(m+n)$ -Schritt-Shift ist, gibt es nach Satz 3.23 einen Graphen G mit Kanten, die mit Blöcken aus $\mathcal{B}_{m+n+1}(Y)$ beschriftet sind, so dass $Y^{[m+n+1]} = X_G$. Wir definieren die Beschriftung \mathcal{L} auf G durch $\mathcal{L}(e) := \Phi(e)$. Wir werden zeigen, dass $\mathcal{G} = (G, \mathcal{L})$ eine Darstellung von X ist und damit beweisen, dass X sofisch ist. Dazu stellen wir zunächst fest, dass das folgende Diagramm kommutiert:

$$\begin{array}{ccc} Y & \xrightarrow[\cong]{\psi} & Y^{[m+n+1]} = X_G \\ \phi \downarrow & \swarrow \mathcal{L}_\infty & \\ X & & \end{array}$$

Ist nämlich $y \in Y$, dann gilt $\phi(y)_{[i]} = \Phi(y_{[i-m, i+n]})$ und

$$\mathcal{L}_\infty(\psi(y))_{[i]} = \mathcal{L}(\psi(y)_{[i]}) = \Phi(y_{[i-m, i+n]}).$$

Da ψ eine Konjugation ist, sind die Bilder von ϕ und \mathcal{L}_∞ identisch. Folglich gilt $X = \phi(Y) = \mathcal{L}_\infty(X_G)$, so dass \mathcal{G} eine Darstellung von X ist. \square

4.12 Korollar: *Ein Faktor eines sofischen Shifts ist sofisch. Folglich ist die Klasse aller sofischen Shifts abgeschlossen unter Faktorbildung.*

Beweis: Sei $\phi : Y \rightarrow X$ ein Faktorcode und sei Y sofisch. Nach dem obigen Satz gibt es einen Faktorcode $\psi : Z \rightarrow Y$ mit einem Shift Z von endlichem Typ.

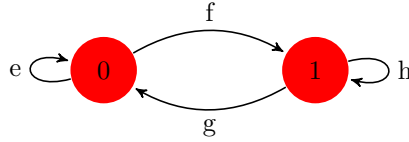


Abbildung 9: Graph für Nachfolgemengen-Beispiel

Da $\phi \circ \psi : Z \rightarrow X$ ein Faktorcode ist (siehe Lösung von Übungsaufgabe 3 auf Blatt 2), folgern wir, dass X sofisch ist. \square

Das nächste Korollar folgt unmittelbar.

4.13 Korollar: *Ist X konjugiert zu einem sofischen Shift, so ist X sofisch.*

Als nächstes leiten wir ein Sprachen-basiertes Kriterium für sofische Shifts her. Dazu zunächst folgende Definition:

4.14 Definition: *Sei X ein Shift und w ein Block in $\mathcal{B}(X)$. Dann definieren wir die Nachfolgemenge von w als*

$$F_X(w) := \{v \in \mathcal{B}(X) : wv \in \mathcal{B}(X)\}.$$

Ferner definieren wir die Menge aller Nachfolgemengen

$$\mathcal{C}_X := \{F_X(w) : w \in \mathcal{B}(X)\}.$$

4.15 Bemerkung: Die Mengen $F_X(w)$ sind in der Regel unendlich, während die Menge \mathcal{C}_X oft endlich ist.

4.16 Beispiel: Sei G der Graph in Abbildung 9 und $X = X_G$ mit Alphabet $\{e, f, g, h\}$. Ist π ein Pfad auf G , dann ist die Nachfolgemenge von π die Menge aller Pfade, die bei $t(\pi)$ beginnen. Sei C_0 die Menge der bei 0 beginnenden und C_1 die Menge der bei 1 beginnenden Pfade. Dann gilt $\mathcal{C}_X = \{C_0, C_1\}$, z.B.

$$\begin{aligned} F_X(e) &= F_X(fg) = F_X(hhgee) = C_0, \\ F_X(f) &= F_X(gef h) = F_X(h^{100}) = C_1. \end{aligned}$$

Dieselbe Überlegung zeigt: Ist G ein wesentlicher Graph mit r Knoten, dann enthält \mathcal{C}_{X_G} genau r Mengen, eine für jeden Knoten.

4.17 Beispiel: Sei $\mathcal{A} = \{a, b, c\}$ und X der kontextfreie Shift, der $ab^m c^k a$ genau dann enthält, wenn $m = k$. Dann ist $c^k a \in F_X(ab^m)$ genau dann, wenn $k = m$. Folglich sind für $m = 0, 1, 2, \dots$ die Nachfolgemengen $F_X(c^k a)$ alle verschieden voneinander und \mathcal{C}_X ist unendlich.

Sei X ein Shift über \mathcal{A} , so dass \mathcal{C}_X endlich ist. Dann konstruieren wir einen beschrifteten Graphen $\mathcal{G} = (G, \mathcal{L})$ wie folgt: Die Knotenmenge ist $\mathcal{V} = \mathcal{C}_X$. Sei $C = F_X(w) \in \mathcal{V}$ und $a \in \mathcal{A}$. Ist $wa \in \mathcal{B}(X)$, so ist $F_X(wa) \in \mathcal{V}$, also $F_X(wa) = C' \in \mathcal{V}$. Zeichne in diesem Fall eine Kante von C nach C' mit der Beschriftung a . Ist $wa \notin \mathcal{B}(X)$, zeichne keine Kante. Der Graph \mathcal{G} entsteht, indem wir für alle $C \in \mathcal{C}_X$ und $a \in \mathcal{A}$ so vorgehen.

Man macht sich leicht klar, dass diese Konstruktion unabhängig von den gewählten Repräsentanten w der Mengen $C \in \mathcal{C}_X$ ist.

4.18 Definition: Sei X ein Shift über \mathcal{A} , so dass \mathcal{C}_X endlich ist. Dann heißt der wie oben definierte beschriftete Graph $\mathcal{G} = (G, \mathcal{L})$ der Nachfolgermengengraph von X .

4.19 Satz: Ist X ein Shift, so dass \mathcal{C}_X endlich ist, dann ist \mathcal{G} eine Darstellung von X . Insbesondere ist X dann sofisch.

Beweis: Um zu zeigen, dass $X = X_{\mathcal{G}}$, reicht es nach Satz 2.9 zu beweisen, dass $\mathcal{B}(X) = \mathcal{B}(X_{\mathcal{G}})$.

Zuerst nehmen wir an, dass $u = a_1a_2 \dots a_n \in \mathcal{B}(X_{\mathcal{G}})$. Dann gibt es einen Pfad π im zugrundeliegenden Graphen G mit $\mathcal{L}(\pi) = u$. Der Pfad π beginne beim Knoten $F_X(w)$. Die Definition von \mathcal{G} zeigt, dass $wa_1 \in \mathcal{B}(X)$ und dann, dass $wa_1a_2 \in \mathcal{B}(X)$, usw., so dass $wa_1a_2 \dots a_n = wu \in \mathcal{B}(X)$. Folglich gilt $u \in \mathcal{B}(X)$ und damit $\mathcal{B}(X_{\mathcal{G}}) \subseteq \mathcal{B}(X)$.

Um die umgekehrte Inklusion zu zeigen, sei $u \in \mathcal{B}(X)$. Indem wir iterativ Satz 2.9(1b) anwenden, sehen wir, dass es einen Block w gibt mit $wu \in \mathcal{B}(X)$ und $|w| > |\mathcal{V}(G)|$. Nach Definition von \mathcal{G} gibt es einen Pfad in G mit der Beschriftung wu . Ein solcher Pfad kann als $\alpha\beta\gamma\pi$ geschrieben werden, wobei β ein geschlossener Pfad ist und π durch u beschriftet ist. Da β beliebig oft wiederholt werden kann (siehe auch Beispiel 4.10) und bei jedem Knoten von G mindestens eine Kante beginnt, folgt, dass π zu einem bi-unendlichen Pfad in G erweitert werden kann und folglich u zu $\mathcal{B}(X_{\mathcal{G}})$ gehört. \square

4.20 Satz: Ein Shift X ist genau dann sofisch, wenn \mathcal{C}_X endlich ist.

Beweis: Eine Beweisrichtung haben wir mit dem vorherigen Satz abgehandelt. Um die andere zu beweisen, zeigen wir, dass \mathcal{C}_X endlich ist, falls X sofisch ist. Sei dazu $\mathcal{G} = (G, \mathcal{L})$ eine Darstellung von X . Für einen Block $w \in \mathcal{B}(X)$ werden wir $F_X(w)$ mit Hilfe des beschrifteten Graphen \mathcal{G} beschreiben. Betrachte dazu alle Pfade in G , die w darstellen und bezeichne mit T die Menge der Knoten, bei denen diese Pfade enden. Dann ist $F_X(w)$ die Menge der Beschriftungen von Pfaden in G , die bei einem Knoten in T beginnen. Daher haben zwei Blöcke mit demselben T dieselbe Nachfolgermenge. Da es nur endlich viele Teilmengen T der Knotenmenge von G gibt, ist \mathcal{C}_X endlich. \square

4.3 Minimale rechtsauflösende Darstellungen

Eine rechtsauflösende Darstellung eines sofischen Shifts ist eine Darstellung, in der alle Kanten, die bei demselben Knoten beginnen, verschiedene Beschriftungen haben. Für einen irreduziblen sofischen Shift gibt es bis auf Isomorphie genau eine rechtsauflösende Darstellung mit einer minimalen Anzahl von Knoten. Auf diese Art können wir einem solchen Shift eine Normalform zuordnen.

4.21 Definition: Eine rechtsauflösende Darstellung eines sofischen Shifts X ist eine Darstellung $\mathcal{G} = (G, \mathcal{L})$, so dass für jeden Knoten I von G die Einschränkung von \mathcal{L} auf \mathcal{E}_I injektiv ist. Das heißt, verschiedene Kanten, die bei I beginnen, haben verschiedene Namen.

Der Nachfolgermengengraph aus dem vorherigen Abschnitt ist nach Definition eine rechtsauflösende Darstellung. Allerdings ist es schwierig, die Nachfolgermengen zu bestimmen. Im Beweis des folgenden Satzes präsentieren wir eine andere Konstruktion einer rechtsauflösenden Darstellung, die sogenannte *Teilmengekongstruktion*.

4.22 Satz: Jeder sofische Shift hat eine rechtsauflösende Darstellung.

Beweis: Sei X ein sofischer Shift über einem Alphabet \mathcal{A} . Dann hat X eine Darstellung $\mathcal{G} = (G, \mathcal{L})$, so dass $X = X_{\mathcal{G}}$.

Wir konstruieren einen neuen beschrifteten Graphen $\mathcal{H} = (H, \mathcal{L}')$ wie folgt. Die Knoten \mathcal{I} von \mathcal{H} sind die nichtleeren Teilmengen der Knotenmenge $\mathcal{V}(G)$. Ist $\mathcal{I} \in \mathcal{V}(H)$ und $a \in \mathcal{A}$, dann bezeichnen wir mit \mathcal{J} die Menge aller Knoten, bei denen Kanten enden, die in \mathcal{I} beginnen und die Beschriftung a tragen. Mit anderen Worten: \mathcal{J} ist die Menge aller von \mathcal{I} aus durch Kanten mit der Beschriftung a erreichbaren Knoten. Ist \mathcal{J} nichtleer, dann ist $\mathcal{J} \in \mathcal{V}(H)$ und wir zeichnen eine Kante in H von \mathcal{I} nach \mathcal{J} mit der Beschriftung a . Ist \mathcal{J} leer, so tun wir nichts. Indem wir so für alle $\mathcal{I} \in \mathcal{V}(H)$ und $a \in \mathcal{A}$ vorgehen, erzeugen wir den beschrifteten Graphen \mathcal{H} . Wir stellen fest, dass es für jeden Knoten \mathcal{I} in \mathcal{H} höchstens eine Kante mit einer gegebenen Beschriftung gibt, die bei \mathcal{I} beginnt. Also ist \mathcal{H} rechtsauflösend.

Nun beweisen wir, dass $X = X_{\mathcal{G}} = X_{\mathcal{H}}$. Dazu reicht es zu zeigen, dass $\mathcal{B}(X_{\mathcal{G}}) = \mathcal{B}(X_{\mathcal{H}})$. Zuerst nehmen wir an, dass $w = a_1 a_2 \dots a_n \in \mathcal{B}(X_{\mathcal{G}})$. Sei $\pi = e_1 e_2 \dots e_n$ ein Pfad in G , der w darstellt. Dann beginnt π bei einem Knoten I in G . Sei $\mathcal{I}_0 = \{I\}$ und für $1 \leq k \leq n$ sei \mathcal{I}_k die Menge der Knoten, bei denen Pfade der Länge k enden, die bei I beginnen und die Beschriftung $a_1 a_2 \dots a_k$ tragen. Dann gilt $t(e_1 e_2 \dots e_k) \in \mathcal{I}_k$, also ist \mathcal{I}_k nichtleer und es gibt eine Kante in \mathcal{H} mit der Beschriftung a_k von \mathcal{I}_{k-1} nach \mathcal{I}_k . Dies zeigt, dass es einen Pfad in \mathcal{H} mit der Beschriftung w gibt, so dass $w \in \mathcal{B}(X_{\mathcal{H}})$, und damit $\mathcal{B}(X_{\mathcal{G}}) \subseteq \mathcal{B}(X_{\mathcal{H}})$.

Um die umgekehrte Inklusion zu zeigen, nehmen wir an, dass $u \in \mathcal{B}(X_{\mathcal{H}})$. Sei τ ein Pfad in H , der u darstellt, bei \mathcal{I} beginnt und bei \mathcal{J} endet. Die Konstruktion von \mathcal{H} zeigt, dass \mathcal{J} die Menge aller Knoten in G ist, die von einem Knoten in

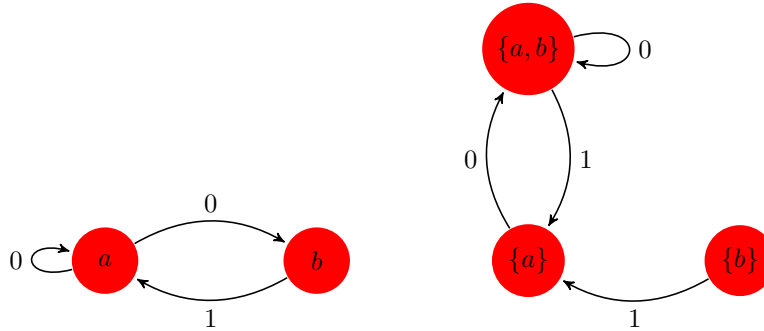


Abbildung 10: Ein Graph und sein Teilmengengraph

\mathcal{I} erreichbar sind mit einem Pfad der Beschriftung u . Da $\mathcal{J} \neq \emptyset$, zeigt dies, dass es einen Pfad in G gibt mit der Beschriftung u , so dass $u \in \mathcal{B}(X_G)$. Folglich gilt $\mathcal{B}(X_{\mathcal{H}}) \subseteq \mathcal{B}(X_G)$. \square

4.23 Beispiel: Sei X der sofische Shift, der von dem linken beschrifteten Graphen \mathcal{G} in Abbildung 10 dargestellt wird. Die Teilmengenkonstruktion liefert den daneben abgebildeten beschrifteten Graphen \mathcal{H} . Dieser ist nicht irreduzibel, obwohl \mathcal{G} irreduzibel ist. Entfernen wir den Knoten $\{b\}$ und die zugehörige Kante, so erhalten wir einen irreduziblen Teilgraphen, der ebenfalls eine Darstellung von X ist.

Die Konstruktion im Beweis von Satz 4.22 liefert einen rechtsauflösenden beschrifteten Graphen \mathcal{H} mit $2^r - 1$ Knoten, falls der Graph \mathcal{G} r Knoten hat. Der Preis, den man für eine solche rechtsauflösende Darstellung bezahlen muss, ist ein gewaltiger Anstieg in der Anzahl der Knoten. Glücklicherweise gibt es häufig einen Teilgraphen von \mathcal{H} , der denselben sofischen Shift darstellt, wie in obigem Beispiel.

4.24 Definition: Eine minimale rechtsauflösende Darstellung eines sofischen Shifts X ist eine rechtsauflösende Darstellung mit einer minimalen Anzahl von Knoten unter allen rechtsauflösenden Darstellungen.

4.25 Beispiel: Ist X der volle 2-Shift, dann ist der erste beschriftete Graph in Abbildung 6 eine minimale rechtsauflösende Darstellung, da er nur einen Knoten hat. Der zweite sowie der vierte rechtsauflösende Graph in derselben Abbildung ist folglich keine minimale rechtsauflösende Darstellung von X . Ist $\mathcal{G} = (G, \mathcal{L})$ eine beliebige minimale rechtsauflösende Darstellung von X , dann darf G nur einen Knoten haben, und folglich ist jede Kante eine Schleife. Außerdem muss \mathcal{G} eine mit 0 beschriftete und eine mit 1 beschriftete Schleife haben. Da \mathcal{G} rechtsauflösend ist, kann es keine weiteren Schleifen geben, so dass, bis auf den Namen des Knotens, \mathcal{G} der erste beschriftete Graph in Abbildung 6 ist. Daher sind zwei

beliebige minimale rechtsauflösende Darstellungen des vollen 2-Shifts isomorph. Wir werden zeigen, dass dies für jeden sofischen Shift so ist.

Ein wichtiges Merkmal minimaler rechtsauflösender Darstellungen ist, dass es für jedes Paar von verschiedenen Knoten einen Beschriftungsblock w gibt, so dass es nur für einen der beiden Knoten einen Pfad mit der Beschriftung w gibt, der bei diesem Knoten beginnt.

4.26 Definition: Sei $\mathcal{G} = (G, \mathcal{L})$ ein beschrifteter Graph und $I \in \mathcal{V}$ ein Knoten von \mathcal{G} . Die Nachfolgermenge $F_{\mathcal{G}}(I)$ von I in \mathcal{G} ist die Menge aller Beschriftungen von Pfaden, die bei I beginnen, also

$$F_{\mathcal{G}}(I) := \{\mathcal{L}(\pi) : \pi \in \mathcal{B}(X_{\mathcal{G}}) \text{ und } i(\pi) = I\}.$$

Wir sagen, dass \mathcal{G} Nachfolger-separiert ist, falls verschiedene Knoten verschiedene Nachfolgermengen haben, d.h. $F_{\mathcal{G}}(I) \neq F_{\mathcal{G}}(J)$, falls $I \neq J$.

Sei \mathcal{G} ein beschrifteter Graph. Zwei Knoten I und J heißen *äquivalent*, falls sie dieselbe Nachfolgermenge haben, d.h. $F_{\mathcal{G}}(I) = F_{\mathcal{G}}(J)$. Diese Relation partitioniert die Knotenmenge von \mathcal{G} in disjunkte Äquivalenzklassen, die wir mit $\mathcal{I}_1, \mathcal{I}_2, \dots, \mathcal{I}_r$ bezeichnen. Definiere einen beschrifteten Graphen \mathcal{H} wie folgt. Die Knotenmenge ist $\{\mathcal{I}_1, \dots, \mathcal{I}_r\}$. Es gibt eine Kante mit der Beschriftung a in \mathcal{H} von \mathcal{I} nach \mathcal{J} genau dann, wenn es Knoten $I \in \mathcal{I}$ und $J \in \mathcal{J}$ und eine Kante von I nach J mit der Beschriftung a im Graphen \mathcal{G} gibt. Wir nennen \mathcal{H} den *aus \mathcal{G} fusionierten Graphen*.

4.27 Lemma: Sei \mathcal{G} ein beschrifteter Graph und \mathcal{H} der aus \mathcal{G} fusionierte Graph. Dann ist \mathcal{H} Nachfolger-separiert und $X_{\mathcal{H}} = X_{\mathcal{G}}$. Falls \mathcal{G} irreduzibel/rechtsauflösend ist, dann ist auch \mathcal{H} irreduzibel/rechtsauflösend.

Beweis: Der Knoten I in \mathcal{G} habe die Äquivalenzklasse \mathcal{I} , ein Knoten in \mathcal{H} . Wir zeigen zuerst, dass $F_{\mathcal{G}}(I) = F_{\mathcal{H}}(\mathcal{I})$. Die Definition von \mathcal{H} zeigt, dass ein Pfad in \mathcal{G} , der bei I beginnt, einen Pfad in \mathcal{H} mit derselben Beschriftung erzeugt, der bei \mathcal{I} beginnt und die Äquivalenzklassen von Knoten auf dem ursprünglichen Pfad durchläuft. Daher gilt $F_{\mathcal{G}}(I) \subseteq F_{\mathcal{H}}(\mathcal{I})$. Falls $F_{\mathcal{G}}(I) \neq F_{\mathcal{H}}(\mathcal{I})$ für ein I , sei $w = a_1 a_2 \dots a_m$ der kürzeste Block in $F_{\mathcal{H}}(\mathcal{I}) \setminus F_{\mathcal{G}}(I)$ unter allen I in \mathcal{G} . Es gibt eine Kante mit der Beschriftung a_1 von \mathcal{I} zu irgendeinem \mathcal{J} , und daher auch ein $I' \in \mathcal{I}$, ein $J \in \mathcal{J}$ und eine Kante in \mathcal{G} mit der Beschriftung a_1 von I' nach J . Aufgrund der Minimalität von m gilt $a_2 \dots a_m \in F_{\mathcal{H}}(\mathcal{J}) \cap \mathcal{A}^{m-1} = F_{\mathcal{G}}(J) \cap \mathcal{A}^{m-1}$. Aber dies liefert einen Pfad in \mathcal{G} , der bei I' beginnt mit der Beschriftung w im Widerspruch zu $w \notin F_{\mathcal{G}}(I') = F_{\mathcal{G}}(I)$. Daher ist $F_{\mathcal{H}}(\mathcal{I}) = F_{\mathcal{G}}(I)$.

Nun ist klar, dass \mathcal{H} Nachfolger-separiert ist, da die Nachfolgermengen in \mathcal{G} für nicht-äquivalente Knoten verschieden sind. Da $\mathcal{B}(X_{\mathcal{G}})$ die Vereinigung der Nachfolgermengen der Knoten in \mathcal{G} ist, und ähnlich für \mathcal{H} , erhalten wir unmittelbar $\mathcal{B}(X_{\mathcal{G}}) = \mathcal{B}(X_{\mathcal{H}})$, so dass $X_{\mathcal{G}} = X_{\mathcal{H}}$.

Als nächstes nehmen wir an, dass \mathcal{G} irreduzibel ist. Seien \mathcal{I}, \mathcal{J} verschiedene Knoten des fusionierten Graphen \mathcal{H} und seien $I \in \mathcal{I}$ und $J \in \mathcal{J}$. Da \mathcal{G} irreduzibel ist, gibt es einen Pfad in \mathcal{G} von I nach J . Indem wir zu den Äquivalenzklassen übergehen, erhalten wir einen Pfad in \mathcal{H} von \mathcal{I} nach \mathcal{J} und dies zeigt, dass \mathcal{H} auch irreduzibel ist.

Schließlich nehmen wir an, dass \mathcal{G} rechtsauflösend ist. Betrachte den Fall einer Kante in \mathcal{H} mit Beschriftung a von \mathcal{I} nach \mathcal{J} . Dann gibt es eine Kante in \mathcal{G} mit Beschriftung a von einem $I \in \mathcal{I}$ zu einem $J \in \mathcal{J}$. Da \mathcal{G} rechtsauflösend ist, ist diese Kante die einzige Kante mit Beschriftung a , die bei I beginnt. Folglich kann man $F_{\mathcal{G}}(J)$ aus $F_{\mathcal{G}}(I)$ erhalten, indem man das erste Symbol aus allen Blöcken in $F_{\mathcal{G}}(I)$ entfernt, die mit a beginnen. Daher bestimmt $F_{\mathcal{G}}(I)$ zusammen mit a die Menge $F_{\mathcal{G}}(J)$. Da $F_{\mathcal{H}}(\mathcal{I}) = F_{\mathcal{G}}(I)$ und $F_{\mathcal{H}}(\mathcal{J}) = F_{\mathcal{G}}(J)$, sehen wir, dass $F_{\mathcal{H}}(\mathcal{I})$ zusammen mit a die Menge $F_{\mathcal{H}}(\mathcal{J})$ bestimmt. Dies bedeutet, dass es höchstens eine Kante mit Beschriftung a in \mathcal{H} gibt, die bei \mathcal{I} beginnt. Also ist \mathcal{H} rechtsauflösend. \square

Die Verschmelzung von Knoten erlaubt es uns, die Haupteigenschaft minimaler rechtsauflösender Darstellungen abzuleiten.

4.28 Satz: *Eine minimale rechtsauflösende Darstellung eines sofischen Shifts ist Nachfolger-separiert.*

Beweis: Sei \mathcal{G} eine minimale rechtsauflösende Darstellung eines sofischen Shifts X . Ist \mathcal{G} nicht Nachfolger-separiert, dann hat der aus \mathcal{G} fusionierte Graph \mathcal{H} weniger Knoten. Nach Lemma 4.27 ist \mathcal{H} eine rechtsauflösende Darstellung von X , was der Minimalität von \mathcal{G} widerspricht. \square

Als nächstes betrachten wir den Zusammenhang zwischen der Irreduzibilität eines sofischen Shifts und der Irreduzibilität seiner Darstellungen. Zunächst stellen wir fest, dass der sofische Shift $X_{\mathcal{G}}$ irreduzibel ist, falls $\mathcal{G} = (G, \mathcal{L})$ irreduzibel ist. Sind nämlich $u, v \in \mathcal{B}(X_{\mathcal{G}})$, dann gibt es Pfade π und τ in G mit $\mathcal{L}(\pi) = u$ und $\mathcal{L}(\tau) = v$. Da G irreduzibel ist, gibt es einen Pfad ω in G von $t(\pi)$ nach $i(\tau)$, so dass $\pi\omega\tau$ ein Pfad in G ist. Sei $w = \mathcal{L}(\omega)$. Dann gilt

$$\mathcal{L}(\pi\omega\tau) = \mathcal{L}(\pi)\mathcal{L}(\omega)\mathcal{L}(\tau) = uwv \in \mathcal{B}(X_{\mathcal{G}}),$$

was die Irreduzibilität von $X_{\mathcal{G}}$ beweist. Also ist ein sofischer Shift, der durch einen irreduziblen Graphen dargestellt wird, irreduzibel.

Die Umkehrung ist jedoch nicht wahr. Sei zum Beispiel X irreduzibel und dargestellt durch \mathcal{G} . Sei \mathcal{H} die Vereinigung zweier disjunkter Kopien von \mathcal{G} . Dann ist \mathcal{H} irreduzibel und ebenfalls eine Darstellung von X . Für minimale rechtsauflösende Darstellungen jedoch gilt die Umkehrung.

4.29 Lemma: *Sei X ein irreduzibler sofischer Shift und sei \mathcal{G} eine minimale rechtsauflösende Darstellung von X . Dann ist \mathcal{G} ein irreduzibler Graph.*

Beweis: Sei $\mathcal{G} = (G, \mathcal{L})$ eine minimale rechtsauflösende Darstellung von X . Wir zeigen zuerst, dass für jeden Knoten $I \in \mathcal{V}(G)$ ein Block $u_I \in \mathcal{B}(X_{\mathcal{G}})$ existiert, so

dass jeder Pfad in G , der u_I darstellt, I enthält. Wir nehmen an, dies sei für ein I nicht der Fall. Entfernen wir dann I und alle Kanten, die I enthalten, so entsteht ein rechtsauflösender Graph \mathcal{H} mit weniger Knoten als \mathcal{G} und $\mathcal{B}(X_{\mathcal{H}}) = \mathcal{B}(X_{\mathcal{G}})$, so dass $X_{\mathcal{H}} = X_{\mathcal{G}}$, im Widerspruch zur Minimalität von \mathcal{G} . Jetzt seien I und J verschiedene Knoten in G und u_I, u_J seien die oben beschriebenen Blöcke. Da X irreduzibel ist, gibt es einen Block w mit $u_I w u_J \in \mathcal{B}(X_{\mathcal{G}})$. Sei π ein Pfad, der $u_I w u_J$ darstellt. Dann ist $\pi = \tau_I \omega \tau_J$ mit $\mathcal{L}(\tau_I) = u_I$, $\mathcal{L}(\omega) = w$ und $\mathcal{L}(\tau_J) = u_J$. Aber τ_I enthält I und τ_J enthält J , so dass es einen Teilpfad von π gibt, der von I nach J verläuft. Dies zeigt, dass \mathcal{G} irreduzibel ist. \square

4.30 Satz: *Ein sofischer Shift ist irreduzibel genau dann, wenn er eine irreduzible Darstellung hat.*

Beweis: Die Überlegungen vor Lemma 4.29 zeigen, dass ein sofischer Shift irreduzibel ist, falls er eine irreduzible Darstellung hat. Das Lemma selbst zeigt, dass eine minimale rechtsauflösende Darstellung (die es aufgrund von Satz 4.22 gibt) irreduzibel ist. \square

Ist $\mathcal{G} = (G, \mathcal{L})$ eine Darstellung eines sofischen Shifts X , so kann es für einen Block $w \in \mathcal{B}(X)$ verschiedene Pfade in \mathcal{G} geben, die w darstellen. Diese Uneindeutigkeit ist bei rechtsauflösenden Darstellungen reduziert, da es bei diesen für jeden Knoten nur einen Pfad geben kann, der bei diesem Knoten beginnt und einen gegebenen Block darstellt. Manchmal gibt es Blöcke w , so dass alle w darstellenden Pfade bei demselben Knoten enden.

4.31 Definition: *Sei $\mathcal{G} = (G, \mathcal{L})$ ein beschrifteter Graph. Ein Block $w \in \mathcal{B}(X_{\mathcal{G}})$ ist ein synchronisierender Block für \mathcal{G} , falls alle Pfade in G , die w darstellen, bei demselben Knoten enden. Ist I dieser Knoten, so sagen wir, dass w auf I ausgerichtet ist.*

4.32 Beispiel: Wir betrachten nochmals die vier Darstellungen des vollen 2-Shifts in Abbildung 6. Im ersten Graphen ist jeder Block synchronisierend, da es nur einen Knoten gibt. Im zweiten und dritten Graphen ist jeweils kein Block synchronisierend. Im vierten Graphen sind 00 und 11 synchronisierend. Daher ist auch jeder Block, der 00 oder 11 enthält, synchronisierend, da der Graph rechtsauflösend ist. Blöcke, in denen 0 und 1 alternierend auftauchen, wie z.B. 0101010, sind nicht synchronisierend.

4.33 Lemma: *Sei \mathcal{G} ein rechtsauflösender beschrifteter Graph und w ein synchronisierender Block für \mathcal{G} . Dann ist auch jeder Block in $\mathcal{B}(X_{\mathcal{G}})$ von der Form wu synchronisierend für \mathcal{G} . Falls w auf I ausgerichtet ist, dann ist $F_{X_{\mathcal{G}}}(w) = F_{\mathcal{G}}(I)$.*

Beweis: Sei w auf I ausgerichtet. Jeder Pfad, der wu darstellt, ist von der Form $\pi\tau$, wobei π bei I endet. Aber es gibt nur einen Pfad mit der Beschriftung u , der bei I beginnt, da \mathcal{G} rechtsauflösend ist. Folglich muss jeder Pfad, der wu

darstellt, bei $t(\tau)$ enden, woraus folgt, dass wu synchronisierend ist. Jeder Block, der auf w folgen kann, wird dargestellt durch einen Pfad, der bei I beginnt und umgekehrt. Daher gilt auch die zweite Aussage. \square

Das nächste Resultat zeigt, dass man für einige beschriftete Graphen synchronisierende Blöcke finden kann, sogar unendlich viele.

4.34 Satz: *Sei \mathcal{G} ein rechtsauflösender beschrifteter Graph, der Nachfolger-separiert ist. Dann kann jeder Block $u \in \mathcal{B}(X_{\mathcal{G}})$ nach rechts zu einem synchronisierenden Block uw erweitert werden.*

Beweis: Für einen beliebigen Block $v \in \mathcal{B}(X_{\mathcal{G}})$ sei $T(v)$ die Menge aller Endknoten von Pfaden, die v darstellen. Besteht $T(u)$ nur aus einem einzelnen Knoten, so ist u synchronisierend und jede Wahl von w mit $uw \in \mathcal{B}(X_{\mathcal{G}})$ ist möglich.

Jetzt nehmen wir an, dass $T(u)$ aus mehr als einem Knoten besteht, und wählen zwei verschiedene Knoten I und J in $T(u)$. Da $F_{\mathcal{G}}(I) \neq F_{\mathcal{G}}(J)$ (nach der Voraussetzung, dass \mathcal{G} Nachfolger-separiert ist), dürfen wir annehmen, dass es einen Block v_1 mit $v_1 \in F_{\mathcal{G}}(I)$ und $v_1 \notin F_{\mathcal{G}}(J)$ gibt (ansonsten vertauschen wir die Rollen von I und J). Da \mathcal{G} rechtsauflösend ist, gibt es für jedes Element von $T(u)$ höchstens einen Pfad mit der Beschriftung v_1 , der dort beginnt. Da $v_1 \notin F_{\mathcal{G}}(J)$, hat $T(uv_1)$ weniger Elemente als $T(u)$. Wir können so fortfahren, um immer kleinere Mengen $T(uv_1), T(uv_1v_2), \dots$ zu erzeugen, wobei wir aufhören, wenn $T(uv_1 \dots v_n)$ nur noch einen Knoten enthält. Dann wählen wir $w := v_1 \dots v_n$. \square

Nun kommen wir zu den Hauptresultaten dieses Abschnitts.

4.35 Satz: *Sind $\mathcal{G} = (G, \mathcal{L})$ und $\mathcal{G}' = (G', \mathcal{L}')$ irreduzible rechtsauflösende Darstellungen eines sofischen Shifts X , die auch Nachfolger-separiert sind, so sind \mathcal{G} und \mathcal{G}' als beschriftete Graphen isomorph.*

Beweis: Wir werden zunächst unter Verwendung, dass \mathcal{G} und \mathcal{G}' rechtsauflösend und Nachfolger-separiert sind, einen Block wählen, der synchronisierend für beide Graphen ist. Dieser Block liefert je einen Knoten in \mathcal{G} und \mathcal{G}' , die wir einander zuordnen. Anschließend werden wir die Irreduzibilität verwenden, um diese Zuordnung zu invertierbaren Abbildungen $\partial\Phi : \mathcal{V}(G) \rightarrow \mathcal{V}(G')$ und $\Phi : \mathcal{E}(G) \rightarrow \mathcal{E}(G')$ fortzusetzen, die die Beschriftungen respektieren. Dabei wird die zweite Aussage in Lemma 4.33 eine Schlüsselrolle spielen, die besagt, dass die Nachfolgermenge eines synchronisierenden Blocks identisch ist mit der Nachfolgermenge des Knotens, auf den dieser Block ausgerichtet ist.

Nach Satz 4.34 können wir einen synchronisierenden Block u_1 für \mathcal{G} und eine Erweiterung $w = u_1u_2$ finden, die synchronisierend für \mathcal{G}' ist. Nach Lemma 4.33 ist w auch synchronisierend für \mathcal{G} . Wir werden jetzt w verwenden, um $\partial\Phi$ zu definieren. Sei I_0 der Knoten in \mathcal{G} und I'_0 der in \mathcal{G}' , auf den w ausgerichtet ist. Wir definieren $\partial\Phi(I_0) := I'_0$. Ist J ein beliebiger Knoten in G , so wählen wir einen Pfad π von I_0 nach J , der existiert, da \mathcal{G} irreduzibel ist. Ist $u := \mathcal{L}(\pi)$,

so ist der Block wu synchronisierend für beide Graphen, und ist auf J in \mathcal{G} und auf ein J' in \mathcal{G}' ausgerichtet. Wir definieren $\partial\Phi(J) := J'$. Diese Definition ist unabhängig von der Wahl des Pfads, da

$$F_{\mathcal{G}}(J) = F_{X_{\mathcal{G}}}(wu) = F_X(wu) = F_{X_{\mathcal{G}'}}(wu) = F_{\mathcal{G}'}(J') \quad (7)$$

und \mathcal{G}' Nachfolger-separiert ist. Genauer: Ist $\tilde{\pi}$ ein weiterer Pfad von I_0 nach J und ist $\tilde{u} = \mathcal{L}(\tilde{\pi})$, so ist $w\tilde{u}$ auf einen Knoten \tilde{J}' in \mathcal{G}' ausgerichtet. Damit folgt

$$F_{\mathcal{G}'}(\tilde{J}') = F_{\mathcal{G}}(J) = F_{\mathcal{G}'}(J'),$$

woraus sich $\tilde{J}' = J'$ ergibt, da \mathcal{G}' Nachfolger-separiert ist. Es folgt aus (7) auch, dass $\partial\Phi$ injektiv ist. Gilt nämlich $\partial\Phi(J_1) = \partial\Phi(J_2)$, so folgt

$$F_{\mathcal{G}}(J_1) = F_{\mathcal{G}'}(\partial\Phi(J_1)) = F_{\mathcal{G}'}(\partial\Phi(J_2)) = F_{\mathcal{G}}(J_2),$$

und damit $J_1 = J_2$, da \mathcal{G} Nachfolger-separiert ist. Wir können eine Inverse zu $\partial\Phi$ konstruieren, indem wir die Rollen von \mathcal{G} und \mathcal{G}' vertauschen. Also ist $\partial\Phi$ invertierbar.

Als nächstes definieren wir Φ . Sei e eine Kante mit der Beschriftung a von I nach J . Sei wu ein synchronisierender Block, der auf I in \mathcal{G} ausgerichtet ist. Dann ist wua auf J ausgerichtet. Wir bezeichnen mit I' und J' die Knoten in \mathcal{G}' , auf die wu und wua ausgerichtet sind. Dann gilt $\partial\Phi(I) = I'$ und $\partial\Phi(J) = J'$. Da $a \in F_{\mathcal{G}}(I) = F_{\mathcal{G}'}(I')$, gibt es eine Kante e' mit der Beschriftung a , die bei I' beginnt. Wir definieren $\Phi(e) := e'$. Da die Graphen rechtsauflösend sind, ist Φ injektiv. Wiederum durch Vertauschung der Rollen von \mathcal{G} und \mathcal{G}' sehen wir, dass Φ invertierbar ist und die Beschriftungen erhält. Also sind \mathcal{G} und \mathcal{G}' isomorph. \square

4.36 Satz: *Je zwei minimale rechtsauflösende Darstellungen eines irreduziblen sofischen Shifts sind isomorph als beschriftete Graphen.*

Beweis: Sind \mathcal{G} und \mathcal{G}' minimale rechtsauflösende Darstellungen desselben irreduziblen sofischen Shifts, so sind sie nach Lemma 4.29 irreduzibel und nach Satz 4.28 Nachfolger-separiert. Damit folgt die Aussage aus Satz 4.35. \square

Nach dem obigen Satz können wir von *der* minimalen rechtsauflösenden Darstellung eines sofischen Shifts X sprechen, die wir mit $\mathcal{G}_X = (G_X, \mathcal{L}_X)$ bezeichnen. Aber wie können wir \mathcal{G}_X mit Hilfe von X beschreiben und konkret konstruieren? Die Antwort auf die erste Frage hat mit dem Nachfolgermengengraphen und „intrinsic synchronisierenden Blöcken“ zu tun. Die Antwort auf die zweite Frage verwendet das untenstehende Korollar 4.38, beginnend mit einer irreduziblen rechtsauflösenden Darstellung, und wird durch Satz 4.50 im nächsten Abschnitt beantwortet.

4.37 Korollar: *Sei X ein irreduzibler sofischer Shift. Eine rechtsauflösende Darstellung von X ist die minimale rechtsauflösende Darstellung genau dann, wenn sie irreduzibel und Nachfolger-separiert ist.*

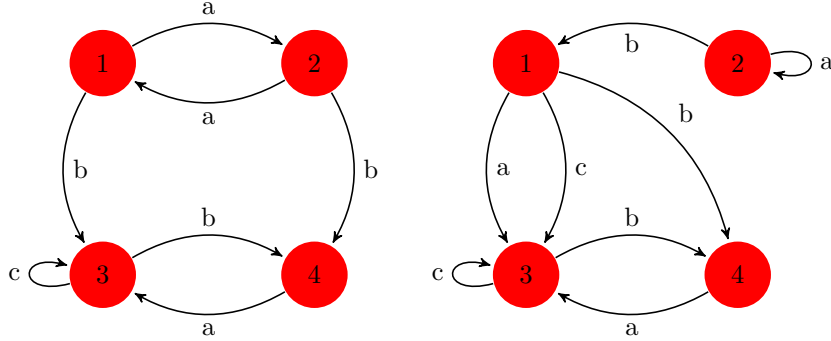


Abbildung 11: Darstellungen eines reduziblen sofischen Shifts

Beweis: Ist \mathcal{G} eine minimale rechtsauflösende Darstellung von X , so ist \mathcal{G} nach Lemma 4.29 irreduzibel und nach Satz 4.28 Nachfolger-separiert. Ist \mathcal{G}' eine rechtsauflösende, irreduzible und Nachfolger-separierte Darstellung von X , so ist \mathcal{G}' nach Satz 4.35 isomorph zu \mathcal{G} . \square

4.38 Korollar: Sei X ein irreduzibler sofischer Shift und sei \mathcal{G} eine irreduzible rechtsauflösende Darstellung von X . Dann ist der aus \mathcal{G} fusionierte Graph die minimale rechtsauflösende Darstellung von X .

Beweis: Dies folgt aus Lemma 4.27 und Korollar 4.37. \square

Das folgende Beispiel zeigt, dass für reduzible sofische Shifts die Eindeutigkeit einer minimalen rechtsauflösenden Darstellung nicht garantiert ist.

4.39 Beispiel: Seien \mathcal{G} und \mathcal{H} die in Abbildung 11 dargestellten beschrifteten Graphen. Ihre symbolischen Übergangsmatrizen sind

$$A_{\mathcal{G}} = \begin{bmatrix} \emptyset & a & b & \emptyset \\ a & \emptyset & \emptyset & b \\ \emptyset & \emptyset & c & b \\ \emptyset & \emptyset & a & \emptyset \end{bmatrix} \quad \text{und} \quad A_{\mathcal{H}} = \begin{bmatrix} \emptyset & \emptyset & a+c & b \\ b & a & \emptyset & \emptyset \\ \emptyset & \emptyset & c & b \\ \emptyset & \emptyset & a & \emptyset \end{bmatrix}.$$

Da jedes Symbol höchstens einmal in jeder Zeile erscheint, sind beide Graphen rechtsauflösend. Wir werden beweisen, dass diese beiden Graphen denselben sofischen Shift X darstellen und dass X keine rechtsauflösende Darstellung mit weniger als vier Knoten hat. Zunächst stellen wir fest, dass \mathcal{G} und \mathcal{H} nicht isomorph sind, da \mathcal{H} eine Schleife mit der Beschriftung a hat, nicht jedoch \mathcal{G} . Also sind \mathcal{G} und \mathcal{H} nicht-isomorphe minimale rechtsauflösende Darstellungen desselben (reduziblen) sofischen Shifts X , was zeigt, dass die Voraussetzung der Irreduzibilität in Satz 4.36 nicht weggelassen werden kann.

Wir zeigen zunächst, dass $\mathcal{B}(X_{\mathcal{G}}) = \mathcal{B}(X_{\mathcal{H}})$, so dass $X_{\mathcal{G}} = X_{\mathcal{H}} = X$. Jeder Pfad in \mathcal{G} kann zu einem Pfad fortgesetzt werden, der zwischen 1 und 2 hin und her

und dann zu Knoten 3 oder 4 läuft und von dort aus weitergeht. Daher besteht $\mathcal{B}(X_{\mathcal{G}})$ aus allen Teilblöcken der (unendlichen) Blöcke

$$\begin{aligned} v_1 &= a^k b c^{m_1} b a c^{m_2} b a c^{m_3} b a \dots, \\ v_2 &= a^k b a c^{n_1} b a c^{n_2} b a c^{n_3} b a \dots, \end{aligned}$$

wobei die Exponenten k, m_i, n_j beliebige ganze Zahlen ≥ 0 sein können. Ganz ähnlich gibt es drei Typen von Pfaden auf \mathcal{H} , abhängig davon, welche Kante den Knoten 1 verlässt. $\mathcal{B}(X_{\mathcal{H}})$ besteht daher aus allen Teilblöcken der folgenden drei Typen von (unendlichen) Blöcken:

$$\begin{aligned} w_1 &= a^k b a c^{p_1} b a c^{p_2} b a c^{p_3} b a \dots, \\ w_2 &= a^k b c c^{q_1} b a c^{q_2} b a c^{q_3} b a \dots, \\ w_3 &= a^k b b a c^{r_1} b a c^{r_2} b a c^{r_3} b a \dots, \end{aligned}$$

wobei die Exponenten wieder beliebige Zahlen ≥ 0 sind. Blöcke vom Typ w_1 entsprechen denen vom Typ v_2 . Der Typ w_2 entspricht dem Typ v_1 mit $m_1 = 1 + q_1 \geq 1$, wohingegen der Typ w_3 dem Typ v_1 mit $m_1 = 0$ entspricht. Dies zeigt, dass $\mathcal{B}(X_{\mathcal{G}}) = \mathcal{B}(X_{\mathcal{H}})$.

Als nächstes wollen wir zeigen, dass kein beschrifteter Graph mit weniger Knoten eine rechtsauflösende Darstellung von X sein kann. Dazu stellen wir fest, dass jede der Nachfolgermengen $F_X(aa)$, $F_X(c)$ und $F_X(cb)$ einen Block enthält, der in keiner der anderen enthalten ist. Zum Beispiel ist $aab \in F_X(aa) \setminus (F_X(c) \cup F_X(cb))$. Hätte also X eine rechtsauflösende Darstellung \mathcal{K} mit nur drei Knoten, dann könnten wir die Blöcke den Knoten 1, 2, 3 zuordnen in dem Sinne, dass

- (1) $F_{\mathcal{K}}(1) \subseteq F_X(aa)$,
- (2) $F_{\mathcal{K}}(2) \subseteq F_X(c)$ und
- (3) $F_{\mathcal{K}}(3) \subseteq F_X(cb)$.

Es ist auch leicht einzusehen, dass $F_X(aab) = F_X(c) \cup F_X(cb)$. Folglich muss es mindestens zwei verschiedene Pfade geben, die aab darstellen. Wenn es nämlich nur einen Pfad gäbe, der aab darstellt, dann endet dieser bei einem der drei Knoten und es würde $F_X(aab) = F_X(c) \cup F_X(cb) \subseteq F_X(aa)$, $F_X(aab) = F_X(c) \cup F_X(cb) \subseteq F_X(c)$ oder $F_X(aab) = F_X(c) \cup F_X(cb) \subseteq F_X(cb)$ gelten. Alle drei Möglichkeiten sind aber ausgeschlossen. Da $aab \notin F_X(c) \cup F_X(cb)$, müssen alle Pfade, die aab darstellen, bei 1 beginnen. Dies widerspricht der Tatsache, dass \mathcal{K} rechtsauflösend ist.

4.4 Konstruktionen und Algorithmen

In diesem Unterabschnitt wollen wir Antworten auf folgende Fragen finden:

- (1) Gegeben zwei sofische Shifts X_1 und X_2 , wie können wir Darstellungen für $X_1 \cup X_2$, $X_1 \cap X_2$ und $X_1 \times X_2$ erhalten?

- (2) Wie können wir feststellen, ob zwei beschriftete Graphen denselben sofi-
schen Shift darstellen?
- (3) Wie finden wir die minimale rechtsauflösende Darstellung eines irredu-
ziblen sofiischen Shifts?
- (4) Wie können wir herausfinden, ob ein sofiischer Shift von endlichem Typ
ist?

Wir beginnen mit verschiedenen Konstruktionen, um aus gegebenen Graphen neue Graphen zu bauen.

Seien $\mathcal{G}_1 = (G_1, \mathcal{L}_1)$ und $\mathcal{G}_2 = (G_2, \mathcal{L}_2)$ beschrifteten Graphen mit Alphabeten \mathcal{A}_1 und \mathcal{A}_2 . Wir definieren $\mathcal{A} := \mathcal{A}_1 \cup \mathcal{A}_2$ und betrachten \mathcal{G}_1 und \mathcal{G}_2 als Graphen über dem Alphabet \mathcal{A} . Indem wir \mathcal{G}_1 und \mathcal{G}_2 nebeneinander zeichnen, erhalten wir einen neuen beschrifteten Graphen, den wir mit $\mathcal{G}_1 \cup \mathcal{G}_2$ bezeichnen und die *disjunkte Vereinigung* von \mathcal{G}_1 und \mathcal{G}_2 nennen. Dies kann auf verschiedene Arten noch präziser gemacht werden, aber uns soll die informelle Definition genügen.

4.40 Satz: Sind \mathcal{G}_1 und \mathcal{G}_2 beschriftete Graphen, so gilt

$$X_{\mathcal{G}_1 \cup \mathcal{G}_2} = X_{\mathcal{G}_1} \cup X_{\mathcal{G}_2}.$$

Also ist die Vereinigung zweier sofiischer Shifts auch sofiisch.

Beweis: \mathcal{G}_1 und \mathcal{G}_2 sind Teilgraphen von $\mathcal{G}_1 \cup \mathcal{G}_2$. Also ist jeder Pfad in \mathcal{G}_k ($k = 1, 2$) auch ein Pfad in \mathcal{G} und es folgt $X_{\mathcal{G}_1} \cup X_{\mathcal{G}_2} \subseteq X_{\mathcal{G}_1 \cup \mathcal{G}_2}$. Da es keine Kante gibt, die \mathcal{G}_1 und \mathcal{G}_2 verbindet, ist jeder Pfad in \mathcal{G} vollständig in \mathcal{G}_1 oder in \mathcal{G}_2 enthalten. Daher gilt $X_{\mathcal{G}} \subseteq X_{\mathcal{G}_1} \cup X_{\mathcal{G}_2}$. Für zwei sofiische Shifts liefert also die disjunkte Vereinigung ihrer Darstellungen eine Darstellung ihrer Vereinigung, die folglich auch sofiisch ist. \square

Als nächstes definieren wir Produkte von Graphen und von beschrifteten Graphen.

4.41 Definition: Seien $G_1 = (\mathcal{V}_1, \mathcal{E}_1)$ und $G_2 = (\mathcal{V}_2, \mathcal{E}_2)$ Graphen. Ihr Graphenprodukt $G_1 \times G_2$ hat die Knotenmenge $\mathcal{V}_1 \times \mathcal{V}_2$ und die Kantenmenge $\mathcal{E}_1 \times \mathcal{E}_2$. Eine Kante (e_1, e_2) beginnt bei $(i(e_1), i(e_2))$ und endet bei $(t(e_1), t(e_2))$.

Seien X_1 und X_2 Shifträume über Alphabeten \mathcal{A}_1 und \mathcal{A}_2 . Ein Element (x, y) von $X_1 \times X_2$ ist ein Paar von Folgen

$$(\dots x_{-1}x_0x_1\dots, \dots y_{-1}y_0y_1\dots).$$

Wir können so ein Paar mit der Folge

$$\dots (x_{-1}, y_{-1}), (x_0, y_0), (x_1, y_1) \dots$$

identifizieren. Wir betrachten daher das Produkt $X_1 \times X_2$ als einen Shiftraum über $\mathcal{A}_1 \times \mathcal{A}_2$.

4.42 Satz: Sind G_1 und G_2 Graphen, so gilt $X_{G_1} \times X_{G_2} = X_{G_1 \times G_2}$.

Beweis: Eine bi-unendliche Folge von Kantenpaaren ist in $X_{G_1 \times G_2}$ enthalten genau dann, wenn die Folge der ersten Komponenten in X_{G_1} und die der zweiten Komponenten in X_{G_2} enthalten ist. \square

Auf ähnliche Weise kann man ein Produkt beschrifteter Graphen definieren.

4.43 Definition: Seien $\mathcal{G}_1 = (G_1, \mathcal{L}_1)$ und $\mathcal{G}_2 = (G_2, \mathcal{L}_2)$ beschriftete Graphen über Alphabeten \mathcal{A}_1 und \mathcal{A}_2 . Ihr Graphenprodukt $\mathcal{G} = \mathcal{G}_1 \times \mathcal{G}_2$ hat als zugrundeliegenden Graphen $G_1 \times G_2$ und die Beschriftungsabbildung $\mathcal{L} = \mathcal{L}_1 \times \mathcal{L}_2 : \mathcal{E}_1 \times \mathcal{E}_2 \rightarrow \mathcal{A}_1 \times \mathcal{A}_2$, definiert durch $\mathcal{L}(e_1, e_2) := (\mathcal{L}_1(e_1), \mathcal{L}_2(e_2))$.

4.44 Satz: Es gilt $X_{\mathcal{G}_1 \times \mathcal{G}_2} = X_{\mathcal{G}_1} \times X_{\mathcal{G}_2}$. Das Produkt zweier sofischer Shifts ist folglich auch sofisch.

Beweis: Seien $\mathcal{G}_1 = (G_1, \mathcal{L}_1)$ und $\mathcal{G}_2 = (G_2, \mathcal{L}_2)$. Nach Satz 4.42 ist $X_{\mathcal{G}_1 \times \mathcal{G}_2} = X_{G_1 \times G_2}$. Da wir ein Paar von Folgen mit einer Folge von Paaren identifiziert haben, gilt

$$\begin{aligned} X_{\mathcal{G}_1 \times \mathcal{G}_2} &= (\mathcal{L}_1 \times \mathcal{L}_2)_\infty(X_{G_1 \times G_2}) = (\mathcal{L}_1 \times \mathcal{L}_2)_\infty(X_{G_1} \times X_{G_2}) \\ &= (\mathcal{L}_1)_\infty(X_{G_1}) \times (\mathcal{L}_2)_\infty(X_{G_2}) = X_{\mathcal{G}_1} \times X_{\mathcal{G}_2}. \end{aligned}$$

Dies liefert eine Darstellung für das Produkt zweier sofischer Shifts, das folglich auch sofisch ist. \square

Die nächste Konstruktion stellt fest, ob Pfade in zwei beschrifteten Graphen dieselbe Beschriftung haben. Wir werden sie verwenden, um den Schnitt zweier sofischer Shifts darzustellen.

4.45 Definition: Seien $\mathcal{G}_1 = (G_1, \mathcal{L}_1)$ und $\mathcal{G}_2 = (G_2, \mathcal{L}_2)$ zwei beschriftete Graphen über demselben Alphabet \mathcal{A} und seien $G_1 = (\mathcal{V}_1, \mathcal{E}_1)$ und $G_2 = (\mathcal{V}_2, \mathcal{E}_2)$ ihre zugrundeliegenden Graphen. Das Beschriftungsprodukt $\mathcal{G} = \mathcal{G}_1 \star \mathcal{G}_2$ hat als zugrundeliegenden Graphen den Graphen G mit Knotenmenge $\mathcal{V}_1 \times \mathcal{V}_2$ und Kantenmenge

$$\mathcal{E} := \{(e_1, e_2) \in \mathcal{E}_1 \times \mathcal{E}_2 : \mathcal{L}_1(e_1) = \mathcal{L}_2(e_2)\}.$$

Die Beschriftung von \mathcal{G} ist definiert durch $\mathcal{L}(e_1, e_2) := \mathcal{L}_1(e_1) = \mathcal{L}_2(e_2)$.

Das Beschriftungsprodukt entfernt aus dem Graphenprodukt alle Paare von Kanten, die verschiedene Beschriftungen haben. Im Gegensatz zum Graphenprodukt hat das Beschriftungsprodukt allerdings dasselbe Alphabet wie die gegebenen Graphen. Haben diese Alphabete \mathcal{A}_1 und \mathcal{A}_2 , so können wir sie uns als beschriftete Graphen mit Alphabet $\mathcal{A} = \mathcal{A}_1 \cup \mathcal{A}_2$ vorstellen und ihr Beschriftungsprodukt betrachten.

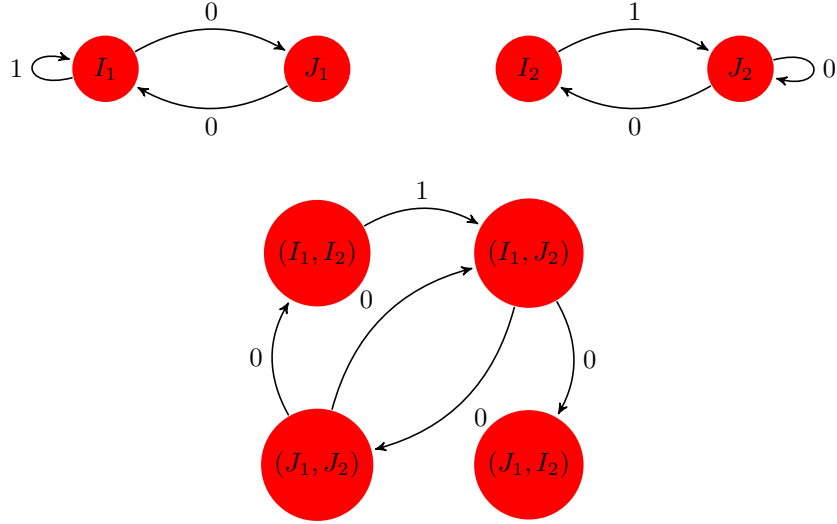


Abbildung 12: Zwei Graphen und ihr Beschriftungsprodukt

4.46 Beispiel: In Abbildung 12 sehen wir zwei Graphen und ihr Beschriftungsprodukt. Insbesondere sehen wir, dass $\mathcal{G}_1 \star \mathcal{G}_2$ gestrandete Knoten haben kann, auch wenn das für \mathcal{G}_1 und \mathcal{G}_2 nicht der Fall ist, und dass $\mathcal{G}_1 \star \mathcal{G}_2$ nicht irreduzibel sein muss, wenn es \mathcal{G}_1 und \mathcal{G}_2 sind.

4.47 Satz: Sind \mathcal{G}_1 und \mathcal{G}_2 zwei beschriftete Graphen, so gilt

$$X_{\mathcal{G}_1} \cap X_{\mathcal{G}_2} = X_{\mathcal{G}_1 \star \mathcal{G}_2}.$$

Insbesondere ist also der Schnitt zweier sofischer Shifts sofisch. Sind \mathcal{G}_1 und \mathcal{G}_2 rechtsauflösend, so auch $\mathcal{G}_1 \star \mathcal{G}_2$.

Beweis: Seien $\mathcal{G}_1 = (G_1, \mathcal{L}_1)$ und $\mathcal{G}_2 = (G_2, \mathcal{L}_2)$ sowie $\mathcal{L} = \mathcal{L}_1 \times \mathcal{L}_2$. Ist $\xi = (\xi_1, \xi_2)$ ein bi-unendlicher Pfad auf dem $\mathcal{G}_1 \star \mathcal{G}_2$ zugrundeliegenden Graphen, dann ist $\mathcal{L}_\infty(\xi) = (\mathcal{L}_1)_\infty(\xi_1) \in X_{\mathcal{G}_1}$ und $\mathcal{L}_\infty(\xi) = (\mathcal{L}_2)_\infty(\xi_2) \in X_{\mathcal{G}_2}$. Es folgt $X_{\mathcal{G}_1 \star \mathcal{G}_2} \subseteq X_{\mathcal{G}_1} \cap X_{\mathcal{G}_2}$.

Ist umgekehrt $x \in X_{\mathcal{G}_1} \cap X_{\mathcal{G}_2}$, dann gibt es einen bi-unendlichen Pfad ξ_1 auf G_1 und ξ_2 auf G_2 mit $(\mathcal{L}_1)_\infty(\xi_1) = x = (\mathcal{L}_2)_\infty(\xi_2)$. Daher hat die i -te Kante in ξ_1 für jedes i dieselbe Beschriftung wie die i -te Kante in ξ_2 , so dass (ξ_1, ξ_2) ein bi-unendlicher Pfad im $\mathcal{G}_1 \star \mathcal{G}_2$ zugrundeliegenden Graphen ist. Damit folgt $X_{\mathcal{G}_1} \cap X_{\mathcal{G}_2} \subseteq X_{\mathcal{G}_1 \star \mathcal{G}_2}$.

Jetzt nehmen wir an, dass \mathcal{G}_1 und \mathcal{G}_2 rechtsauflösend sind und dass (e_1, e_2) und (e'_1, e'_2) verschiedene Kanten in $\mathcal{G}_1 \star \mathcal{G}_2$ sind, die bei demselben Knoten beginnen und dieselbe Beschriftung tragen. Falls $e_1 \neq e'_1$, würde $\mathcal{L}_1(e_1) = \mathcal{L}(e_1, e_2) = \mathcal{L}(e'_1, e'_2) = \mathcal{L}_1(e'_1)$ folgen im Widerspruch zur Annahme, dass \mathcal{G}_1

rechtsauflösend ist. Ein ähnliches Argument zeigt, dass $e_2 = e'_2$. Folglich ist $\mathcal{G}_1 \star \mathcal{G}_2$ rechtsauflösend. \square

Unser nächstes Ziel ist es, mittels eines Algorithmus festzustellen, ob zwei beschriftete Graphen denselben sofischen Shift darstellen. Um zu garantieren, dass der Algorithmus terminiert, benötigen wir die Aussagen des folgenden Lemmas.

4.48 Lemma: *Sei G ein Graph mit einer r -elementigen Knotenmenge \mathcal{V} . Sei ferner $I \in \mathcal{V}$ und \mathcal{S} eine s -elementige Teilmenge von \mathcal{V} , die I nicht enthält. Gibt es einen Pfad von I zu einem Knoten in \mathcal{S} , so ist die Länge des kürzesten solchen Pfades $\leq r - s$.*

Wenn A die Übergangsmatrix von G ist und $B = A + A^2 + \dots + A^{r-s}$, dann gibt es einen Pfad von I zu einem Knoten in \mathcal{S} genau dann, wenn $B_{IJ} > 0$ für ein $J \in \mathcal{S}$.

Beweis: Sei $\pi = \pi_1 \dots \pi_n$ ein kürzester Pfad von I zu einem Knoten in \mathcal{S} . Dann ist $i(e_j) \notin \mathcal{S}$ für ein $1 \leq j \leq n$, da es andernfalls einen kürzeren Pfad von I nach \mathcal{S} geben würde. Die Knoten $i(e_j)$ müssen alle verschieden voneinander sein, da π sonst einen geschlossenen Teilpfad enthalten würde, dessen Entfernung zu einem kürzeren Pfad von I nach \mathcal{S} führen würde. Also enthält $\mathcal{V} \setminus \mathcal{S}$ n verschiedene Elemente, so dass

$$|\pi| = n \leq |\mathcal{V} \setminus \mathcal{S}| = r - s.$$

Die zweite Behauptung folgt nun, da nach Satz 3.19 die Anzahl der Pfade der Länge n von I nach J durch $(A^n)_{IJ}$ gegeben ist. \square

4.49 Satz: *Es gibt einen Algorithmus, der für zwei beschriftete Graphen \mathcal{G}_1 und \mathcal{G}_2 bestimmt, ob $X_{\mathcal{G}_1} = X_{\mathcal{G}_2}$.*

Beweis: Für jedes \mathcal{G}_k konstruieren wir zunächst einen Hilfsgraphen $\hat{\mathcal{G}}_k$, der verwendet wird, um alle Blöcke zu bestimmen, die *nicht* in $X_{\mathcal{G}_k}$ vorkommen. Danach bilden wir das Beschriftungsprodukt $\hat{\mathcal{G}} = \hat{\mathcal{G}}_1 \star \hat{\mathcal{G}}_2$. In $\hat{\mathcal{G}}$ gibt es einen Knoten \mathcal{I} und eine Teilmenge \mathcal{S} der Knotenmenge, so dass $X_{\mathcal{G}_1} \neq X_{\mathcal{G}_2}$ genau dann, wenn es einen Pfad von \mathcal{I} nach \mathcal{S} gibt. Lemma 4.48 zeigt, wie man feststellt, ob es einen solchen Pfad gibt.

Indem wir die Vereinigung der Alphabete nehmen, können wir annehmen, dass beide beschrifteten Graphen dasselbe Alphabet \mathcal{A} haben. Sei \mathcal{V}_k die Knotenmenge von \mathcal{G}_k . Wir werden zunächst \mathcal{G}_k vergrößern zu einem beschrifteten Graphen \mathcal{G}'_k wie folgt. Wir fügen einen Knoten K_k zur Knotenmenge von \mathcal{G}_k hinzu, den wir uns als "Terminationsknoten" vorstellen. Für jedes I in \mathcal{V}_k und $a \in \mathcal{A}$, falls es keine Kante mit der Beschriftung a gibt, die bei I beginnt, füge eine Kante mit Beschriftung a hinzu von I nach K_k . Andernfalls tue nichts. Zusätzlich füge für jedes $a \in \mathcal{A}$ eine Schleife bei K_k mit der Beschriftung a hinzu. Damit sind genügend viele Kanten hinzugefügt worden, dass für jeden Knoten I in \mathcal{G}'_k und jeden Block w über \mathcal{A} mindestens ein Pfad bei I beginnt, der w darstellt. Insbesondere ist $X_{\mathcal{G}'_k}$ der volle \mathcal{A} -Shift. Der Terminationsknoten K_k ist ein "schwarzes

Loch" für Pfade in \mathcal{G}'_k . Wenn sie K_k einmal erreicht haben, können sie nie mehr von dort wegkommen.

Als nächstes wenden wir die Teilmengenkonstruktion auf \mathcal{G}'_k an, die wir im Beweis von Satz 4.22 beschrieben haben, um den rechtsauflösenden Graphen $\hat{\mathcal{G}}_k$ zu erzeugen. Die Knoten von $\hat{\mathcal{G}}_k$ sind die nichtleeren Teilmengen von $\mathcal{V}'_k = \mathcal{V}_k \cup \{K_k\}$ und es gibt eine Kante von \mathcal{I} nach \mathcal{J} mit Beschriftung a genau dann, wenn \mathcal{J} die Menge der Knoten ist, die von einem Knoten in \mathcal{I} mittels einer Kante mit der Beschriftung a erreicht werden können.

Sei \mathcal{I}_k der Knoten in $\hat{\mathcal{G}}_k$, der die Menge der Knoten im ursprünglichen Graphen \mathcal{G}_k ist, d.h. $\mathcal{I}_k = \mathcal{V}_k$. Die Menge $\{K_k\}$ werde mit \mathcal{K}_k bezeichnet. Für jeden Block w über \mathcal{A} gibt es einen eindeutigen Pfad in $\hat{\mathcal{G}}_k$, der bei \mathcal{I}_k beginnt und w darstellt. Die entscheidende Eigenschaft des Graphen $\hat{\mathcal{G}}_k$ ist, dass dieser Pfad bei \mathcal{K}_k endet genau dann, wenn $w \notin \mathcal{B}(X_{\mathcal{G}_k})$. Das Komplement der Sprache von $X_{\mathcal{G}_k}$ ist also die Menge der Beschriftungen von Pfaden von \mathcal{I}_k nach \mathcal{K}_k . Diese Eigenschaft folgt aus der Beobachtung, dass $w \in \mathcal{B}(X_{\mathcal{G}_k})$ genau dann, wenn es mindestens einen Pfad in \mathcal{G}_k mit der Beschriftung w gibt, und dies genau dann der Fall ist, wenn der eindeutige Pfad in $\hat{\mathcal{G}}_k$ mit der Beschriftung w bei einer Menge endet, die einen Knoten aus \mathcal{G}_k enthält.

Setze $\hat{\mathcal{G}} := \hat{\mathcal{G}}_1 \star \hat{\mathcal{G}}_2$, $\mathcal{I} := (\mathcal{I}_1, \mathcal{I}_2)$ und

$$\mathcal{S}_1 := \{(\mathcal{J}, \mathcal{K}_2) : \mathcal{J} \neq \mathcal{K}_1\}.$$

Dann ist ein Block w in $X_{\mathcal{G}_1} \setminus X_{\mathcal{G}_2}$ genau dann, wenn der eindeutige Pfad mit der Beschriftung w in $\hat{\mathcal{G}}$, der bei \mathcal{I} beginnt, bei einem Knoten in \mathcal{S}_1 endet. Folglich gilt $X_{\mathcal{G}_1} \subseteq X_{\mathcal{G}_2}$ genau dann, wenn es keinen Pfad in $\hat{\mathcal{G}}$ von \mathcal{I} nach \mathcal{S}_1 gibt. Ein ähnliches Argument zeigt, dass $X_{\mathcal{G}_2} \subseteq X_{\mathcal{G}_1}$ genau dann, wenn es keinen Pfad in $\hat{\mathcal{G}}$ von \mathcal{I} nach $\mathcal{S}_2 := \{(\mathcal{J}, \mathcal{K}_1) : \mathcal{J} \neq \mathcal{K}_2\}$ gibt.

Insgesamt folgt: Es gilt $X_{\mathcal{G}_1} = X_{\mathcal{G}_2}$ genau dann, wenn es keinen Pfad in $\hat{\mathcal{G}}$ von \mathcal{I} nach $\mathcal{S}_1 \cup \mathcal{S}_2$ gibt. Lemma 4.48 zeigt, wie man die Übergangsmatrix von $\hat{\mathcal{G}}$ verwenden kann, um dies herauszufinden. \square

Das Theorem zeigt, dass es prinzipiell möglich ist, mit Hilfe eines Algorithmus zu entscheiden, ob zwei Graphen denselben Shift darstellen. Allerdings ist der Algorithmus hoffnungslos kompliziert. Hat \mathcal{G}_k r_k Knoten, so hat \mathcal{G}'_k $r_k + 1$ Knoten und die Teilmengenkonstruktion liefert einen Graphen $\hat{\mathcal{G}}_k$ mit $2^{r_k+1} - 1$ Knoten, so dass $\hat{\mathcal{G}}$ $(2^{r_1+1} - 1)(2^{r_2+1} - 1)$ Knoten hat. Wenn \mathcal{G}_k nur 9 Knoten hätte, so hätte $\hat{\mathcal{G}}$ bereits über eine Million Knoten! Für irreduzible rechtsauflösende Graphen jedoch werden wir bald eine viel effizientere Methode kennenlernen.

Das nächste Verfahren dient zur Bestimmung einer minimalen rechtsauflösenden Darstellung.

4.50 Satz: *Es gibt einen Algorithmus, der auf der Basis einer irreduziblen rechtsauflösenden Darstellung die minimale rechtsauflösende Darstellung bestimmt.*

Beweis: Sei \mathcal{G} eine irreduzible rechtsauflösende Darstellung eines sofischen Shifts. Wir erinnern daran, dass zwei Knoten I und J äquivalent heißen, wenn sie dieselbe Nachfolgermenge haben, d.h. $F_{\mathcal{G}}(I) = F_{\mathcal{G}}(J)$. Dies partitioniert die Knotenmenge in Äquivalenzklassen $\mathcal{I}_1, \dots, \mathcal{I}_r$, die die Knoten des aus \mathcal{G} fusionierten Graphen \mathcal{H} sind. Nach Korollar 4.38 ist \mathcal{H} die minimale rechtsauflösende Darstellung von \mathcal{G} . Sobald die Knoten von \mathcal{H} bestimmt sind, zeigt die Definition von \mathcal{H} , wie die Kanten von \mathcal{H} und ihre Beschriftungen am Graphen \mathcal{G} abgelesen werden können.

Wir müssen also eine Methode finden, um zu entscheiden, wann zwei Knoten äquivalent sind. Wir tun dies durch Konstruktion eines neuen Graphen \mathcal{G}' , ähnlich wie im Beweis von Satz 4.49. Wir fügen der Knotenmenge \mathcal{V} von \mathcal{G} einen Terminationsknoten K hinzu. Gibt es für einen Knoten I und ein $a \in \mathcal{A}$ keine Kante mit der Beschriftung a , die bei I beginnt, so fügen wir eine derartige Kante hinzu, die bei K endet. Für jeden Buchstaben $a \in \mathcal{A}$ fügen wir außerdem eine Schleife bei K mit Beschriftung a hinzu. Der so entstehende Graph \mathcal{G}' ist offensichtlich rechtsauflösend. Sei $\mathcal{G}' \star \mathcal{G}'$ das Beschriftungsprodukt und

$$\mathcal{S} := (\mathcal{V} \times \{K\}) \cup (\{K\} \times \mathcal{V})$$

die Menge der Knoten in $\mathcal{G}' \star \mathcal{G}'$, von denen genau eine Komponente nicht der Terminationsknoten ist. Für $I \neq J$ in \mathcal{V} behaupten wir, dass $F_{\mathcal{G}}(I) \neq F_{\mathcal{G}}(J)$ genau dann, wenn es einen Pfad in \mathcal{G}' von (I, J) nach \mathcal{S} gibt. Ist nämlich $w \in F_{\mathcal{G}}(I) \setminus F_{\mathcal{G}}(J)$, so muss der eindeutige Pfad in \mathcal{G}' mit der Beschriftung w , der bei I beginnt, bei einem Knoten in \mathcal{V} enden, während der, der bei J beginnt, bei K enden muss. Dies würde einen Pfad in $\mathcal{G}' \star \mathcal{G}'$ von (I, J) nach $\mathcal{V} \times \{K\}$ liefern. Wir stellen fest, dass die Beschriftung irgendeines Pfades von (I, J) nach \mathcal{S} in einer der Nachfolgermengen enthalten ist, aber nicht in der anderen, woraus dann die Behauptung folgt.

Also sind I und J äquivalent genau dann, wenn es keinen Pfad in $\mathcal{G}' \star \mathcal{G}'$ von (I, J) nach \mathcal{S} gibt. Lemma 4.48 liefert eine Methode um festzustellen, ob es einen solchen Pfad gibt. Diese Methode kann verwendet werden, um die Äquivalenzklassen $\mathcal{I}_1, \dots, \mathcal{I}_r$ zu bestimmen und damit \mathcal{H} zu konstruieren. \square

In obigem Beweis haben wir gezeigt, dass $F_{\mathcal{G}}(I) \neq F_{\mathcal{G}}(J)$ genau dann, wenn es einen Pfad in $\mathcal{G}' \star \mathcal{G}'$ von (I, J) nach \mathcal{S} gibt. Sei r die Anzahl der Knoten in \mathcal{G} . Da es höchstens $(r+1)^2$ Knoten in $\mathcal{G}' \star \mathcal{G}'$ gibt, folgt aus $F_{\mathcal{G}}(I) \neq F_{\mathcal{G}}(J)$, dass es einen Block der Länge $(r+1)^2$ gibt, der zu einer der Mengen $F_{\mathcal{G}}(I), F_{\mathcal{G}}(J)$ aber nicht zur anderen gehört. Tatsächlich kann man zeigen, dass zwei verschiedene Nachfolgermengen durch einen Block der Länge r unterschieden werden können.

Das praktische Problem, das zu lösen ist, wenn man die minimale rechtsauflösende Darstellung bestimmen möchte, besteht darin, zu bestimmen, welche Knoten äquivalent zueinander sind. Eine gebräuchliche und effiziente Art, dies zu tun, wird in folgendem Beispiel beschrieben.

4.51 Beispiel: Sei \mathcal{G} der beschriftete Graph in Abbildung 13. Wie man leicht sieht, ist \mathcal{G} irreduzibel und rechtsauflösend. Unser Ziel ist es, herauszufinden,

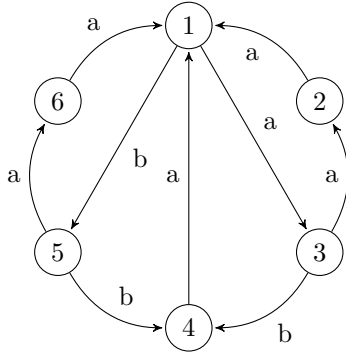


Abbildung 13: Eine zu minimierende Darstellung

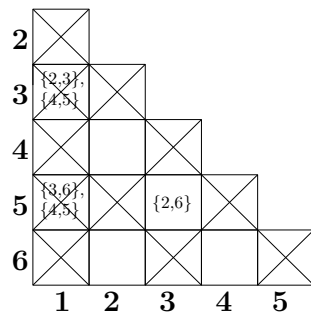


Abbildung 14: Zustandsminimierung: Tabelle

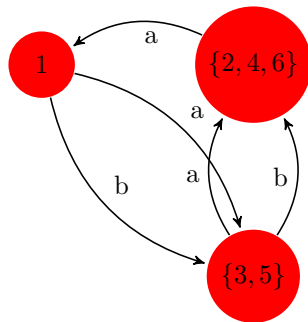


Abbildung 15: Zustandsminimierung: Fusionierter Graph

welche Knoten äquivalent zueinander sind und diese dann zu verschmelzen, um die minimale rechtsauflösende Darstellung von X_G zu bestimmen.

Die Tabelle, die in Abbildung 14 dargestellt ist, hat eine Box für jedes Paar $\{I, J\}$ von verschiedenen Knoten. Die Anordnung der Knoten ist egal. Deshalb müssen wir nur Paare (I, J) mit $I < J$ auflisten. Unser Algorithmus wird am Ende ein „X“ in die Boxen für $\{I, J\}$ setzen genau dann, wenn I nicht äquivalent zu J ist. Die übrig bleibenden Boxen zeigen uns dann, welche Knoten äquivalent sind.

Der Algorithmus durchläuft ein Paar nach dem anderen. Dabei wird jede Spalte von oben nach unten durchlaufen und die Spalten von links nach rechts abgearbeitet. Zu dem Zeitpunkt, wenn wir das Paar $\{I, J\}$ betrachten, wird jede Box für ein bereits zuvor betrachtetes Paar leer sein, mit einem „X“ markiert sein oder eine Liste von Paaren enthalten.

Betrachte ein Knotenpaar $\{I, J\}$. Wenn es ein Symbol c gibt (in diesem Beispiel entweder a oder b), so dass es eine Kante mit Beschriftung c gibt, die bei einem der beiden Knoten beginnt, aber nicht bei dem anderen, dann sind I und J nicht äquivalent und daher setzen wir ein „X“ in die Box (I, J) und auch in jede Box, die $\{I, J\}$ in ihrer Liste enthält. Diese Vorgehensweise wiederholen wir rekursiv für alle Boxen, die dabei ein „X“ erhalten. Andernfalls erstellen wir eine Liste für $\{I, J\}$ wie folgt. Für jedes Symbol c , falls es eine Kante mit Beschriftung c von I nach I' gibt, und auch eine von J nach J' , und falls $I' \neq J'$, füge $\{I', J'\}$ der Liste hinzu. Falls keine Paare der Liste hinzugefügt werden, lassen wir die Box $\{I, J\}$ leer. Dann schreiten wir fort zur nächsten Box.

Wir wollen dieses Verfahren nun auf unser Beispiel anwenden. Das erste Paar ist $\{1, 2\}$. Es gibt eine Kante mit Beschriftung b , die bei 1 beginnt, aber keine beim Knoten 2, also setzen wir ein „X“ in die Box $\{1, 2\}$. Für $\{1, 3\}$ hat jeder Knoten eine Kante für jedes Symbol, also erstellen wir eine Liste. Das Symbol a liefert das Paar $\{3, 2\} = \{2, 3\}$ und b liefert $\{5, 4\} = \{4, 5\}$, also setzen wir die Liste mit diesen beiden Paaren in die Box $\{1, 3\}$.

Als nächstes betrachten wir b , um zu sehen, dass die Box $\{1, 4\}$ ein „X“ erhält. Da $\{1, 4\}$ auf keiner der vorherigen Listen erscheint, geschieht sonst nichts. Die Box $\{1, 5\}$ erhält die dargestellte Liste, und das Symbol b zeigt, dass die Box $\{1, 6\}$ ein „X“ erhält. Das nächste Paar ist $\{2, 3\}$, das ein „X“ erhält (verwende b). $\{2, 3\}$ taucht in der Liste für Box $\{1, 3\}$ auf, also wird diese Box auch markiert. Die neu markierte Box $\{1, 3\}$ ist auf keiner der Listen, also endet die rekursive Markierung hier. Die Box $\{2, 4\}$ hat keine Elemente in ihrer Liste, bleibt also leer. Die restlichen Boxen werden ähnlich behandelt. Die vollständige Tabelle ist in Abbildung 14 dargestellt.

Die Tabelle zeigt, dass die Knoten 2, 4 und 6 äquivalent sind, und dass die Knoten 3 und 5 auch äquivalent sind (obwohl $\{3, 5\}$ nicht leer ist). Das Verschmelzen der äquivalenten Knoten liefert die minimale rechtsauflösende Darstellung, die in Abbildung 15 dargestellt ist.

Unser letzter Algorithmus entscheidet, ob ein sofischer Shift von endlichem Typ ist.

4.52 Satz: Sei \mathcal{G} ein rechtsauflösender beschrifteter Graph, so dass alle Blöcke in $\mathcal{B}_N(X_{\mathcal{G}})$ synchronisierend für \mathcal{G} sind. Dann ist $X_{\mathcal{G}}$ ein N -Schritt-Shift.

Umgekehrt gilt: Sei $\mathcal{G} = (G, \mathcal{L})$ die minimale rechtsauflösende Darstellung eines irreduziblen sofischen Shifts X und r die Anzahl der Knoten in \mathcal{G} . Ist X ein N -Schritt-Shift, so sind alle Blöcke in $\mathcal{B}_N(X)$ synchronisierend für \mathcal{G} und \mathcal{L}_{∞} ist eine Konjugation. Ist X von endlichem Typ, so ist X ein $(r^2 - r)$ -Schritt-Shift.

Beweis: Wir nehmen an, dass \mathcal{G} rechtsauflösend ist und dass alle Blöcke in $\mathcal{B}_N(X_{\mathcal{G}})$ synchronisierend für \mathcal{G} sind. Nach Satz 3.5 reicht es zu zeigen, dass aus $uw, vw \in \mathcal{B}(X_{\mathcal{G}})$ mit $|v| \geq N$ folgt, dass $uvw \in \mathcal{B}(X_{\mathcal{G}})$. Sei dazu $\pi\tau_1$ ein uw darstellender und $\tau_2\omega$ ein vw darstellender Pfad. Da $|v| \geq N$, ist v synchronisierend für \mathcal{G} . Folglich enden τ_1 und τ_2 bei demselben Knoten und $\pi\tau_1\omega$ ist ein Pfad in \mathcal{G} mit $uvw = \mathcal{L}(\pi\tau_1\omega) \in \mathcal{B}(X_{\mathcal{G}})$.

Um den zweiten Teil des Satzes zu beweisen, sei \mathcal{G} die minimale rechtsauflösende Darstellung eines irreduziblen sofischen Shifts X und wir nehmen an, dass X ein N -Schritt-Shift ist. Angenommen, es gäbe einen Block w der Länge N , der nicht synchronisierend für \mathcal{G} ist. Dann würde es Pfade π und τ in \mathcal{G} geben, die w darstellen und bei verschiedenen Knoten enden, sagen wir $t(\pi) = I \neq J = t(\tau)$. Da \mathcal{G} minimal ist, ist \mathcal{G} nach Satz 4.28 Nachfolger-separiert. Indem wir π und τ , falls nötig, vertauschen, können wir einen Block v in $F_{\mathcal{G}}(I) \setminus F_{\mathcal{G}}(J)$ finden. Nach Satz 4.34 gibt es einen synchronisierenden Block für \mathcal{G} , der, unter Verwendung der Irreduzibilität von \mathcal{G} und Lemma 4.33, erweitert werden kann zu einem synchronisierenden Block u , der auf $i(\tau)$ ausgerichtet ist. Dann gilt $uw, uv \in \mathcal{B}(X_{\mathcal{G}})$, aber wir behaupten, dass $uvw \notin \mathcal{B}(X_{\mathcal{G}})$. Dies würde Satz 3.5 widersprechen und zeigen, dass X kein N -Schritt-Shift ist.

Um unsere Behauptung zu beweisen, nehmen wir an, dass uvw von $\pi = \pi_1\pi_2\pi_3$ dargestellt wird. Da u synchronisierend ist, gilt $t(\pi_1) = i(\pi_2) = i(\tau)$. Nun folgt $\mathcal{L}(\pi_2) = w = \mathcal{L}(\tau)$ und es gibt einen eindeutigen Pfad mit Beschriftung w , der bei $i(\pi_2) = i(\tau)$ beginnt, da \mathcal{G} rechtsauflösend ist. Es folgt $\pi_2 = \tau$. Daher gilt $v \in F_{\mathcal{G}}(t(\tau)) = F_{\mathcal{G}}(J)$ im Widerspruch zu $v \notin F_{\mathcal{G}}(J)$.

Da alle Blöcke der Länge N synchronisierend sind, ist \mathcal{L}_{∞} eine Konjugation (ähnliches Argument wie in Übungsaufgabe 3 auf Blatt 6). Sei schließlich r die Anzahl der Knoten in der minimalen rechtsauflösenden Darstellung \mathcal{G} . Da X von endlichem Typ ist, ist X ein N -Schritt-Shift für ein $N \geq 1$, so dass alle Blöcke in $\mathcal{B}_N(X_{\mathcal{G}})$ synchronisierend sind. Sei $\hat{\mathcal{G}}$ der beschriftete Graph, der aus dem Beschriftungsprodukt $\mathcal{G} \star \mathcal{G}$ entsteht, indem wir die r Knoten der Form (I, I) zusammen mit den Kanten, die sie enthalten, entfernen. Dann hat $\hat{\mathcal{G}}$ $r^2 - r$ Knoten und die Beschriftungen von Pfaden in $\hat{\mathcal{G}}$ sind genau die nicht synchronisierenden Blöcke für \mathcal{G} . Wenn $\hat{\mathcal{G}}$ einen geschlossenen Pfad enthalten würde, dann hätte \mathcal{G} beliebig lange nicht synchronisierende Blöcke, was nicht sein kann, da alle Blöcke in $\mathcal{B}_N(X_{\mathcal{G}})$ synchronisierend sind. Der Graph $\hat{\mathcal{G}}$ hat

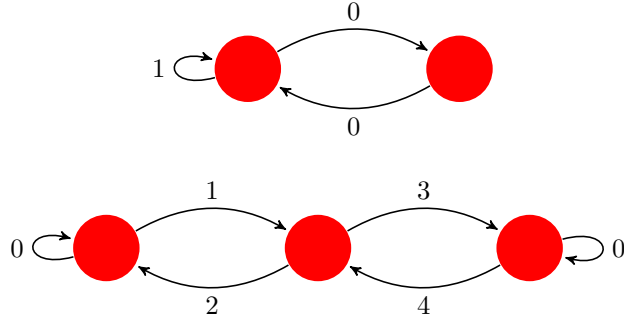


Abbildung 16: Zwei beschriftete Graphen

$r^2 - r$ Knoten, also würde jeder Pfad der Länge $r^2 - r$ einen Knoten wiederholen und damit einen geschlossenen Teilpfad enthalten. Daher haben alle Pfade auf $\hat{\mathcal{G}}$ und damit auch alle nicht synchronisierenden Blöcke eine Länge $\leq r^2 - r - 1$. Wir folgern, dass alle Blöcke in $\mathcal{B}_{r^2-r}(X_{\mathcal{G}})$ synchronisierend sind, woraus nach dem ersten Teil des Satzes folgt, dass X ein $(r^2 - r)$ -Schritt-Shift ist. \square

4.53 Bemerkung: In der Tat kann man zeigen, dass $X_{\mathcal{G}}$ in obigem Satz ein $(r^2 - r)/2$ -Schritt-Shift ist (siehe Aufgabe 2 auf Blatt 8).

4.54 Beispiel: Jeder der beschrifteten Graphen in Abbildung 16 ist irreduzibel, rechtsauflösend und Nachfolger-separiert. Nach Korollar 4.37 ist also jeder eine minimale rechtsauflösende Darstellung eines sofischen Shifts. Für jedes $N \geq 1$ ist 0^N nicht synchronisierend für jeden der Graphen. Mit Satz 4.52 können wir folgern, dass keiner der sofischen Shifts von endlichem Typ ist.

4.55 Beispiel: In Beispiel 4.39 haben wir gezeigt, dass die zwei beschrifteten Graphen in Abbildung 11 denselben sofischen Shift X darstellen und dass beide minimale rechtsauflösende Darstellungen sind. Sei \mathcal{G}_2 der rechts abgebildete Graph in der Abbildung. Dann ist jeder Block in $\mathcal{B}_3(X)$ synchronisierend für \mathcal{G}_2 . Denn c, aa, ba und bb sind offensichtlich synchronisierend und jeder Block in $\mathcal{B}_3(X)$ enthält mindestens einen dieser Blöcke. Aus Satz 4.52 folgt dann, dass X ein 3-Schritt-Shift ist.

Sei \mathcal{G}_1 der links abgebildete Graph in Abbildung 11. Wir beobachten, dass für $N \geq 1$ der Block a^N nicht synchronisierend für \mathcal{G}_1 ist. Obwohl also X ein 3-Schritt-Shift ist und \mathcal{G}_1 eine minimale rechtsauflösende Darstellung von X , gibt es beliebig lange nicht synchronisierende Blöcke für \mathcal{G}_1 . Dies zeigt, warum es im zweiten Teil von Satz 4.52 notwendig ist voranzusetzen, dass X irreduzibel ist.

5 Entropie

Die Entropie einer Abbildung (aufgefasst als dynamisches System) ist ein Maß für die Komplexität ihrer Orbitstruktur. Für Shifts misst die Entropie außerdem die „Informationskapazität“ oder die Fähigkeit, Nachrichten zu übermitteln.

Die Entropie eines Shifts ist eine wichtige Größe, da sie invariant unter Konjugation ist, für eine große Klasse von Shifts berechnet werden kann und sich gut mit Standardoperationen wie Faktorcodes und Produkten verträgt.

5.1 Definitionen und elementare Eigenschaften

Wir erinnern daran, dass das charakteristische Polynom einer $n \times n$ -Matrix A definiert ist als $\chi_A(t) = \det(tI - A)$, wobei I die $n \times n$ -Einheitsmatrix bezeichnet. Die Eigenwerte von A sind die Nullstellen von χ_A . Ein zu einem Eigenwert λ gehöriger Eigenvektor ist ein Vektor $v \neq 0$ mit $Av = \lambda v$.

Eine (nicht notwendigerweise quadratische) Matrix A heißt *positiv*, falls all ihre Einträge positiv sind. In diesem Fall schreiben wir $A > 0$. Analog sagen wir, dass A *nichtnegativ* ist, falls alle Einträge von A nichtnegativ sind und schreiben $A \geq 0$. Wir schreiben $A > B$ bzw. $A \geq B$, falls A und B dieselben Dimensionen haben und $A - B > 0$ bzw. $A - B \geq 0$. Da Vektoren spezielle Matrizen sind, können diese Definitionen genauso auf Vektoren angewandt werden.

Sei X ein Shiftraum. Die Anzahl $|\mathcal{B}_n(X)|$ der erlaubten Blöcke der Länge n gibt uns eine gewisse Idee von der Komplexität von X . Je größer die Anzahl der n -Blöcke, desto komplizierter ist der Shiftraum. Wie wir später sehen werden, wächst $|\mathcal{B}_n(X)|$ annähernd wie eine Exponentialfunktion 2^{cn} und die Zahl c ist die Entropie von X . Diese Zahl sollte ungefähr gleich $(1/n) \log_2 |\mathcal{B}_n(X)|$ sein für hinreichend große n . Wir werden im Folgenden stets die Basis 2 für Logarithmen verwenden und nur \log statt \log_2 schreiben.

5.1 Definition: Die Entropie eines Shifts X ist definiert als

$$h(X) := \lim_{n \rightarrow \infty} \frac{1}{n} \log |\mathcal{B}_n(X)|.$$

Dass der Limes in obiger Definition existiert, werden wir weiter unten zeigen. Zunächst stellen wir fest, dass $|\mathcal{B}_n(X)| \leq |\mathcal{A}|^n$, wenn X ein Shift über dem Alphabet \mathcal{A} ist, so dass

$$\frac{1}{n} \log |\mathcal{B}_n(X)| \leq \log |\mathcal{A}|,$$

woraus $h(X) \leq \log |\mathcal{A}|$ folgt. Falls $X \neq \emptyset$, gilt $|\mathcal{B}_n(X)| \geq 1$ für alle n . Insgesamt folgt $0 \leq h(X) < \infty$. Da $\mathcal{B}_n(\emptyset) = \emptyset$ für alle n , gilt $h(\emptyset) = -\infty$.

5.2 Beispiel: Sei $X = X_{[r]}$ der volle r -Shift. Dann gilt $|\mathcal{B}_n(X)| = r^n$, also $h(X) = \log r$.



Abbildung 17: Eine Darstellung des Goldenen-Schnitt-Shifts als Knotenshift

5.3 Beispiel: Sei G ein Graph, so dass bei jedem Knoten genau r Kanten beginnen. Die Anzahl der Pfade der Länge n , die bei einem Knoten in diesem Graphen beginnen, ist folglich r^n . Hat G k Knoten, so gilt also $|\mathcal{B}_n(X_G)| = k \cdot r^n$. Damit folgt

$$h(X_G) = \lim_{n \rightarrow \infty} \left(\frac{1}{n} \log k + \log r \right) = \log r.$$

5.4 Beispiel: Sei $X = X_{\mathcal{F}}$ über $\mathcal{A} = \{0, 1\}$ mit $\mathcal{F} = \{11\}$. Wir erinnern daran, dass X der Knotenshift \hat{X}_G für den Graphen in Abbildung 17 ist. Für $n \geq 2$ gibt es eine Bijektion zwischen $\mathcal{B}_n(X) = \mathcal{B}_n(\hat{X}_G)$ und $\mathcal{B}_{n-1}(X_G)$ und wir können Satz 3.19 verwenden, um die Anzahl der $(n-1)$ -Blöcke in X_G zu zählen. Die Übergangsmatrix von G ist

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}.$$

Nach Satz 3.19 gilt $|\mathcal{B}_m(X_G)| = \sum_{I, J=1}^2 (A^m)_{IJ}$. Wir können die Einträge von A^m berechnen, indem wir A diagonalisieren. Die Eigenwerte von A sind die Nullstellen des charakteristischen Polynoms von A , also

$$\chi_A(t) = \det(tI - A) = (t-1)t - 1 = t^2 - t - 1,$$

für die wir die Werte

$$\lambda = \frac{1 + \sqrt{5}}{2} \quad \text{und} \quad \mu = \frac{1 - \sqrt{5}}{2}$$

erhalten. Die zugehörigen Eigenvektoren ergeben sich zu

$$v_1 = \begin{bmatrix} \lambda \\ 1 \end{bmatrix} \quad \text{und} \quad v_2 = \begin{bmatrix} \mu \\ 1 \end{bmatrix}.$$

Indem wir also

$$P := \begin{bmatrix} \lambda & \mu \\ 1 & 1 \end{bmatrix}$$

setzen, so dass

$$P^{-1} = \frac{1}{\sqrt{5}} \begin{bmatrix} 1 & -\mu \\ -1 & \lambda \end{bmatrix},$$

erhalten wir

$$P^{-1}AP = \begin{bmatrix} \lambda & 0 \\ 0 & \mu \end{bmatrix}.$$

Es folgt

$$P^{-1}A^mP = (P^{-1}AP)^m = \begin{bmatrix} \lambda^m & 0 \\ 0 & \mu^m \end{bmatrix}.$$

Damit erhalten wir

$$\begin{aligned} A^m &= P \begin{bmatrix} \lambda^m & 0 \\ 0 & \mu^m \end{bmatrix} P^{-1} = \frac{1}{\sqrt{5}} \begin{bmatrix} \lambda^{m+1} - \mu^{m+1} & \lambda^m - \mu^m \\ \lambda^m - \mu^m & \lambda^{m-1} - \mu^{m-1} \end{bmatrix} \\ &= \begin{bmatrix} f_{m+1} & f_m \\ f_m & f_{m-1} \end{bmatrix}, \end{aligned}$$

wobei $f_m := (1/\sqrt{5})(\lambda^m - \mu^m)$. Da $\lambda^2 = \lambda + 1$, folgt $\lambda^{m+2} = \lambda^{m+1} + \lambda^m$, und analog für μ , so dass $f_{m+2} = f_{m+1} + f_m$. Wir können daher f_m rekursiv aus den ersten beiden Einträgen $f_1 = 1$ und $f_2 = 1$ bestimmen und erhalten die Fibonacci-Folge $1, 1, 2, 3, 5, 8, 13, \dots$. Dies zeigt zugleich, dass das m -te Folgenglied der Fibonacci-Folge gegeben ist durch

$$f_m = \frac{1}{\sqrt{5}} \left(\left(\frac{1 + \sqrt{5}}{2} \right)^m - \left(\frac{1 - \sqrt{5}}{2} \right)^m \right).$$

Wir folgern, dass

$$|\mathcal{B}_m(X_G)| = f_{m+1} + f_m + f_m + f_{m-1} = f_{m+2} + f_{m+1} = f_{m+3},$$

woraus

$$|\mathcal{B}_n(X)| = |\mathcal{B}_{n-1}(X_G)| = f_{n+2} = \frac{1}{\sqrt{5}}(\lambda^{n+2} - \mu^{n+2})$$

folgt. Da $\lambda \approx 1.61803$ und $\mu \approx -0.61803$, ist die Wachstumsrate von $|\mathcal{B}_n(X)|$ durch λ bestimmt. Genauer gilt: Da $\mu^n/\lambda^n \rightarrow 0$, ist

$$\begin{aligned} h(X) &= \lim_{n \rightarrow \infty} \frac{1}{n} \log |\mathcal{B}_n(X)| = \lim_{n \rightarrow \infty} \frac{1}{n} \log \frac{1}{\sqrt{5}}(\lambda^{n+2} - \mu^{n+2}) \\ &= \lim_{n \rightarrow \infty} \left[\frac{1}{n} \log \frac{1}{\sqrt{5}} + \frac{n+2}{n} \log \lambda + \frac{1}{n} \log \left(1 - \frac{\mu^{n+2}}{\lambda^{n+2}} \right) \right] \\ &= \lim_{n \rightarrow \infty} \left(1 + \frac{2}{n} \right) \log \lambda = \log \lambda. \end{aligned}$$

Die Zahl $\lambda = (1 + \sqrt{5})/2$ ist bekannt als der *goldene Schnitt*. Deshalb bezeichnen wir $X_{\mathcal{F}}$ auch als *Goldenen-Schnitt-Shift*.

Um zu zeigen, dass der Limes in der Definition von $h(X)$ existiert, benötigen wir folgendes Lemma.

5.5 Lemma: Sei $(a_n)_{n=1}^{\infty}$ eine Folge nichtnegativer reeller Zahlen, so dass

$$a_{n+m} \leq a_n + a_m \quad \text{für alle } n, m \geq 1.$$

Dann existiert der Limes $\lim_{n \rightarrow \infty} a_n/n$ und stimmt mit $\inf_{n \geq 1} a_n/n$ überein.

Beweis: Sei $\alpha := \inf_{n \geq 1} a_n/n$. Nach Definition gilt $a_n/n \geq \alpha$ für alle $n \geq 1$. Sei $\varepsilon > 0$. Um die Aussage des Lemmas zu beweisen, müssen wir zeigen, dass $a_n/n \leq \alpha + \varepsilon$ für alle hinreichend großen n . Dazu wählen wir ein k mit $a_k/k < \alpha + \varepsilon/2$. Für $0 \leq j < k$ und $m \geq 1$ gilt dann

$$\begin{aligned} \frac{a_{mk+j}}{mk+j} &\leq \frac{a_{mk}}{mk+j} + \frac{a_j}{mk+j} \leq \frac{a_{mk}}{mk} + \frac{a_j}{mk} \\ &\leq \frac{ma_k}{mk} + \frac{ja_1}{mk} \leq \frac{a_k}{k} + \frac{a_1}{m} < \alpha + \frac{\varepsilon}{2} + \frac{a_1}{m}. \end{aligned}$$

Falls $n = mk + j$ groß genug ist, gilt $a_1/m < \varepsilon/2$, so dass $a_n/n < \alpha + \varepsilon$. \square

5.6 Satz: Sei X ein Shiftraum. Dann existiert der Limes

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log |\mathcal{B}_n(X)|$$

und ist identisch mit

$$\inf_{n \geq 1} \frac{1}{n} \log |\mathcal{B}_n(X)|.$$

Beweis: Für beliebige $m, n \geq 1$ ist ein $(n+m)$ -Block eindeutig bestimmt durch seinen initialen n -Block und den darauffolgenden m -Block. Da es $|\mathcal{B}_n(X)| \cdot |\mathcal{B}_m(X)|$ Kombinationen von n -Blöcken und m -Blöcken gibt, folgt $|\mathcal{B}_{n+m}(X)| \leq |\mathcal{B}_n(X)| \cdot |\mathcal{B}_m(X)|$ und damit

$$\log |\mathcal{B}_{n+m}(X)| \leq \log |\mathcal{B}_n(X)| + \log |\mathcal{B}_m(X)|.$$

Also können wir Lemma 5.5 anwenden. \square

5.7 Satz: Ist Y ein Faktor von X , so gilt $h(Y) \leq h(X)$.

Beweis: Sei $\phi : X \rightarrow Y$ ein Faktorcode, induziert von einer Blockabbildung Φ , so dass $\phi = \Phi^{[-m,k]}$. Jeder Block in $\mathcal{B}_n(Y)$ ist das Bild eines Blocks in $\mathcal{B}_{n+m+k}(X)$ unter Φ (erweiterte Blockabbildung). Daraus folgt $|\mathcal{B}_n(Y)| \leq |\mathcal{B}_{n+m+k}(X)|$ für alle n , also

$$\begin{aligned} h(Y) &= \lim_{n \rightarrow \infty} \frac{1}{n} \log |\mathcal{B}_n(Y)| \leq \lim_{n \rightarrow \infty} \frac{1}{n} \log |\mathcal{B}_{n+m+k}(X)| \\ &= \lim_{n \rightarrow \infty} \left(\frac{n+m+k}{n} \right) \frac{1}{n+m+k} \log |\mathcal{B}_{n+m+k}(X)| = h(X). \end{aligned}$$

\square

Aus dem obigen Satz folgt, dass $h(X)$ unter Konjugation erhalten bleibt.

5.8 Korollar: Sind X und Y konjugierte Shifträume, so gilt $h(X) = h(Y)$.

Beweis: Da X und Y konjugiert sind, ist X ein Faktor von Y und umgekehrt. Also gilt $h(X) \leq h(Y) \leq h(X)$. \square

5.9 Beispiel: Der volle 2-Shift ist nicht konjugiert zum vollen 3-Shift, da die Entropien nicht übereinstimmen. Der Goldene-Schnitt-Shift ist nicht konjugiert zu einem vollen Shift, da der Goldene Schnitt keine ganze Zahl ist.

5.10 Satz: Wenn eine Einbettung von Y in X existiert, dann gilt $h(Y) \leq h(X)$.

Beweis: Sei $\phi : Y \rightarrow X$ eine Einbettung. Dann ist $\phi(Y)$ nach Satz 2.29 ein Shiftraum und ϕ ist eine Konjugation von Y nach $\phi(Y)$ nach Satz 2.30. Daher gilt $h(Y) = h(\phi(Y))$, da die Entropie unter Konjugation erhalten bleibt. Zudem gilt $h(\phi(Y)) \leq h(X)$, da $\mathcal{B}_n(\phi(Y)) \subseteq \mathcal{B}_n(X)$ für alle n , woraus unmittelbar die Behauptung folgt. \square

5.11 Satz: Sei $\mathcal{G} = (G, \mathcal{L})$ ein rechtsauflösender beschrifteter Graph. Dann gilt $h(X_{\mathcal{G}}) = h(X_G)$.

Beweis: Die Abbildung \mathcal{L}_{∞} ist ein 1-Block-Faktorcode von X_G nach $X_{\mathcal{G}}$, so dass nach Satz 5.7 gilt, dass $h(X_{\mathcal{G}}) \leq h(X_G)$. Der Graph G habe k Knoten. Dann hat jeder Block in $\mathcal{B}_n(X_{\mathcal{G}})$ höchstens k Darstellungen, da \mathcal{G} rechtsauflösend ist. Es folgt $|\mathcal{B}_n(X_{\mathcal{G}})| \geq (1/k)|\mathcal{B}_n(X_G)|$, woraus sich $h(X_{\mathcal{G}}) \geq h(X_G)$ ergibt. \square

Als nächstes wollen wir die Wachstumsrate der Anzahl n -periodischer Punkte in einem Shift untersuchen.

5.12 Definition: Für einen Shiftraum X sei $p_n(X)$ die Anzahl der periodischen Punkte in X mit Periode n .

Um die Wachstumsrate von $p_n(X)$ einzuführen, erinnern wir an die Definition des Limes Superior einer Folge $(b_n)_{n \in \mathbb{N}}$ reeller Zahlen:

$$b = \limsup_{n \rightarrow \infty} b_n := \lim_{n \rightarrow \infty} \sup_{m \geq n} b_m.$$

Äquivalent dazu können wir b als den größten Häufungswert der Folge $(b_n)_{n \in \mathbb{N}}$ definieren. Falls $(b_n)_{n \in \mathbb{N}}$ konvergiert, gilt $\lim_{n \rightarrow \infty} b_n = \limsup_{n \rightarrow \infty} b_n$.

5.13 Satz: Sei X ein Shiftraum. Dann gilt

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log p_n(X) \leq h(X).$$

Beweis: Jeder Punkt x mit Periode n ist eindeutig bestimmt durch $x_{[0, n-1]} \in \mathcal{B}_n(X)$. Daraus folgt $p_n(X) \leq |\mathcal{B}_n(X)|$, woraus sich die Behauptung ergibt, wenn wir auf beiden Seiten den Logarithmus nehmen, durch n teilen und den \limsup für $n \rightarrow \infty$ betrachten. \square

5.14 Beispiel: Die Folge der Zahlen $(1/n) \log p_n(X)$ konvergiert nicht immer. Zum Beispiel sei

$$A = \begin{bmatrix} 0 & 2 \\ 3 & 0 \end{bmatrix}$$

und $X = X_A$. Dann ist $p_n(X) = \text{Spur } A^n$ und daher

$$p_n(X) = \begin{cases} 2 \cdot 6^{n/2} & \text{für gerade } n, \\ 0 & \text{für ungerade } n. \end{cases}$$

Denn für gerade n gilt

$$A^n = (A^2)^{n/2} = \begin{bmatrix} 6 & 0 \\ 0 & 6 \end{bmatrix}^{n/2} = \begin{bmatrix} 6^{n/2} & 0 \\ 0 & 6^{n/2} \end{bmatrix}$$

und folglich

$$A^{n+1} = \begin{bmatrix} 6^{n/2} & 0 \\ 0 & 6^{n/2} \end{bmatrix} \begin{bmatrix} 0 & 2 \\ 3 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 2 \cdot 6^{n/2} \\ 3 \cdot 6^{n/2} & 0 \end{bmatrix}.$$

5.2 Perron-Frobenius-Theorie

Sei G ein Graph mit $r \times r$ -Übergangsmatrix A . Die Anzahl der n -Blöcke in X_G ist gegeben durch

$$|\mathcal{B}_n(X_G)| = \sum_{I,J=1}^r (A^n)_{IJ}.$$

Um also die Entropie von X_G zu berechnen, müssen wir die Wachstumsrate der Einträge von A^n bestimmen. Die Perron-Frobenius-Theorie ist genau das richtige Werkzeug, um dies zu tun, wenn G irreduzibel ist. Insbesondere zeigt die Theorie, dass die Wachstumsrate durch den größten Eigenwert von A bestimmt ist. Allgemeine Graphen werden wir in Unterabschnitt 5.4 behandeln.

Sei $A \neq 0$ eine quadratische nichtnegative Matrix. Wir nehmen zunächst an, dass A einen positiven Eigenvektor v hat. Dann gilt $(Av)_I > 0$ für ein I . Ist also λ der Eigenwert für v , so gilt $\lambda v_I = (Av)_I > 0$. Da $v_I > 0$, folgt $\lambda > 0$. Indem wir A auf $Av = \lambda v$ anwenden, erhalten wir $A^2v = \lambda Av = \lambda^2v$ und allgemein $A^n v = \lambda^n v$ für alle $n \geq 1$. Daher gilt für jedes I , dass

$$\sum_{J=1}^r (A^n)_{IJ} v_J = \lambda^n v_I.$$

Sei

$$c := \min\{v_1, v_2, \dots, v_r\} \quad \text{und} \quad d := \max\{v_1, v_2, \dots, v_r\}.$$

Dann gilt

$$c \sum_{J=1}^r (A^n)_{IJ} \leq \sum_{I=1}^r (A^n)_{IJ} v_J = \lambda^n v_I \leq d \lambda^n.$$

Indem wir durch c teilen und über I summieren, erhalten wir

$$\sum_{I,J=1}^r (A^n)_{IJ} \leq \sum_{I=1}^r \frac{d}{c} \lambda^n = \left(\frac{rd}{c}\right) \lambda^n = d_0 \lambda^n,$$

wobei $d_0 := rd/c > 0$.

Um $\sum_{I,J}(A^n)_{IJ}$ nach unten abzuschätzen, beobachten wir, dass für jedes I gilt:

$$c\lambda^n \leq \lambda^n v_I = \sum_{J=1}^r (A^n)_{IJ} v_J \leq d \sum_{J=1}^r (A^n)_{IJ} \leq d \sum_{I,J=1}^r (A^n)_{IJ}.$$

Setzen wir $c_0 := c/d > 0$, so können wir folgern, dass

$$c_0 \lambda^n \leq \sum_{I,J=1}^r (A^n)_{IJ}.$$

Wir können unsere Ergebnisse folgendermaßen zusammenfassen:

5.15 Satz: Sei $A \neq 0$ eine nichtnegative Matrix mit einem positiven Eigenvektor v . Dann ist der zugehörige Eigenwert λ positiv und es gibt positive Konstanten c_0 und d_0 , so dass

$$c_0 \lambda^n \leq \sum_{I,J=1}^n (A^n)_{IJ} \leq d_0 \lambda^n. \quad (8)$$

Ist daher G ein Graph mit Übergangsmatrix A , so gilt $h(X_G) = \log \lambda$.

Beweis: Wir haben bereits alles außer der letzten Aussage bewiesen. Dazu stellen wir fest, dass $|\mathcal{B}_n(X_G)|$ identisch ist mit dem mittleren Term in (8). Indem wir den Logarithmus nehmen, durch n teilen und n gegen unendlich gehen lassen, folgt die Aussage. \square

Unsere Annahme, dass A einen positiven Eigenvektor hat, hat zwei zusätzliche Konsequenzen. Sei zunächst u ein anderer positiver Eigenvektor für A und θ der zugehörige Eigenwert. Unsere vorherige Annahme kann auch auf u angewendet werden, so dass $h(X_G) = \log \theta$, woraus $\theta = \lambda$ folgt. Daher ist λ der einzige Eigenwert von A , der zu einem positiven Eigenvektor gehört.

Nun sei μ ein anderer Eigenwert von A und w ein Eigenvektor zum Eigenwert μ . (Beachte, dass sowohl μ als auch w komplex sein können.) Wir definieren die *Summennorm* von w als $\|w\| := \sum_{I=1}^r |w_I|$ (komplexer Betrag). Da $A^n w = \mu^n w$ und $A \geq 0$, folgt

$$\begin{aligned} |\mu|^n \|w\| &= \|\mu^n w\| = \|A^n w\| \\ &= \left| \sum_{J=1}^r (A^n)_{1J} w_J \right| + \cdots + \left| \sum_{J=1}^r (A^n)_{rJ} w_J \right| \\ &\leq \sum_{J=1}^r (A^n)_{1J} |w_J| + \cdots + \sum_{J=1}^r (A^n)_{rJ} |w_J| \\ &\leq \left(\max_{1 \leq J \leq r} |w_J| \right) \sum_{I,J=1}^r (A^n)_{IJ} \leq \|w\| d_0 \lambda^n. \end{aligned}$$

Da $\|w\| \neq 0$, folgt $|\mu| \leq d_0^{1/n} \lambda$ und für $n \rightarrow \infty$ folgt $|\mu| \leq \lambda$, da $\lim_{n \rightarrow \infty} d_0^{1/n} = 1$. Daher ist λ der betragsgrößte Eigenwert von A .

Die oben verwendeten Argumente hängen stark von der Tatsache ab, dass A nichtnegative Einträge hat. Jedoch hängen sie nicht von der Tatsache ab, dass die Einträge ganzzahlig sind. In der Tat lässt sich die Perron-Frobenius-Theorie auf nichtnegative (reelle) Matrizen anwenden, obwohl wir sie hauptsächlich verwenden, wenn die Einträge nichtnegative ganze Zahlen sind.

5.16 Definition: Eine nichtnegative Matrix A heißt irreduzibel, falls für jedes Paar (I, J) von Indizes ein $n \geq 0$ existiert mit $(A^n)_{IJ} > 0$. Wir verwenden die Konvention, dass $A^0 = I$, und folglich ist die 1×1 -Matrix $[0]$ irreduzibel. Eine nichtnegative Matrix A ist wesentlich, falls keine ihrer Zeilen oder Spalten identisch Null ist.

5.17 Bemerkung: Man kann leicht feststellen, dass ein Graph irreduzibel ist genau dann, wenn seine Übergangsmatrix irreduzibel ist (siehe Aufgabe (1c) auf Blatt 4). Wir betonen, dass die Irreduzibilität von A nicht $A^n > 0$ für irgendein n impliziert. Außerdem gilt: Ein Graph G ist wesentlich genau dann, wenn seine Übergangsmatrix wesentlich ist. Da die 0-te Potenz jeder Matrix die Einheitsmatrix ist, ist die 1×1 -Matrix $[0]$ irreduzibel. Mit dieser einen Ausnahme kann eine irreduzible Matrix keine Nullzeilen oder Nullspalten enthalten (d.h. sie ist wesentlich).

Die Perron-Frobenius-Theorie zeigt, dass eine irreduzible Matrix immer einen positiven Eigenvektor hat, so dass die obigen Argumente anwendbar sind. Für den restlichen Unterabschnitt sei A stets eine nichtnegative Matrix. Ein reeller Eigenwert λ von A ist *geometrisch einfach*, falls der zugehörige Eigenraum eindimensional ist und *algebraisch einfach*, falls er eine einfache Nullstelle des charakteristischen Polynoms von A ist. Wir erinnern daran, dass algebraische Einfachheit geometrische Einfachheit impliziert, aber nicht umgekehrt.

Der folgende Satz ist der Satz von Perron-Frobenius. Im Buch [1] von Marcus und Lind wird er nur in Dimension 2 bewiesen. Wir werden jedoch einen für alle Dimensionen gültigen Beweis geben.

5.18 Satz: Sei $A \neq 0$ eine irreduzible Matrix. Dann hat A einen positiven Eigenvektor v_A mit zugehörigem Eigenwert $\lambda_A > 0$, der sowohl geometrisch als auch algebraisch einfach ist. Ist μ ein anderer Eigenwert von A , dann gilt $|\mu| \leq \lambda_A$. Jeder positive Eigenvektor von A ist ein positives Vielfaches von v_A .

Für eine irreduzible Matrix A nennen wir λ_A den *Perron-Eigenwert* von A und v_A einen *Perron-Eigenvektor* von A . Der Satz zeigt, dass v_A eindeutig bis auf Multiplikation mit einer positiven Zahl ist.

Beweis: Der Beweis ist in drei Schritte unterteilt.

Schritt 1: Wir stellen zunächst fest, dass die Matrix $I + A$ die Eigenschaft hat, dass eine ihrer Potenzen ausschließlich positive Einträge hat, da

$$(I + A)^k = \sum_{j=0}^k \binom{k}{j} A^j = I + kA + \frac{k(k-1)}{2} A^2 + \cdots + A^k,$$

und die Matrix A irreduzibel ist. Nun sei k fest gewählt, so dass $P := (I + A)^k > 0$. Es gilt

$$v \leq w, v \neq w \Rightarrow Pv < Pw, \quad (9)$$

da

$$[P(v - w)]_I = \sum_{J=1}^r P_{IJ}(v - w)_J > 0.$$

Wir definieren den positiven $Q \subset \mathbb{R}^r$ durch

$$Q := \{x \in \mathbb{R}^r : x \geq 0, x \neq 0\}$$

und bezeichnen mit C den Schnitt von Q mit der Einheitssphäre $\{x \in \mathbb{R}^r : \|x\| = 1\}$, wobei $\|\cdot\|$ die Summennorm bezeichnet. Wir definieren

$$L(z) := \max \{s \geq 0 : sz \leq Az\} \stackrel{(*)}{=} \min_{1 \leq I \leq r, z_I \neq 0} \frac{(Az)_I}{z_I}$$

für jedes $z \in Q$. Die Gleichheit $(*)$ gilt aus folgendem Grund: Zum einen gilt

$$\left(\min_{1 \leq I \leq r, z_I \neq 0} \frac{(Az)_I}{z_I} \right) z_J \leq \frac{(Az)_J}{z_J} z_J = (Az)_J$$

für jedes $J \in \{1, \dots, r\}$ mit $z_J > 0$. Zum anderen gilt: Ist $\min_{1 \leq I \leq r, z_I \neq 0} (Az)_I / z_I = (Az)_J / z_J$, so folgt

$$\left(\delta + \min_{1 \leq I \leq r, z_I \neq 0} \frac{(Az)_I}{z_I} \right) z_J = \delta z_J + (Az)_J > (Az)_J$$

für jedes $\delta > 0$. Da $L(rz) = L(z)$ für alle $r > 0$, hängt $L(z)$ nur von dem Strahl durch z ab. Mit (9) folgt aus $sz \leq Az$, dass $sPz \leq PAz = APz$ und damit $L(Pz) \geq L(z)$. Falls $L(z)z \neq Az$, folgt mit (9), dass $L(z)Pz < APz$. Folglich ist $L(Pz) > L(z)$, außer z ist ein Eigenvektor von A zum Eigenwert $L(z)$. Dies legt es nahe, nach einem positiven Vektor zu suchen, der L maximiert. Betrachte dazu das Bild von C unter P . Dies ist eine kompakte Menge, da P stetig ist und C kompakt. Da P positiv ist, ist zudem jeder Vektor in $P(C)$ positiv. Die Einschränkung von L auf $P(C)$ ist eine stetige Funktion und nimmt daher ein Maximum L_{\max} auf $P(C)$ an. Da $L(z) \leq L(Pz)$, ist dies sogar ein Maximum auf Q . Da $L(Pz) > L(z)$ außer wenn z ein Eigenvektor ist, folgt, dass L sein Maximum bei einem Eigenvektor $v_A > 0$ von A annimmt. Wir können daher Satz 5.15 anwenden. Aufgrund der auf diesen Satz folgenden Überlegungen ist $\lambda_A := L_{\max} > 0$ der betragsgrößte Eigenwert von A .

Schritt 2: Wir beweisen, dass aus $0 \leq B \leq A$ und $B \neq A$ folgt, dass $|\sigma| < \lambda_A$ für jeden Eigenwert σ von B gilt. Es gelte $Bz = \sigma z$, $z \neq 0$. Dann folgt

$$A|z| \geq B|z| \geq |\sigma||z|, \quad (10)$$

wobei $|z|$ der Vektor mit Komponenten $|z_1|, \dots, |z_r|$ ist. Daraus folgt

$$|\sigma| \leq L_{\max}(A) = \lambda_A.$$

Falls $|\sigma| = \lambda_A$, folgt $L(|z|) = L_{\max}(A)$. Daraus ergibt sich $|z| > 0$ und mit (10), dass $|z|$ ein Eigenvektor von A und B zum Eigenwert λ_A ist. Aber dann gilt $(A - B)|z| = 0$, was wegen $|z| > 0$ unmöglich ist, außer wenn $A = B$. Indem wir die i -te Zeile und die i -te Spalte von A Null setzen, erhalten wir eine Matrix A_i mit $A_i \leq A$ und $A_i \neq A$ (wegen der Irreduzibilität kann keine Zeile und keine Spalte von A identisch Null sein.) Also sind die Eigenwerte von A_i alle betragsmäßig kleiner als λ_A .

Schritt 3: Sei C eine beliebige $r \times r$ -Matrix und $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_r)$. Indem wir $\det(\Lambda - C)$ nach der i -ten Zeile entwickeln, sehen wir, dass

$$\frac{\partial}{\partial \lambda_i} \det(\Lambda - C) = \det(\Lambda_{(i)} - C_{(i)}),$$

wobei der Index (i) bedeutet, dass wir die i -te Zeile und die i -te Spalte aus der jeweiligen Matrix entfernt haben. Wenn wir $\lambda_i = \lambda$ setzen und die Kettenregel anwenden,⁸ folgt

$$\frac{d}{d\lambda} \det(\lambda I - C) = \sum_i \det(\lambda I - C_{(i)}).$$

Jede der Matrizen $\lambda_A I - A_{(i)}$ hat eine positive Determinante, denn die Determinante ist das Produkt der Eigenwerte, also $\prod_{\sigma} (\lambda_A - \sigma)$, wobei σ die Eigenwerte von $A_{(i)}$ durchläuft. Die Eigenwerte von $A_{(i)}$ sind auch Eigenwerte von A_i und deshalb gilt $|\sigma| < \lambda_A$ nach Schritt 2. Ist σ reell, so gilt $\lambda_A - \sigma > 0$. Ist σ echt komplex, so ist auch die zu σ komplex konjugierte Zahl $\bar{\sigma}$ ein Eigenwert von $A_{(i)}$ mit derselben Vielfachheit. Dies führt nur zu Faktoren der Form

$$(\lambda_A + \sigma)(\lambda_A + \bar{\sigma}) = |\lambda_A - \sigma|^2 > 0$$

in $\prod_{\sigma} (\lambda_A - \sigma)$ und zeigt deshalb, dass $|\det(\lambda_A I - A_{(i)})| > 0$. Folglich ist

$$\left. \frac{d}{d\lambda} \right|_{\lambda=\lambda_A} \det(\lambda I - A) > 0,$$

woraus folgt, dass λ_A keine Nullstelle der Ableitung des charakteristischen Polynoms ist. Damit hat λ_A die algebraische und geometrische Vielfachheit 1. Da, wie wir bereits weiter oben gesehen haben, jeder positive Eigenvektor ein Eigenvektor zum Eigenwert λ_A sein muss, folgt, dass jeder positive Eigenvektor ein Vielfaches von v_A ist. \square

⁸Wir betrachten die Komposition der differenzierbaren Abbildungen $\lambda \mapsto (\lambda, \dots, \lambda) \in \mathbb{R}^r$ und $(\lambda_1, \dots, \lambda_r) \mapsto \det(\text{diag}(\lambda_1, \dots, \lambda_r) - C)$.

5.3 Berechnung der Entropie

Ist X ein irreduzibler Shift von endlichem Typ, so ist er konjugiert zu einem Kantenshift mit irreduzibler Übergangsmatrix A . Aus den Ergebnissen der vorherigen zwei Unterabschnitte können wir folgern, dass $h(X) = \log \lambda_A$. Wir werden außerdem die Entropie eines irreduziblen sofischen Shifts berechnen und zeigen, dass die Wachstumsrate der Anzahl n -periodischen Punkte in diesem Fall mit der Entropie übereinstimmt. Reduzible Shifts werden wir im nächsten Unterabschnitt behandeln.

5.19 Satz: *Es gelten folgende Aussagen:*

- (1) *Ist G ein irreduzibler Graph, so gilt $h(X_G) = \log \lambda_{A(G)}$.*
- (2) *Ist X ein irreduzibler M -Schritt-Shift von endlichem Typ und G der wesentliche Graph mit $X^{[M+1]} = X_G$, dann gilt $h(X) = \log \lambda_{A(G)}$.*

Beweis: Da $A = A(G)$ irreduzibel ist, zeigt der Satz von Perron-Frobenius, dass A einen positiven Eigenvektor zu einem positiven Eigenwert λ_A hat. Daher kann Satz 5.15 angewendet werden, der $h(X_G) = \log \lambda_{A(G)}$ liefert.

Um die zweite Aussage zu beweisen, stellen wir fest, dass $X \cong X^{[M+1]}$ und die Entropie invariant unter Konjugation ist. Also folgt $h(X) = h(X^{[M+1]}) = h(X_G)$. Da X irreduzibel ist, ist es auch $A(G)$ und wir können Aussage (1) anwenden. \square

5.20 Beispiel: Sei $\mathcal{A} = \{0, 1\}$, $\mathcal{F} = \{111\}$ und $X = X_{\mathcal{F}}$. Dann ist X die Menge aller Punkte im vollen 2-Shift, in denen keine drei Einsen aufeinander folgen. Dies ist ein 2-Schritt-Shift. Der Graph G mit $X^{[3]} = X_G$ hat die Knoten 00, 01, 10 und 11 und die irreduzible Übergangsmatrix

$$A = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

(Der Übergang von 11 nach 11 ist verboten, da der Block 111 verboten ist.) Folglich gilt $h(X) = \log \lambda_A$. Das charakteristische Polynom von A ist $\chi_A(t) = t^4 - t^3 - t^2 - t$, dessen Nullstellen $t = 0$, $t \approx -0.41964 \pm 0.60629i$ und $t \approx 1.83929$ sind. Daher gilt $h(X) \approx \log 1.83929 \approx 0.87915$.

Als nächstes betrachten wir sofische Shifts.

5.21 Satz: *Sei X ein irreduzibler sofischer Shift und $\mathcal{G} = (G, \mathcal{L})$ eine irreduzible rechtsauflösende Darstellung von X . Dann gilt $h(X) = \log \lambda_{A(G)}$.*

Beweis: Nach Satz 5.11 gilt $h(X) = h(X_{\mathcal{G}}) = h(X_G)$ und da G irreduzibel ist, folgt aus dem vorherigen Satz, dass $h(X_G) = \log \lambda_{A(G)}$. \square

5.22 Bemerkung: Natürlich kann man den vorherigen Satz auch auf irreduzible Shifts von endlichem Typ anwenden. Also liefern Satz 5.19(2) und Satz 5.21 zwei verschiedene Methoden, um die Entropie eines irreduziblen Shifts von endlichem Typ zu berechnen.

5.23 Satz: *Ist X ein irreduzibler sofischer Shift, dann gilt*

$$h(X) = \limsup_{n \rightarrow \infty} \frac{1}{n} \log p_n(X).$$

Beweis: Die Ungleichung

$$h(X) \geq \limsup_{n \rightarrow \infty} \frac{1}{n} \log p_n(X)$$

gilt für jeden Shift X nach Satz 5.13.

Wir beweisen die umgekehrte Ungleichung zunächst für irreduzible Shifts X von endlichem Typ. Sei G ein irreduzibler Graph, so dass $X_G \cong X$ und folglich $p_n(X) = p_n(X_G)$ und $h(X) = h(X_G)$. Da G irreduzibel ist, gibt es eine Zahl $N \in \mathbb{N}$, so dass es für jedes Paar I, J von Knoten in G einen Pfad von I nach J mit Länge $\leq N$ gibt. Gilt folglich $\pi \in \mathcal{B}_n(X_G)$, so gibt es einen Pfad τ der Länge $\leq N$ von $t(\pi)$ nach $i(\pi)$. Dann ist $\pi\tau$ ein geschlossener Pfad und $(\pi\tau)^\infty$ hat Periode $n + |\tau| \in \{n, n+1, \dots, n+N\}$. Auf diese Art erzeugt jeder Block in $\mathcal{B}_n(X_G)$ einen periodischen Punkt in X_G , dessen Periode zwischen n und $n+N$ liegt, wobei verschiedene Blöcke verschiedene Punkte erzeugen. Daher gilt

$$|\mathcal{B}_n(X_G)| \leq p_n(X_G) + p_{n+1}(X_G) + \dots + p_{n+N}(X_G).$$

Für jedes n sei $k(n)$ so gewählt, dass $p_{n+k(n)}(X_G)$ die größte unter den Zahlen $p_n(X_G), \dots, p_{n+N}(X_G)$ ist. Dann folgt

$$|\mathcal{B}_n(X_G)| \leq (N+1)p_{n+k(n)}(X_G)$$

und daher

$$\begin{aligned} h(X) = h(X_G) &= \lim_{n \rightarrow \infty} \frac{1}{n} \log |\mathcal{B}_n(X_G)| \\ &\leq \limsup_{n \rightarrow \infty} \left[\frac{1}{n} \log(N+1) + \frac{1}{n} \log p_{n+k(n)}(X_G) \right] \\ &\leq \limsup_{m \rightarrow \infty} \frac{1}{m} \log p_m(X_G), \end{aligned}$$

womit das Resultat für irreduzible Shifts von endlichem Typ bewiesen ist.

Nun sei X ein irreduzibler sofischer Shift. Sei $\mathcal{G} = (G, \mathcal{L})$ eine irreduzible rechtsauflösende Darstellung von X . Satz 5.11 liefert $h(X) = h(X_{\mathcal{G}}) = h(X_G)$. G habe r Knoten. Da \mathcal{G} rechtsauflösend ist, hat jeder Punkt in X_G höchstens r Darstellungen in $X_{\mathcal{G}}$. Da \mathcal{L}_∞ n -periodische Punkte in X_G auf n -periodische Punkte in $X_{\mathcal{G}}$ abbildet, folgt

$$p_n(X_{\mathcal{G}}) \geq \frac{1}{r} p_n(X_G).$$

Damit folgt

$$\begin{aligned} \limsup_{n \rightarrow \infty} \frac{1}{n} \log p_n(X_G) &\geq \limsup_{n \rightarrow \infty} \left[\frac{1}{n} \log \frac{1}{r} + \frac{1}{n} \log p_n(X_G) \right] \\ &= \limsup_{n \rightarrow \infty} \frac{1}{n} \log p_n(X_G) = h(X_G) = h(X). \end{aligned}$$

□

Dasselbe Resultat gilt für die Wachstumsrate der Punkte mit Primperiode n .

5.24 Definition: Für einen Shiftraum X sei $q_n(X)$ die Anzahl der periodischen Punkte in X mit Primperiode n .

5.25 Korollar: Ist X ein irreduzibler sofischer Shift, so gilt

$$h(X) = \limsup_{n \rightarrow \infty} \frac{1}{n} \log q_n(X).$$

Beweis: Nach Satz 5.13 gilt

$$0 \leq \limsup_{n \rightarrow \infty} \frac{1}{n} \log q_n(X) \leq \limsup_{n \rightarrow \infty} \frac{1}{n} \log p_n(X) \leq h(X).$$

Ist $h(X) = 0$, so ist nichts mehr zu tun. Also dürfen wir annehmen, dass $h(X) > 0$. Nach Satz 5.23 gibt es eine Teilfolge $n_i \rightarrow \infty$ mit

$$\lim_{i \rightarrow \infty} \frac{1}{n_i} \log p_{n_i}(X) = h(X).$$

Sei $\varepsilon > 0$. Dann gilt für hinreichend große i , dass

$$p_{n_i}(X) > \exp(n_i h(X) - n_i \varepsilon).$$

Nach Satz 5.23 gilt außerdem, dass für hinreichend große m und alle $j \leq m$ gilt:

$$p_j(X) < \exp(mh(X) + m\varepsilon).$$

Würde dies nämlich nicht gelten, so könnten wir eine Folge $m_i \rightarrow \infty$ finden und eine weitere Folge $j_i \leq m_i$ mit

$$\frac{1}{j_i} \log p_{j_i}(X) \geq \frac{m_i}{j_i} (h(X) + \varepsilon) \geq h(X) + \varepsilon.$$

Da $h(X) > 0$, können wir die Folge j_i so wählen, dass $j_i \rightarrow \infty$. Dann folgt

$$\limsup_{i \rightarrow \infty} \frac{1}{j_i} \log p_{j_i}(X) \geq h(X) + \varepsilon$$

im Widerspruch zu Satz 5.23. Aber es gilt auch

$$q_n(X) \geq p_n(X) - \sum_{j=1}^{m(n)} p_j(X), \quad m(n) := \lfloor n/2 \rfloor,$$

da auf der rechten Seite mindestens die Anzahl aller n -periodischen Punkte mit kleinerer Primperiode als n abgezogen wird. Für hinreichend große i folgt

$$\begin{aligned}
q_{n_i}(X) &\geq p_{n_i}(X) - \sum_{j=1}^{\lfloor n_i/2 \rfloor} p_j(X) \\
&> \exp(n_i h(X) - n_i \varepsilon) - \lfloor n_i/2 \rfloor \exp(\lfloor n_i/2 \rfloor h(X) + \lfloor n_i/2 \rfloor \varepsilon) \\
&\geq \exp(n_i h(X) - n_i \varepsilon) [1 - \lfloor n_i/2 \rfloor \exp(-(n_i/2 + 1)h(X) + n_i(3/2)\varepsilon)] \\
&= \exp(n_i h(X) - n_i \varepsilon) [1 - \lfloor n_i/2 \rfloor \exp(-h(X) - n_i(1/2h(X) + 3/2\varepsilon))].
\end{aligned}$$

Für hinreichend kleines ε konvergiert der Faktor in eckigen Klammern gegen 1, wenn $i \rightarrow \infty$, da $h(X) > 0$. Damit folgt

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log q_n(X) \geq \limsup_{i \rightarrow \infty} \frac{1}{n_i} \log q_{n_i}(X) \geq h(X) - \varepsilon,$$

womit der Beweis abgeschlossen ist, da $\varepsilon > 0$ beliebig gewählt war. \square

5.4 Irreduzible Komponenten

Sei G ein beliebiger Graph. Wir werden in diesem Unterabschnitt zeigen, dass G gewisse irreduzible Teilgraphen G_i enthält, die sogenannten *irreduziblen Komponenten* von G , und dass die Untersuchung des Shifts X_G sich im Wesentlichen auf die Untersuchung der irreduziblen Shifts X_{G_i} reduziert. Zum Beispiel werden wir zeigen, dass $h(X_G) = \max_i h(X_{G_i})$ und wir wissen, dass $h(X_{G_i}) = \log \lambda_i$, wobei λ_i der Perron-Eigenwert von $A(G_i)$ ist.

Wir werden wie zuvor Matrizen mit nichtnegativen Einträgen betrachten. Sei A eine solche $m \times m$ -Matrix. Unser Ziel ist es, A durch eine Umnummerierung der Indizes als Block-Dreiecksmatrix darzustellen. Die quadratischen Matrizen A_i entlang der Diagonalen werden die irreduziblen Komponenten von A sein, und das Wachstum von A^n wird kontrolliert durch das Wachstum der A_i^n .

Hat A nichtnegative ganzzahlige Einträge, so haben wir A bereits einen Graphen $G(A) = G_A$ zugeordnet, denjenigen Graphen mit Knotenmenge $\{1, \dots, r\}$ und A_{IJ} verschiedenen Kanten von I nach J . Für eine beliebige nichtnegative Matrix sei $\lceil A \rceil$ die Matrix, die man erhält, indem man jeden Eintrag A_{IJ} von A durch $\lceil A_{IJ} \rceil$ (d.h. der kleinsten ganzen Zahl $\geq A_{IJ}$) ersetzt. Dies ist eine nichtnegative Matrix mit ganzzahligen Einträgen und wir definieren $G(A) = G_A := G(\lceil A \rceil)$.

Für Knoten I, J schreiben wir $I \rightsquigarrow J$, falls $(A^n)_{IJ} > 0$ für ein $n \geq 0$. Also gilt $I \rightsquigarrow J$, falls ein Pfad in G von I nach J existiert. Es gilt stets $I \rightsquigarrow I$, da $(A^0)_{II} = 1$, oder äquivalent dazu, da der leere Pfad ε_I von I nach I verläuft. Wir sagen, dass I mit J *kommuniziert*, falls $I \rightsquigarrow J$ und $J \rightsquigarrow I$. Dies ist eine Äquivalenzrelation und partitioniert deshalb die Knoten von G in Äquivalenzklassen, die wir *kommunizierende Klassen* nennen. Eine kommunizierende Klasse ist daher eine maximale Teilmenge der Knotenmenge, so dass darin jeder Knoten mit jedem kommuniziert.

5.26 Beispiel: Wir betrachten die Matrix

$$A = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \end{matrix} & \begin{pmatrix} 0 & 0 & 2 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \end{matrix}.$$

Die kommunizierenden Klassen sind $C_1 = \{4\}$, $C_2 = \{2, 5\}$, $C_3 = \{7\}$ und $C_4 = \{1, 3, 6\}$, denn

$$\begin{aligned} 2 &\rightsquigarrow 5 \rightsquigarrow 2, \\ 6 &\rightsquigarrow 1 \rightsquigarrow 3 \rightsquigarrow 6, \end{aligned}$$

$\{4\}$ und $\{7\}$ sind offensichtlich Äquivalenzklassen und von 5 aus ist nur die Menge $\{2, 4, 5\}$ erreichbar.

Wir konstruieren einen Graphen \mathcal{H} , dessen Knoten die kommunizierenden Klassen von G sind und für den eine Kante von der Klasse C zur Klasse D verläuft genau dann, wenn $C \neq D$ und wenn es eine Kante in G von einem Knoten in C zu einem Knoten in D gibt. Beachte, dass es in \mathcal{H} keine geschlossenen Pfade gibt, da ein geschlossener Pfad eine größere kommunizierende Klasse erzeugen würde. Hätte jeder Knoten in \mathcal{H} eine ausgehende Kante, so würde man, diesen Kanten sukzessive folgend, einen geschlossenen Pfad erhalten. Daher muss \mathcal{H} mindestens einen Knoten ohne ausgehende Kanten haben, eine *Senke*. Ein ähnliches Argument zeigt, dass \mathcal{H} auch Knoten ohne eingehende Kanten haben muss (*Quellen*).

Indem wir eine Senke nach der anderen entfernen, können wir die kommunizierenden Klassen in eine Ordnung C_1, C_2, \dots, C_k bringen, so dass es nur dann einen Pfad von C_j nach C_i geben kann, wenn $j > i$. Diese Ordnung der Knoten führt zu einer Darstellung von A in Block-Dreiecksform:

$$A = \begin{bmatrix} A_1 & 0 & 0 & \dots & 0 \\ \star & A_2 & 0 & \dots & 0 \\ \star & \star & A_3 & \dots & 0 \\ \vdots & \dots & \vdots & \ddots & \vdots \\ \star & \star & \star & \dots & A_k \end{bmatrix},$$

wobei die \star -Einträge Untermatrizen mit nichtnegativen Einträgen entsprechen. Sei G_i der Teilgraph von G mit Knotenmenge C_i , dessen Kanten diejenigen von G sind, die in C_i beginnen und enden. Dann ist jedes G_i irreduzibel und hat Übergangsmatrix A_i . Wir nennen die G_i die *irreduziblen Komponenten* von G und die A_i die *irreduziblen Komponenten* von A . Manchmal werden wir die Shifts X_{G_i} als die irreduziblen Komponenten von X_G bezeichnen.

Indem wir in Beispiel 5.26 die Ordnung C_1, C_2, C_3, C_4 verwenden, erhalten wir die Matrix

$$\begin{array}{c} 4 \quad 2 \quad 5 \quad 7 \quad 1 \quad 3 \quad 6 \\ \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 2 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 & 0 \end{pmatrix} \end{array}.$$

Das charakteristische Polynom einer Matrix bleibt bei einer simultanen Permutation von Zeilen und Spalten unverändert. Also können wir das charakteristische Polynom mit Hilfe der obigen Block-Dreiecksform berechnen. Dies ergibt

$$\chi_A(t) = \chi_{A_1}(t)\chi_{A_2}(t) \cdots \chi_{A_k}(t).$$

Daher sind die Eigenwerte von A die Eigenwerte der A_i . Insbesondere der betragsmäßig größte Eigenwert λ_{A_i} ist auch ein Eigenwert von A . Dies legt folgende Definition nahe.

5.27 Definition: Sei A eine nichtnegative Matrix mit irreduziblen Komponenten A_1, \dots, A_k . Der Perron-Eigenwert λ_A von A ist $\lambda_A := \max_{1 \leq i \leq k} \lambda_{A_i}$.

5.28 Lemma: Für eine beliebige nichtnegative Matrix A ist der Perron-Eigenwert λ_A der betragsmäßig größte Eigenwert von A .

Beweis: Die Menge der Eigenwerte von A ist die Vereinigung der Eigenwertmengen der Komponenten A_i . Ist also μ ein Eigenwert von A , so ist $|\mu| \leq \lambda_{A_i}$ für ein i , woraus unmittelbar die Behauptung folgt. \square

Ist C eine beliebige (komplexe) $n \times n$ -Matrix, so ist der *Spektralradius* von C der Betrag des betragsgrößten Eigenwerts von C . Das obige Lemma zeigt deshalb, dass der Spektralradius einer nichtnegativen Matrix ihr Perron-Eigenwert ist.

Das folgende Resultat zeigt, dass der Perron-Eigenwert das exponentielle Wachstum von A^n bestimmt. Dies ist eine Verallgemeinerung von Satz 5.19(1) auf allgemeine Graphen.

5.29 Satz: Sei G ein Graph mit Übergangsmatrix A . Dann gilt $h(X_G) = \log \lambda_A$.

Beweis: Seien A_1, A_2, \dots, A_k die irreduziblen Komponenten von A . Dann ist $\lambda_A = \lambda_{A_q}$ für ein q . Da $|\mathcal{B}_n(X_G)| \geq |\mathcal{B}_n(X_{G_q})|$, gilt

$$h(X_G) \geq h(X_{G_q}) = \log \lambda_{A_q} = \log \lambda_A.$$

Beachte, dass aus $\lambda_A = 0$ folgt, dass jedes A_i die 1×1 -Matrix $[0]$ ist (da nach dem Satz von Perron-Frobenius eine irreduzible Matrix ungleich $[0]$ einen positiven

Eigenwert hat). Dann gibt es keine bi-unendlichen Pfade auf G und folglich gilt $X_G = \emptyset$. In diesem Fall gilt also das Resultat. Im Folgenden nehmen wir $\lambda_A > 0$ an.

Wir werden zeigen, dass $h(X_G) \leq \log \lambda_A$, indem wir die Anzahl der Pfade der Länge n in G abschätzen. So ein Pfad lässt sich zerlegen in Teilpfade, die in den kommunizierenden Klassen verlaufen und verbunden sind durch Kanten zwischen den Klassen. Folglich hat jedes $\pi \in \mathcal{B}_n(X_G)$ die Form

$$\pi = \pi_1 e_1 \pi_2 e_2 \dots \pi_{j-1} e_{j-1} \pi_j, \quad (11)$$

wobei π_i ein Pfad in einem Teilgraphen $G_{q(i)}$ ist und e_i eine Kante von einem Knoten in $G_{q(i)}$ zu einem Knoten in $G_{q(i+1)}$. Also gilt $q(1) > q(2) > \dots > q(j)$, insbesondere $j \leq k$. Sei M die Gesamtzahl der Übergangskanten zwischen kommunizierenden Klassen. Dann gibt es höchstens M Möglichkeiten für jedes e_i in (11) und höchstens n Stellen, wo es auftauchen kann. Deshalb ist die Anzahl der möglichen Anordnungen für die Übergangskanten höchstens $(Mn)^k$. Wenn wir so eine Anordnung festgelegt haben, dann ist die Anzahl der Möglichkeiten π_i zu wählen höchstens $|\mathcal{B}_{n(i)}(X_{G_{q(i)}})|$, wobei $n(i) = |\pi_i|$. Nach Satz 5.15 und dem Satz von Perron-Frobenius gibt es ein $d > 0$, so dass für jedes G_q gilt:

$$|\mathcal{B}_m(X_{G_q})| \leq d \lambda_{A_q}^m \leq d \lambda_A^m.$$

Die Anzahl der Möglichkeiten, die Teilpfade π_i einzufügen, ist daher höchstens

$$\prod_{i=1}^j |\mathcal{B}_{n(i)}(X_{G_{q(i)}})| \leq d^j \lambda_A^{n(1)+n(2)+\dots+n(j)} \leq d^k \lambda_A^n.$$

Fügen wir alle Abschätzungen zusammen, so erhalten wir

$$|\mathcal{B}_n(X_G)| \leq (Mn)^k d^k \lambda_A^n.$$

Daher gilt

$$\begin{aligned} h(X_G) &= \lim_{n \rightarrow \infty} \frac{1}{n} \log |\mathcal{B}_n(X_G)| \\ &\leq \lim_{n \rightarrow \infty} \left[\frac{1}{n} \log(M^k d^k) + k \frac{\log n}{n} + \log \lambda_A \right] = \log \lambda_A, \end{aligned}$$

womit der Beweis abgeschlossen ist. \square

5.30 Beispiel: Sei $A = \begin{bmatrix} 2 & 0 \\ 1 & 2 \end{bmatrix}$. Dann ist A bereits in Dreiecksform und

$$A^n = \begin{bmatrix} 2^n & 0 \\ n2^{n-1} & 2^n \end{bmatrix},$$

so dass $|\mathcal{B}_n(X_A)| = (n/2 + 2)2^n$. Beachte, dass hier keine Konstante d auftaucht, so dass $|\mathcal{B}_n(X_A)| \leq d2^n$, dass aber der polynomiale Faktor $n/2 + 2$ keine Auswirkung auf die Wachstumsrate $h(X_A) = \log \lambda_A = \log 2 = 1$ hat.

Das obige Theorem zeigt auch, wie man die Entropie eines (möglicherweise reduzierbaren) sofischen Shifts berechnen kann. Ist nämlich $\mathcal{G} = (G, \mathcal{L})$ eine rechtsauflösende Darstellung von $X = X_{\mathcal{G}}$, so ist nach Satz 5.11 $h(X_{\mathcal{G}}) = h(X_G) = \log \lambda_{A(G)}$. Dies zeigt auch, dass jeder sofische Shift einen irreduzierbaren sofischen Shift von maximaler Entropie enthält.

Wie können wir die Entropie eines beliebigen Shifts X berechnen? Eine Idee ist, Shifts von endlichem Typ als äußere Approximationen von X wie folgt zu verwenden. Sei $\mathcal{F} = \{w_1, w_2, w_3, \dots\}$ eine (abzählbar) unendliche Menge verbotener Blöcke, für die $X = X_{\mathcal{F}}$ gilt und sei $\mathcal{F}_k := \{w_1, w_2, \dots, w_k\}$. Dann sind $X_{\mathcal{F}_1} \supseteq X_{\mathcal{F}_2} \supseteq X_{\mathcal{F}_3} \supseteq \dots$ Shifts von endlichem Typ und

$$\bigcap_{k=1}^{\infty} X_{\mathcal{F}_k} = X_{\bigcup_{k=1}^{\infty} \mathcal{F}_k} = X_{\mathcal{F}} = X.$$

Wir können $h(X_{\mathcal{F}_k})$ für jedes \mathcal{F}_k berechnen, indem wir das obige Theorem auf den Kantenshift X_{G_k} anwenden, wobei G_k der einer rechtsauflösenden Darstellung von $X_{\mathcal{F}_k}$ zugrundeliegende Graph ist. Das nächste Resultat zeigt, dass $h(X_{\mathcal{F}_k})$ gegen $h(X)$ konvergiert.

5.31 Satz: Seien $X_1 \supseteq X_2 \supseteq X_3 \supseteq \dots$ Shifträume, deren Durchschnitt wir mit X bezeichnen. Dann gilt $h(X_k) \rightarrow h(X)$ für $k \rightarrow \infty$.

Beweis: Sei $\varepsilon > 0$. Wir zeigen, dass $h(X) \leq h(X_k) \leq h(X) + \varepsilon$ für alle hinreichend großen k . Da $X \subseteq X_k$, gilt $h(X) \leq h(X_k)$ für alle k . Da $(1/n) \log |\mathcal{B}_n(X)|$ gegen $h(X)$ konvergiert, gibt es ein N , so dass

$$\frac{1}{N} \log |\mathcal{B}_N(X)| < h(X) + \varepsilon.$$

Da $X = \bigcap_{k=1}^{\infty} X_k$, gibt es ein K , so dass $\mathcal{B}_N(X_k) = \mathcal{B}_N(X)$ für alle $k \geq K$. Andernfalls würde es eine Folge $k_l \rightarrow \infty$ und eine Folge von N -Blöcken $w_l \in \mathcal{B}_N(X_{k_l}) \setminus \mathcal{B}_N(X)$ geben. Da es nur endlich viele N -Blöcke über einem endlichen Alphabet gibt, existiert ein N -Block w in $\bigcap_{l=1}^{\infty} \mathcal{B}_N(X_{k_l})$ mit $w \notin \mathcal{B}_N(X)$. Sei x_l ein Punkt in X_{k_l} mit $(x_l)_{[0, N-1]} = w$. Dann gibt es eine unendliche Menge $S_0 \subseteq \mathbb{N}$, so dass die 0-ten Komponenten $(x_l)_{[0]}$ für alle $l \in S_0$ übereinstimmen. Ferner gibt es eine unendliche Menge $S_1 \subseteq S_0$, so dass die zentralen 3-Blöcke $(x_l)_{[-1, 1]}$ für alle $l \in S_1$ übereinstimmen. So fortfahrend, erhalten wir unendliche Mengen $S_0 \supseteq S_1 \supseteq S_2 \supseteq \dots$, so dass die Blöcke $(x_l)_{[-k, k]}$ für alle $l \in S_k$ identisch sind. Wir definieren einen Punkt x durch $x_{[-k, k]} := (x_l)_{[-k, k]}$ für alle $k \geq 0$ und $l \in S_k$. Nach Konstruktion ist x wohldefiniert. Ferner gilt $x_{[0, N-1]} = w$ und $x \in \bigcap_{l=1}^{\infty} X_{k_l} = \bigcap_{k=1}^{\infty} X_k = X$, da jeder Block u , der in x vorkommt, in einem zentralen Block $x_{[-k, k]}$ mit hinreichend großem k vorkommt und daraus $u \in \mathcal{B}(X_{k_l})$ für alle $l \in S_k$, also für beliebig große l , folgt. Dies ist ein Widerspruch zu $w \notin \mathcal{B}_N(X)$.

Nach Satz 5.6 gilt für alle $k \geq K$, dass

$$\begin{aligned} h(X_k) &= \inf_{n \geq 1} \frac{1}{n} \log |\mathcal{B}_n(X_k)| \leq \frac{1}{N} \log |\mathcal{B}_N(X_k)| \\ &= \frac{1}{N} \log |\mathcal{B}_N(X)| < h(X) + \varepsilon, \end{aligned}$$

womit der Beweis abgeschlossen ist. \square

Leider hat die Berechnung von $h(X_{\mathcal{F}})$ als Grenzwert der $h(X_{\mathcal{F}_k})$ zwei ernsthafte Nachteile. Der erste ist ein praktischer: Die Größe der Übergangsmatrix des Graphen, der $X_{\mathcal{F}_k}$ darstellt, kann rapide anwachsen, was es immer schwerer macht, den Perron-Eigenwert mit ausreichender Genauigkeit zu berechnen. Der zweite ist ein theoretischer: Wir wissen nicht, wie groß k gewählt werden muss, so dass $h(X_{\mathcal{F}_k})$ die Entropie von X mit einer gegebenen Toleranz approximiert.

Eine Möglichkeit mit dem zweiten Problem umzugehen ist es, „innere“ sofische Approximationen $Y \subseteq X_{\mathcal{F}}$ zu bestimmen. Ist nämlich $h(X_{\mathcal{F}_k}) - h(Y)$ klein, so haben wir den Wert von $h(X)$ eingegrenzt, da $h(Y) \leq h(X) \leq h(X_{\mathcal{F}_k})$. Solche inneren Approximationen zu finden ist allerdings nicht immer einfach. Zum Beispiel gibt es einen Shift X mit $h(X) > 0$ ohne periodische Punkte. Ein solcher Shift enthält keine nichtleeren sofischen Shifts, seine Entropie kann also nicht von innen approximiert werden.

Als nächstes überlegen wir uns, wie Änderungen in der Übergangsmatrix die Entropie beeinflussen. Sind A und B nichtnegative Matrizen mit $B \leq A$, dann ist der Graph $G(B)$ ein Teilgraph von $G(A)$, also gilt $\lambda_B \leq \lambda_A$. Wir beweisen als nächstes eine schärfere Aussage, indem wir zeigen, dass $\lambda_B < \lambda_A$, falls A irreduzibel ist und B mindestens einen Eintrag hat, der echt kleiner als der entsprechende Eintrag in A ist.

5.32 Satz: Sei A irreduzibel und $0 \leq B \leq A$. Ferner sei $B_{KL} < A_{KL}$ für ein Indexpaar (K, L) . Dann ist $\lambda_B < \lambda_A$.

Beweis: Da A irreduzibel ist, gibt es ein N , so dass für jeden Knoten I ein ausgehender Pfad der Länge N existiert, der eine Kante von K nach L enthält. Also gibt es für jedes I ein J , so dass $(B^N)_{IJ} < (A^N)_{IJ}$.

Sei $v = v_A > 0$ der Perron-Eigenvektor von A und λ_A der Perron-Eigenwert. Dann ist $B^N v < A^N v = \lambda_A^N v$ und folglich gibt es ein $\varepsilon > 0$ mit

$$B^N v \leq (1 - \varepsilon) \lambda_A^N v.$$

Folglich gilt für jedes $k \geq 1$, dass

$$(B^N)^k v \leq [(1 - \varepsilon) \lambda_A^N]^k v.$$

Ein zu Satz 5.15 analoges Argument zeigt die Existenz einer Konstanten $d > 0$ mit

$$\sum_{I,J} (B^{Nk})_{IJ} \leq d [(1 - \varepsilon) \lambda_A^N]^k = d [(1 - \varepsilon)^{1/N} \lambda_A]^{Nk}. \quad (12)$$

Nun sei C eine irreduzible Komponente von B mit $\lambda_C = \lambda_B$. Dann hat C einen positiven Perron-Eigenvektor und nach Satz 5.15 gibt es eine Konstante $c > 0$ mit

$$c\lambda_B^{Nk} \leq \sum_{I,J} (C^{Nk})_{IJ} \leq \sum_{I,J} (B^{Nk})_{IJ}. \quad (13)$$

Vergleichen wir (12) und (13), ziehen die (Nk) -te Wurzel und lassen k gegen unendlich gehen, so erhalten wir

$$\lambda_B \leq (1 - \varepsilon)^{1/N} \lambda_A < \lambda_A,$$

womit der Beweis abgeschlossen ist. \square

5.33 Beispiel: Seien $A = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$ und $B = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$. Dann gilt $B \leq A$ und $B_{11} < A_{11}$, aber dennoch $\lambda_B = \lambda_A = 2$. Dies zeigt, dass die Irreduzibilitätsannahme in obigem Satz notwendig ist.

5.34 Korollar: *Ist X ein irreduzibler sofischer Shift und Y ein echter Teilshift von X , dann gilt $h(Y) < h(X)$.*

Beweis: Wir nehmen zuerst an, dass X von endlichem Typ ist. Wir dürfen annehmen, dass X ein Kantenshift ist. Da $\mathcal{B}(Y) \neq \mathcal{B}(X)$, gibt es einen Block $w = a_1 \dots a_N$, der in X aber nicht in Y auftaucht. Sei $\beta_N : X \rightarrow X^{[N]}$ die Konjugation vom Shift X zu seinem N -ten höheren Blockshift. Nach Satz 3.27 ist $X^{[N]}$ auch ein Kantenshift. Wir bezeichnen seine Übergangsmatrix mit A . Sei B die Matrix, die aus A entsteht, indem die 1 für den Übergang von $a_1 \dots a_{N-1}$ nach $a_2 \dots a_N$ durch eine 0 ersetzt wird. Da A irreduzibel ist, zeigt der vorherige Satz, dass $\lambda_B < \lambda_A$. Daher gilt

$$h(Y) = h(\beta_N(Y)) \leq h(X_B) = \log \lambda_B < \log \lambda_A = h(X_A) = h(X).$$

Jetzt nehmen wir an, dass X sofisch ist und bezeichnen mit $\mathcal{G} = (G, \mathcal{L})$ eine irreduzible rechtsauflösende Darstellung von X . Dann gilt $h(X) = h(X_{\mathcal{G}}) = h(X_G)$ nach Satz 5.11. Da $\mathcal{L}_{\infty} : X_G \rightarrow X_{\mathcal{G}}$ ein Faktorcode ist, ist $\tilde{Y} := \mathcal{L}_{\infty}^{-1}(Y)$ ein echter Teilshift von X_G , woraus nach obiger Argumentation $h(\tilde{Y}) < h(X_G)$ folgt. Aber die Entropie kann bei Anwendung eines Faktorcodes nicht größer werden, so dass

$$h(Y) \leq h(\tilde{Y}) < h(X_G) = h(X_{\mathcal{G}}) = h(X).$$

\square

5.5 Zyklische Struktur

In diesem Abschnitt untersuchen wir die zyklische Struktur irreduzibler Matrizen. Wir werden zeigen, dass jede irreduzible Matrix A eine charakteristische

Periode p hat und die Knoten des zugehörigen Graphen in Gruppen unterteilt werden können, die zyklisch mit Periode p permutiert werden. Matrizen mit Periode 1 heißen *aperiodisch*. Wir werden zeigen, dass die Untersuchung irreduzibler Matrizen in gewisser Weise auf die Untersuchung aperiodischer Matrizen reduziert werden kann.

5.35 Beispiel: Sei

$$A = \begin{bmatrix} 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 \end{bmatrix}. \quad (14)$$

Wir stellen fest, dass man von den Knoten 1 und 2 nur zu 3 und 4 gelangen kann, und von diesen nur zu 5, von wo aus man nur zurück nach 1 und 2 gelangen kann. Daher können wir die Knotenmenge in die Teilmengen $D_0 = \{1, 2\}$, $D_1 = \{3, 4\}$ und $D_2 = \{5\}$ zerlegen, so dass nur Übergänge der Form $D_0 \rightarrow D_1 \rightarrow D_2 \rightarrow D_0$ auftauchen können. Das heißt, die Mengen D_i werden zyklisch permutiert mit Periode 3. Darüberhinaus hat

$$A^3 = \begin{bmatrix} 2 & 2 & 0 & 0 & 0 \\ 2 & 2 & 0 & 0 & 0 \\ 0 & 0 & 2 & 2 & 0 \\ 0 & 0 & 2 & 2 & 0 \\ 0 & 0 & 0 & 0 & 4 \end{bmatrix}.$$

Blockdiagonalform, wobei jede Komponente keine weitere Zerlegung mehr zulässt. Dies ist ein Beispiel für die zyklische Struktur, nach der wir suchen.

5.36 Definition: Sei A eine nichtnegative Matrix. Die Periode des Knotens I , die wir mit $\text{per}(I)$ bezeichnen, ist definiert als der größte gemeinsame Teiler derjenigen Zahlen $n \geq 1$ mit $(A^n)_{II} > 0$. Gibt es keine solchen Zahlen, so definieren wir $\text{per}(I) := \infty$. Die Periode $\text{per}(A)$ der Matrix A ist definiert als der größte gemeinsame Teiler derjenigen Zahlen $\text{per}(I)$, die endlich sind, oder als ∞ , falls $\text{per}(I) = \infty$ für alle I gilt. Eine Matrix heißt *aperiodisch*, falls sie Periode 1 hat. Die Periode $\text{per}(G)$ eines Graphen G ist die Periode seiner Übergangsmatrix.

5.37 Satz: Ist A irreduzibel, so haben alle Knoten dieselbe Periode, also ist $\text{per}(A)$ die Periode aller Knoten.

Beweis: Sei I ein Knoten und $p := \text{per}(I)$. Ist $p = \infty$, so gilt $A = [0]$, da A irreduzibel ist, und wir sind fertig. Sei also $p < \infty$. Sei J ein anderer Knoten. Dann gibt es r, s , so dass $(A^r)_{IJ} > 0$ und $(A^s)_{JI} > 0$. Sei n so gewählt, dass $(A^n)_{JJ} > 0$. Dann folgt

$$(A^{r+s})_{II} \geq (A^r)_{IJ}(A^s)_{JI} > 0$$

und

$$(A^{r+n+s})_{II} \geq (A^r)_{IJ}(A^n)_{II}(A^s)_{IJ} > 0.$$

Also teilt p sowohl $r+s$ als auch $r+n+s$, und daher auch ihre Differenz n . Dies zeigt, dass p alle n mit $(A^n)_{JJ} > 0$ teilt, so dass p ihren größten gemeinsamen Teiler $\text{per}(J)$ teilt. Indem wir die Rollen von I und J vertauschen, folgt, dass $\text{per}(J)$ ein Teiler von $\text{per}(I)$ ist, also $\text{per}(I) = \text{per}(J)$. \square

Es gibt einen verwandten Periodenbegriff für Shifträume, der auf periodischen Punkten basiert.

5.38 Definition: Sei X ein Shiftraum. Die Periode von X , die wir mit $\text{per}(X)$ bezeichnen, ist der größte gemeinsame Teiler aller Zahlen $n \geq 1$, so dass $p_n(X) > 0$, oder ∞ , falls es keine solchen Zahlen gibt.

Gilt $X \cong Y$, so ist $p_n(X) = p_n(Y)$ für alle $n \geq 1$, also auch $\text{per}(X) = \text{per}(Y)$. Folglich ist $\text{per}(X)$ eine Invariante unter Konjugation. Da ein Shift von endlichem Typ konjugiert zu einem Kantenshift ist, können wir seine Periode bestimmen durch die Berechnung der Periode eines Kantenshifts.

5.39 Satz: Ist G ein Graph, so gilt $\text{per}(X_G) = \text{per}(G)$.

Beweis: Die Perioden der periodischen Punkte in X_G sind die Längen der geschlossenen Pfade in G . \square

Sei $A \neq [0]$ eine irreduzible Matrix, so dass $p = \text{per}(A) < \infty$. Unser nächstes Ziel ist es, zu zeigen, dass die Knotenmenge von $G(A)$ in p Klassen aufgeteilt werden kann, die zyklisch permutiert werden.

Wir sagen, dass I und J *Perioden-äquivalent* sind, in Zeichen $I \sim J$, falls es einen Pfad in $G(A)$ von I nach J gibt, dessen Länge durch p teilbar ist. Um zu zeigen, dass \sim eine Äquivalenzrelation ist, bemerken wir zunächst, dass ε_I ein Pfad der Länge 0 von I nach I ist und da 0 durch p teilbar ist, folgt $I \sim I$. Ist $I \sim J$, so gibt es einen Pfad π von I nach J , so dass $|\pi|$ durch p teilbar ist. Da A irreduzibel ist, gibt es einen Pfad τ von J nach I . Dann ist $\pi\tau$ ein geschlossener Pfad bei I , so dass $p = \text{per}(A) = \text{per}(I)$ die Zahl $|\pi\tau| = |\pi| + |\tau|$ teilt. Also ist p ein Teiler von $|\tau|$, so dass $J \sim I$. Gilt schließlich $I \sim J$ und $J \sim K$, so sehen wir durch die Verkettung entsprechender Pfade, dass $I \sim K$. Also ist \sim eine Äquivalenzrelation. Wir nennen die Äquivalenzklassen *Periodenklassen*. In Beispiel 5.35 sind $D_0 = \{1, 2\}$, $D_1 = \{3, 4\}$ und $D_2 = \{5\}$ die Periodenklassen.

5.40 Satz: Sei $A \neq [0]$ eine irreduzible Matrix mit Periode p . Dann gibt es genau p Periodenklassen, die in eine Ordnung D_0, D_1, \dots, D_{p-1} gebracht werden können, so dass jede Kante, die in D_i beginnt, in D_{i+1} endet bzw. in D_0 , falls $i = p - 1$.

Beweis: Beginne mit einer beliebigen Periodenklasse D_0 und definiere D_{i+1} induktiv als die Menge aller Endknoten von Kanten, die in D_i beginnen. Den

Beweis, dass die D_i Periodenklassen sind, dass sie die Menge der Knoten partitionieren und dass jede in D_i beginnende Kante in D_{i+1} endet, wobei $D_p = D_0$, verlagern wir in eine Übungsaufgabe. \square

Durch die Verwendung der Periodenklassen einer irreduziblen Matrix A mit Periode p zur Indizierung der Knoten kann die Matrix in die Form

$$\begin{array}{c} D_0 \\ D_1 \\ \vdots \\ D_{p-2} \\ D_{p-1} \end{array} \begin{pmatrix} D_0 & D_1 & D_2 & \dots & D_{p-1} \\ 0 & B_0 & 0 & \dots & 0 \\ 0 & 0 & B_1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & B_{p-2} \\ B_{p-1} & 0 & 0 & \dots & 0 \end{pmatrix}$$

gebracht werden. Dann hat A^p die Blockdiagonalform

$$\begin{array}{c} D_0 \\ D_1 \\ D_2 \\ \vdots \\ D_{p-1} \end{array} \begin{pmatrix} D_0 & D_1 & D_2 & \dots & D_{p-1} \\ A_0 & 0 & 0 & \dots & 0 \\ 0 & A_1 & 0 & \dots & 0 \\ 0 & 0 & A_2 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & A_{p-1} \end{pmatrix} \quad (15)$$

mit $A_i = B_i \dots B_{p-1} B_0 \dots B_{i-1}$.

Ist $I \sim J$, dann gibt es einen Pfad in $G(A)$ von I nach J , dessen Länge durch p teilbar ist, so dass jeder der Diagonalblöcke in (15) irreduzibel ist. Außerdem ist jedes A_i aperiodisch, da sonst die Periode von A größer wäre als p .⁹ Irreduzible aperiodische Matrizen sind die grundlegenden „Bausteine“ nichtnegativer Matrizen.

5.41 Definition: Eine Matrix heißt primitiv, falls sie irreduzibel und aperiodisch ist. Ein Graph heißt primitiv, falls seine Übergangsmatrix primitiv ist.

Die Diagonalblöcke in (15) sind daher primitiv, so wie jede strikt positive Matrix.

5.42 Satz: Sei A eine nichtnegative Matrix. Dann sind folgende Aussagen äquivalent:

- (1) A ist primitiv.
- (2) $A^N > 0$ für ein $N \geq 1$.
- (3) $A^N > 0$ für alle hinreichend großen N .

⁹Ist $\text{ggT}(n : (A_i^n)_{II} > 0) =: d \geq 2$, so teilt d jede der Zahlen np mit $(A^{np})_{II} > 0$. Dies sind alle Zahlen m mit $(A^m)_{II} > 0$, also folgt $\text{per}(A) = \text{ggT}(m : (A^m)_{II} > 0) \geq dp > p$.

Beweis: „(2) \Rightarrow (1)“: Sei $A^N > 0$ für ein $N \geq 1$. Dann gibt es einen Pfad der Länge N in $G(A)$ von I nach J für alle I, J . Also ist A irreduzibel. Insbesondere hat A keine Nullzeilen, woraus $A^{N+1} = AA^N > 0$ folgt. Also teilt $\text{per}(A)$ sowohl N als auch $N + 1$, so dass A Periode 1 hat und damit primitiv ist.

„(1) \Rightarrow (3)“: Sei A primitiv. Wir zeigen zuerst, dass für jeden Knoten I ein $N_I \geq 1$ existiert mit $(A^n)_{II} > 0$ für alle $n \geq N_I$. Sei dazu $R_I := \{n \geq 1 : (A^n)_{II} > 0\}$. Nach Definition ist $\text{per}(I) = \text{per}(A) = 1$ der größte gemeinsame Teiler der Zahlen in R_I . Also gibt es Zahlen $n_i, m_j \in R_I$ und $a_i, b_j \in \mathbb{N}$ mit

$$1 = \sum_{i=1}^k a_i m_i - \sum_{j=1}^l b_j n_j.$$

(siehe Übungsaufgabe 2 auf Blatt 11). Sei K die erste und L die zweite Summe, so dass $1 = K - L$. Sei $N_I := L^2$. Falls $n \geq N_I$, dann gilt $n = cL + d$, wobei $0 \leq d < L$ und $c \geq L$ (Teilen mit Rest). Daher gilt

$$n = cL + d = cL + d(K - L) = (c - d)L + dK,$$

wobei $c - d \geq L - d > 0$ und $d \geq 0$. Folglich ist

$$n = \sum_{j=1}^l [(c - d)b_j] n_j + \sum_{i=1}^k [da_i] m_i$$

eine Summe von Zahlen in R_I , und damit selbst in R_I . Dies zeigt, dass R_I alle $n \geq N_I$ enthält.

Um den Beweis zu vervollständigen, verwenden wir die Irreduzibilität von A , um ein M so zu wählen, dass für jedes Paar I, J von Knoten ein Pfad von I nach J der Länge $\leq M$ existiert. Definiere $N := M + \max_I N_I$. Dann gibt es für beliebige I, J einen Pfad der Länge $l \leq M$ von I nach J , und da $N - l \geq N_I$, gibt es eine Schleife der Länge $N - l$ bei J , daher auch einen Pfad der Länge N von I nach J . Dies zeigt, dass $A^N > 0$. Da aber A primitiv ist, kann keine Zeile von A eine Nullzeile sein und es gilt $A^n > 0$ für alle $n \geq N$.

„(3) \Rightarrow (2)“: offensichtlich. □

Die irreduziblen Graphen entsprechen den irreduziblen Shifts von endlichem Typ in folgendem Sinne: (1) für einen wesentlichen Graphen G ist der Kantenshift X_G irreduzibel genau dann, wenn G irreduzibel ist (Satz 3.21), und (2) jeder irreduzible Shift von endlichem Typ kann rekodiert werden zu einem irreduziblen Kantenshift (Satz 3.23 und Satz 3.21). Dies führt zur Frage, welche Klasse der Shifts von endlichem Typ den primitiven Graphen entsprechen.

5.43 Definition: *Ein Shift X heißt mischend, falls für jedes geordnete Paar $u, v \in \mathcal{B}(X)$ ein N existiert, so dass für jedes $n \geq N$ ein Block $w \in \mathcal{B}_n(X)$ existiert mit $uwv \in \mathcal{B}(X)$.*

5.44 Satz: *Es gelten folgende Aussagen:*

- (1) Die Mischungseigenschaft ist invariant unter Konjugation (in der Tat ist ein Faktor eines mischenden Shifts auch mischend).
- (2) Ist G ein wesentlicher Graph, dann ist der Kantenshift X_G mischend genau dann, wenn G primitiv ist.
- (3) Ein Shift von endlichem Typ ist mischend genau dann, wenn er konjugiert zu einem Kantenshift X_G mit primitivem Graphen G ist.
- (4) Ein Shift von endlichem Typ ist mischend genau dann, wenn er irreduzibel ist und $\text{per}(X) = 1$, d.h. der größte gemeinsame Teiler der Perioden seiner periodischen Punkte ist 1.

Wir verlagern den Beweis dieses Satzes in eine Übungsaufgabe.

Das Perron-Frobenius-Theorem sagt uns, dass $|\mu| \leq \lambda_A$ für alle Eigenwerte μ einer irreduziblen Matrix A . Ist A primitiv, so kann man noch mehr aussagen.

5.45 Satz: Ist A eine primitive Matrix und $\mu \neq \lambda_A$ ein Eigenwert von A , so gilt $|\mu| < \lambda_A$.

Beweis: Sei $\lambda := \lambda_A$ und μ ein Eigenwert von A mit $|\mu| = \lambda$. Wir zeigen, dass dann $\mu = \lambda$. Dazu sei w ein zu μ gehöriger Eigenvektor. Sei $|w|$ der Vektor, den man aus w erhält, indem man jeden Eintrag durch seinen komplexen Betrag ersetzt. Da $Aw = \mu w$ und $A \geq 0$, folgt durch Anwendung der Dreiecksungleichung in jeder Komponente, dass

$$\lambda|w| = |\mu w| = |Aw| \leq A|w|. \quad (16)$$

Als nächstes zeigen wir, dass diese Ungleichung in der Tat eine Gleichheit ist, d.h. $A|w| = \lambda|w|$. Wäre das nicht so, dann wäre $A|w| - \lambda|w|$ nichtnegativ und hätte mindestens einen positiven Eintrag. Da A primitiv ist, gilt $A^N > 0$ für ein $N \geq 1$. Daher gilt $A^N(A|w| - \lambda|w|) > 0$, so dass $u > 0$ und $Au > \lambda u$ für $u := A^N|w|$. Deshalb gibt es ein $\varepsilon > 0$, so dass $Au \geq (\lambda + \varepsilon)u$, und folglich $A^n u \geq (\lambda + \varepsilon)^n u$ für alle $n \geq 1$. Da $u > 0$, folgt dann wie im Beweis von Satz 5.15, dass es ein $c > 0$ gibt mit

$$c(\lambda + \varepsilon)^n \leq \sum_{I,J=1}^r (A^n)_{IJ}.$$

Aber daraus folgt $h(X_A) \geq \log(\lambda + \varepsilon)$, ein Widerspruch.

Also gilt Gleichheit in (16), so dass $|w| = tv_A$ für ein $t > 0$ und $|A^N w| = A^N|w|$. Folglich gilt für jedes I , dass

$$\left| \sum_{J=1}^r (A^N)_{IJ} w_J \right| = \sum_{J=1}^r (A^N)_{IJ} |w_J|. \quad (17)$$

Die Dreiecksungleichung gilt mit Gleichheit nur unter sehr speziellen Voraussetzungen: sind nämlich z_1, \dots, z_r komplexe Zahlen, dann gilt

$$|z_1 + \dots + z_r| = |z_1| + \dots + |z_r|$$

genau dann, wenn es eine reelle Zahl θ mit $z_k = e^{i\theta}|z_k|$ für alle k gibt. Da $A^N > 0$, folgern wir aus (17), dass $w = e^{i\theta}|w| = te^{i\theta}v_A$, woraus $\mu = \lambda_A$ folgt. \square

Ist A irreduzibel mit Periode > 1 und gilt $\lambda_A > 1$, dann oszilliert $(A^n)_{IJ}$ zwischen 0 und exponentiell großen Zahlen, wenn $n \rightarrow \infty$ (siehe Beispiel 5.14). Wir werden das vorherige Theorem verwenden, um zu zeigen, dass für eine primitive Matrix A die Einträge von A^n beständig anwachsen mit einer vorhersagbaren Rate.

Bevor wir unser Resultat formulieren, führen wir eine neue Terminologie ein. Wir werden sowohl den Perron-Eigenvektor w_A für A^T als auch den Perron-Eigenvektor v_A für A verwenden müssen. Um die Transpositionsnotation $(\cdot)^T$ zu vermeiden, werden wir w_A als Zeilenvektor und v_A als Spaltenvektor betrachten. Wir nennen w_A den *Links-Perron-Eigenvektor*, um ihn von v_A zu unterscheiden, den wir den *Rechts-Perron-Eigenvektor* nennen.

5.46 Satz: Sei A eine primitive Matrix mit Perron-Eigenwert λ . Sei v der Rechts- und w der Links-Perron-Eigenvektor von A , d.h. $v, w > 0$ und $Av = \lambda v$, $wA = \lambda w$. Außerdem nehmen wir an, dass $w \cdot v = 1$.¹⁰ Dann gilt

$$(A^n)_{IJ} = [(v_I w_J) + \rho_{IJ}(n)] \lambda^n$$

für alle I und J , wobei $\rho_{IJ}(n) \rightarrow 0$ für $n \rightarrow \infty$.

Beweis: Sei W der eindimensionale Unterraum von Zeilenvektoren, der von w erzeugt wird und sei $U := \{u : u \cdot v = 0\}$ der Unterraum von Zeilenvektoren, die senkrecht auf v stehen. Schreibe $WA = \{xA : x \in W\}$ und $UA := \{uA : u \in U\}$. Dann gilt $WA \subseteq W$, da w ein Eigenvektor ist, und $UA \subseteq U$, da aus $u \in U$ folgt, dass

$$(uA) \cdot v = u \cdot (\lambda v) = \lambda(u \cdot v) = 0.$$

Wir beobachten, dass $\lambda^{-1}A$ den Perron-Eigenwert 1 hat mit Links-Perron-Eigenvektor w , während der vorherige Satz zeigt, dass die Eigenwerte der Einschränkung von $\lambda^{-1}A$ auf U alle Betrag < 1 haben. Die Potenzen jeder Matrix mit Eigenwerten < 1 konvergieren gegen die Nullmatrix (Übungsaufgabe). Daher konvergiert die Einschränkung von $(\lambda^{-1}A)^n$ auf U (betrachtet als eine Matrix bzgl. einer beliebigen Basis) gegen die Nullmatrix. Sei m die Dimension von A . Wähle eine Basis $\{x^{(1)}, \dots, x^{(m)}\}$ von \mathbb{R}^m , die wir uns als bestehend aus Zeilenvektoren denken, so dass $x^{(1)} = w$ und $\{x^{(2)}, \dots, x^{(m)}\}$ eine Basis von U

¹⁰Wir können w und v so normieren, dass $w \cdot v = 1$ gilt, da $w \cdot Av = \lambda(w \cdot v) \neq 0$ und damit $w \cdot v \neq 0$.

ist. Folglich können wir jeden Vektor $z \in \mathbb{R}^m$ darstellen als

$$z = \sum_{i=1}^m a_i x^{(i)},$$

und $z(\lambda^{-1}A)^n$ konvergiert gegen $a_1 x^{(1)} = a_1 w$. Nach Definition von U und der Annahme, dass $w \cdot v = 1$ folgt, dass $a_1 = z \cdot v$. Also konvergiert $z(\lambda^{-1}A)^n$ gegen $(z \cdot v)w$.

Sei e_I der I -te Standard-Basisvektor (als Zeilen- oder Spaltenvektor betrachtet). Dann gilt

$$(\lambda^{-n}A^n)_{IJ} = e_I [(\lambda^{-1}A)^n] \cdot e_J \rightarrow (e_I \cdot v)w \cdot e_J = v_I w_J.$$

Es folgt $\rho_{IJ}(n) := (\lambda^{-n}A^n)_{IJ} - v_I w_J \rightarrow 0$ für $n \rightarrow \infty$. \square

Das folgende Resultat ist eine Verschärfung von Satz 5.23 und Korollar 5.25 für primitive Matrizen.

5.47 Korollar: *Es gelten folgende Aussagen:*

(1) *Für eine primitive ganzzahlige Matrix A existieren die folgenden Grenzwerte und sind identisch mit $\log \lambda_A$:*

- (a) $\lim_{n \rightarrow \infty} (1/n) \log p_n(X_A)$,
- (b) $\lim_{n \rightarrow \infty} (1/n) \log q_n(X_A)$,
- (c) $\lim_{n \rightarrow \infty} (1/n) \log (A^n)_{IJ}$ für alle I, J .

(2) *Für einen mischenden sofischen Shift X gilt*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log p_n(X) = \lim_{n \rightarrow \infty} \frac{1}{n} \log q_n(X) = h(X).$$

Beweis: Aussage (1) folgt unmittelbar aus Satz 5.46. Ist X ein mischender Shift von endlichem Typ, so können wir Satz 5.44(3) verwenden, um X zu einem Kantenshift X_G mit primitiver Matrix A zu rekodieren. Um den sofischen Fall zu behandeln, betrachten wir die minimale rechtsauflösende Darstellung (siehe Übungsaufgabe 1 auf Blatt 12). \square

Wir erinnern daran, dass der Winkel $\theta_{x,y}$ zwischen zwei Vektoren $x, y \in \mathbb{R}^m \setminus \{0\}$ durch die Gleichung

$$\cos(\theta_{x,y}) = \frac{x \cdot y}{\|x\|_2 \|y\|_2} \tag{18}$$

bestimmt ist, wobei $\|\cdot\|_2$ die Euklidische Norm bezeichnet: $\|x\|_2 = \sqrt{x_1^2 + \dots + x_m^2}$. Das folgende Resultat ist eine geometrische Interpretation von Satz 5.46.

5.48 Korollar: Sei A eine primitive $m \times m$ -Matrix. Sei $(\mathbb{R}^m)^+ = \{z \in \mathbb{R}^m : z \geq 0\}$ der positive Orthant von \mathbb{R}^m , dessen Elemente wir als Zeilenvektoren betrachten. Sei w ein Links-Perron-Eigenvektor für A . Dann konvergiert $\theta_{zA^n, w}$ für jeden Vektor $z \in (\mathbb{R}^m)^+ \setminus \{0\}$ gegen Null. Außerdem ist die Konvergenz gleichmäßig in dem Sinne, dass für jedes $\varepsilon > 0$ ein $n_0 \in \mathbb{N}$ existiert, so dass $\theta_{zA^n, w} < \varepsilon$ für alle $n \geq n_0$ und $z \in (\mathbb{R}^m)^+ \setminus \{0\}$.

Mit anderen Worten, jede primitive Matrix „drückt“ den positiven Orthanten schließlich in Richtung des Strahls, der von w erzeugt wird.

Beweis: Nach (18) reicht es, zu zeigen, dass $(zA^n)w / (\|zA^n\|_2 \|w\|_2)$ gleichmäßig gegen 1 konvergiert. Schreiben wir $B = \lambda^{-1}A$, so ist dies äquivalent zur Aussage, dass

$$\frac{(zB^n)w}{\|zB^n\|_2 \|w\|_2} \rightarrow 1 \text{ gleichmäßig.}$$

Nach Satz 5.46 konvergiert zB^n gegen $(z \cdot v)w$, wobei v ein Rechts-Perron-Eigenvektor ist, so dass $w \cdot v = 1$. Dann folgt

$$\frac{(zB^n)w}{\|zB^n\|_2 \|w\|_2} \rightarrow \frac{(z \cdot v)w \cdot w}{\|(z \cdot v)w\|_2 \|w\|_2} = 1.$$

Die Gleichmäßigkeit der Konvergenz verlagern wir in eine Übungsaufgabe. \square

Wir wollen das Material der letzten zwei Unterabschnitte nochmals rekapitulieren: Um $h(X_A)$ zu berechnen, müssen wir die Wachstumsraten der Einträge von A^n kennen. Aus diesem Gesichtspunkt heraus sind primitive Matrizen am besten, da nach dem letzten Satz alle Einträge von A^n regulär mit derselben Rate wachsen. Am nächstbesten sind irreduzible Matrizen B , da alle Knoten mit derselben Periode p permutiert werden und B^p eine Blockdiagonalmatrix mit primitiven Blöcken ist. Dies reduziert das Studium von irreduziblen Kantenshifts auf diejenigen mit primitiver Übergangsmatrix (siehe Übungsaufgabe 2 auf Blatt 12). Schließlich kann eine allgemeine Übergangsmatrix C in Blockdreiecksform gebracht werden mit irreduziblen Komponenten C_1, C_2, \dots, C_k entlang der Diagonalen. Dann können die bi-unendlichen Pfade in X_C niemals von C_i nach C_j verlaufen, wenn $j > i$. Jeder bi-unendliche Pfad zerfällt also in höchstens k Teilpfade, jeder auf einer irreduziblen Komponente, mit Übergangskanten dazwischen. Dies reduziert das Studium von X_C auf das Studium der Komponenten X_{C_i} , die wir bereits verstanden haben.

6 Finite-State-Codes

6.1 Straßenfärbungen und rechtsabschließende Beschriftungen

Wir erinnern uns, dass ein beschrifteter Graph rechtsauflösend heißt, falls die Kanten, die bei einem Knoten beginnen, alle unterschiedlich beschriftet sind. Der

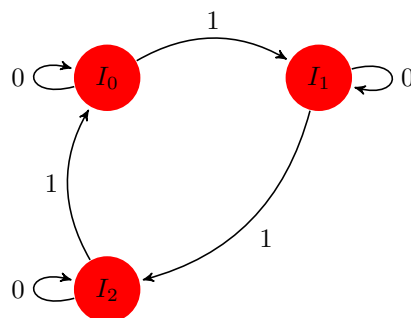


Abbildung 18: Eine Straenfrbung eines Graphen

Begriff eines Finite-State-Codes verwendet zwei Varianten dieser Eigenschaft, eine strkere, die wir *Straenfrbung*, und eine schwchere, die wir *rechtsabschlieend* nennen.

Eine Straenfrbung ist eine rechtsauflsende Darstellung, bei der zustzlich gefordert wird, dass bei jedem Knoten alle Symbole verwendet werden, um die ausgehenden Kanten zu beschriften.

6.1 Definition: Sei G ein Graph mit Kantenmenge \mathcal{E} . Eine Straenfrbung \mathcal{C} von G ist eine Beschriftung $\mathcal{C} : \mathcal{E} \rightarrow \mathcal{A}$, die fr jeden Knoten I von G eine eins-zu-eins-Korrespondenz zwischen \mathcal{A} und den bei I beginnenden Kanten von G herstellt. Ein Graph ist straen-frbbar, falls er eine Straenfrbung zulsst.

6.2 Beispiel: Die Beschriftung in Abbildung 18 ist eine Straenfrbung des zugrundeliegenden Graphen G , die das Alphabet $\mathcal{A} = \{0, 1\}$ verwendet. Es gibt auch andere Straenfrbungen. In der Tat knnen die zwei ausgehenden Kanten von jedem der drei Knoten auf zwei verschiedene Arten beschriftet werden, so dass es bei Verwendung des Alphabets \mathcal{A} insgesamt $2 \times 2 \times 2 = 8$ Straenfrbungen von G gibt.

Ein Graph ist straen-frbbar genau dann, wenn er dieselbe Anzahl von ausgehenden Kanten bei jedem Knoten hat, d.h. wenn er konstanten *Ausgangsgrad* hat. Die quivalente Bedingung an seine bergangsmatrix ist, dass alle Zeilensummen identisch sind. Hat ein Graph r Knoten und konstanten Ausgangsgrad n , dann gibt es $(n!)^r$ mgliche Straenfrbungen fr ein gegebenes Alphabet mit n Symbolen.

Sei \mathcal{C} eine Straenfrbung von G unter Verwendung des Alphabets \mathcal{A} . Sei $w = a_1 a_2 \dots a_m$ ein Block ber \mathcal{A} und I ein beliebiger Knoten von G . Stehen wir bei $I_0 = I$, so bestimmt das Symbol a_1 eine eindeutige Kante e_1 mit Beschriftung a_1 . Denken wir uns a_1 als eine Anweisung, e_1 zu durchlaufen, so laufen wir zu einem Knoten $I_1 = t(e_1)$. Das nchste Symbol a_2 bestimmt eine von I_1 ausgehende Kante mit Beschriftung a_2 . Wir durchlaufen e_2 , kommen zu einem Knoten I_2 ,

usw. Auf diese Art und Weise können wir jeden Block w als eine Folge von Anweisungen verstehen, die zu einem eindeutigen Pfad $\pi_{\mathcal{C}}(I, w)$ führt, der bei I beginnt und die Beschriftung w trägt. Insbesondere, da jeder Block über \mathcal{A} die Beschriftung eines Pfades ist, sehen wir, dass (G, \mathcal{C}) eine Darstellung des vollen \mathcal{A} -Shifts ist.

6.3 Definition: *Ein beschrifteter Graph heißt rechtsabschließend mit Verzögerung D , falls Folgendes gilt: Wenn zwei Pfade der Länge $D + 1$ bei demselben Knoten beginnen und dieselbe Beschriftung haben, dann müssen sie auch dieselbe Anfangskante haben. Ein beschrifteter Graph ist rechtsabschließend, falls er rechtsabschließend für irgendein $D \geq 0$ ist.*

Gemäß obiger Definition ist ein beschrifteter Graph rechtsauflösend genau dann, wenn er rechtsabschließend mit Verzögerung 0 ist. Grob gesprochen ist ein beschrifteter Graph rechtsabschließend, falls wir für jeden fest gewählten Anfangsknoten sagen können, welche Kante in einem Pfad auftaucht, indem wir uns die Symbole der darauffolgenden Kanten anschauen, d.h. er ist „schließlich rechtsauflösend“. Manchmal werden wir eine Beschriftung als *rechtsabschließend* bezeichnen, wenn der resultierende beschriftete Graph rechtsabschließend ist.

6.4 Beispiel: Betrachte die beschrifteten Graphen in Abbildung 19. Der erste Graph hat Verzögerung $D = 1$. Um dies zu sehen, betrachte den Knoten I . Da es zwei Kanten mit Beschriftung a gibt, die bei I beginnen, ist der Graph nicht rechtsauflösend. Auf die Schleife bei I kann nur eine Kante mit Beschriftung a folgen, während auf die Kante von I nach J nur eine Kante mit Beschriftung b oder c folgen kann. Folglich bestimmt die Beschriftung eines Pfades der Länge 2, der bei I beginnt, die Anfangskante. Beim Knoten J ist es ähnlich und beim Knoten K ist es sogar einfacher, da die Kanten, die dort beginnen, verschieden beschriftet sind. Auch der zweite Graph hat Verzögerung $D = 1$, wie man sich leicht überlegt. Im dritten Graphen gibt es zwei Pfade mit verschiedenen Anfangskanten, die bei I beginnen und mit ab beschriftet sind, also hat dieser Graph nicht Verzögerung 1. Das nächste Symbol bestimmt jedoch, welcher Pfad genommen wird, also ist der Graph rechtsabschließend mit Verzögerung $D = 2$. Im vierten Graphen gibt es für jedes m zwei Pfade, die bei J beginnen und den Block ab^m darstellen. Da diese Pfade verschiedene Anfangskanten haben, ist dieser Graph nicht rechtsabschließend.

Ist ein beschrifteter Graph $\mathcal{G} = (G, \mathcal{L})$ rechtsabschließend mit Verzögerung D , dann stimmen Pfade der Länge $D + 1$, die bei demselben Knoten beginnen und dieselbe Beschriftung haben, in der ersten Kante überein. Wir können diese Eigenschaft auf längere Pfade in G anwenden, um eine Übereinstimmung in längeren Teilpfaden zu erreichen. Seien $k \geq D$ und $e_0 e_1 \dots e_k, f_0 f_1 \dots f_k$ Pfade in G mit demselben Anfangsknoten und derselben Beschriftung. Indem wir die Definition auf die initialen $(D + 1)$ -Blöcke anwenden, erhalten wir $e_0 = f_0$. Daher beginnen $e_1 e_2 \dots e_{D+1}$ und $f_1 f_2 \dots f_{D+1}$ beide bei $t(e_0) = t(f_0)$ und haben dieselbe Beschriftung, so dass $e_1 = f_1$. Indem wir so fortfahren, schließen

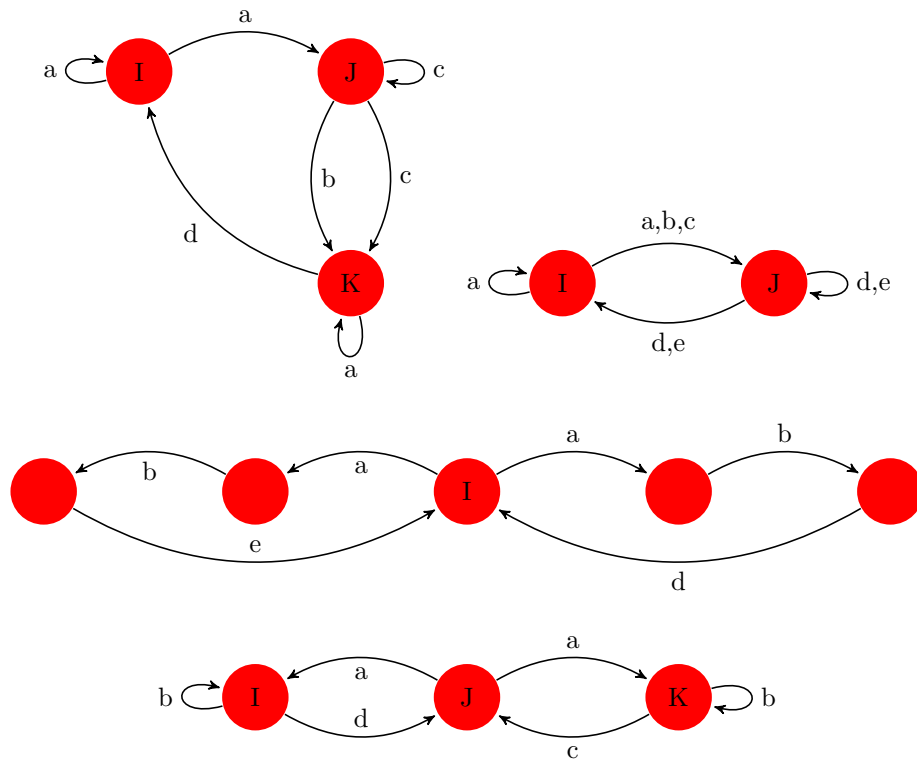


Abbildung 19: Beispiele von rechtsabschließenden und nicht rechtsabschließenden Graphen

wir sukzessive, dass $e_0 = f_0, e_1 = f_1, \dots, e_{k-D} = f_{k-D}$. Mit anderen Worten, gegeben zwei Pfade der Länge $\geq D$ mit demselben Anfangsknoten und derselben Beschriftung, wirkt die Eigenschaft „rechtsabschließend“ wie ein Reißverschluss, der die Pfade bis auf die letzten D Kanten zusammenzieht.

Indem wir diese Idee auf rechtsseitig unendliche Pfade anwenden, sehen wir, dass zwei solche Pfade identisch sein müssen, wenn sie denselben Anfangsknoten und dieselbe Beschriftung haben. Um dies präzise zu machen, führen wir etwas Notation ein.

Sei X ein Shiftraum über \mathcal{A} . Wir definieren den *einseitigen Shift*

$$X^+ := \{x|_{[0,\infty)} : x \in X\},$$

die Menge der rechtsseitig unendlichen Folgen, die in X auftauchen. Ist G ein Graph, dann ist X_G^+ die Menge der rechtsseitig unendlichen Pfade auf G . Für einen Knoten I in G sei $X_{G,I}^+$ die Menge derjenigen rechtsseitig unendlichen Pfade in X_G^+ , die bei I beginnen. Ein typisches Element von $X_{G,I}^+$ ist von der Form $e_0e_1e_2\dots$ mit $i(e_0) = I$. Ist \mathcal{L} eine Beschriftung von G , dann sei \mathcal{L}^+ die Erweiterung von \mathcal{L} auf X_G^+ , d.h.

$$\mathcal{L}^+(e_0e_1e_2\dots) = \mathcal{L}(e_0)\mathcal{L}(e_1)\mathcal{L}(e_2)\dots$$

Definieren wir $\mathcal{G} := (G, \mathcal{L})$, so ist $\mathcal{L}^+ : X_G^+ \rightarrow X_{\mathcal{G}}^+$ surjektiv. Das folgende Resultat charakterisiert rechtsabschließende Beschriftungen mit Hilfe von einseitigen Folgen.

6.5 Satz: *Ein beschrifteter Graph (G, \mathcal{L}) ist rechtsabschließend genau dann, wenn für jeden Knoten I von G die Einschränkung von \mathcal{L}^+ auf $X_{G,I}^+$ injektiv ist.*

Beweis: Wir nehmen zuerst an, dass (G, \mathcal{L}) rechtsabschließend ist, sagen wir mit Verzögerung D . Seien $x = e_0e_1e_2\dots$ und $y = f_0f_1f_2\dots$ in $X_{G,I}^+$, so dass $\mathcal{L}^+(x) = \mathcal{L}^+(y)$. Dann gilt $i(e_0) = i(f_0) = I$ und $\mathcal{L}(e_0e_1\dots e_D) = \mathcal{L}(f_0f_1\dots f_D)$, also zeigt die Definition von rechtsabschließenden Graphen, dass $e_0 = f_0$. Daher haben $e_1e_2\dots e_{D+1}$ und $f_1f_2\dots f_{D+1}$ denselben Anfangsknoten und dieselbe Beschriftung. Also gilt $e_1 = f_1$. Wir können beliebig so fortfahren, um zu zeigen, dass $e_k = f_k$ für alle $k \geq 0$, so dass $x = y$. Dies zeigt, dass \mathcal{L}^+ injektiv auf $X_{G,I}^+$ ist.

Umgekehrt nehmen wir nun an, dass (G, \mathcal{L}) nicht rechtsabschließend ist. Dann muss es für jedes $n \geq 1$ einen Knoten I_n geben mit der Eigenschaft, dass es Pfade π_n und τ_n der Länge n gibt, bei I_n beginnend, mit derselben Beschriftung aber verschiedenen Anfangskanten. Sei $n \geq |\mathcal{V}(G)|^2$ fest gewählt. Da es nur $|\mathcal{V}(G)|^2$ verschiedene Knotenpaare gibt, können wir

$$\pi_n = \alpha\beta\gamma, \quad \tau_n = \alpha'\beta'\gamma'$$

schreiben mit $|\alpha| = |\alpha'|$, $|\beta| = |\beta'|$, $|\gamma| = |\gamma'|$, wobei β und β' geschlossene Pfade sind. Dann sind $\alpha(\beta)^\infty$ und $\alpha'(\beta')^\infty$ rechtsseitig unendliche Pfade, die bei I_n

beginnen mit derselben Beschriftung. Da sie aber verschiedene Anfangskanten haben, ist \mathcal{L}^+ nicht injektiv auf X_{G,I_n}^+ . \square

6.6 Satz: Sei $\mathcal{G} = (G, \mathcal{L})$ ein beschrifteter Graph, so dass der 1-Blockcode $\mathcal{L}_\infty : X_G \rightarrow X_G$ eine Konjugation ist. Hat dann \mathcal{L}_∞^{-1} Antizipation n , so ist \mathcal{L} rechtsabschließend mit Verzögerung n .

Beweis: Wir dürfen annehmen, dass G wesentlich ist. Seien $e_0 e_1 \dots e_n$ und $f_0 f_1 \dots f_n$ Pfade in G , die bei demselben Knoten I beginnen und dieselbe Beschriftung haben. Wir müssen zeigen, dass $e_0 = f_0$.

Sei m das Gedächtnis von \mathcal{L}_∞^{-1} , so dass $\mathcal{L}_\infty^{-1} = \Phi_\infty^{[-m, n]}$ für eine Blockabbildung $\Phi : \mathcal{B}_{m+n+1}(X_G) \rightarrow \mathcal{E}(G)$. Da G wesentlich ist, gibt es einen bei I endenden Pfad π in G der Länge m . Dann gilt $\mathcal{L}(\pi e_0 e_1 \dots e_n) = \mathcal{L}(\pi f_0 f_1 \dots f_n)$ und die Anwendung von Φ zeigt, dass $e_0 = f_0$. \square

Wir erinnern an den Begriff der Zustandsaufspaltung aus Unterabschnitt 3.4.

6.7 Satz: Sei \mathcal{G} rechtsabschließend mit Verzögerung D . Jeder beschriftete Graph, den man aus \mathcal{G} durch eine Zustandsaufspaltung erhält, ist rechtsabschließend mit Verzögerung $D + 1$.

Beweis: Sei $\mathcal{G} = (G, \mathcal{L})$ und sei H ein Graph, der aus G durch eine Zustandsaufspaltung entsteht. Wir erinnern uns daran, dass H die Knoten I^i und Kanten e^i hat und dass die Amalgamationsabbildung $e^i \mapsto e$ eine 1-Block-Konjugation von X_H nach X_G induziert. Die Beschriftung \mathcal{L} induziert eine Beschriftung \mathcal{L}' von H durch $\mathcal{L}'(e^i) := \mathcal{L}(e)$. Wir zeigen, dass (H, \mathcal{L}') rechtsabschließend mit Verzögerung $D + 1$ ist.

Dazu nehmen wir an, dass $e_0^{i_0} e_1^{i_1} \dots e_{D+1}^{i_{D+1}}$ und $f_0^{j_0} f_1^{j_1} \dots f_{D+1}^{j_{D+1}}$ Pfade in H sind, die bei I^i beginnen und dieselbe \mathcal{L}' -Beschriftung haben. Indem wir die Amalgamationsabbildung anwenden, sehen wir, dass $e_0 e_1 \dots e_{D+1}$ und $f_0 f_1 \dots f_{D+1}$ Pfade mit derselben \mathcal{L} -Beschriftung sind, die bei I beginnen. Daher gilt $e_0 e_1 = f_0 f_1$. Wir erinnern uns, dass i_0 dadurch bestimmt ist, welches Partitionselement e_1 enthält und j_0 entsprechend durch f_1 bestimmt ist. Es folgt $e^{i_0} = f^{j_0}$, womit gezeigt ist, dass \mathcal{L}' rechtsabschließend mit Verzögerung $D + 1$ ist. \square

Obigen Satz werden wir im Beweis des *Finite-State-Coding-Theorems* verwenden, um zu zeigen, dass ein Graph, der durch k aufeinanderfolgende Zustandsaufspaltungen aus einem rechtsauflösenden Graphen entsteht, rechtsabschließend mit Verzögerung k ist.

Wir erinnern an Satz 5.11, der sagt, dass rechtsauflösende Darstellungen die Entropie erhalten. Dasselbe gilt für rechtsabschließende Darstellungen:

6.8 Satz: Sei \mathcal{G} rechtsabschließend mit zugrundeliegendem Graphen G . Dann gilt $h(X_{\mathcal{G}}) = h(X_G)$.

Beweis: Da $X_{\mathcal{G}}$ ein Faktor von X_G ist, gilt $h(X_{\mathcal{G}}) \leq h(X_G)$. Sei \mathcal{G} rechtsabschließend mit Verzögerung D . Dann bestimmt jeder Knoten $I \in \mathcal{V}$ zusammen

mit einem Block $w \in \mathcal{B}_{n+D}(X_G)$ höchstens einen Pfad der Länge n in G und alle Pfade in $\mathcal{B}_n(X_G)$ entstehen auf diese Weise. Daher gilt

$$|\mathcal{B}_n(X_G)| \leq |\mathcal{V}| \cdot |\mathcal{B}_{n+D}(X_G)|,$$

so dass

$$\begin{aligned} h(X_G) &= \lim_{n \rightarrow \infty} \frac{1}{n} \log |\mathcal{B}_n(X_G)| \\ &\leq \lim_{n \rightarrow \infty} \frac{1}{n} \log |\mathcal{V}| + \lim_{n \rightarrow \infty} \frac{1}{n} \log |\mathcal{B}_{n+D}(X_G)| = h(X_G), \end{aligned}$$

womit der Beweis abgeschlossen ist. \square

Unser letztes Resultat in diesem Unterabschnitt zeigt, dass sich eine rechtsabschließende zu einer rechtsauflösenden Beschriftung rekodieren lässt.

6.9 Satz: Sei Φ eine rechtsabschließende Beschriftung auf einem Graphen G mit Verzögerung D . Dann gibt es einen Graphen H und Beschriftungen Ψ, Θ von H , so dass Ψ rechtsauflösend ist, $\Theta_\infty : X_H \rightarrow X_G$ eine Konjugation ist und folgendes Diagramm kommutiert:

$$\begin{array}{ccc} X_G & \xleftarrow[\cong]{\Theta_\infty \circ \sigma^D} & X_H \\ \Phi_\infty \downarrow & & \swarrow \Psi_\infty \\ & & \Phi_\infty(X_G) \end{array}$$

Beweis: Ist $D = 0$, so ist Φ bereits rechtsauflösend und wir können $H = G$, $\Psi = \Phi$ und $\Theta = \text{id}$ wählen. Daher nehmen wir im Folgenden an, dass $D \geq 1$.

Wir definieren den Graphen H wie folgt. Die Knoten von H sind die Paare der Form $(I, w_1 w_2 \dots w_D)$, wobei I ein Knoten in G ist und $w_1 w_2 \dots w_D$ ein Block, der von einem bei I beginnenden Pfad dargestellt wird. Sei a ein Symbol, so dass $w_1 w_2 \dots w_D a$ von einem Pfad dargestellt wird, der bei I beginnt. Die Anfangskante aller solcher Pfade ist dieselbe, sagen wir e . Dann fügen wir in H eine Kante mit dem Namen $(I, w_1 w_2 \dots w_D a)$ ein, die bei $(I, w_1 w_2 \dots w_D)$ beginnt und bei $(t(e), w_2 w_3 \dots w_D a)$ endet. Auf diese Art wird der Graph H definiert. Die Abbildung Θ beschriftet diese Kante mit e .

Nach unserer Konstruktion ist Θ ein Graphenhomomorphismus von H nach G . Die zugehörige Knotenabbildung ist durch $\partial\Theta(I, w_1 w_2 \dots w_D) := I$ definiert. Durch diese Konstruktion ist außerdem garantiert, dass aus

$$\Theta((I^{(1)}, w^{(1)}) \dots (I^{(D+1)}, w^{(D+1)})) = e_1 \dots e_{D+1}$$

folgt, dass $I^{(1)}$ der Anfangsknoten von e_1 ist und $w^{(1)}$ die Φ -Beschriftung von $e_1 \dots e_{D+1}$. Folglich ist Θ_∞^{-1} ein gleitender Blockcode mit Gedächtnis 0 und Antizipation D , und daher Θ_∞ eine Konjugation.

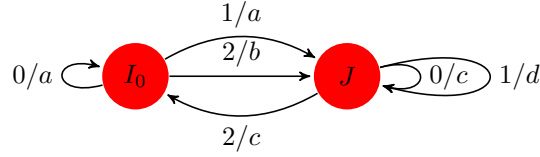


Abbildung 20: Ein Finite-State-Code

Definiere die Beschriftung Ψ auf H durch $\Psi(I, w_1 \dots w_D a) := a$. Diese Beschriftung ist rechtsauflösend, da der Knoten $(I, w_1 \dots w_D)$ und die Beschriftung a die Kante $(I, w_1 \dots w_D a)$ bestimmen.

Es bleibt zu zeigen, dass das Diagramm kommutiert. Es gilt

$$\Phi \circ \Theta(I, w_1 \dots w_D a) = w_1 \quad \text{und} \quad \Psi(I, w_1 \dots w_D a) = a.$$

Daher reicht es, Folgendes zu zeigen: Ist

$$(I^{(1)}, w^{(1)}) \dots (I^{(D+1)}, w^{(D+1)})$$

ein Pfad in H , so ist das letzte Symbol von $w^{(1)}$ das Anfangssymbol von $w^{(D+1)}$. Dies folgt aber unmittelbar aus der Definition der Kanten von H . \square

6.2 Finite-State-Codes

Wir werden in diesem Unterabschnitt eine Methode vorstellen, Symbolfolgen ohne Beschränkungen in Symbolfolgen mit Beschränkungen zu transformieren, und ein theoretisches Resultat beweisen, das zeigt, wann dies möglich ist. Der Beweis dieses Resultats wird die nächsten zwei Unterabschnitte einnehmen, in denen wir außerdem einen praktischen Algorithmus vorstellen werden, um Finite-State-Codes zu konstruieren.

6.10 Definition: *Ein Finite-State-Code ist ein Tripel $(G, \mathcal{I}, \mathcal{O})$, wobei G ein Graph, genannt Kodierungsgraph, \mathcal{I} eine Straßenfärbung von G , genannt Input-Beschriftung und \mathcal{O} eine rechtsabschließende Beschriftung von G , genannt Output-Beschriftung, ist. Hat G konstanten Ausgangsgrad n und ist X ein Shiftraum mit $\mathcal{O}_\infty(X_G) \subseteq X$, so nennen wir $(G, \mathcal{I}, \mathcal{O})$ einen Finite-State- (X, n) -Code.*

Wir werden Finite-State-Codes zeichnen, indem wir eine Kante e mit a/b beschriften, wobei $a = \mathcal{I}(e)$ die Input-Beschriftung und $b = \mathcal{O}(e)$ die Output-Beschriftung von e ist (wie beispielsweise in Abbildung 20 dargestellt).

Sei $(G, \mathcal{I}, \mathcal{O})$ ein Finite-State- (X, n) -Code. Wir können diesen verwenden, um Folgen aus dem vollen n -Shift in Folgen aus X zu transformieren, und zwar wie

folgt. Fixiere einen Knoten I_0 von G . Sei $x = x_0x_1x_2 \dots \in X_{[n]}^+$ eine Folge im einseitigen vollen n -Shift. Da \mathcal{I} eine Straenfrbung ist, gibt es genau einen rechtsseitig unendlichen Pfad $e_0e_1e_2 \dots \in X_G^+$ mit \mathcal{I} -Beschriftung $x_0x_1x_2 \dots$, der bei I_0 beginnt. Wir wenden \mathcal{O} auf diesen Pfad an, um die Output-Beschriftung $y = \mathcal{O}(e_0)\mathcal{O}(e_1)\mathcal{O}(e_2) \dots \in X^+$ zu erhalten.

Wir knnen uns vorstellen, dass die Kodierung von einer Maschine durchgefhrt wird, deren interne Zustnde den Knoten des Kodierungsgraphen entsprechen. Die Maschine startet im Zustand I_0 . Das Eingangssymbol x_0 veranlasst die Maschine dazu, die eindeutige Kante $e_0 \in \mathcal{E}_{I_0}$ mit Eingangsbeschriftung x_0 zu finden, deren Ausgangsbeschriftung $y_0 = \mathcal{O}(e_0)$ auszugeben und dann zum Zustand $I_1 = t(e_0)$ zu wechseln. Danach liest sie das nchste Eingangssymbol x_1 , das e_1 bestimmt, gibt $y_1 = \mathcal{O}(e_1)$ aus und wechselt zum Zustand $I_2 = t(e_1)$. Dieser Prozess wird fortgefhrt, solange es Eingangssymbole gibt.

Natrlich wollen wir in der Lage sein, den Eingang $x_0x_1x_2 \dots$ aus dem Ausgang $y_0y_1y_2 \dots$ zu ermitteln. Die Folge $x_0x_1x_2 \dots$ bestimmt einen eindeutigen Pfad $e_0e_1e_2 \dots$ in X_{G,I_0}^+ und Satz 6.5 sagt uns, dass \mathcal{O}_∞ , eingeschrnkt auf X_{G,I_0}^+ , eine eins-zu-eins-Abbildung ist, da \mathcal{O} rechtsabschlieend ist. Dies zeigt, dass wir den Eingang aus dem Ausgang rekonstruieren knnen.

Um prziser zu sein, nehmen wir an, dass (G, \mathcal{O}) rechtsabschlieend mit Verzgerung D ist. Dann bestimmen I_0 und $y_0y_1 \dots y_D$ die Kante e_0 und daher I_1 und $x_0 = \mathcal{I}(e_0)$. Die Kenntnis von I_1 und $y_1y_2 \dots y_{D+1}$ bestimmt e_1 und daher I_2 und $x_1 = \mathcal{I}(e_1)$ usw. Folglich knnen wir aus dem Ausgang $y_0y_1 \dots y_k$ alle auer die letzten D Eingangssymbole $x_0x_1 \dots x_{k-D}$ bestimmen.

Da wir es in der Praxis nur mit endlichen Folgen zu tun haben, ist der Verlust der letzten D Symbole alarmierend. Eine einfache Lsung ist es, dem Input D „Fllsymbole“ hinzuzufgen. Dies macht eine zustzliche Kodierung notwendig, aber in der Regel ist D mikroskopisch im Vergleich mit der Input-Lnge und folglich ist die zustzliche Arbeit relativ gering.

Wann existieren Finite-State-Codes? Das nchste Resultat, das Haupttheorem dieses Abschnitts, liefert fr sofische Shifts eine vollstndige Antwort auf diese Frage.

6.11 Satz: *Ist X ein sofischer Shift und $n \geq 1$, so gibt es einen Finite-State- (X, n) -Code genau dann, wenn $h(X) \geq \log n$.*

Die Notwendigkeit der Bedingung $h(X) \geq \log n$ ist eine einfache Schlussfolgerung aus dem, was wir bereits wissen. Sei $(G, \mathcal{I}, \mathcal{O})$ nmlich ein (X, n) -Code. Da rechtsabschlieende Beschriftungen nach Satz 6.8 die Entropie erhalten, gilt

$$h(\mathcal{I}_\infty(X_G)) = h(X_G) = h(\mathcal{O}_\infty(X_G)).$$

Da (G, \mathcal{I}) den vollen n -Shift darstellt und $\mathcal{O}_\infty(X_G) \subseteq X$, erhalten wir

$$\log n = h(X_{[n]}) = h(\mathcal{I}_\infty(X_G)) = h(\mathcal{O}_\infty(X_G)) \leq h(X).$$

Um zu zeigen, dass $h(X) \geq \log n$ auch hinreichend ist, benötigen wir einige neue Ideen. Um zu sehen, wie diese Ideen verwendet werden, werden wir kurz den Beweis skizzieren. Wir beginnen mit einer rechtsauflösenden Darstellung \mathcal{G} von X . Mit Hilfe eines Vektors, genannt *approximativer Eigenvektor*, werden wir eine Reihe von Zustandsaufspaltungen durchführen, um am Ende eine Darstellung \mathcal{H} von X zu erhalten, so dass jeder Knoten von \mathcal{H} Ausgangsgrad $\geq n$ hat. Nach Satz 6.7 ist \mathcal{H} rechtsabschließend. Wir können Kanten auf \mathcal{H} entfernen, um einen Graphen mit konstantem Ausgangsgrad n zu erhalten, der einen in X enthaltenen sofischen Shift darstellt. Dann liefert jede Straßenfärbung dieses Graphen den Finite-State- (X, n) -Code, den wir suchen.

Als nächstes überlegen wir uns, wie man das Finite-State-Coding-Theorem anwenden kann, auch wenn die Entropiebedingung nicht direkt erfüllt ist. Wir betrachten eine typische Situation, in der wir darauf aus sind, Folgen im vollen 2-Shift in Folgen eines sofischen Shifts über $\{0, 1\}$ zu transformieren. Dann ist $h(X) < \log 2$ (siehe Korollar 5.34) und folglich ist die Entropiebedingung verletzt. Um dieses Problem zu umgehen, transformieren wir Blöcke von Symbolen statt einzelne Symbole. Die höheren Potenzshifts aus Abschnitt 2 sind genau das richtige Hilfsmittel, um dies zu tun.

Wir zerlegen die Eingangsfolge in Blöcke der Länge p und die Ausgangsfolge in Blöcke der Länge q . Dies ersetzt den Input-Shift $X_{[2]}$ durch seinen höheren Potenzshift $X_{[2]}^p \cong X_{[2^p]}$ und den Output-Shift X durch X^q . Die Entropiebedingung für diese Ersatzshifts ist

$$qh(X) = h(X^q) \geq \log 2^p = p \log 2 = p,$$

oder

$$h(X) \geq \frac{p}{q}. \quad (19)$$

Indem wir also Brüche p/q nahe bei, aber kleiner als $h(X)$ finden, können wir Finite-State-Codes konstruieren, die p -Blöcke aus $X_{[2]}$ in q -Blöcke aus X transformieren. Wir sagen, dass ein solcher Code die *Kodierungsrate* p/q hat.

Die Kodierungsrate p/q drückt die Anzahl der Input-Bits aus, die von jedem Output-Bit dargestellt werden. Ist X eine echte Teilmenge des vollen 2-Shifts, so wird diese Rate stets kleiner als 1 sein. Je höher aber die Rate ist, desto effizienter ist die Kodierung. Also sagt uns (19), was die maximale Effizienz ist, die für eine Finite-State-Kodierung von binären Daten in Daten aus X möglich ist.

6.12 Beispiel: Sei $X = X(0, 1)$ der Shift über $\{0, 1\}$ mit folgenden zwei Eigenschaften: In jedem Punkt $x \in X$ kommt die Eins unendlich oft in beiden Richtungen vor und zwischen zwei aufeinanderfolgenden Einsen steht höchstens eine Null. Indem wir Einsen und Nullen vertauschen, sehen wir dass X zum Goldenen-Schnitt-Shift konjugiert ist, dessen Entropie wir in Beispiel 5.4 zu $h(X) = \log[(1 + \sqrt{5})/2] \approx 0.69424$ berechnet haben. Also wählen wir $p = 2$ und $q = 3$. Das Input-Alphabet ist dann die Menge der 2-Blöcke aus $X_{[2]}$,

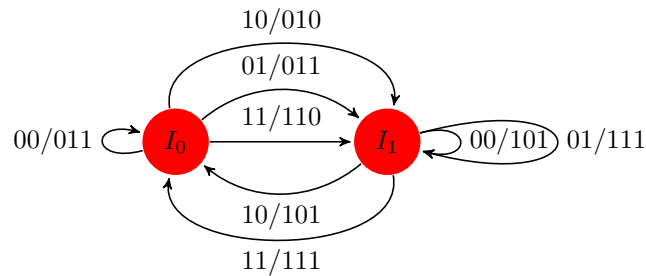


Abbildung 21: Der Finite-State-Code aus Beispiel 6.12

d.h. $\{00, 01, 10, 11\}$, während das Output-Alphabet die Menge der 3-Blöcke aus X ist, also $\{010, 011, 101, 110, 111\}$. Abbildung 21 zeigt einen Finite-State $(X^3, 2^2)$ -Code.

Um die Verwendung dieses Finite-State-Codes zu illustrieren, betrachten wir die Input-Folge 000111010011. Zuerst zerlegen wir sie in 2-Blöcke, so dass wie 00 01 11 01 00 11 aussieht. Dann verwenden wir diese 2-Blöcke als Input-Symbole, wobei wir beim Knoten I_0 beginnen. Dies erzeugt die Output-Tripel

$$011 \ 011 \ 111 \ 011 \ 101 \ 111,$$

die eine Folge in X^3 darstellen.

Wir beenden diesen Unterabschnitt mit ein paar allgemeinen Bemerkungen über Kodierung: Es scheint besser zu sein, zu fordern, dass die Output-Beschriftung rechtsauflösend ist, was die Kodierungsverzögerung auf 0 reduzieren würde. Leider ist es nicht immer möglich so einen Finite-State-Code zu finden, auch wenn die Entropiebedingung erfüllt ist.

Ein anderer Ansatz, um beliebige Folgen in Folgen mit Beschränkungen zu transformieren, ist es, eine Einbettung zu verwenden, deren Inverse der Dekodierer ist. Die Entropiebedingung ist hier sicherlich notwendig, aber es gibt zwei ernstzunehmende Schwierigkeiten, eine theoretische und eine praktische. Hat der sofische Shift X weniger als n Fixpunkte, so gibt es keine Einbettung von $X_{[n]}$ in X . Auch wenn Fixpunkte oder periodische Punkte kein Problem darstellen würden, führen die bekannten Konstruktionen von Einbettungen zu Blockabbildungen mit enormen Fenstergrößen und die resultierenden Codes sind bei weitem zu kompliziert, um sie zu verwenden.

6.3 Approximative Eigenvektoren

Sei X ein sofischer Shift, der die notwendige Bedingung $h(X) \geq \log n$ aus dem Finite-State-Coding-Theorem erfüllt. Unser Ziel ist es, einen Finite-State- (X, n) -Code zu finden. Indem wir die Teilmengenkonstruktion aus Satz 4.22

verwenden, bestimmen wir zunächst eine rechtsauflösende Darstellung (H, \mathcal{O}) von X . Nehmen wir an, wir haben Glück, und der Ausgangsgrad jedes Knotens in H ist mindestens n . Indem wir Kanten aus H auswählen, können wir einen Teilgraphen G mit konstantem Ausgangsgrad n bilden und \mathcal{O} auf G einschränken, so dass $\mathcal{O}_\infty(X_G) \subseteq X$. Ist \mathcal{I} eine beliebige Straßenfärbung von G , dann ist $(G, \mathcal{I}, \mathcal{O})$ der Finite-State- (X, n) -Code, den wir suchen.

Es bezeichne A die Übergangsmatrix von H . Unsere Annahme über die Ausgangsgrade von H drückt sich darin aus, dass alle Zeilensummen von A mindestens n sind. Dies können wir mit Hilfe des Spaltenvektors $\mathbf{1} = [1 \ 1 \ \dots \ 1]^T$ ausdrücken durch die Vektorungleichung

$$A\mathbf{1} \geq n\mathbf{1}. \quad (20)$$

Üblicherweise werden wir nicht so viel Glück haben mit unserer ersten Darstellung von X und einige Zeilensummen werden $< n$ sein. Um mit dieser Situation umzugehen, werden wir in diesem Unterabschnitt zeigen, dass eine Ungleichung ähnlich zu (20) immer noch erfüllt ist, aber mit $\mathbf{1}$ ersetzt durch einen nichtnegativen ganzzahligen Vektor v , den wir einen approximativen Eigenvektor nennen. Dieses v wird im nächsten Unterabschnitt verwendet, um eine Reihe von Zustandsaufspaltungen zu definieren, die so gemacht sind, dass sie die ursprüngliche Darstellung von X in eine andere umformen, für die (20) gilt. Die Überlegungen am Anfang dieses Unterabschnitts zeigen dann, wie man die modifizierte Darstellung verwenden kann, um einen Finite-State- (X, n) -Code zu konstruieren.

Wir beginnen damit approximative Eigenvektoren zu definieren.

6.13 Definition: Sei $A \geq 0$ eine Matrix mit ganzzahligen Einträgen und $n \in \mathbb{N}$. Ein (A, n) -approximativer Eigenvektor ist ein Vektor $v \neq 0$ mit nichtnegativen ganzzahligen Komponenten, der $Av \geq nv$ erfüllt.

6.14 Beispiel: Sei $A = \begin{bmatrix} 1 & 3 \\ 6 & 1 \end{bmatrix}$ und $n = 5$. Dann ist $v = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$ ein $(A, 5)$ -approximativer Eigenvektor, da

$$Av = \begin{bmatrix} 11 \\ 15 \end{bmatrix} \geq \begin{bmatrix} 10 \\ 15 \end{bmatrix} = 5v.$$

Es ist nützlich, sich vorzustellen, dass ein approximativer Eigenvektor jedem Knoten I ein nichtnegatives ganzzahliges „Gewicht“ v_I zuordnet. Die Vektorungleichung $Av \geq nv$ entspricht dann der Bedingung, dass für jeden Knoten I , wenn wir die Gewichte am Ende aller bei I beginnenden Kanten aufsummieren, die Summe mindestens n mal das Gewicht bei I ergibt. In Beispiel 6.14 gibt es vier Kanten, die bei dem ersten Knoten beginnen, eine, die dort endet (und damit das Gewicht $v_1 = 2$ hat), und drei, die bei dem anderen Knoten enden (jede mit Gewicht $v_2 = 3$). Ihre Summe ist $2 + 3 + 3 + 3 = 11$ und dies ist mindestens

$5v_1 = 10$, womit die approximative Eigenvektor-Bedingung beim ersten Knoten erfüllt ist.

Es gibt noch eine andere nützliche Art, die Bedeutung von approximativen Eigenvektoren zu verstehen. Wir nehmen an, dass A eine $r \times r$ -Matrix ist. Dann ist die Vektorungleichung $Av \geq nv$ äquivalent zu r skalaren Ungleichungen

$$\sum_{J=1}^r A_{IJ}v_J \geq nv_I \quad (I = 1, 2, \dots, r).$$

Jede dieser Ungleichungen bestimmt einen Halbraum im r -dimensionalen Raum, und ihr Schnitt wird typischerweise ein polyedrischer Kegel sein (eventuell von kleinerer Dimension). Dann ist jeder nichtnegative ganzzahlige Vektor in diesem Kegel ein (A, n) -approximativer Eigenvektor.

Bevor wir zum Hauptresultat dieses Unterabschnitts kommen, zeigen wir zuerst, wie man die allgemeine Situation auf die eines irreduziblen Graphen und eines positiven approximativen Eigenvektors reduzieren kann.

6.15 Lemma: *Sei G ein Graph und v ein (A_G, n) -approximativer Eigenvektor. Dann gibt es einen irreduziblen Teilgraphen H von G , so dass Folgendes gilt: Wenn wir w aus v gewinnen, indem wir v auf die Knoten in H einschränken, dann gilt $w > 0$ und w ist ein (A_H, n) -approximativer Eigenvektor.*

Beweis: Zuerst bilden wir einen Teilgraphen K von G durch die Entfernung aller Knoten I mit $v_I = 0$ zusammen mit allen Kanten, die bei solchen Knoten beginnen oder enden. Wie wir in Unterabschnitt 5.4 festgestellt haben, hat K mindestens eine irreduzible Komponente H , die eine Senke ist, d.h. H enthält die Endknoten aller Kanten, die in H beginnen. Ist w die Einschränkung von v auf H , dann gilt $w > 0$ nach Konstruktion und mit Hilfe der „Gewichtsinterpretation“ sehen wir, dass w ein (A_H, n) -approximativer Eigenvektor ist. \square

Der folgende Satz sagt uns, wann approximative Eigenvektoren existieren.

6.16 Satz: *Sei A eine nichtnegative ganzzahlige Matrix und $n \geq 1$. Dann gibt es einen (A, n) -approximativen Eigenvektor genau dann, wenn $\lambda_A \geq n$. Nehmen wir zusätzlich an, dass A irreduzibel ist, dann gibt es einen positiven (A, n) -approximativen Eigenvektor.*

Beweis: Zuerst nehmen wir an, dass es einen (A, n) -approximativen Eigenvektor v gibt. Nach dem vorherigen Lemma können wir annehmen, dass A irreduzibel ist und $v > 0$ gilt. Dann gilt $Av \geq nv$, so dass $A^2v \geq n(Av) \geq n^2v$ und folglich $A^k v \geq n^k v$ für alle $k \geq 1$. Da $v > 0$, zeigt dasselbe Argument, das wir verwendet haben um Satz 5.15 zu beweisen, dass es ein $c > 0$ gibt, so dass

$$cn^k \leq \sum_{I,J=1}^r (A^k)_{IJ}$$

für alle $k \geq 1$ gilt. Satz 5.15 zeigt auch, dass es eine Konstante $d > 0$ gibt mit

$$\sum_{I,J=1}^r (A^k)_{IJ} \leq d\lambda_A^k.$$

Es folgt $(c/d)^{1/k}n \leq \lambda_A$ für alle $k \geq 1$. Lassen wir k gegen unendlich gehen, so folgt $\lambda_A \geq n$.

Nun nehmen wir umgekehrt an, dass $\lambda_A \geq n$. Indem wir zu einer irreduziblen Komponente mit maximalem Spektralradius übergehen, können wir annehmen, dass A irreduzibel ist.¹¹ Dann gilt $v_A > 0$ für den Perron-Eigenvektor und $Av_A = \lambda_A v_A \geq nv_A$. Leider hat v_A sehr selten ganzzahlige Einträge, also müssen wir noch etwas arbeiten. Wir unterscheiden zwei Fälle.

Zuerst nehmen wir an, dass $\lambda_A = n$, also suchen wir einen exakten Eigenvektor. Wir können die Gleichung $Av = nv$ durch den Gauß-Algorithmus lösen. Da A und n ganzzahlig sind, wird die Lösung ein Vektor v mit rationalen Einträgen sein. Da A irreduzibel ist, folgt aus dem Perron-Frobenius-Theorem, dass v ein Vielfaches von v_A ist. Wenn wir v , falls notwendig, mit -1 multiplizieren, können wir $v > 0$ annehmen. Dann können wir v mit einer positiven ganzen Zahl multiplizieren, so dass ein ganzzahliger Vektor entsteht und das Resultat ist ein positiver (A, n) -approximativer Eigenvektor (siehe auch Übungsaufgabe (1b) auf Blatt 10).

Als nächstes nehmen wir an, dass $\lambda_A > n$, so dass $Av_A = \lambda_A v_A > nv_A$. Wir wählen einen positiven rationalen Vektor v , der so nahe bei v_A liegt, dass immer noch $Av > nv$ gilt. Indem wir v mit einer geeigneten ganzen Zahl multiplizieren, erhalten wir einen positiven (A, n) -approximativen Eigenvektor. \square

Obwohl uns obiger Satz sagt, wann approximative Eigenvektoren existieren, liefert er uns keine gute Methode, um sie zu finden. Glücklicherweise gibt es einen effizienten Algorithmus, um kleine approximative Eigenvektoren zu finden, die wesentlich sind für die Praktikabilität einer Finite-State-Kodierung.

Wir benötigen etwas Notation, um diesen Algorithmus zu formulieren. Sind u und v Vektoren, so sei $w = \min\{u, v\}$ das komponentenweise Minimum, so dass $w_I = \min\{u_I, v_I\}$ für jedes I . Für eine reelle Zahl s sei $\lfloor s \rfloor$ die größte ganze Zahl $\leq s$. Für einen Vektor v ist $\lfloor v \rfloor$ der Vektor, der aus v durch komponentenweise Anwendung von $\lfloor \cdot \rfloor$ entsteht.

6.17 Satz: Sei A eine nichtnegative ganzzahlige Matrix, $n \geq 1$ und $z \neq 0$ ein nichtnegativer ganzzahliger Vektor. Berechne

$$z' := \min \left\{ z, \left\lfloor \frac{1}{n} Az \right\rfloor \right\}.$$

Ist $z' = z$, gib z' aus und beende das Programm. Ist $z' \neq z$, ersetze z durch z' und wiederhole den letzten Schritt. Dieser Prozess terminiert nach endlich vielen

¹¹Wenn wir einen approximativen Eigenvektor v für eine irreduzible Komponente gefunden haben, erhalten wir einen approximativen Eigenvektor für die ganze Matrix, indem wir v durch Nullen in den restlichen Komponenten erweitern.

Schritten und gibt entweder einen (A, n) -approximativen Eigenvektor oder 0 aus.

Beweis: Das Programm endet nach endlich vielen Schritten, da die berechneten Vektoren nichtnegativ, ganzzahlig und (nicht notwendigerweise strikt) monoton fallend in jeder Komponente sind. Falls der Output z ist, dann gilt $z \geq 0$ und

$$z = \min \left\{ z, \left\lfloor \frac{1}{n} Az \right\rfloor \right\} \leq \frac{1}{n} Az,$$

so dass $Az \geq nz$. Folglich ist entweder $z = 0$ oder z ist ein (A, n) -approximativer Eigenvektor. \square

Eine typische Anwendung des obigen Algorithmus beginnt mit dem Vektor $z = [1 \ 1 \ \dots \ 1]^T$. Ist der Output $\neq 0$, so sind wir fertig. Andernfalls versuchen wir es mit $z = [2 \ 2 \ \dots \ 2]^T$. Ist der Output $\neq 0$, so sind wir fertig. Andernfalls versuchen wir es mit $z = [3 \ 3 \ \dots \ 3]^T$ usw. So erhalten wir schließlich den approximativen Eigenvektor mit kleinstmöglichem maximalen Eintrag (Übungsaufgabe).

6.18 Beispiel: Sei $A = \begin{bmatrix} 1 & 3 \\ 6 & 1 \end{bmatrix}$ und $n = 5$ wie in Beispiel 6.14. Beginne mit $z = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$. Die Vektoren, die der Algorithmus berechnet, sind

$$\begin{bmatrix} 1 \\ 1 \end{bmatrix} \rightarrow \begin{bmatrix} 0 \\ 1 \end{bmatrix} \rightarrow \begin{bmatrix} 0 \\ 0 \end{bmatrix} \rightarrow \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

so dass diese Wahl von z nicht funktioniert. Als nächstes wählen wir $z = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$, woraus

$$\begin{bmatrix} 2 \\ 2 \end{bmatrix} \rightarrow \begin{bmatrix} 1 \\ 2 \end{bmatrix} \rightarrow \begin{bmatrix} 1 \\ 1 \end{bmatrix} \rightarrow \begin{bmatrix} 0 \\ 1 \end{bmatrix} \rightarrow \begin{bmatrix} 0 \\ 0 \end{bmatrix} \rightarrow \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

resultiert. Aber beim dritten Versuch ergibt sich

$$\begin{bmatrix} 3 \\ 3 \end{bmatrix} \rightarrow \begin{bmatrix} 2 \\ 3 \end{bmatrix} \rightarrow \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

und wir haben den kleinsten $(A, 5)$ -approximativen Eigenvektor gefunden.

6.4 Konstruktion von Codes

In diesem Unterabschnitt wollen wir den Beweis des Finite-State-Coding-Theorems vervollständigen und die Ideen des Beweises verwenden, um in der Praxis verwendbare Codes zu konstruieren.

Wir erinnern an die Aussage des Theorems: Sei X ein sofscher Shift und $n \geq 1$. Dann existiert ein Finite-State- (X, n) -Code genau dann, wenn $h(X) \geq \log n$.

Die Notwendigkeit dieser Bedingung haben wir bereits bewiesen. Für die andere Beweisrichtung beginnen wir mit einer rechtsauflösenden Darstellung (G, \mathcal{L}) und verwenden eine Folge von elementaren Zustandsaufspaltungen, wobei wir mit G beginnen und bei einem Graphen H enden, dessen Ausgangsgrad $\geq n$ ist. Der folgende Satz liefert eine präzise Formulierung des Zustandsaufspaltungsarguments.

6.19 Satz: Sei G ein irreduzibler Graph mit Übergangsmatrix A . Ist $\lambda_A \geq n$, dann gibt es eine Folge

$$G = G_0, G_1, \dots, G_m = H,$$

so dass G_{j+1} aus G_j für $1 \leq j \leq m$ durch eine elementare Zustandsaufspaltung entsteht und H minimalen Ausgangsgrad n hat. Ist v ein positiver (A, n) -approximativer Eigenvektor und k die Summe seiner Komponenten, so gibt es eine derartige Folge von Zustandsaufspaltungen, so dass

$$m \leq k - |\mathcal{V}(G)| \quad \text{und} \quad |\mathcal{V}(H)| \leq k.$$

Wir werden diesen Satz später beweisen. Zuerst verwenden wir ihn, um den Beweis des Finite-State-Coding-Theorems abzuschließen.

Beweis (von Satz 6.11): Sei X ein sofischer Shift und $n \geq 1$, so dass $h(X) \geq \log n$. Wir dürfen annehmen, dass $X = X_{\mathcal{G}}$, wobei $\mathcal{G} = (G, \mathcal{L})$ eine irreduzible und rechtsauflösende Darstellung ist. Eine solche Darstellung erhalten wir mit Hilfe einer beliebigen Darstellung, auf die wir die Teilmengekonstruktion aus dem Beweis von Satz 4.22 anwenden und anschließend mit Hilfe von Satz 5.29 eine irreduzible Komponente \mathcal{K} auswählen, so dass $h(X_{\mathcal{G}}) = h(X_{\mathcal{K}}) = h(X) \geq \log n$. Dann gilt $\lambda_A \geq n$ für $A = A_G$. Nach Satz 6.19 gibt es eine Folge elementarer Zustandsaufspaltungen $G = G_0, G_1, \dots, G_m = H$, so dass H minimalen Ausgangsgrad n hat. Die Beschriftung \mathcal{L} setzt sich bei diesen Zustandsaufspaltungen fort zu einer Beschriftung \mathcal{L}' von H mit $\mathcal{L}'_{\infty}(X_H) = \mathcal{L}_{\infty}(X_G) = X$. Indem wir Satz 6.7 m mal anwenden, sehen wir, dass (H, \mathcal{L}') rechtsabschließend mit Verzögerung m ist.

Wir wählen einen Teilgraphen \hat{H} von H mit konstanten Ausgangsgrad n . Ist \mathcal{O} die Einschränkung von \mathcal{L}' auf \hat{H} und \mathcal{I} eine beliebige Straßenfärbung von \hat{H} , dann ist (\hat{H}, \mathcal{O}) weiterhin rechtsabschließend und $\mathcal{O}_{\infty}(X_{\hat{H}}) \subseteq \mathcal{L}'_{\infty}(X_H) = X$, so dass $(\hat{H}, \mathcal{I}, \mathcal{O})$ ein Finite-State- (X, n) -Code ist. \square

Wir werden Satz 6.19 beweisen durch wiederholte Anwendung des folgenden Resultats.

6.20 Satz: Sei G ein irreduzibler Graph und $v > 0$ ein (A_G, n) -approximativer Eigenvektor. Der minimale Ausgangsgrad von G sei $< n$. Dann gibt es einen Knoten I , eine elementare Zustandsaufspaltung G' von G bei I mit resultierenden Knoten I^1 und I^2 sowie einen $(A_{G'}, n)$ -approximativen Eigenvektor $v' > 0$, so dass $v'_J = v_J$ für $J \neq I$ und $v'_{I^1} + v'_{I^2} = v_I$.

Der Beweis des obigen Satzes basiert auf folgendem Lemma.

6.21 Lemma: Seien k_1, k_2, \dots, k_n positive ganze Zahlen. Dann gibt es eine Teilmenge $S \subseteq \{1, 2, \dots, n\}$, so dass $\sum_{s \in S} k_s$ durch n teilbar ist.

Beweis: Die Teilsummen $k_1, k_1 + k_2, k_1 + k_2 + k_3, \dots, k_1 + k_2 + \dots + k_n$ sind entweder alle voneinander verschieden (modulo n), oder zwei davon sind gleich (modulo n). Im ersten Fall ist mindestens eine Teilsumme 0 (modulo n). Im zweiten Fall gibt es $1 \leq m < p \leq n$, so dass

$$k_1 + k_2 + \dots + k_m = k_1 + k_2 + \dots + k_p \pmod{n}.$$

Daher gilt $k_{m+1} + k_{m+2} + \dots + k_p = 0 \pmod{n}$. □

Nun beweisen wir Satz 6.20.

Beweis (von Satz 6.20): Sei $p := \max_I v_I$. Zuerst wählen wir einen Knoten I mit $v_I = p$, so dass es eine Kante e mit Anfangsknoten I gibt und $v_{t(e)} < p$. Um zu sehen, dass so ein Knoten existiert, beginne mit einem Knoten J , für den $v_J = p$ gilt. Dann muss es ein K mit $v_K < p$ geben, da andernfalls $\mathbf{1}$ ein (A_G, n) -approximativer Eigenvektor wäre und somit G minimalen Ausgangsgrad $\geq n$ hätte. Da G irreduzibel ist, gibt es einen Pfad von J nach K und der letzte Knoten I dieses Pfads, für den $v_I = p$ gilt, ist der gesuchte Knoten.

Nun müssen wir die Menge \mathcal{E}_I der bei I beginnenden Kanten partitionieren. Die I -te Komponente der Vektorungleichung $A_G v \geq n v$ ist

$$\sum_{e \in \mathcal{E}_I} v_{t(e)} \geq n v_I = n p.$$

Jede der Zahlen $v_{t(e)}$ mit $e \in \mathcal{E}_I$ ist höchstens p und mindestens eine dieser Zahlen ist echt kleiner als p . Daher gilt

$$\sum_{e \in \mathcal{E}_I} v_{t(e)} < p |\mathcal{E}_I|,$$

so dass $m := |\mathcal{E}_I| > n$. Sei $\mathcal{E}_I = \{e_1, \dots, e_n, \dots, e_m\}$, wobei e_1 die Kante e ist mit $v_{t(e)} < p$. Indem wir Lemma 6.21 auf die n positiven Zahlen $v_{t(e_1)}, v_{t(e_2)}, \dots, v_{t(e_n)}$ anwenden, erhalten wir eine Teilmenge \mathcal{E}_I^1 von $\{e_1, e_2, \dots, e_n\}$ mit

$$\sum_{e \in \mathcal{E}_I^1} v_{t(e)} = 0 \pmod{n}. \quad (21)$$

Wir definieren $\mathcal{E}_I^2 := \mathcal{E}_I \setminus \mathcal{E}_I^1$. Diese Menge ist nichtleer, da $n < m$. Sei G' die aus der Partition $\{\mathcal{E}_I^1, \mathcal{E}_I^2\}$ resultierende Zustandsaufspaltung von G .

Nach (21) ist

$$q := \frac{1}{n} \sum_{e \in \mathcal{E}_I^1} v_{t(e)}$$

eine positive ganze Zahl. Definiere v' durch

$$v'_J := \begin{cases} q & \text{falls } J = I^1, \\ v_I - q & \text{falls } J = I^2, \\ v_J & \text{andernfalls.} \end{cases}$$

Wir vervollständigen den Beweis, indem wir zeigen, dass $v' > 0$ und dass v' ein $(A_{G'}, n)$ -approximativer Eigenvektor ist. Da q und die Einträge von v positive ganze Zahlen sind, ist die einzige Komponente von v' , über die ein Zweifel besteht, die mit Index $J = I^2$. Aber

$$q = \frac{1}{n} \sum_{e \in \mathcal{E}_I^1} v_{t(e)} \leq \frac{1}{n} \sum_{j=1}^n v_{t(e_j)} < p,$$

da $v_{t(e_j)} \leq p$ für alle j und $v_{t(e_1)} < p$. Folglich gilt $v_I - q = p - q > 0$, womit $v' > 0$ gezeigt ist.

Um zu zeigen, dass v' ein approximativer Eigenvektor ist, müssen wir beweisen, dass für jeden Knoten J von G' gilt, dass $\sum_{e \in \mathcal{E}_J} v'_{t(e)} \geq nv_J$. Dazu beobachten wir zuerst, dass jede bei I endende Kante e in e^1 und e^2 aufgespalten wird, aber dass die Summe $v'_{I^1} + v'_{I^2}$ der Gewichte ihrer Endknoten gleich dem Gewicht v_I des Endknotens von e ist. Ist daher J nicht I^1 oder I^2 , dann bleiben die Summen in der approximativen Eigenvektorgleichung unverändert und die Ungleichung bleibt bestehen. Ist $J = I^1$, so gilt

$$\sum_{e \in \mathcal{E}_{I^1}} v'_{t(e)} = \sum_{e \in \mathcal{E}_I^1} v_{t(e)} = nq = nv'_{I^1},$$

während für $J = I^2$ gilt, dass

$$\begin{aligned} \sum_{e \in \mathcal{E}_{I^2}} v'_{t(e)} &= \sum_{e \in \mathcal{E}_I \setminus \mathcal{E}_I^1} v_{t(e)} = \left(\sum_{e \in \mathcal{E}_I} v_{t(e)} \right) - nq \\ &\geq nv_I - nq = n(v_I - q) = nv'_{I^2}. \end{aligned}$$

□

Nun können wir die Beweise in diesem Unterabschnitt abschließen.

Beweis (von Satz 6.19): Sei v ein positiver (A_G, n) -approximativer Eigenvektor und k die Summe seiner Komponenten. Hat G minimalen Ausgangsgrad $\geq n$, dann sind wir fertig. Wenn nicht, dann können wir gemäß Satz 6.20 G aufspalten und dadurch einen irreduziblen Graphen G' mit einem Zustand mehr bilden und einen $(A_{G'}, n)$ -approximativen Eigenvektor $v' > 0$ finden, dessen Komponenten sich ebenfalls zu k aufsummieren. Hat G' minimalen Ausgangsgrad n , so sind wir fertig. Andernfalls wenden wir Satz 6.20 nochmals an. Um zu sehen, dass dieser Prozess schließlich zu einem Ende kommt, beobachten wir, dass die approximativen Eigenvektoren v, v', v'', \dots immer mehr Komponenten

haben, während die positiven ganzen Zahlen in jedem Vektor sich stets zu k aufsummieren. Die positiven ganzzahligen Gewichte können sich nur begrenzt verteilen und wir müssen nach höchstens $k - |\mathcal{V}(G)|$ Schritten aufhören, was zu einem Graphen mit höchstens k Knoten führt. \square

Die Ideen aus diesem Beweis können in konkreten Situationen verwendet werden, um Codes zu konstruieren. Um einen Code mit Kodierungsrate p/q zu konstruieren, können wir nach dem Gezeigten wie folgt vorgehen:

- (1) Finde eine rechtsauflösende Darstellung \mathcal{K} von X mit zugrundeliegendem Graphen K (die minimale rechtsauflösende Darstellung in Satz 4.50 ist vielleicht der einfachste Ausgangspunkt).
- (2) Berechne $h(X) = \log \lambda_{A(K)}$.
- (3) Wähle ganze Zahlen p und q , so dass $h(X) \geq p/q$.
- (4) Konstruiere \mathcal{K}^q . (Beachte, dass $h(X_{\mathcal{K}^q}) \geq \log 2^p$.)
- (5) Wende den approximativen Eigenvektor-Algorithmus an, um einen $(A(K^q), 2^p)$ -approximativen Eigenvektor v zu finden. Entferne alle Knoten I mit $v_I = 0$ und betrachte die Einschränkung auf eine Senkenkomponente \mathcal{G} dieses Graphen (siehe Lemma 6.15).
- (6) Wende die Zustandsaufspaltungsmethode aus Satz 6.20 so lange an, bis eine andere Darstellung \mathcal{H} von X mit minimalem Ausgangsgrad $\geq 2^p$ entstanden ist. Satz 6.19 garantiert, dass dies funktioniert. Beachte, dass \mathcal{H} rechtsabschließend ist.
- (7) Entferne, falls notwendig, Kanten von \mathcal{H} , um $\widehat{\mathcal{H}} = (\widehat{H}, \mathcal{O})$ mit konstantem Ausgangsgrad 2^p zu erhalten. Verwende als Input-Beschriftung \mathcal{I} eine Straßenfärbung von G unter Verwendung binärer p -Blöcke. Dann ist $(\widehat{H}, \mathcal{I}, \mathcal{O})$ der gesuchte Finite-State-Code.

6.22 Bemerkung:

- (a) Sei X ein sofischer Shift mit $h(X) \geq \log n$ und A die Übergangsmatrix einer rechtsauflösenden Darstellung von X . Satz 6.19 liefert eine obere Schranke für die Größe des kleinsten Finite-State- (X, n) -Codes: Es gibt so einen Code mit höchstens $\min_v \sum_i v_i$ Zuständen (Knoten), wobei das Minimum über alle (A, n) -approximativen Eigenvektoren v genommen wird. Satz 6.19 liefert auch eine obere Schranke für die kleinste Verzögerung eines beliebigen Finite-State- (X, n) -Codes: $\min_v \sum_i (v_i - 1)$, wobei wiederum das Minimum über alle (A, n) -approximativen Eigenvektoren v genommen wird.
- (b) Im Beweis von Satz 6.19 haben wir Finite-State-Codes konstruiert durch die Verwendung eines positiven approximativen Eigenvektors v . Es kann jedoch einfacher sein, einen approximativen Eigenvektor mit ein paar

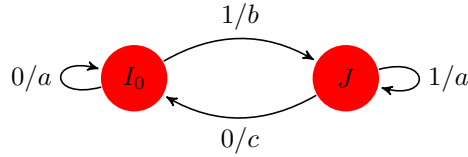


Abbildung 22: Eine schlechte Input-Beschriftung

verschwindenden Komponenten zu verwenden und entsprechende Knoten (wie in Schritt 5) zu entfernen.

- (c) Man muss sich nicht auf elementare Zustandsaufspaltungen beschränken, wenn man Satz 6.20 anwendet. Es ist eventuell effizienter, viele Knoten auf einmal in viele Teilknoten aufzuspalten. Man spaltet Zustände auf, indem man Partitionen $\mathcal{P}_I = \{\mathcal{E}_I^1, \dots, \mathcal{E}_I^{m(I)}\}$ von \mathcal{E}_I findet und Zerlegungen $v_I = \sum_{i=1}^{m(I)} v'_{I,i}$ in positive ganze Zahlen, so dass für $i = 1, \dots, m(I)$ gilt:

$$\sum_{e \in \mathcal{E}_I^i} v_{t(e)} \geq n v'_{I,i}.$$

Dies führt zu einer Aufspaltung von Knoten I in Knoten $I^1, \dots, I^{m(I)}$ mit neuen approximativen Eigenvektorkomponenten $v'_{I,1}, \dots, v'_{I,m(I)}$.

6.5 Gleitende Block-Dekodierer

Angenommen, wir verwenden einen Finite-State-Code $(G, \mathcal{I}, \mathcal{O})$, um eine Input-Folge $x = x_0 x_1 x_2 \dots$ in eine Output-Folge $y = y_0 y_1 y_2 \dots$ zu transformieren. Um x wiederzugewinnen, muss der Dekodierer genau verfolgen, in welchem Zustand von G er sich in jedem Schritt befindet. Ein fehlerhaftes Symbol in y könnte dafür sorgen, dass der Dekodierer den aktuellen Zustand nicht mehr kennt. Im Folgenden sehen wir ein einfaches Beispiel für ein solches Verhalten.

6.23 Beispiel: Betrachte den Finite-State-Code in Abbildung 22. Mit I_0 als Anfangsknoten wird der Eingang $00000 \dots$ kodiert zu $aaaaa \dots$. Aber es können stets Fehler auftreten. Zum Beispiel könnte das erste Symbol als ein b erfasst werden. In diesem Fall würde der Dekodierer den Input $11111 \dots$ rekonstruieren, der in jedem Bit falsch ist.

Diese Art von unendlicher Fehlerfortpflanzung kann verhindert werden durch die Verwendung eines Finite-State-Codes mit einem gleitenden Block-Dekodierer.

6.24 Definition: Sei $(G, \mathcal{I}, \mathcal{O})$ ein Finite-State- (X, n) -Code. Ein gleitender Block-Dekodierer für $(G, \mathcal{I}, \mathcal{O})$ ist ein gleitender Blockcode $\phi : X \rightarrow X_{[n]}$, so

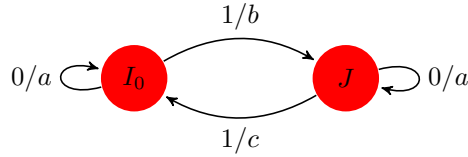
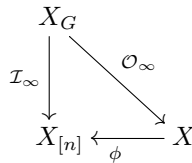


Abbildung 23: Eine gute Input-Beschriftung

dass $\phi \circ \mathcal{O}_\infty = \mathcal{I}_\infty$, d.h. das folgende Diagramm kommutiert:



In dieser Definition muss ϕ nur auf dem Bild von \mathcal{O}_∞ definiert sein. Allerdings kann jede solche Abbildung nach ganz X erweitert werden (Übungsaufgabe).

6.25 Beispiel: Wir betrachten nochmals den Finite-State-Code in Abbildung 22. \mathcal{O}_∞ bildet beide Fixpunkte in $X_{[2]}$ auf a^∞ ab, während \mathcal{I}_∞ sie unterscheidet. Folglich kann es für diesen Code keinen gleitenden Block-Dekodierer ϕ mit $\phi \circ \mathcal{O}_\infty = \mathcal{I}_\infty$ geben.

Nun verändern wir die Input-Beschriftung auf \mathcal{E}_J , so dass der Finite-State-Code in 23 entsteht. Dieser neue Finite-State-Code hat einen gleitenden Block-Dekodierer, nämlich den 1-Blockcode, der von der 1-Blockabbildung Φ induziert wird, die durch $\Phi(a) := 0$ und $\Phi(b) = \Phi(c) := 1$ definiert ist.

Bis jetzt war die Wahl der Input-Straßenfärbung beliebig. Dieses Beispiel zeigt, dass manche Input-Straßenfärbungen anderen vorzuziehen sind.

Hat ein Finite-State-Code $(G, \mathcal{I}, \mathcal{O})$ einen gleitenden Block-Dekodierer ϕ , so gibt es eine Dekodierungsmethode, die den aktuellen Zustand auf G nicht verfolgen muss. Um dies zu zeigen, nehmen wir an, dass $\phi = \Phi_\infty^{[-m,n]}$ und dass $y = y_0y_1y_2\dots$ eine Output-Folge ist. Für jedes $k \geq m$ gibt es einen Pfad $e_{k-m}e_{k-m+1}\dots e_{k+n}$ auf G mit

$$\mathcal{O}(e_{k-m}e_{k-m+1}\dots e_{k+n}) = y_{k-m}y_{k-m+1}\dots y_{k+n}.$$

Da $\phi \circ \mathcal{O}_\infty = \mathcal{I}_\infty$, folgt

$$x_k = \mathcal{I}(e_k) = \Phi(y_{k-m}y_{k-m+1}\dots y_{k+n}).$$

Wir können folglich alle außer die ersten m Input-Symbole wiedergewinnen, indem wir Φ auf Output-Blöcke anwenden. Der entscheidende Vorteil der Verwendung von Φ ist, dass ein fehlerhaftes Output-Symbol die Dekodierung nur

dann beeinträchtigen kann, wenn es in dem Fenster für Φ auftaucht. Solche Fehler können sich also nicht unendlich fortpflanzen.

Da Finite-State-Codes mit gleitenden Block-Dekodierern so wünschenswert sind, stellt sich die Frage, wie wir sie finden können. Der folgende Satz zeigt, dass dies kein Problem ist, wenn X von endlichem Typ ist.

6.26 Satz: *Sei X ein Shift von endlichem Typ mit $h(X) \geq \log n$. Dann gibt es einen Finite-State- (X, n) -Code mit gleitendem Block-Dekodierer.*

Beweis: Indem wir zu einem irreduziblen Teilshift mit maximaler Entropie übergehen, können wir annehmen, dass X irreduzibel ist. Sei $\mathcal{G} = (G, \mathcal{L})$ eine beliebige Darstellung von X , so dass $\mathcal{L}_\infty : X_G \rightarrow X_{\mathcal{G}}$ eine Konjugation ist. Zum Beispiel könnte \mathcal{G} eine minimale rechtsauflösende Darstellung von X sein (siehe Satz 4.52). Da Zustandsaufspaltungen Konjugationen induzieren, erhalten wir im Beweis des Finite-State-Coding-Theorems, wenn wir \mathcal{G} als Anfangsdarstellung verwenden, am Ende einen beschrifteten Graphen (H, \mathcal{L}') , so dass $\mathcal{L}'_\infty : X_H \rightarrow X$ eine Konjugation ist. Indem wir Kanten aus H entfernen, erhalten wir einen Graphen \hat{H} mit konstantem Ausgangsgrad n und einen Finite-State-Code $(\hat{H}, \mathcal{I}, \mathcal{O})$, wobei \mathcal{O} die Einschränkung von \mathcal{L}' auf \hat{H} ist. Nach den Sätzen 2.29 und 2.30 ist \mathcal{O}_∞^{-1} ein gleitender Blockcode. Dann ist auch $\phi := \mathcal{I}_\infty \circ \mathcal{O}_\infty^{-1}$ ein gleitender Blockcode mit $\phi \circ \mathcal{O}_\infty = \mathcal{I}_\infty$, so dass ϕ ein gleitender Block-Dekodierer ist. \square

6.27 Bemerkung: Obiger Satz kann auf sofische Shifts X mit $h(X) > \log n$ verallgemeinert werden. Für sofische Shifts mit $h(X) = \log n$ muss es keine Finite-State- (X, n) -Codes mit gleitenden Block-Dekodierern geben.

Eine Folgerung von Satz 6.26 ist, dass unter allen Shifts von endlichem Typ mit $h(X) \geq \log n$ der volle n -Shift in dem Sinne der einfachste ist, dass er ein Faktor von jedem solchen Shift ist.

6.28 Korollar: *Sei X ein Shift von endlichem Typ mit $h(X) \geq \log n$. Dann gibt es einen Faktorcode $\phi : X \rightarrow X_{[n]}$.*

Beweis: Satz 6.26 liefert einen Finite-State- (X, n) -Code $(G, \mathcal{I}, \mathcal{O})$ mit einem gleitenden Block-Dekodierer ϕ . Dann gilt

$$X_{[n]} = \mathcal{I}_\infty(X_G) = \phi(\mathcal{O}_\infty(X_G)) \subseteq \phi(X) \subseteq X_{[n]},$$

so dass ϕ surjektiv und damit ein Faktorcode ist. \square

7 Übungsaufgaben

Blatt 0 (Präsenzübungsblatt)

Aufgabe 1:

Sei \mathcal{A} ein Alphabet. Bestimmen Sie die Anzahl der

- (a) Fixpunkte von σ in $\mathcal{A}^{\mathbb{Z}}$.
- (b) n -periodischen Punkte von σ in $\mathcal{A}^{\mathbb{Z}}$.
- (c) periodischen Punkte mit Primperiode 12 von σ in $\mathcal{A}^{\mathbb{Z}}$.

Aufgabe 2:

Wir betrachten die Abbildung

$$\phi : \{0, 1\}^{\mathbb{Z}} \rightarrow \{0, 1\}^{\mathbb{Z}}.$$

mit

$$\phi(x)_i = x_i + x_{i+1} \bmod 2.$$

Dies ist ein stationärer Code (kommutiert mit σ).

- (a) Zeigen Sie, dass ϕ surjektiv ist.
- (b) Bestimmen Sie die Anzahl der Punkte x mit $\phi^n(x) = 0^\infty$ für jedes $n \geq 1$.

Aufgabe 3:

Sei X ein Shiftraum und $N \geq 1$. Zeigen Sie, dass es eine Menge \mathcal{F} von Blöcken der Länge mindestens N gibt, so dass $X = X_{\mathcal{F}}$.

Aufgabe 4:

Sei $S \subset \mathbb{N}_0 = \{0, 1, 2, \dots\}$ eine unendliche Teilmenge. Zeigen Sie, dass

$$X = \{\dots 10^{n-1} 10^{n_0} 10^{n_1} 1 \dots \mid n_j \in S \forall j \in \mathbb{Z}\}$$

kein Shiftraum ist.

Aufgabe 5:

Sei S eine nichtleere Teilmenge von \mathbb{N}_0 . Ist S endlich, so definieren wir $X = X(S)$ als die Menge aller binären Sequenzen, in denen 1 unendlich oft in beiden Richtungen vorkommt und die Anzahl der Nullen zwischen zwei aufeinanderfolgenden Einsen in S ist. Also ist

$$x = \dots 10^{n-1}10^{n_0}10^{n_1}1\dots$$

mit $n_j \in S$ ein typischer Punkt in X . Ist S unendlich, so lassen wir die Forderung, dass die 1 unendlich oft in beiden Richtungen vorkommt, weg. Zeigen Sie, dass $X(S)$ ein Shiftraum ist.

Aufgabe 6:

Zeigen Sie, dass es überabzählbar unendlich viele Shifts in $\{0, 1\}^{\mathbb{Z}}$ gibt.

Blatt 1**Aufgabe 1: Volle Shifts**

- Bestimmen Sie für den vollen $\{+1, -1\}$ -Shift und $k \geq 1$ die Anzahl der k -Blöcke mit der Eigenschaft, dass die Summe der Symbole 0 ergibt.
- Beschreiben Sie diejenigen Paare von Blöcken u und v über einem Alphabet \mathcal{A} , so dass $uv = vu$.

Aufgabe 2: Lauflängen-beschränkte Shifts

Sei X die Menge aller binären Folgen, in denen die 1 unendlich oft in beiden Richtungen vorkommt, so dass die Anzahl von Nullen zwischen aufeinanderfolgenden Einsen entweder 1, 2 oder 3 ist. Zeigen Sie, dass X ein Shiftraum ist, indem Sie eine Menge \mathcal{F} von Blöcken angeben, so dass $X = X_{\mathcal{F}}$.

Aufgabe 3: Shifträume I

- Geben Sie eine Menge \mathcal{F} von Blöcken über $\{0, 1\}$ an, so dass $X_{\mathcal{F}} = \emptyset$.
- Sei X die Teilmenge von $\{0, 1\}^{\mathbb{Z}}$, die aus allen Folgen besteht, die genau eine Eins enthalten. Zeigen Sie, dass $X \cup \{0^\infty\}$ ein Shiftraum ist.
- Geben Sie zwei Mengen \mathcal{F}_1 und \mathcal{F}_2 über $\mathcal{A} = \{0, 1\}$ an mit $X_{\mathcal{F}_1} = X_{\mathcal{F}_2} \neq \emptyset$, so dass \mathcal{F}_1 endlich und \mathcal{F}_2 unendlich ist.
- Zeigen Sie, dass $X_{\mathcal{F}_1} \cap X_{\mathcal{F}_2} = X_{\mathcal{F}_1 \cup \mathcal{F}_2}$. Zeigen Sie außerdem, dass beliebige Schnitte von Shifträumen wieder Shifträume sind.

Aufgabe 4: Shifträume II

Sei X der volle \mathcal{A} -Shift über einem Alphabet \mathcal{A} .

- (a) Zeigen Sie, dass aus $X = X_1 \cup X_2$ für Shifträume X_1 und X_2 folgt, dass $X = X_1$ oder $X = X_2$.
- (b) Erweitern Sie Ihr Argument in (a) um zu zeigen, dass aus $X = \bigcup_{\alpha \in A} X_\alpha$ für Shifträume X_α (A beliebige nichtleere Indexmenge) folgt, dass $X = X_\alpha$ für ein $\alpha \in A$.
- (c) Erklären Sie, warum diese Aussagen nicht gelten, wenn wir lediglich annehmen, dass X ein Shiftraum ist.

Blatt 2

Aufgabe 1: Sprachen I

Bestimmen Sie die Sprachen der folgenden Shifträume:

- (a) $X =$ Menge aller Punkte in $\{0, 1\}^{\mathbb{Z}}$, so dass die Anzahl der Nullen zwischen je zwei Einsen gerade ist.
- (b) $X =$ Menge aller Punkte in $\{0, 1\}^{\mathbb{Z}}$, in denen die 1 unendlich oft in beiden Richtungen vorkommt, so dass die Anzahl der Nullen zwischen aufeinanderfolgenden Einsen entweder 1, 2 oder 3 ist.
- (c) $X =$ Menge aller Punkte in $\{+1, -1\}^{\mathbb{Z}}$, so dass für jeden Block w , der in einem $x \in X$ vorkommt, die algebraische Summe s der Symbole in w die Ungleichungen $-c \leq s \leq c$ erfüllt. Hier ist c eine fest gewählte natürliche Zahl.

Zeigen Sie für die Menge X in (c) auch, dass es sich um einen Shiftraum handelt.

Aufgabe 2: Sprachen II

Im Folgenden verstehen wir unter einer Sprache eine Menge von Blöcken, die die Eigenschaften aus Satz 2.9(1) erfüllt.

- (a) Seien \mathcal{L}_1 und \mathcal{L}_2 Sprachen über demselben Alphabet. Zeigen Sie, dass auch $\mathcal{L}_1 \cup \mathcal{L}_2$ eine Sprache ist. Folgern Sie, dass die Vereinigung zweier Shifträume wieder ein Shiftraum ist.
- (b) Nun seien \mathcal{L}_n , $n \in \mathbb{N}$, Sprachen über demselben Alphabet. Zeigen Sie, dass $\bigcup_{n=1}^{\infty} \mathcal{L}_n$ eine Sprache ist. Warum können Sie daraus nicht folgern, dass die Vereinigung abzählbar unendlich vieler Shifträume über demselben Alphabet wieder ein Shiftraum ist?

Aufgabe 3: Konjugationen und Faktorcodes

- (a) Zeigen Sie, dass die Konjugationsrelation \cong für Shifträume eine Äquivalenzrelation ist, d.h. (i) $X \cong X$, (ii) aus $X \cong Y$ folgt $Y \cong X$ und (iii) aus $X \cong Y$, $Y \cong Z$ folgt $X \cong Z$.
- (b) Ist der volle 2-Shift konjugiert zum vollen 3-Shift?
- (c) Bestimmen Sie einen Faktorcode vom vollen 3-Shift auf den vollen 2-Shift, d.h. einen surjektiven gleitenden Blockcode $\phi : \{0, 1, 2\}^{\mathbb{Z}} \rightarrow \{0, 1\}^{\mathbb{Z}}$.
- (d) Gibt es einen Faktorcode vom vollen 2-Shift auf den vollen 3-Shift?
- (e) Sei $X = \{0, 1\}^{\mathbb{Z}}$ und $\Phi : \{0, 1\} \rightarrow \{0, 1\}$ die 1-Blockabbildung gegeben durch $\Phi(0) = 1$ und $\Phi(1) = 0$. Zeigen Sie, dass der induzierte gleitende Blockcode $\phi = \Phi_{\infty} : X \rightarrow X$ eine Konjugation ist.

Aufgabe 4: Abbildungen zwischen Shifträumen

Geben Sie ein Beispiel für eine Abbildung $\phi : X \rightarrow Y$ zwischen Shifträumen X, Y an, die *kein* gleitender Blockcode ist, aber dennoch $\phi \circ \sigma_X = \sigma_Y \circ \phi$ erfüllt.

Blatt 3

Aufgabe 1: Shifts von endlichem Typ

- (a) Seien X, Y Shifts von endlichem Typ über einem Alphabet \mathcal{A} . Zeigen Sie, dass auch $X \cap Y$ von endlichem Typ ist. Gilt dies auch für $X \cup Y$?
- (b) Charakterisieren Sie die Teilmengen $S \subseteq \mathbb{N}_0$, so dass der Shift $X(S)$ (definiert in Aufgabe 5, Präsenzübungsblatt) von endlichem Typ ist.

Aufgabe 2: Übergangsmatrix eines Graphen

- (a) Zeigen Sie, dass für jede $r \times r$ -Matrix A mit nichtnegativen ganzzahligen Einträgen gilt, dass $A = A(G(A))$. Zeigen Sie außerdem, dass für jeden Graphen G gilt, dass $G \cong G(A(G))$.
- (b) Sei G ein Graph mit Übergangsmatrix A . Zeigen Sie, dass die folgenden Aussagen äquivalent sind:
 - (i) Es gibt ein $n \in \mathbb{N}$, so dass die Länge aller Pfade in G kleiner als n ist.
 - (ii) A ist nilpotent, d.h. $A^n = 0$ für ein $n \in \mathbb{N}$.
 - (iii) $X_G = \emptyset$.

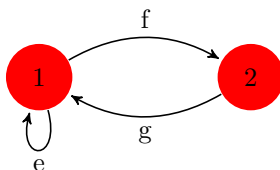


Abbildung 24: Der Graph G

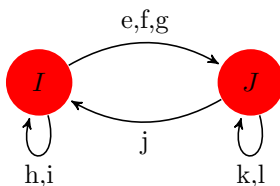


Abbildung 25: Der Graph H

Aufgabe 3: Kantenshifts

Zeigen Sie, dass die Kantenshifts zweier isomorpher Graphen konjugiert zueinander sind.

Aufgabe 4: Graphen

Betrachten Sie die abgebildeten Graphen G und H .

- Zeigen Sie, dass G isomorph zu einem Teilgraphen von H ist.
- Bestimmen Sie für $1 \leq n \leq 6$ die Anzahl der Pfade¹² auf G der Länge n .
- Bestimmen Sie für $1 \leq n \leq 6$ die Anzahl der periodischen Punkte in X_G mit Periode n . Was könnte die Anzahl für beliebiges n sein?

Blatt 4

Aufgabe 1: Irreduzible und wesentliche Graphen

- Zeigen Sie, dass mit genau einer Ausnahme jeder irreduzible Graph wesentlich ist und finden Sie die Ausnahme.

¹²Ein Pfad $\pi = e_1 e_2 \dots e_m$ auf einem Graphen G ist eine endliche Folge von Kanten e_i aus G , so dass $t(e_i) = i(e_{i+1})$ für $1 \leq i \leq m-1$. Die Länge von $\pi = e_1 e_2 \dots e_m$ ist $|\pi| = m$.

- (b) Seien G und H wesentliche Graphen mit $X_G = X_H$. Zeigen Sie, dass $G = H$.
- (c) Wie erkennt man an der Übergangsmatrix eines Graphen, ob der Graph irreduzibel ist?

Aufgabe 2: Höhere Kantengraphen und Potenzgraphen

- (a) Sei G der Graph des vollen 2-Shifts, der nur einen Knoten und genau zwei Schleifen bei diesem Knoten hat. Zeichnen Sie die Graphen $G^{[3]}$ und G^3 .
- (b) Sei X ein Shift von endlichem Typ. Zeigen Sie, dass auch $X^{[N]}$ und X^N von endlichem Typ sind.
- (c) Sei X ein Shift, so dass $X^{[N]}$ von endlichem Typ ist. Ist dann auch X von endlichem Typ? Ist X von endlichem Typ, falls X^N von endlichem Typ ist?

Aufgabe 3: Graphen und höhere Blockshifts

Sei $\mathcal{A} = \{0, 1\}$, $\mathcal{F} = \{000, 111\}$ und $X = X_{\mathcal{F}}$.

- (a) Konstruieren Sie einen Graphen G mit $X_G = X_{\mathcal{F}}^{[3]}$.
- (b) Verwenden Sie die Übergangsmatrix von G , um die Anzahl der Punkte mit Periode 5 in X zu berechnen.

Aufgabe 4: Algorithmus zur Blockzuordnung

Formulieren Sie einen Algorithmus, der auf der Basis einer endlichen Liste \mathcal{F} verbotener Blöcke und eines gegebenen Blocks w entscheidet, ob w in $\mathcal{B}(X_{\mathcal{F}})$ vorkommt.

Blatt 5

Aufgabe 1: Zustandsaufspaltung I

Sei G ein Graph und \mathcal{P} eine Partition der Kanten von G in nichtleere Teilmengen. Ferner sei $H = G^{[\mathcal{P}]}$ der zugehörige Zustandsaufspaltungsgraph.

- (a) Zeigen Sie, dass H irreduzibel ist, falls G irreduzibel ist.

- (b) Beschreiben Sie unter Verwendung der Divisions- und Kantenmatrizen für diese Aufspaltung, wie A_H aus A_G entsteht, indem Sie jede Zeile von A_G in eine Summe von Zeilen zerlegen und Kopien der zugehörigen Spalten einfügen.

Aufgabe 2: Zustandsaufspaltung II

Zeigen Sie, dass sich Satz 3.14 im Skript folgendermaßen erweitern lässt: Seien G und H wesentliche Graphen. Dann ist H eine Zustandsaufspaltung von G genau dann, wenn eine Divisionsmatrix D (d.h. eine nicht-quadratische 0-1-Matrix mit mindestens einer 1 in jeder Zeile und genau einer 1 in jeder Spalte) und eine nicht-quadratische Matrix E mit nichtnegativen ganzzahligen Einträgen existieren, so dass

$$A_G = DE \quad \text{und} \quad A_H = ED.$$

Aufgabe 3: Anzahl der sofischen Shifts

Beweisen Sie, dass es nur abzählbar unendlich viele sofische Shifts im vollen 2-Shift gibt.

Aufgabe 4: Sofische Shifts

Wir betrachten den Shift $X = X_{\mathcal{F}}$ über $\mathcal{A} = \{+1, -1\}$, definiert durch (siehe Aufgabe (1c) auf Blatt 2)

$$\mathcal{F} = \left\{ u : u = x_0 \dots x_n \text{ Block über } \{+1, -1\} \text{ mit } \left| \sum_{i=0}^n x_i \right| > 3 \right\}.$$

Zeigen Sie, dass X durch einen beschrifteten Graphen mit symbolischer Übergangsmatrix

$$\begin{bmatrix} \emptyset & 1 & \emptyset & \emptyset \\ -1 & \emptyset & 1 & \emptyset \\ \emptyset & -1 & \emptyset & 1 \\ \emptyset & \emptyset & -1 & \emptyset \end{bmatrix}$$

dargestellt wird. Zeigen Sie außerdem, dass X strikt sofisch ist.

Blatt 6

Aufgabe 1: Konjugiertheit sofischer Shifts

Seien \mathcal{G} und \mathcal{H} die beschrifteten Graphen mit symbolischen Übergangsmatrizen

$$\begin{bmatrix} a & b \\ b & \emptyset \end{bmatrix} \quad \text{und} \quad \begin{bmatrix} a & a \\ b & \emptyset \end{bmatrix}.$$

Zeigen Sie, dass $X_{\mathcal{G}}$ und $X_{\mathcal{H}}$ nicht konjugiert zueinander sind.

Aufgabe 2: Nachfolgermengen

Sei \mathcal{G} ein beschrifteter Graph mit r Knoten. Beweisen Sie, dass $X_{\mathcal{G}}$ höchstens $2^r - 1$ verschiedene Nachfolgermengen hat.

Aufgabe 3: Nachfolgermengengraph

Sei X ein sofischer Shift mit Nachfolgermengengraph (G, \mathcal{L}) . Zeigen Sie, dass X ein Shift von endlichem Typ ist genau dann, wenn \mathcal{L}_{∞} eine Konjugation ist.

Blatt 7

Aufgabe 1: Fusionierter Graph

Konstruieren Sie den aus \mathcal{G} fusionierten Graphen, wobei \mathcal{G} der beschriftete Graph mit folgender symbolischer Übergangsmatrix ist:

$$\begin{bmatrix} \emptyset & b & a \\ b & \emptyset & a \\ a & \emptyset & \emptyset \end{bmatrix}.$$

Aufgabe 2: Minimale rechtsauflösende Darstellungen

Bestimmen Sie die minimale rechtsauflösende Darstellung des Shifts $X = X_{\mathcal{F}}$ über $\{0, 1\}$ mit $\mathcal{F} = \{10^{2k+1}1 : k \geq 0\}$ und des Shifts aus Aufgabe 4 auf Blatt 5.

Aufgabe 3: Synchronisierende Blöcke

Sei X ein irreduzibler sofischer Shift, \mathcal{G} seine minimale rechtsauflösende Darstellung und $w \in \mathcal{B}(X)$. Zeigen Sie, dass w ein synchronisierender Block für \mathcal{G} ist genau dann, wenn aus $uw, vw \in \mathcal{B}(X)$ folgt, dass $uvw \in \mathcal{B}(X)$.

Aufgabe 4: Darstellungen

Finden Sie eine Darstellung des sofischen Shifts aus Beispiel 4.39 mit nur drei Knoten. Warum widerspricht dies nicht der Minimalität der dort angegebenen Darstellungen?

Blatt 8

Aufgabe 1: Minimale rechtsauflösende Darstellungen

Bestimmen Sie die minimalen rechtsauflösenden Darstellungen der sofischen Shifts, die von den beschrifteten Graphen mit folgenden symbolischen Übergangsmatrizen dargestellt werden:

(a)

$$\begin{bmatrix} a & \emptyset & c & b \\ a & c & \emptyset & b \\ a & c & \emptyset & b \\ c & a & \emptyset & b \end{bmatrix}$$

(b)

$$\begin{bmatrix} \emptyset & \emptyset & a & b & \emptyset & c \\ a & \emptyset & \emptyset & b & \emptyset & c \\ \emptyset & a & \emptyset & \emptyset & b & c \\ \emptyset & \emptyset & \emptyset & \emptyset & \emptyset & d \\ \emptyset & \emptyset & \emptyset & \emptyset & \emptyset & d \\ \emptyset & c & a & \emptyset & d & \emptyset \end{bmatrix}$$

Aufgabe 2: Minimale rechtsauflösende Darstellung für Shifts von endlichem Typ

Sei \mathcal{G} die minimale rechtsauflösende Darstellung eines irreduziblen sofischen Shifts X und sei r die Anzahl der Knoten in \mathcal{G} . Zeigen Sie: Ist X von endlichem Typ, dann ist X ein N -Schritt-Shift für $N = (r^2 - r)/2$.

Aufgabe 3: Existenz synchronisierender Blöcke

Sei $\mathcal{G} = (G, \mathcal{L})$ ein rechtsauflösender irreduzibler beschrifteter Graph. Sei \mathcal{H} der (beschriftete) Teilgraph, den man aus $\mathcal{G} \star \mathcal{G}$ erhält, indem man alle Knoten der Form (I, I) mit $I \in \mathcal{V}(G)$ und die zugehörigen Kanten entfernt. Zeigen Sie, dass es einen synchronisierenden Block für \mathcal{G} gibt genau dann, wenn $X_{\mathcal{G}} \neq X_{\mathcal{H}}$.

Aufgabe 4: Bestimmung des Nachfolgermengengraphs

Geben Sie einen Algorithmus an, der auf der Basis einer gegebenen Darstellung eines sofischen Shifts den Nachfolgermengengraphen bestimmt.

Hinweis: Verwenden Sie die Teilmengenkonstruktion und die Methode, die wir in Satz 4.50 verwendet haben, um die minimale rechtsauflösende Darstellung zu bestimmen.

Blatt 9

Aufgabe 1: Faktorcodes zwischen vollen Shifts

Beweisen Sie, dass es einen Faktorcode von $X_{[m]}$ nach $X_{[n]}$ gibt genau dann, wenn $m \geq n$. (Hier ist $X_{[r]}$ der volle r -Shift.)

Aufgabe 2: Eigenschaften der Entropie

Seien X und Y Shifträume und $N \geq 1$. Beweisen Sie:

- (a) $h(X^{[N]}) = h(X)$.
- (b) $h(X^N) = N \cdot h(X)$.
- (c) $h(X \times Y) = h(X) + h(Y)$.
- (d) $h(X \cup Y) = \max\{h(X), h(Y)\}$.

Aufgabe 3: Nicht konjugierte Shifts mit gleicher Entropie

Finden Sie zwei irreduzible Shifts von endlichem Typ mit derselben Entropie, die nicht zueinander konjugiert sind.

Aufgabe 4: Perron-Eigenwert

- (a) Berechnen Sie den Perron-Eigenwert und -eigenvektor der folgenden Matrizen:

$$A = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & 3 \\ 2 & 0 \end{bmatrix}.$$

- (b) Sei A eine irreduzible $r \times r$ -Matrix mit nichtnegativen reellen Einträgen. Sei s_{\min} die kleinste und s_{\max} die größte Zeilensumme von A . Zeigen Sie, dass $s_{\min} \leq \lambda_A \leq s_{\max}$ für den Perron-Eigenwert λ_A gilt.

Blatt 10

Aufgabe 1: Perron-Eigenwert

- (a) Sei A eine irreduzible nichtnegative reelle Matrix. Zeigen Sie, dass für den Perron-Eigenwert λ_A Folgendes gilt:

$$\begin{aligned} \lambda_A &= \inf \{ \lambda \geq 0 : Av \leq \lambda v \text{ für ein } v > 0 \} \\ &= \sup \{ \lambda \geq 0 : Av \geq \lambda v \text{ für ein } v > 0 \}. \end{aligned}$$

- (b) Sei A eine irreduzible Matrix mit nichtnegativen ganzzahligen Einträgen. Zeigen Sie, dass es einen Perron-Eigenvektor v_A mit ganzzahligen Einträgen gibt genau dann, wenn λ_A ganzzahlig ist.

Aufgabe 2: Berechnung der Entropie eines 2-Schritt-Shifts

Sei $\mathcal{A} = \{a, b, c\}$ und

$$\mathcal{F} = \{aac, aba, acb, baa, bbb, bca, cac, cba, cca\}.$$

Berechnen Sie $h(X_{\mathcal{F}})$.

Aufgabe 3: Kriterium für Verschwinden der Entropie

Sei G ein irreduzibler Graph. Zeigen Sie, dass $h(X_G) = 0$ genau dann, wenn G nur aus einem einzigen geschlossenen Pfad besteht.

Aufgabe 4: Mögliche Werte der Entropie

Welche der folgenden Zahlen ist die Entropie eines irreduziblen Shifts von endlichem Typ?

$$(a) \log \sqrt{2}, \quad (b) \log \frac{3}{2}, \quad (c) \log(3 - \sqrt{2}), \quad (d) \log \pi.$$

Blatt 11

Aufgabe 1: Periodenklassen

Vervollständigen Sie den Beweis von Satz 5.40.

Aufgabe 2: Größter gemeinsamer Teiler

Sei S eine Menge positiver ganzer Zahlen mit größtem gemeinsamen Teiler d . Zeigen Sie, dass für ein $k \geq 1$ Zahlen $s_1, \dots, s_k \in S$ und ganze Zahlen a_1, \dots, a_k existieren mit $d = a_1 s_1 + \dots + a_k s_k$.

Hinweis: Betrachten Sie die kleinste positive Zahl in der Untergruppe von $(\mathbb{Z}, +)$, die von S erzeugt wird.

Aufgabe 3: Primitivität

Beweisen Sie Satz 5.44 der Vorlesung.

Blatt 12

Aufgabe 1: Mischende sofische Shifts

Sei X ein irreduzibler sofischer Shift mit minimaler rechtsauflösender Darstellung (G_X, \mathcal{L}_X) .

- (a) Zeigen Sie, dass X mischend ist genau dann, wenn G_X primitiv ist. (Hinweis: Verwenden Sie einen synchronisierenden Block für (G_X, \mathcal{L}_X)).
- (b) Vervollständigen Sie den Beweis von Korollar 5.47(2).

Aufgabe 2: Konjugiertheit der Perioden-Komponenten

Sei A irreduzibel mit Periode p und seien A_1, A_2, \dots, A_p die primitiven Blöcke von A^p . Beweisen Sie, dass die Kantenshifts X_{A_i} alle zueinander konjugiert sind.

Aufgabe 3: Wirkung primitiver Matrizen auf dem positiven Orthanten

Vervollständigen Sie den Beweis von Korollar 5.48.

Aufgabe 4: Konvergenz von A^n

Sei A eine Matrix, deren Eigenwerte alle Absolutbetrag < 1 haben. Zeigen Sie, dass A^n für $n \rightarrow \infty$ gegen die Nullmatrix konvergiert.

Literatur

- [1] Lind, Douglas; Marcus, Brian. *An Introduction to Symbolic Dynamics and Coding*. Cambridge University Press, Cambridge, 1995.