Inhaltsverzeichnis

Kapitel 1

Schnell mischende Markov-Ketten

Betrachten wir eine Irrfahrt auf dem n-dimensionalen Hyperwürfel: Ein Teilchen bewegt sich in jedem Schritt zufällig von seinem derzeitigen Standpunkt zu einem seiner n Nachbarn oder bleibt wo es ist. Bezeichnen wir mit X_t seinen Aufenthaltsort nach t Schritten, so sind die X_t offensichtlich Zufallsvariablen, deren Verteilung, zumindest für kleine t, stark von dem Startpunkt X_0 der Irrfahrt abhängt. Intuitiv sollte andererseits einleuchten, daß bei zunehmender Dauer der Irrfahrt der Einfluß des Startpunktes immer mehr zurückgeht. Nach "sehr vielen" Schritten sind alle Punkte "in etwa" gleichwahrscheinlich. Wie aber läßt sich "sehr viele" konkretisieren: $O(\log n)$ viele oder $O(n)$ oder gar $O(2^n)$ viele Schritte? Und was bedeutet "in etwa gleichwahrscheinlich"? In diesem Kapitel werden wir Methoden entwickeln, um diese Fragen zu beantworten.

1.1 Markov-Ketten

Obiges Beispiel einer Irrfahrt ist das klassische Beispiel der Wahrscheinlichkeitstheorie für eine homogene Markov-Kette mit endlichem Zustandsraum. Während in der Wahrscheinlichkeitstheorie jedoch Verallgemeinerungen dieses einfachen Beispiels (wie zum Beispiel der Übergang vom diskreten zum kontinuierlichen Zeitmodell, unendliche Zustandsräume, etc.) eine zentrale Rolle spielen, können wir uns für unsere Zwecke auf diesen einfachen Fall beschränken. Insbesondere werden wir daher auch keine Kenntnisse über Markov-Ketten voraussetzen, sondern entwickeln und motivieren im folgenden die für uns wesentlichen Begriffe.

Was sind die charakteristischen Merkmale einer Irrfahrt und damit einer Markov-Kette? Zum einen nehmen die Zufallsvariablen X_t nur Werte eines vorgegebenen Raumes $\mathcal{X}$ an, den wir ab jetzt Zustandsraum nennen werden. In unserem Beispiel entspricht also $\mathcal{X}$ den Punkten des n-dimensionalen Würfels.

Viel wichtiger aber: Der Zustand (Aufenthaltsort) zum Zeitpunkt t, d.h. nach t Schritten, hängt nur vom Zustand zum Zeitpunkt $t-1$, nicht aber von den früheren Zuständen ab. Es gilt also

$$Prob(X_t = x \mid X_{t-1} = y_{t-1}, \ldots, X_0 = y_0) = Prob(X_t = x \mid X_{t-1} = y_{t-1})$$
$$\text{für alle } t \in \mathbb{N} \text{ und } x, y_0 \ldots, y_{t-1} \in \mathcal{X}. \quad (1.1.1)$$

Ein drittes wesentliches Merkmal ist, daß die Wahrscheinlichkeit mit der das Teilchen von einem Zustand x zu einem Zustand y wechselt, nicht vom Zeitpunkt, zu dem dieser Übergang stattfindent, abhängt. Mit anderen Worten, es gibt Konstanten p_{xy}, so daß

$$Prob(X_t = x \mid X_{t-1} = y) = p_{xy} \qquad \text{für alle } t \in \mathbb{N} \text{ und } x, y \in \mathcal{X}. \qquad (1.1.2)$$

Jede Folge von Zufallsvariablen $\{X_t, t \in \mathbb{N}_0\}$, die Werte in einer Menge $\mathcal{X}$ annehmen und die die Bedingungen (1.1.1) und (1.1.2) erfüllen, heißt eine *homogene Markov-Kette mit Zustandsraum* $\mathcal{X}$. Da wir hier nur homogene Ketten mit endlichen Zustandsräumen betrachtet werden, meinen wir im folgenden immer eine solche – auch wenn wir nur kurz von einer "Markov-Kette" sprechen.

Die durch (1.1.2) definierte Konstanten p_{xy} einer Markov-Kette nennt man *Übergangswahrscheinlichkeiten* und faßt sie üblicherweise in einer Matrix $P = (p_{xy})$, der *Übergangsmatrix* der Markov-Kette, zusammen.

Stationäre Verteilung einer Markov-Kette

Entsprechend den Übergangswahrscheinlichkeiten p_{xy} einer Markov-Kette definiert man die *Übergangswahrscheinlichkeiten k-ter Stufe* durch

$$p_{xy}^{(k)} = Prob(X_{t+k} = y \mid X_t = x) \qquad \text{für alle } t \in \mathbb{N}_0.$$

(Der Leser sollte sich davon überzeugen, daß aus den Markov-Eigenschaften (1.1.1) und (1.1.2) zusammen mit Induktion sofort folgt, daß diese Wahrscheinlichkeiten wohldefiniert sind.)

Den einfachen Beweis der folgenden Eigenschaft k-stufiger Übergangswahrscheinlichkeiten, die auch unter dem Namen *Chapman-Kolmogorow-Gleichungen* bekannt sind, überlassen wir dem Leser:

$$p_{xy}^{(k+r)} = \sum_z p_{xz}^{(k)} p_{zy}^{(r)} \qquad \text{für alle } k, r \in \mathbb{N} \text{ und } x, y \in \mathcal{X}. \qquad (1.1.3)$$

Die Übergangswahrscheinlichkeiten $p_{xy}^{(k)}$ sind also genau die Elemente der k-ten Potenz P^k der Übergangsmatrix P. Insbesondere haben wir damit ein Verfahren gefunden, um die Verteilung der Markov-Kette zum Zeitpunkt t in Abhängigkeit von einer gegebenen Startverteilung π_0 zu berechnen: Bezeichnen wir mit π_t die Verteilung von X_t so gilt offensichtlich

$$\pi_t = (P^t)^T \pi_0 \qquad \text{für alle } t \in \mathbb{N}.$$

Kehren wir zu unserer Ausgangsfrage zurück: Wann ist die Verteilung von X_t für "große" t von der Startverteilung π_0 unabhängig? Offenbar dann, wenn es eine Verteilung π gibt, so daß

$$\lim_{t\to\infty} (P^t)^T \pi_0 = \pi \qquad \text{für alle Startverteilungen } \pi_0. \tag{1.1.4}$$

Nehmen wir für einen Moment an, die Eigenschaft (1.1.4) wäre erfüllt. Was können wir dann über die Verteilung π sagen? Setzt man in (1.1.4) $\pi_0 = \pi$, so rechnet man leicht nach, daß gelten muß $P^T\pi = \pi$. (D.h., π ist ein Linkseigenvektor von P zum Eigenwert 1.) Eine Verteilung π, für die $P^T\pi = \pi$ gilt, nennt man ein *stationäre Verteilung* der Markov-Kette. Man beachte, daß eine Markov-Kette, die (1.1.4) erfüllt, nur eine eindeutig bestimmte stationäre Verteilung haben kann.

Ergodische Markov-Ketten

Ist π die stationäre Verteilung einer Markov-Kette und ist x ein Zustand mit $\pi_x > 0$, so kann man leicht zeigen, daß eine in x startende Markov-Kette mit Wahrscheinlichkeit Eins auch wieder nach x zurückkehren wird. Stationäre Verteilungen π mit $\pi > 0$ spielen daher in den Anwendungen eine besondere Rolle.

Markov-Ketten, die eine stationäre Verteilung π besitzen, für die $\pi > 0$ gilt und die darüber hinaus auch noch (1.1.4) erfüllt, nennt man *ergodisch*. Als nächstes werden wir notwendige Bedingungen dafür herleiten, daß eine Markov-Kette ergodisch ist.

Betrachten wir dazu noch einmal unsere Irrfahrt auf dem n-dimensionalen Würfel. Diesmal jedoch nehmen wir an, daß das Teilchen sich in jedem Schritt zu einem Nachbarpunkt bewegt. Was ändert sich? Offenbar ist die Summe der Koordinaten der Punkte auf denen sich das Teilchen befindet, abwechselnd gerade und ungerade. Die Parität des Aufenthaltsortes zum Zeitpunkt t, und damit auch die überhaupt in Frage kommenden Aufenthaltsorte, hängt in diesem Fall für alle $t \in \mathbb{N}$ vom Startpunkt ab. Diese Schwierigkeiten werden wir umgehen, in dem wir nur aperiodische Markov-Ketten betrachten.

Eine Markov-Kette heißt *aperiodisch*, falls gilt

$$\mathrm{ggT}\{t : p_{xy}^{(t)} > 0\} = 1 \qquad \text{für alle } x, y \in \mathcal{X}. \tag{1.1.5}$$

(Hier bezeichnet ggT(A) den größten gemeinsamen Teiler der Zahlen aus A und wir vereinbaren, daß ggT($\emptyset$) $= 1$ ist.) Eine hinreichende Bedingung für die Aperiodizität einer Markov-Kette ist offensichtlich, daß $p_{xx} > 0$ für alle $x \in \mathcal{X}$. Gilt sogar $p_{xx} \geq \frac{1}{2}$ für alle $x \in \mathcal{X}$, so heißt die Markov-Kette *stark aperiodisch.*

Durch einen einfachen Trick läßt sich jede Markov-Kette in eine stark aperiodische Markov-Kette transformieren. Dazu modifiziert man die Übergangswahrscheinlichkeiten so, daß die Kette mit einer gewissen Wahrscheinlichkeit, sagen wir 1/2, immer in dem Zustand bleibt, in dem sie sich gerade befindet.

Proposition 1.1.1. *Ist $\mathcal{MC}$ eine Markov-Kette mit Übergangsmatrix P, so bildet die Matrix $\frac{1}{2}(P+I)$ die Übergangsmatrix einer stark aperiodischen Markov-Kette $\mathcal{MC}'$.*

■

Neben der Aperiodizität werden wir als zusätzliche Eigenschaft von den von uns betrachteten Markov-Ketten fordern, daß jeder Zustand von jedem anderen aus erreicht werden kann. Es soll also gelten

$$\text{zu jedem Paar } x, y \in \mathcal{X}, x \neq y \text{ gibt es ein } t \in \mathbb{N} \text{ mit } p_{xy}^{(t)} > 0. \tag{1.1.6}$$

Eine Markov-Kette, die (1.1.6) erfüllt, heißt *irreduzibel.*

Bemerkung. Unsere Beschränkung auf irreduzible Ketten ist keine wesentliche Einschränkung, da sich beliebige Markov-Ketten in der Regel leicht in zwei oder mehrere irreduzible Ketten zerlegen lassen. (Als Veranschaulichung dieser Aussage betrachte man zum Beispiel eine Irrfahrt auf einem unzusammenhängenden Graphen.)

Es gilt nun die folgende Charakterisierung.

Satz 1.1.2. *Ist $\mathcal{MC}$ eine irreduzible Markov-Kette, so sind die folgenden Aussagen äquivalent:*

(i) *$\mathcal{MC}$ ist aperiodisch,*

(ii) *$\mathcal{MC}$ besitzt genau eine stationäre Verteilung, und*

(iii) *es gibt eine Verteilung $\pi > 0$, so daß $\mathcal{MC}$ die Bedingung* (1.1.4) *erfüllt.*

Der Beweis dieses Satzes ist zwar nicht schwierig, aber technisch. Da er uns zudem keine neuen Einsichten vermittelt, verweisen wir den interessierten Leser auf die einschlägige Literatur der Wahrscheinlichkeitstheorie. Ein Beweis findet sich beispielsweise in [**?**]1968.

Bemerkung. Ist eine Markov-Kette (mit endlichem Zustandsraum) zwar aperiodisch, aber nicht irreduzibel, so besitzt auch sie eine stationäre Verteilung π. Für diese gilt aber zum einen nicht notwendigerweise $\pi > 0$ und zum anderen kann es auch mehrere stationäre Verteilungen geben. (Man veranschauliche sich dies an Hand einer Irrfahrt mit mehreren absorbierenden Zuständen, d.h. Zuständen x mit $p_{xx} = 1$.)

Beispiele reversibler Markov-Ketten

Satz 1.1.2 liefert uns ein notwendiges und hinreichendes Kriterium dafür, daß eine irreduzible Markov-Kette eine stationäre Verteilung besitzt. Aber wie finden wir eine solche? Das folgende Lemma wird sich dabei als sehr nützlich erweisen.

Lemma 1.1.3. *Ist $\mathcal{MC}$ eine aperiodische, irreduzible Markov-Kette und π eine Verteilung mit*

$$p_{xy}\pi_x = p_{yx}\pi_y \qquad \text{für alle } x, y \in \mathcal{X}, \tag{1.1.7}$$

so ist π die (eindeutig bestimmte) stationäre Verteilung.

Beweis. Aus Satz 1.1.2 wissen wir, daß $\mathcal{MC}$ genau eine stationäre Verteilung hat. Daß diese durch π gegeben ist, rechnet man mit Hilfe von (1.1.7) leicht nach:

$$(P^T\pi)_x = \sum_y p_{yx}\pi_y = \sum_y p_{xy}\pi_x = \pi_x \sum_y p_{xy} = \pi_x.$$

∎

Erfüllt die stationäre Verteilung π einer irreduziblen Markov-Kette die Bedingung (1.1.7), so heißt die Markov-Kette (*zeitlich*) *reversibel.* (Als Begründung für diesen Namen mache man sich klar, daß die linke bzw. rechte Seite in (1.1.7) gerade die Wahrscheinlichkeit darstellt, mit der in einer sich im stationären Zustand befindlichen Markov-Kette im nächsten Schritt ein Übergang vom Zustand x in den Zustand y bzw. vom Zustand y in den Zustand x stattfindet.)

Wir beenden unsere kurze Einführung in die Theorie der Markov-Ketten mit fünf Beispielen, auf die wir auch in den folgenden Abschnitten wiederholt Bezug nehmen werden.

Beispiel 1: Irrfahrt auf dem n-dimensionalen Hyperwürfel. Für eine natürliche Zahl $n \in \mathbb{N}$ bezeichne $\mathcal{B}(n) = \{0,1\}^n$ die Menge der Worte der Länge n über dem Alphabet $\{0,1\}$. Mit $\mathcal{MC}_1(n)$ bezeichnen wir die Markov-Kette mit Zustandsraum $\mathcal{B}(n)$ und Übergängen wie folgt: in jedem Zustand $x = (x_1, \ldots, x_n)$ wähle man zunächst zufällig und gleichverteilt einen Index $i \in \{1, \ldots, n\}$ und vertausche dann mit Wahrscheinlichkeit $\frac{1}{2}$ den Wert des i-ten Buchstabens. Da in der üblichen Definition des n-dimensionalen Hyperwürfels zwei Knoten genau dann verbunden sind, wenn sich ihre Koordinaten in genau einer Stelle unterscheiden, entspricht die Markov-Kette $\mathcal{MC}_1(n)$ einer Irrfahrt auf dem n-dimensionalen Würfel mit den Übergangswahrscheinlichkeiten

$$p_{xy} = \begin{cases} \frac{1}{2n} & \text{falls } x \neq y \text{ und } x,\, y \text{ benachbart sind,} \\ 0 & \text{falls } x \neq y \text{ und } x,\, y \text{ nicht benachbart sind,} \\ \frac{1}{2} & \text{falls } x = y. \end{cases}$$

Die Markov-Kette $\mathcal{MC}_1(n)$ ist offensichtlich stark aperiodisch und irreduzibel. Zudem prüft man leicht nach, daß die Bedingung (1.1.7) bezüglich der Verteilung $\pi \equiv 1/2^n$ erfüllt ist. Die Gleichverteilung ist also die eindeutig bestimmte stationäre Verteilung der Irrfahrt.

Beispiel 2: Irrfahrt auf dem $n \times n$ Gitter. Als $n \times n$ Gitter bezeichnet man den Graphen G mit Knotenmenge $V = \{(x_1, x_2) \mid 1 \leq x_1, x_2 \leq n\}$ in dem zwei Knoten $x = (x_1, x_2)$ und $y = (y_1, y_2)$ genau dann durch eine Kante verbunden sind, falls gilt $|x_1 - y_1| + |x_2 - y_2| = 1$. Als Markov-Kette $\mathcal{MC}_2(n)$ bezeichnen wir die Irrfahrt auf dem $n \times n$ Gitter mit den Übergangswahrscheinlichkeiten

$$p_{xy} = \begin{cases} \frac{1}{8} & \text{falls } x \neq y \text{ und } x,\ y \text{ benachbart sind,} \\ 0 & \text{falls } x \neq y \text{ und } x,\ y \text{ nicht benachbart sind,} \\ 1 - \sum\limits_{z \neq x} p_{xz} & \text{falls } x = y. \end{cases}$$

Wieder sieht man mit Hilfe von Lemma 1.1.3 sofort ein, daß die Markov-Kette $\mathcal{MC}_2(n)$ ergodisch ist und die Gleichverteilung als stationäre Verteilung besitzt.

Beispiel 3: Bernoulli-Laplacesches Diffusionsmodell. Um den Austausch zwischen zwei nicht komprimierbaren Flüssigkeiten zu untersuchen, führte Daniel Bernoulli 1769 das folgende Urnenmodell ein. Ausgehend von zwei Urnen, die n rote beziehungsweise n schwarze Kugeln enthalten, werden in jedem Schritt aus jeder Urne eine Kugel ausgewählt und die beiden Kugeln ausgetauscht.

Wir verallgemeinern dieses Modell etwas und betrachten stattdessen eine Markov-Kette $\mathcal{MC}_3(n, k)$, auf den k-elementigen Teilmengen von $\{1, \ldots, n\}$. In jedem Schritt wählen wir zufällig ein Element der k-elementigen Teilmenge und eines aus dem Rest und tauschen beide aus. Bezeichnen wir mit $\triangle$ die symmetrische Differenz zweier Mengen, also $A \triangle B = (A \setminus B) \cup (B \setminus A)$, so sind die Übergangswahrscheinlichkeiten der Markov-Kette $\mathcal{MC}_3(n, k)$ daher gegeben durch

$$p_{AB} = \begin{cases} \frac{1}{k(n-k)} & \text{falls } |A \triangle B| = 2, \\ 0 & \text{sonst.} \end{cases}$$

Setzen wir weiter $\pi \equiv 1/\binom{n}{k}$, so folgt aus Lemma 1.1.3 wieder, daß in der stationären Verteilung alle Zustände gleichwahrscheinlich sind – was unserer Vorstellung vom Mischen zweier Flüssigkeiten entspricht.

Beispiel 4: Matchings in Graphen. Für einen beliebigen Graphen $G = (V, E)$ definieren wir eine Markov-Kette $\mathcal{MC}_4(G)$ wie folgt. Als Zustandsraum wählen wir die Menge $\mathcal{M}(G)$ aller Matchings von G. In jedem Zustand $M \in$

$\mathcal{M}(G)$ wählen wir eine Kante $e \in E$ zufällig und gleichwahrscheinlich und gehen mit Wahrscheinlichkeit $\frac{1}{2}$ zum Zustand $\tilde{M}$ über, der definiert ist durch

$$\tilde{M} = \begin{cases} M - e & \text{falls } e \in M, \\ M + e & \text{falls } e \notin M \text{ und } M + e \in \mathcal{M}(G), \\ M & \text{sonst.} \end{cases}$$

Man überzeugt sich leicht davon, daß die Übergangswahrscheinlichkeiten der Markov-Kette $\mathcal{MC}_4(G)$ symmetrisch sind, d.h. daß für alle $M, \tilde{M} \in \mathcal{M}(G)$ gilt $p_{M\tilde{M}} = p_{\tilde{M}M}$. Eine abermalige Anwendung von Lemma 1.1.3 zeigt daher, daß auch diese Markov-Kette ergodisch ist und alle Matchings in der stationären Verteilung gleichwahrscheinlich sind.

Beispiel 5: Spannende Bäume in Graphen. Auch unsere letzte Markov-Kette $\mathcal{MC}_5(G)$ ist für Graphen $G = (V, E)$ definiert, die jedoch zusammenhängend sein sollen. Als Zustandsraum wählen wir diesmal die Menge $\mathcal{S}(G)$ aller spannenden Bäume von G. In jedem Zustand $S \in \mathcal{S}(G)$ wählen wir wie oben zwei Kanten e_1 und e_2 zufällig und gleichwahrscheinlich und gehen zum Zustand T über, der definiert ist durch

$$T = \begin{cases} S - e_1 + e_2 & \text{falls } e_1 \in S,\ e_2 \notin S \text{ und } S - e_1 + e_2 \in \mathcal{S}(G), \\ S & \text{sonst.} \end{cases}$$

Wir überlassen es dem Leser, analog wie in der vorigen Beispielen zu zeigen, daß die Markov-Kette $\mathcal{MC}_5(G)$ ergodisch ist und die Gleichverteilung als stationäre Verteilung besitzt.

1.2 Konvergenz zur stationären Verteilung

Betrachten wir einen beliebigen Graphen $G = (V_n, E)$ auf n Knoten. Aus der Gesamtmenge der Matchings von G möchten wir gerne ein Matching zufällig auswählen. Mit anderen Worten, wir suchen einen uniformen Erzeuger für die Klasse der Matchings von G. Aus Praktikabilitätsgründen sollte dessen Laufzeit natürlich polynomiell in der Größe des Graphens, d.h. in n beschränkt sein.

Geben wir uns bereits mit einem randomisierten uniformen Erzeuger zufrieden – was in Anbetracht der Tatsache, daß das entsprechende Zählproblem $\#P$-vollständig ist (vgl. c:perm) sinnvoll erscheint – so suggeriert der Markov-Ketten Ansatz des letzten Abschnittes eine Lösung. Wir haben dort gesehen, daß die Markov-Kette $\mathcal{MC}_4(G)$ unabhängig von der Startkonfiguration zur einer stationären Verteilung konvergiert, in der alle Matchings gleichwahrscheinlich sind.

Wir brauchen also nur mit einem beliebigen Matching, beispielsweise dem leeren, zu starten, die Markov-Kette genügend lange zu simulieren – und schon haben wir ein Matching zufällig erzeugt. Hier liegt aber auch bereits der Haken: unsere bisherigen Kenntnisse über stationäre Verteilungen erlauben uns nur die Aussage, daß im Limes die stationäre Verteilung erreicht wird – wir suchen aber einen Erzeuger, der in polynomieller Zeit ein zufälliges Matching findet. Der Zustandsraum der Markov-Kette $\mathcal{MC}_4(G)$ besteht im allgemeinen aus exponentiell vielen Matchings (exponentiell in der Knotenanzahl n). Simulieren wir die Kette daher nur für eine polynomielle Anzahl von Schritten, durchläuft sie in dieser Zeit nur einen verschwindend kleinen Teil des Zustandsraumes. Können wir da erwarten, daß sie sich in dieser Zeit trotzdem bereits der stationären Verteilung angenähert hat? Wie wir sehen werden ist dies tatsächlich der Fall. Man nennt die Markov-Kette $\mathcal{MC}_4(G)$ daher auch "schnell mischend".

Schnell mischende Markov-Ketten

Um das Phänomen schnell mischender Markov-Ketten genauer beschreiben zu können, benötigen wir noch einige Definitionen. Jeder Markov-Kette $\mathcal{MC}$ ordnen wir wie folgt eine Familie von randomisierten Algorithmen zu: Für $t \in \mathbb{N}$ und $x \in \mathcal{X}$ bezeichnen wir mit $\mathcal{A}_{MC}(t,x)$, denjenigen Algorithmus, der ausgehend vom Zustand x genau t Schritte der Markov-Kette simuliert und den dann erreichten Zustand ausgibt.

Als ϵ-Konvergenzzeit der Markov-Kette bezeichnen wir diejenige Schrittanzahl, ab der sichergestellt ist, daß die Wahrscheinlichkeit mit der ein Zustand y ausgegeben wird um höchstens ϵ von der Wahrscheinlichkeit π_y von y im stationären Zustand abweicht – und das unabhängig vom Startzustand x. Formal ist die ϵ-Konvergenzzeit definiert durch

$$\tau(\epsilon) = \min\{t : (1-\epsilon)\pi_y \leq Prob(\mathcal{A}_{MC}(t,x) = y) \leq (1+\epsilon)\pi_y \text{ für alle } x, y \in \mathcal{X}\}.$$

Als Motivation für diese Definition betrachte man eine konvergente Markov-Kette, in der alle Zustände in der stationären Verteilung gleich wahrscheinlich sind. Die Algorithmen $\mathcal{A}_{MC}(t,x)$ sind für hinreichend großes t und geeignetes x also (randomisierte) ϵ-uniforme Erzeuger für die Elemente des Zustandsraumes. Es gilt nun

$$\tau(\epsilon) = \min\{t : \text{für alle } x \in \mathcal{X} \text{ ist } \mathcal{A}_{MC}(t,x) \text{ ein } \epsilon\text{-uniformer Erzeuger}\}.$$

Bemerkung. Auch wenn $\tau(\epsilon)$ sehr groß ist kann es einzelne Zustände x geben, von denen aus die stationäre Verteilung wesentlich schneller erreicht wird. Da jedoch im allgemeinen nicht klar ist, wie man diese Zustände findet, interessieren wir uns hier nur für Schranken, die für jeden Startzustand eine entsprechend schnelle Konvergenz garantieren.

Kommen wir zurück zu unserer Frage, wann ist eine Markov-Kette "schnell mischend"? Intuitiv sicherlich dann, wenn $\tau(\epsilon)$ für alle $\epsilon > 0$ "klein" ist. Aber klein im Verhältnis wozu? Eine Möglichkeit wäre, eine Beziehung zur Größe des Zustandsraumes zu fordern. Da wir aber, wie beispielsweise auch in den Beispielen des letzten Abschnittes, immer Familien von Markov-Ketten betrachten werden, hat es sich als vorteilhafter erwiesen "schnell mischend" als Eigenschaft von solchen Familien zu definieren: Ist $\mathcal{MC}(I)$ eine Familie von Markov-Ketten, indiziert durch Probleminstanzen I, so nennen wir $\mathcal{MC}(I)$ *schnell mischend*, falls sich die Konvergenzzeit $\tau(\epsilon)$ jeder Markov-Kette durch ein Polynom in der Eingabelänge $|I|$ und $\ln \epsilon^{-1}$ beschränken läßt.

Wir veranschaulichen diese Definition wieder an Hand von konvergenten Markov-Ketten mit der Gleichverteilung als stationärer Verteilung. Sei also $\mathcal{MC}(I)$ eine Familie solcher Markov-Ketten. Betrachten wir zunächst die ϵ-Konvergenzzeit der Markov-Ketten für ein festes ϵ. Gibt es ein Polynom $p(n)$, so daß $\tau(\epsilon) \leq p(|I|)$ für alle Instanzen I gilt, so bilden die entsprechenden $\mathcal{A}_{MC}$-Algorithmen randomisierte ϵ-uniforme Erzeuger. Um daraus ein vollständig polynomielles Schema von randomisierten uniformen Erzeugern (ein FPRUG), zu gewinnen benötigt man genau die in der Definition von schnell mischenden Markov-Ketten geforderte zusätzliche Polynomialität der Konvergenzzeit in $\ln \epsilon^{-1}$.

Der Leitwert einer Markov-Kette

Wie kann man einer Familie von Markov-Ketten ansehen, ob sie schnell mischend ist? Bevor wir diese Frage beantworten können, müssen wir zunächst geeignete Hilfsmittel entwickeln, um die ϵ-Konvergenzzeit $\tau(\epsilon)$ einer Markov-Kette abschätzen zu können.

Vergleichen wir dazu zunächst noch einmal die beiden Irrfahrten auf dem n-dimensionalen Hyperwürfel (Beispiel 1) und auf dem $n \times n$ Gitter (Beispiel 2) des letzten Abschnitts. Um unsere Überlegungen zu vereinfachen sei im folgenden ϵ immer eine zwar beliebige aber fest vorgegebene Konstante, die insbesondere also nicht von n abhängen soll. Der Hyperwürfel hat 2^n verschiedene Zustände, trotzdem wird, wie wir gleich zeigen werden, die Markov-Kette $\mathcal{MC}_1(n)$ bereits nach $O(n)$ vielen Schritten sehr nahe an der stationären Verteilung sein. Für die Markov-Kette $\mathcal{MC}_1(n)$ gilt also $\tau(\epsilon) = O(n)$.

Die Markov-Kette $\mathcal{MC}_2(n)$ der Irrfahrt in dem $n \times n$ Gitter andererseits hat nur n^2 viele Zustande. Können wir hier also erwarten, daß $\tau(\epsilon) = O(\log n)$ gilt? Sicherlich nicht: um von der linken unteren Ecke in die obere rechte Ecke zu wandern, braucht man mindestens $2(n-1)$ Schritte. Für alle hinreichend kleinen $\epsilon > 0$ muß also gelten $\tau(\epsilon) > 2(n-1)$. (Wir werden später zeigen, daß auch für die Markov-Kette $\mathcal{MC}_2(n)$ gilt $\tau(\epsilon) = O(n)$.)

Woran läßt sich der Unterschied dieser beiden Irrfahrten festmachen? Zum einen natürlich – so haben wir eben argumentiert – am Durchmesser der beiden Graphen. Im Hyperwürfel ist der Durchmesser $O(\log|\mathcal{X}|)$, im Gitter hingegen $O(|\mathcal{X}|)$. Zum anderen, etwas allgemeiner, gilt für jede Menge $S \subseteq \mathcal{X}$ mit $|S| \leq |\mathcal{X}|/2$ im Hyperwürfel, der sich zum Zeitpunkt t im stationären Zustand befindet (vgl. Aufgabe 1.3)

$$Prob(X_{t+1} \notin S \mid X_t \in S) \geq \frac{1}{4}, \tag{1.2.1}$$

während es im Gitter Mengen S gibt, für die die entsprechende Wahrscheinlichkeit $o(1)$ ist (beispielsweise bilden die Punkte der untere Hälfte des Gitters eine solche Menge S).

Abbildung 1.1 veranschaulicht diesen Effekt: läßt sich der Zustandsraum $\mathcal{X}$ so in zwei Teile S und $\bar{S} = \mathcal{X} \setminus S$ partitionieren, daß ein Übergang von einem Teil zum anderen recht unwahrscheinlich ist, kann man nicht erwarten, daß die Kette schnell mischend ist. Genauer kann man sogar vermuten, daß die "engste" Stelle (minimiert über alle Mengen S) umgekehrt proportional zur Konvergenzgeschwindigkeit ist.

Abbildung 1.1: Partition des Zustandsraums

[?]1989 gelang es nun mit diesem Ansatz, die Größenordung der ϵ-Konvergenzzeit für eine große Klasse von Markov-Ketten zu charakterisieren. Wir ordnen dazu jeder ergodischen Markov-Kette $\mathcal{MC}$ mit stationärer Verteilung π einen gerichteten Graphen $\mathcal{G}_{MC} = (V, A)$ mit Knotenmenge $V = \mathcal{X}$, Kantenmenge

$$A = \{(x, y) \mid x, y \in \mathcal{X}, x \neq y \text{ und } p_{xy} > 0\}$$

zu und definieren eine Gewichtsfunktion c auf $\mathcal{X} \times \mathcal{X}$ durch

$$c(x, y) = \begin{cases} \pi_x p_{xy} & \text{falls } (x, y) \in A, \\ 0 & \text{sonst.} \end{cases}$$

Für eine Menge $S \subseteq V$ von Zuständen schreiben wir abkürzend $\pi(S)$ für die Summe $\sum_{x \in S} \pi(x)$. Entsprechend bezeichnet $c(S, \bar{S})$ die Summe $\sum_{x \in S} \sum_{y \notin S} c(x, y)$.

20.6.2001

Der *Leitwert* der Markov-Kette $\mathcal{MC}$ ist dann definiert durch

$$\Phi_{MC} = \min_{\substack{S\subset\mathcal{X} \\ 0<\pi(S)\leq\frac{1}{2}}} \frac{c(S,\bar{S})}{\pi(S)}.$$

Die inhaltliche Bedeutung des Leitwertes einer Markov-Kette veranschaulichen wir uns an Hand der folgenden Eigenschaft, deren einfachen Beweis wir dem Leser überlassen (Aufgabe 1.4).

> Befindet sich die Markov-Kette zum Zeitpunkt t im stationären Zustand, so gilt $\Phi_{MC} \leq \min_{\substack{S\subseteq\mathcal{X} \\ 0<\pi(S)\leq\frac{1}{2}}} Prob(X_{t+1} \notin S | X_t \in S).$ (1.2.2)

Bezeichnen wir mit $\pi_{\min} = \min_{x\in\mathcal{X}} \pi_x$ die minimale Wahrscheinlichkeit, mit der ein Zustand in der stationärer Verteilung π auftritt, so besagt der Satz von Sinclair und Jerrum, daß sich die ϵ-Konvergenzzeit ergodischer, reversibler Markov-Ketten durch ein Polynom in Φ_{MC}^{-1}, in $\ln\epsilon^{-1}$ und in $\ln\pi_{min}^{-1}$ beschränken läßt.

Satz 1.2.1. *Ist $\mathcal{MC}$ eine stark aperiodische, ergodische und reversible Markov-Kette, so gilt für die ϵ-Konvergenzzeit*

$$\tau(\epsilon) \leq \frac{2}{\Phi_{MC}} \left(\ln\epsilon^{-1} + \ln\pi_{min}^{-1}\right).$$

Ausgehend von Satz 1.2.1 ist es nicht schwer, eine notwendige Bedingung dafür anzugeben, daß eine Familie von Markov-Ketten schnell-mischend ist. Dafür benötigen wir allerdings noch eine Definition. Ist $\mathcal{MC}(I)$ eine Familie von Markov-Ketten und α_{MC} ein Parameter, der für jede Kette der Familie definiert ist, so heißt α_{MC} *polynomiell beschränkt*, falls es ein Polynom $p(n)$ gibt, so daß $\alpha_{MC} \leq p(|I|)$ für jedes Element der Familie gilt.

Nach Satz 1.2.1 ist daher eine Familie von Markov-Ketten sicherlich dann schnell mischend, wenn neben $\ln\pi_{min}^{-1}$ auch das Reziproke des Leitwerts polynomiell beschränkt ist. Überraschenderweise kann man auch zeigen, daß diese Bedingung sogar notwendig ist.

Satz 1.2.2. *Ist $\mathcal{MC}(I)$ eine Familie von stark aperiodischen, ergodischen und reversiblen Markov-Ketten für die $\ln\pi_{\min}^{-1}$ polynomiell beschränkt ist. Dann ist $\mathcal{MC}(I)$ genau dann schnell mischend, wenn Φ_{MC}^{-1} polynomiell beschränkt ist.*

Wir verschieben die etwas aufwendigeren Beweise von Satz 1.2.1 und 1.2.2 auf den Abschnitt 1.4 und beschränken uns vorläufig darauf, mit ihrer Hilfe die schnelle Konvergenz verschiedener Markov-Ketten nachzuweisen.

20.6.2001

Mehrgüterflüsse in Markov-Ketten

Während Satz 1.2.2 eine mathematisch attraktive Charakterisierung schnell mischender Markov-Ketten ist, stellt sich bei seiner Anwendung die Frage, wie sich der Leitwert einer Markov-Kette berechnen läßt. Als ein sehr wertvolles Hilfsmittel hat sich hierbei die Einführung von Mehrgüter-Flußsystemen erwiesen.

Auf dem gewichteten Graphen $\mathcal{G}_{MC}$ definieren wir folgendes Mehrgüter Flußproblem: Für jedes (geordnete) Paar (x, y) von Zuständen soll von dem Gut g_{xy} die Menge $\pi_x \pi_y$ von x nach y transportiert werden. Mit anderen Worten, wir suchen Flüsse $f_{xy} : A \to \mathbb{R}_{\geq 0}$, so daß

$$\sum_u f_{xy}(u, v) = \sum_u f_{xy}(v, u) \qquad \text{für alle } v \neq x, y$$

und

$$\sum_u f_{xy}(x, u) = \sum_u f_{xy}(u, y) = \pi_x \pi_y.$$

Zu gegebenen Flüssen f_{xy} definieren wir den Gesamtfluß f, meist nur kurz *Fluß* genannt, durch

$$f(e) = \sum_{x \neq y} f_{xy}(e).$$

Für einen Fluß f bezeichnen wir mit

$$\rho(f) = \max_{e \in A} \frac{f(e)}{c(e)}$$

die relative Kantenauslastung.

Intuitiv sollte klar sein, daß die Kantenauslastung $\rho(f)$ nur dann klein sein kann, wenn der Leitwert des Graphen groß ist. (Der Leser betrachte noch einmal die Abbildung 1.1.) Das nächsten Lemma formalisiert diesen Zusammenhang.

Lemma 1.2.3. *Für jede Markov-Kette $\mathcal{MC}$ mit Fluß f gilt*

$$\Phi_{MC} \geq \frac{1}{2\rho(f)}.$$

Beweis. Es sei $S_0 \subseteq \mathcal{X}$ eine Menge von Zuständen mit $0 < \pi(S_0) \leq \frac{1}{2}$ und

$$\Phi_{MC} = \frac{c(S_0, \bar{S}_0)}{\pi(S_0)}.$$

Da nach Definition $\rho(f) \geq \frac{f(e)}{c(e)}$ für alle Kanten e gilt, folgt daraus

$$c(S_0, \bar{S}_0) \;=\; \sum_{\substack{x \in S_0 \\ y \notin S_0}} c(x, y) \geq \frac{1}{\rho(f)} \sum_{\substack{x \in S_0 \\ y \notin S_0}} f(x, y) \geq \frac{1}{\rho(f)} \sum_{\substack{x \in S_0 \\ y \notin S_0}} \sum_{u,v} f_{uv}(x, y)$$

$$\geq \frac{1}{\rho(f)} \sum_{\substack{u \in S_0 \\ v \notin S_0}} \underbrace{\sum_{\substack{x \in S_0 \\ y \notin S_0}} f_{uv}(x,y)}_{\geq \pi_u \pi_v} \geq \frac{1}{\rho(f)} \pi(S_0)\pi(\bar{S}_0) \geq \frac{1}{2\rho(f)} \pi(S_0).$$

■

Zusammen mit Satz 1.2.1 folgt damit sofort, daß eine Familie $\mathcal{MC}(I)$ von Markov-Ketten sicherlich auch dann schnell mischend ist, wenn statt dem Reziproken des Leitwertes die Kantenauslastung $\rho(f)$ geeignet gewählter Flüsse f polynomiell beschränkt ist. Wie wir in Abschnitt 1.4 sehen werden, ist auch diese Bedingung bereits hinreichend.

Satz 1.2.4. *Ist $\mathcal{MC}(I)$ eine Familie von stark aperiodischen, ergodischen und reversiblen Markov-Ketten für die $\ln \pi_{\min}^{-1}$ polynomiell beschränkt ist, dann ist $\mathcal{MC}(I)$ genau dann schnell mischend, wenn es Flüsse f gibt, so daß $\rho(f)$ polynomiell beschränkt ist.*

Beispiele schnell mischender Markov-Ketten

In Abschnitt 1.1 haben wir fünf Beispiele von ergodischen, reversiblen Markov-Ketten, genauer von Familien solcher Ketten, kennengelernt. Wir haben dort bereits gezeigt, daß die stationäre Verteilung dieser Markov-Ketten jeweils die Gleichverteilung ist. Wie wir jetzt zeigen werden, sind diese Familien auch schnell mischend.

Beispiel 1: Irrfahrt auf dem n-dimensionalen Hyperwürfel. Wegen $\pi \equiv 1/2^n$ ist $\ln \pi_{\min}^{-1} = n$ polynomiell beschränkt. Nach Satz 1.2.4 genügt es daher zu zeigen, daß es Flüsse f gibt, so daß $\rho(f)$ polynomiell beschränkt ist. Diese definieren wir wie folgt. Für zwei Zustände $x \neq y$ verteilen wir den zu verschickenden Fluß von $\pi_x \pi_y$ gleichmäßig auf alle kürzesten Wege, die x und y verbinden. (Haben x und y den Abstand d, gibt es davon $d!$ viele.)

Wie groß ist nun $\rho(f)$? Ad hoc erscheint die Berechnung von $f(e)$ recht schwierig. Ein einfacher Trick hilft uns jedoch weiter: aus Symmetriegründen muß der Fluß auf jeder Kante gleich groß sein. Es genügt daher den "Gesamtfluß" $\sum_{e \in A} f(e)$ zu berechnen und durch die Anzahl Kanten des Graphens zu teilen. Der Gesamtfluß wiederum läßt sich einfach berechnen: es gibt $2^n \cdot \binom{n}{d}$ Paare (x, y) mit Abstand d und die für solche Paare in der Definition von f_{xy} verwendeten Wege haben alle die Länge d. Es gilt daher

$$\sum_{e \in A} f_{xy}(e) = d \cdot \pi_x \pi_y.$$

Für den Fluß $f(e)$ gilt daher

$$\begin{aligned} f(e) &= \frac{1}{|A|}\sum_{d=1}^{n} 2^n \binom{n}{d} \cdot d \cdot \pi_x \pi_y \\ &= \frac{1}{n2^n}\sum_{d=1}^{n} 2^n \binom{n-1}{d-1} \cdot \frac{n}{2^{2n}} \\ &= \frac{1}{2^{2n}}\sum_{d=0}^{n-1} \binom{n-1}{d} \\ &= \frac{2^{n-1}}{2^{2n}} \end{aligned}$$

und somit wegen $c(e) \equiv (2n2^n)^{-1}$ also $\rho(f) = n$.

Beispiel 2: Irrfahrt auf dem $n \times n$ Gitter. Wieder überzeugt man sich leicht, daß es wegen Satz 1.2.4 genügt, Flüsse f anzugeben, für die $\rho(f)$ polynomiell beschränkt ist. Diese konstruieren und analysieren wir wie folgt. Den Fluß zwischen zwei Punkten verteilen wir gleichmäßig auf alle kürzesten Wege mit höchstens einem Knick (je nach Lage der Punkte also auf ein oder zwei Wege). Betrachten wir eine Kante e, so gibt es höchstens n^3 Paare, deren Fluß diese Kante verwendet. Zusammen mit $c(e) \equiv (8n^2)^{-1}$ gilt daher

$$\rho(f) = \max_e \frac{f(e)}{c(e)} \leq \frac{n^3 \cdot \frac{1}{n^4}}{\frac{1}{8n^2}} = 8n.$$

Beispiel 3: Bernoulli-Laplacesches Diffusionsmodell. Auch hier führt das Schema aus Beispiel 1 zum Ziel. Den Fluß verteilen wir wieder gleichmäßig auf alle kürzesten Wege. Für zwei Mengen A und B mit $|A \triangle B| = 2d$ haben diese die Länge d. Es gilt also für alle Kanten $e = (A, B)$

$$\begin{aligned} f(e) &= \frac{1}{|E(\mathcal{G}_{MC})|}\sum_{d=1}^{k} \binom{n}{k}\binom{k}{d}\binom{n-k}{d} \cdot \frac{d}{\pi_A \pi_B} \\ &= \frac{1}{\binom{n}{k} k(n-k)}\sum_{d=1}^{k} \binom{n}{k} \cdot k \cdot \binom{k-1}{d-1}\binom{n-k}{d}\binom{n}{k}^{-2} \\ &= \frac{1}{(n-k)\binom{n}{k}^2}\sum_{d=0}^{k-1} \binom{k-1}{d}\binom{n-k}{n-k-d-1} \\ &= \frac{\binom{n-1}{k}}{(n-k)\binom{n}{k}^2} \end{aligned}$$

und damit wegen $c(e) \equiv [\binom{n}{k}k(n-k)]^{-1}$ für die Kantenauslastung $\rho(f)$:

$$\rho(f) = \frac{f(e)}{c(e)} = \frac{k \cdot \binom{n-1}{k}}{\binom{n}{k}} = \frac{k(n-k)}{n} \leq k.$$

Die Familien $\mathcal{MC}_4(G)$ und $\mathcal{MC}_5(G)$ sind ebenfalls schnell mischend, allerdings ist der Nachweis hier etwas aufwendiger. Für $\mathcal{MC}_4(G)$ werden wir ihn im nächsten Abschnitt führen, für $\mathcal{MC}_5(G)$ verweisen wir den interessierten Leser auf die Originalliteratur ([**?**]1989 und [**?**, **?**]1992).

1.3 Approximation der Permanenten

Die *Permanente* einer $n \times n$ Matrix $A = (a_{ij})$ ist definiert durch

$$\mathrm{per}(A) = \sum_{\sigma} \prod_{i=1}^{n} a_{i\sigma(i)},$$

wobei die Summe über alle Permutationen σ von $\{1,\ldots,n\}$ läuft. Im Gegensatz zu der ähnlich definierten Determinaten einer Matrix kennt man für die Permanente kein effizientes Berechnungsschema: die Komplexität aller bekannten Algorithmen wächst exponentiell mit n.

1979 gelang es Valiant die inherente Schwierigkeit der Berechnung der Permanenten daran festzumachen, daß er zeigte, daß bereits die Berechnung der Permanenten von 0-1 Matritzen $\#P$-vollständig ist. Dies ist umso überraschender, als die Permanente einer 0-1 Matrix genau der Anzahl perfekter Matchings in einem bipartiten Graphen entspricht: Wie wir wissen, ist hierfür das entsprechende Entscheidungsproblem sogar für beliebige Graphen in polynomieller Zeit lösbar.

In diesem Abschnitt zeigen wir, für verschiedene Klassen von Markov-Ketten, daß sie schnell mischend sind. Anschließend folgern wir daraus, unter anderem, die Existenz vollständig polynomieller Approximationsschemata zum einen für die Anzahl aller Matchings eines Graphens und zum anderen für die Anzahl perfekter Matchings in hinreichend dichten Graphen.

Schnell mischende Markov-Ketten

Für einen Graphen $G = (V_n, E)$ auf n Knoten und natürlichen Zahlen $1 \leq k_1 \leq k_2 \leq \frac{n}{2}$ bezeichnen wir mit $\mathcal{M}_k(G)$ die Menge aller Matchings in G mit Kardinalität k und mit $\mathcal{M}(G; k_1, k_2)$ die Menge aller Matchings in G mit Kardinalität mindestens k_1 und höchstens k_2, also

$$\mathcal{M}(G; k_1, k_2) = \bigcup_{k=k_1}^{k_2} \mathcal{M}_k(G).$$

Setzen wir weiter $\mathcal{M}(G) = \mathcal{M}(G; 1, \lfloor n/2 \rfloor)$, so enthält $\mathcal{M}(G)$ also alle Matchings von G.

In Abschnitt 1.1 haben wir bereits eine Markov-Kette mit Zustandsraum $\mathcal{M}(G)$ kennengelernt. Diese verallgemeinern wir jetzt zu Markov-Ketten $\mathcal{MC}(G; k_1, k_2)$ für alle natürlichen Zahlen $1 \leq k_1 < k_2 \leq \frac{n}{2}$. Der Zustandsraum der Markov-Kette $\mathcal{MC}(G; k_1, k_2)$ ist die Menge $\mathcal{M}(G; k_1, k_2)$. In jedem Zustand $M \in \mathcal{M}(G; k_1, k_2)$ wählen wir eine Kante $e \in E$ zufällig und gleichwahrscheinlich und gehen mit Wahrscheinlichkeit $\frac{1}{2}$ zum Zustand $\tilde{M}$ über, der definiert ist durch

$$\tilde{M} = \begin{cases} M - e & \text{falls } e \in M \text{ und } M - e \in \mathcal{M}(G; k_1, k_2), \\ M + e & \text{falls } e \notin M \text{ und } M + e \in \mathcal{M}(G; k_1, k_2), \\ M - f + e & \text{falls } e \notin M, f \in M,\ |e \cap f| = 1 \text{ und } M - f + e \in \mathcal{M}(G), \\ M & \text{sonst.} \end{cases}$$

Man überzeugt sich wieder leicht davon, daß die Übergangswahrscheinlichkeiten der Markov-Kette $\mathcal{MC}(G; k_1, k_2)$ symmetrisch sind. Da man ebenfalls leicht einsieht, daß $\mathcal{MC}(G; k_1, k_2)$ stark aperiodisch und irreduzibel ist, wissen wir daher insbesondere aus Lemma 1.1.3, daß in der stationären Verteilung alle Matchings gleichwahrscheinlich sind.

Satz 1.3.1. *Es sei $q(n)$ ein beliebiges Polynom und seien $k_1 = k_1(n)$ und $k_2 = k_2(n)$ beliebige ganzzahlige Funktionen mit $0 \leq k_1 < k_2 \leq \frac{n}{2}$. Dann gilt: Für alle Graphen $G = (V_n, E)$ auf n Knoten mit $\mathcal{M}_{k_1}(G)/\mathcal{M}_{k_1+1}(G) \leq q(n)$ ist die Markov-Kette $\mathcal{MC}(G; k_1, k_2)$ schnell mischend.*

Für den Beweis von Satz 1.3.1 stellen wir noch zwei Propositionen bereit, deren einfache Beweise wir dem Leser überlasssen (Aufgabe 1.5).

Proposition 1.3.2. *Im q-dimensionalen Hyperwürfel sind 2 Punkte mit maximaler Distanz durch genau $q!$ paarweise verschiedene kürzeste Pfade verbunden. Verbindet man alle 2^q (geordneten) Paare solcher Punkte jeweils durch die $q!$ verschiedenen Pfade, so laufen über jede Kante genau $q!$ Pfade.*

■

Proposition 1.3.3. *Für zwei natürliche Zahlen $k_1 \leq k_2$ bezeichne $G(k_1, k_2)$ den Graphen mit Knotenmenge $V = \{A \subseteq \{1, \ldots, k_1 + k_2\} \mid k_1 \leq |A| \leq k_2\}$ in dem* *zwei Mengen A und B genau dann durch eine gerichtete Kante verbunden sind, wenn $|A \triangle B| = 2$ und $|A| = |B|$ oder wenn $|A \triangle B| = 1$ gilt. Kanten zwischen zwei Mengen gleicher Kardinalität nennen wir Austauschkanten, die übrigen Kanten erhöhende Kanten. Ist $A \in V$ eine Menge der Kardinalität k_1, so sind A und $\bar{A}$ genau durch $k_1! \cdot k_2!$ paarweise verschiedene kürzeste Pfade verbunden, in denen keine Austauchkante auf eine erhöhende Kante folgt. Verbindet man alle $\binom{k_1+k_2}{k_1}$ solcher Paare $(A, \bar{A})$ jeweils durch die $k_1! \cdot k_2!$ verschiedenen solcher kürzesten Pfade, so laufen über jede erhöhende Kante höchstens $k_1!(k_2-1)!$ Pfade und über jede Austauschkante genau $k_1!(k_2-1)!$ Pfade.*

■

Beweis von Satz 1.3.1. Wir definieren einen Fluß f für dessen Kantenauslastung $\rho(f)$ gilt:

$$\rho(f) \leq 2|E| \cdot \left(2 + \left\lceil \frac{\mathcal{M}_{k_1}(G)}{\mathcal{M}_{k_1+1}(G)} \right\rceil \right)^2 \tag{1.3.1}$$

Zusammen mit Satz 1.2.4 und der Annahme, daß $\mathcal{M}_{k_1}(G)/\mathcal{M}_{k_1+1}(G)$ polynomiell beschränkt ist, folgt daraus dann sofort die Aussage des Satzes.

Statt direkt Flüsse f_{AB} für alle Paare (A, B) von Matchings aus $\mathcal{M}(G; k_1, k_2)$ zu konstruieren, zeigen wir zunächst, daß es ausreicht, Flüsse f_{AB} für eine geeignet gewählte Teilmenge aller Paare von Matchings anzugeben. Dazu setzen wir

$$\mathcal{F} = \{(A, B) \mid A, B \in \mathcal{M}(G; k_1, k_2), |A| \leq |B| \text{ und } k_1 < |B|\}.$$

Wir werden zeigen: für die Paare $(A, B) \in \mathcal{F}$ gibt es Flüsse f_{AB}, so daß gilt

$$\max_{e \in E(\mathcal{G}_{MC})} \sum_{(A,B) \in \mathcal{F}} f_{AB}(e) \leq |\mathcal{M}(G; k_1, k_2)|^{-1}. \tag{1.3.2}$$

Da der zwischen zwei Matchings A und B zu verschickende Fluß für alle Paare den gleichen Betrag hat, nämlich $|f| = |\mathcal{M}(G; k_1, k_2)|^{-2}$, läßt sich aus der Gültigkeit von (1.3.2) leicht auch die von (1.3.1) folgern. Um dies einzusehen, wählen wir als Fluß f_{AB} für Paare (A, B) mit $|A| > |B|$ und $|A| > k_1$ den Fluß f_{BA} in umgekehrter Richtung. Für die Konstruktion der Flüsse f_{AB} für Paare (A, B) mit $|A| = |B| = k_1$ fixieren wir zunächst eine beliebige Abbildung σ der Matchings der Kardinalität k_1 auf die Matchings der Kardinalität k_1+1 mit der Eigenschaft,

$$|\sigma^{-1}(M)| \leq \left\lceil \frac{\mathcal{M}_{k_1}(G)}{\mathcal{M}_{k_1+1}(G)} \right\rceil \qquad \text{für alle } M \in \mathcal{M}_{k_1+1}(G).$$

Als Fluß f_{AB} zwischen zwei Matchings A und B mit $|A| = |B| = k_1$ wählen wir dann die Summe der drei Flüsse $f_{A\sigma(B)}$, $f_{\sigma(B)\sigma(A)}$ und $f_{\sigma(A)B}$. Man sieht leicht ein, daß dann für die so gewählten Flüsse gilt

$$\begin{aligned}\max_{e\in E(\mathcal{G}_{MC})} f(e) &\le \left(2+2\left\lceil\frac{\mathcal{M}_{k_1}(G)}{\mathcal{M}_{k_1+1}(G)}\right\rceil+\left\lceil\frac{\mathcal{M}_{k_1}(G)}{\mathcal{M}_{k_1+1}(G)}\right\rceil^2\right)\cdot\max_{e\in E(\mathcal{G}_{MC})}\sum_{(A,B)\in\mathcal{F}} f_{AB}(e)\\ &\le \left(2+\left\lceil\frac{\mathcal{M}_{k_1}(G)}{\mathcal{M}_{k_1+1}(G)}\right\rceil\right)^2\cdot\max_{e\in E(\mathcal{G}_{MC})}\sum_{(A,B)\in\mathcal{F}} f_{AB}(e)\end{aligned}$$

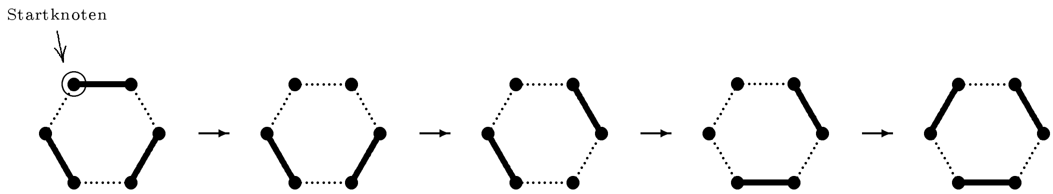
und somit also wegen (1.3.2) und $c(e)\equiv(|\mathcal{M}(G;k_1,k_2)|\cdot 2|E|)^{-1}$ für alle Kanten aus $\mathcal{G}_{MC}$) insbesondere auch (1.3.1).

Für jedes Paar $(A,B)\in\mathcal{F}$ ist $A\triangle B$ die Vereinigung von geraden Kreisen und von geraden und ungeraden Pfaden. Da Kanten in $\mathcal{G}_{MC}(G)$ genau den drei Operation a) Entfernen einer Matchingkante, b) Hinzufügen einer Matchingkante und c) Austausch zweier benachbarter Kanten entsprechen, lassen sich im Fall, daß $A\triangle B$ nur aus genau einem Kreis oder Pfad besteht, alle kürzesten Pfade von A nach B in $\mathcal{G}_{MC}(G)$ leicht bestimmen: Ausgehend von A tauscht man Matchingkanten entlang des Kreises bzw. Pfades aus, wie in Abbildung 1.2 angedeutet. Während es bei geraden Pfaden nur einen möglichen Startknoten für diese Austauschoperation gibt (den in A gematchten Endkonten des Pfades), kommen bei geraden Kreisen alle Knoten als Startknoten in Betracht, bei ungeraden Pfaden beide Endknoten. Wir verabreden daher, daß wir in diesen Fällen jeweils den kleinsten (bezüglich einer beliebigen, aber fest vorgegbenen Ordnung der Knotenmenge) der in Frage kommenden Knoten als Startknoten wählen. Den gesamten Betrag $|f|$ des Flusses f_{AB} schicken wir nun entlang des so ausgewählten Pfades in $\mathcal{G}_{MC}(G)$ von A nach B.

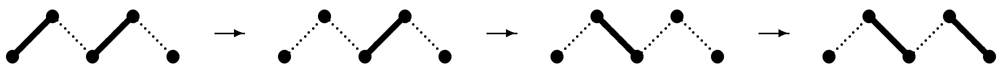
Besteht andererseits $A\triangle B$ aus mehr als nur einer Komponente, so konstruieren wir den Fluß f_{AB} wie folgt. In einem ersten Schritt vertauschen wir entlang ungerader Pfade und zwar, solange wie möglich, abwechselnd entlang ungerader Pfade, die mehr bzw. weniger Kanten aus A enthalten, wobei wir, falls möglich, mit einem Pfad, der weniger Kanten aus A enthält beginnen. Den Fluß verteilen wir hierbei gleichmäßig auf alle zulässigen Reihenfolgen. In einem zweiten Schritt vertauschen wir dann entlang gerader Kreise und gerader Pfade. Den Fluß verteilen wir dabei wieder gleichmäßig auf alle möglichen Reihenfolgen.

Wie groß ist nun der Gesamtfluß $f(e)$ durch eine Kante $e\in\mathcal{G}_{MC}(G)$? Um diesen zu bestimmen, partitionieren wir die Menge $\mathcal{F}$ in Äquivalenzklassen und berechnen den Fluß separat für jede Klasse. Wir setzen $(A,B)\approx(A',B')$ genau dann, wenn $A\cap B=A'\cap B'$, $A\triangle B=A'\triangle B'$ und entweder B auf den ungeraden Pfaden von $A\triangle B$ mit B' übereinstimmt, oder A auf den geraden Kreisen von $A\triangle B$ mit A' übereinstimmt.

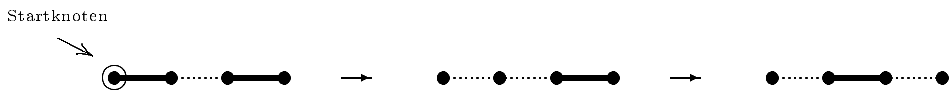
Für ein gegebenes Paar (A,B) aus $\mathcal{F}$ sei $A\triangle B$ eine Vereinigung von g_1 geraden Kreisen, g_2 geraden Pfaden, u_A ungeraden Pfaden mit mehr Kanten aus



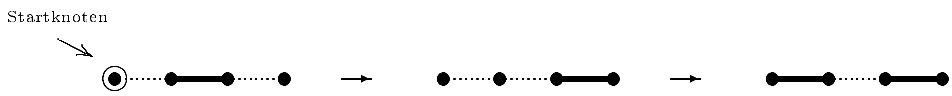
a) Austausch entlang gerader Kreise



b) Austausch entlang gerader Pfade



c) Austausch entlang ungerader Pfade, die mehr Kanten aus A enthalten



d) Austausch entlang ungerader Pfade, die weniger Kanten aus A enthalten

Abbildung 1.2: Austausch von Matchings entlang von Kreisen und Pfaden

A und u_B ungeraden Pfaden mit mehr Kanten aus B. Benutzt der Fluß f_{AB} die Kante $e = (M, M')$ aus $\mathcal{G}_{MC}(G)$, so gilt für $C = M \triangle M'$: $C \subseteq A \triangle B$. Wir unterscheiden zwei Fälle:

1.) C ist in einem geraden Kreis oder einem geraden Pfad aus $A \triangle B$ enthalten. Dann muß B auf den ungeraden Pfaden mit M übereinstimmen. In der Äquivalenzklasse von (A, B) gibt es genau $2^{g_1+g_2}$ viele Paare (A', B'), für die B' auf den ungeraden Pfaden mit M übereinstimmt: Für die geraden Kreise und Pfade von $A \triangle B$ hat man jeweils zwei Möglichkeiten die Kanten einem A- beziehungsweise B-Matching zuzuordnen. Bildet man daher diese Paare isomorph auf einen $(g_1 + g_2)$-dimensionalen Hyperwürfel ab, so zeigt Proposition 1.3.2, daß der Gesamtbeitrag der Flüsse von Paaren (A', B') aus der Äquivalenzklasse von (A, B) zu $f(e)$ genau $(g_1 + g_2)! \cdot \frac{|f|}{(g_1+g_2)!} = |f|$ ist.

2.) C ist in einem ungeraden Pfad aus $A \triangle B$ enthalten. Dann muß A auf den geraden Kreisen und Pfaden mit M übereinstimmen. Analog wie unter 1.) zeigt hier Proposition 1.3.3, daß der Gesamtbeitrag der Flüsse von Paaren (A', B') aus der Äquivalenzklasse von (A, B) zu $f(e)$ höchstens $u_A!(u_B - 1)! \cdot \frac{|f|}{u_A! u_B!} \leq |f|$ ist

Betrachten wir nun eine Kante $e = (M, M')$. Benutzt der Fluß f_{AB} zwischen zwei Matchings A und B die Kante e, so gilt nach Konstruktion des Flusses:

$$M_{AB} = A \triangle B \triangle M \text{ ist ein Matching aus } \mathcal{M}(G; k_1, k_2).$$

Insbesondere gilt aber auch: Sind (A, B) und (A', B') zwei Paare von Matchings aus der selben Äquivalenzklasse, so gilt $M_{AB} = M_{A'B'}$. D.h., die Anzahl Äquivalenzklassen ist sicherlich durch die Kardinalität von $\mathcal{M}(G; k_1, k_2)$ beschränkt.

Zusammen folgt daraus sofort

$$\max_{e \in E(\mathcal{G}_{MC})} \sum_{(A,B) \in \mathcal{F}} f_{AB}(e) \leq |\mathcal{M}(G; k_1, k_2)| \cdot |f| = |\mathcal{M}(G; k_1, k_2)|^{-1},$$

also (1.3.2). ■

Approximationsalgorithmen

In diesem Abschnitt zeigen wir, wie aus Satz 1.3.1 die Exiztenz verschiedener Approximationsalgorithmen folgt. Wir beginnen mit der Existenz eines Approximationsschemas für die Anzahl Matchings eines Graphens.

Satz 1.3.4. *Für alle Graphen G auf n Knoten gibt es für die Gesamtanzahl aller Matchings von G ein vollständig polynomielles randomisiertes Approximationsschemata* (FPAS).

20.6.2001

Beweis. Da in jedem Graphen die Anzahl Matchings der Kardinalität 2 durch $|E(G)|^2$ beschränkt ist, folgt aus Satz 1.3.1 sofort die Existenz eines vollständig polynomiellen randomisierten ϵ-uniformen Erzeugers (FPRUG) für $\mathcal{M}(G)$. Da andererseits die Relation R, die jedem Graphen die Menge aller seiner Matchings zuordnet, selbstreduzibel ist, folgt die Existenz eines FPAS daraus mit Hilfe von Satz $\boxed{\text{zve:sr}}$. ■

Will man statt aller Matchings nur Matchings einer vorgegebenen Kardinalität erzeugen oder zählen, beispielsweise nur perfekte Matchings, läßt sich Satz 1.3.1 nicht unmittelbar anwenden. Selbst wenn man die Vereinigung von fastperfekten und perfekten Matchings betrachtet, läßt sich Satz 1.3.1 nur für Graphen $G = (V_{2n}, E)$ anwenden, für die $|\mathcal{M}_{n-1}(G)|/|\mathcal{M}_n(G)|$ polynomiell beschränkt ist. Dies trifft leider jedoch bei weitem nicht immer zu. Abbildung 1.3 zeigt ein Beispiel von Umesh Vazirani mit nur zwei perfekten Matchings, aber exponentiell vielen fastperfekten Matchings.

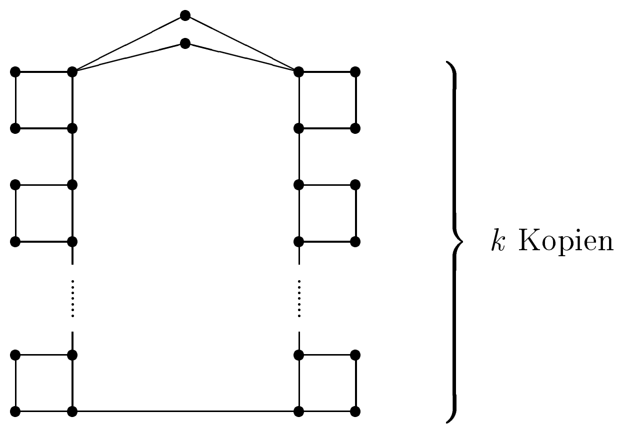


Abbildung 1.3: Beispiel von Vazirani

Ist andererseits $|\mathcal{M}_{n-1}(G)|/|\mathcal{M}_n(G)|$ polynomiell beschränkt ist, so lassen sich aus Satz 1.3.1 Approximationsschemata für jede Menge $\mathcal{M}_k(G)$ gewinnen.

Satz 1.3.5. *Es sei $q(n)$ ein beliebiges Polynom. Dann gilt: für alle Graphen $G = (V_{2n}, E)$ auf $2n$ Knoten mit $|\mathcal{M}_{n-1}(G)|/|\mathcal{M}_n(G)| \leq q(n)$ gibt es für jede Menge $\mathcal{M}_k(G)$ aller Matchings in G der Kardinalität k, $1 \leq k \leq n$, ein vollständig polynomielles randomisiertes Approximationsschemata* (FPAS).

Beweis. Wir setzen $m_k = |\mathcal{M}_k(G)$ für alle $1 \leq k \leq n$ und zeigen zunächst, daß aus der polynomiellen Schranke für $\frac{m_{n-1}}{m_n}$ auch eine solche für $\frac{m_{k-1}}{m_k}$ für alle $k \leq n$ folgt.

Für jedes Paar $A \in \mathcal{M}_{k-1}(G)$ und $B \in \mathcal{M}_n(G)$ enthält $A \triangle B$ genau $n-k+1$ ungerade Pfade, die mehr Kanten aus B als aus A enthalten. Für jeden solchen

Pfad P setzten wir $A' = A \triangle P$ und $B' = B \triangle P$. Dann gilt sicherlich $A' \in \mathcal{M}_k(G)$ und $B' \in \mathcal{M}_{n-1}(G)$. Da aber für $A' \in \mathcal{M}_k(G)$ und $B' \in \mathcal{M}_{n-1}(G)$ die symmetrische Differenz $A' \triangle B'$ höchstens einen ungeraden Pfad P mit mehr Kanten aus A als aus B enthält, ist umgekehrt ist jedes Paar (A', B') aus $\mathcal{M}_k(G) \times \mathcal{M}_{n-1}(G)$ das Bild höchstens einer solchen Abbildung. Für alle $1 \leq k \leq n$ gilt daher $(n-k+1) \cdot m_{k-1} \cdot m_n \leq m_k \cdot m_{n-1}$ und somit insbesondere auch $\frac{m_{k-1}}{m_k} \leq \frac{m_{n-1}}{m_n} \leq q(n)$.

Die Existenz von FPRUGs für die Menge $\mathcal{M}(G; k-1, k)$, $1 < k \leq n$, folgt daher sofort aus Satz 1.3.1. Durch Target-Shooting (Korollar $\boxed{\text{ap:ts}}$) erhält man daraus ein FPAS für $\frac{m_k}{m_{k-1}}$. (Man beachte, daß $\frac{m_k}{m_{k-1}}$ in jedem Graphen trivialerweise durch $|E|$ beschränkt ist.)

Sind nun $\mathcal{A}(k, \epsilon)$ Approximationsalgorithmen, die $\frac{m_k}{m_{k-1}}$ mit Toleranz $1+\epsilon$ approximieren, so ist $\mathcal{B}(k, \epsilon) = \mathcal{A}(k, \frac{\epsilon}{n}) \cdot \dots \cdot \mathcal{A}(2, \frac{\epsilon}{n})$ ein Approximationsalgorithmus, der $\frac{m_k}{m_{k-1}} \cdot \dots \cdot \frac{m_2}{m_1} = \frac{m_k}{|E|}$ sicherlich mit Toleranz $(1+\frac{\epsilon}{n})^{\frac{n}{2}} \leq 1+\epsilon$ approximiert. D.h., aus der Existenz der Approximationsschemata für $\frac{m_k}{m_{k-1}}$ folgt auch die Existenz eines solchen für m_k. ■

Bemerkung. Es ist nicht schwer zu zeigen daß es für die Graphen, die die Voraussetzung von Satz 1.3.5 erfüllen auch FPRUGs für die Mengen $\mathcal{M}_k(G)$ gibt. Da aber die Relation, die jedem Graphen die Menge seiner Matchings der Kardinalität k zuordnet, unter der oben genannten Einschränkung an die Graphenklasse nicht selbstreduzibel ist, läßt sich die Existenz eines FPAS für $|\mathcal{M}_k(G)|$ hier nicht aus Satz $\boxed{\text{zve:sr}}$ ableiten.

Korollar 1.3.6. *Für die Anzahl perfekter Matchings jedes Graphen $G = (V_{2n}, E)$ auf $2n$ Knoten mit minimalem Grad mindestens n gibt es ein vollständig polynomielles randomisiertes Approximationsschemata.*

Beweis. Sei $M \in \mathcal{M}_{n-1}(G)$ und seien u, v die beiden Knoten, die durch M nicht überdeckt werden. Wegen der Gradvoraussetzung an jeden Knoten sind dann entweder u und v in G durch eine Kante verbunden oder es gibt eine Kante $\{x, y\}$ in M, so daß u Nachbar von x ist und v Nachbar von y ist. Ordnet man daher M das perfekte Matching $M \cup \{u, v\}$ beziehungsweise $(M \setminus \{x, y\}) \cup \{\{u, x\}, \{v, y\}\}$ zu, folgt daraus unmittelbar

$$|\mathcal{M}_{n-1}(G)| \leq 4n^2 |\mathcal{M}_n(G)|,$$

da jedes perfekte Matching höchstens $4n^2$ vielen fastperfekten Matchings zugeordnet werden kann. ■

Mit einem ähnlichen Argument läßt sich auch zeigen (Aufgabe 1.6), daß es für die Anzahl perfekter Matchings in bipartiten Graphen $(A \cup B, E)$ mit $|A| = |B| =$

n mit minimalem Grad mindestens $\frac{n}{2}$ ein FPAS gibt. Dies zeigt dann auch, daß es für die Permanente von $n \times n$ 0-1 Matrizen mit mindestens $\frac{n}{2}$ Einsen in jeder Zeile und Spalte ein vollständig polynomielles randomisiertes Approximationsschema gibt.

1.4 Charakterisierung schnell mischender Markov-Ketten

Ziel dieses Abschnittes ist es, die Charakterisierungssätze 1.2.1, 1.2.2 und 1.2.4 schnell mischender Markov-Ketten zu beweisen. Dabei werden zwei weitere Parameter einer Markov-Kette eine wichtige Rolle spielen. Zum einen ist dies der zweitgrößte Eigenwert der Übergangsmatrix P und zum anderen die Variationsdistanz.

Eigenwerte einer reversiblen Markov-Kette

Eine wichtige Eigenschaft reversibler Markov-Ketten ist, daß die zugehörige Übergangsmatrix P nur reelle Eigenwerte besitzt.

Lemma 1.4.1. *Ist $\mathcal{MC}$ eine reversible Markov-Kette mit $|\mathcal{X}| = N$, Übergangsmatrix P und stationärer Verteilung π, so hat P die Eigenwerte $1 = \lambda_1 > \lambda_2 \geq \ldots \geq \lambda_N \geq -1$ und es gilt*

$$P = D^{-1/2} U \Lambda U^T D^{1/2},$$

wobei $D = diag(\pi_1, \ldots, \pi_N)$, $\Lambda = diag(\lambda_1, \ldots, \lambda_N)$ und U eine orthogonale Matrix ist.

Beweis. Wir zeigen zunächst, daß die Matrix $D^{1/2} P D^{-1/2}$ symmetrisch ist. Setzen wir $(a_{ij}) = D^{1/2} P D^{-1/2}$, so ist a_{ij} gegeben durch

$$a_{ij} = \sqrt{\pi_i} \cdot (PD^{-1/2})_{ij} = \sqrt{\pi_i} \cdot p_{ij} \cdot \frac{1}{\sqrt{\pi_j}}. \tag{1.4.1}$$

Da wir $\mathcal{MC}$ als reversibel vorausgesetzt haben, ergibt sich aus (1.4.1) sofort die behauptete Symmetrie:

$$a_{ij} = \frac{1}{\sqrt{\pi_i \pi_j}} \cdot p_{ij} \cdot \pi_i = \frac{1}{\sqrt{\pi_i \pi_j}} \cdot p_{ji} \cdot \pi_j = a_{ji}.$$

Aus der linearen Algebra wissen wir daher: $D^{1/2} P D^{-1/2}$ hat nur reelle Eigenwerte $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_N$ und läßt sich schreiben als $U \Lambda U^T$, wobei U orthogonal ist und der $i$te Spaltenvektor u_i von U ein Eigenvektor des $i$ten Eigenwertes λ_i ist. Aus $\lambda_i u_i = D^{1/2} P D^{-1/2} u_i$ folgt aber $\lambda_i(D^{-1/2}u_i) = P(D^{-1/2}u_i)$, d.h. λ_i ist auch Eigenwert von P mit Eigenvektor $u_i' = D^{-1/2}u_i$.

Um $|\lambda_i| \leq 1$ zu zeigen, wählen wir zu gegebenem $1 \leq i \leq N$ einen Index j_0 der den Absolutbetrag der Komponenten des Eigenvektors u_i' maximiert, d.h. für den gilt $|u'_{j_0 i}| = \max_{1\leq j\leq N} |u'_{ji}|$. Aus $Pu_i' = \lambda_i u_i'$ folgt damit insbesondere

$$\lambda_i u'_{j_0 i} = \sum_k p_{j_0 k} u'_{ki}$$

und somit auch

$$|\lambda_i| \cdot |u'_{j_0 i}| \leq \sum_k p_{j_0 k} |u'_{ki}| \leq |u'_{j_0 i}| \cdot \sum_k p_{j_0 k} = |u'_{j_0 i}|, \tag{1.4.2}$$

also $|\lambda_i| \leq 1$. Weiter beachte man, daß in (1.4.2) Gleichheit nur gilt, falls $u_i' = u'_{j_0 i} \cdot \mathbb{1}$. Der Eigenwert 1 hat also die Vielfachheit 1. ∎

Die Variationsdistanz

Ist $\mathcal{MC}$ eine ergodische Markov-Kette mit stationärer Verteilung π, so ist die *Variationsdistanz zum Zeitpunkt t* definiert als

$$\Delta(t) = \max_x \frac{1}{2} \sum_y |p_{xy}^{(t)} - \pi(y)|.$$

Lemma 1.4.2. *Für jede ergodische, reversible Markov-Kette mit stationärer Verteilung π gilt*

$$\Delta(t) \leq \frac{\lambda_{max}^t}{\pi_{min}},$$

wobei $\lambda_{max} = \max\{\lambda_2, |\lambda_N|\}$.

Beweis. Wir betrachten die Variationsdistanz separat für jeden Startpunkt x. D.h., wir setzen $\Delta_x(t) = \frac{1}{2}\sum_y |p_{xy}^{(t)} - \pi_y|$ und zeigen

$$[2\Delta_x(t)]^2 \leq \frac{1}{\pi_x} \cdot \lambda_{max}^{2t} \qquad \text{für alle } x \in \mathcal{X}. \tag{1.4.3}$$

Offensichtlich folgt aus (1.4.3) sofort die Aussage des Lemmas.

Unter Verwendung der Cauchy-Schwarz Ungleichung läßt sich die linke Seite von (1.4.3) schreiben als

$$[2\Delta_x(t)]^2 = \left[\sum_y \sqrt{\pi_y} \cdot \frac{|p_{xy}^{(t)} - \pi_y|}{\sqrt{\pi_y}}\right]^2 \leq \left[\left(\sum_y \pi_y\right)^{1/2} \cdot \left(\sum_y \frac{(p_{xy}^{(t)} - \pi_y)^2}{\pi_y}\right)^{1/2}\right]^2$$

$$= \sum_y \left(\frac{[p_{xy}^{(t)}]^2}{\pi_y} - 2p_{xy}^{(t)} + \pi_y \right) = \sum_y \frac{[p_{xy}^{(t)}]^2}{\pi_y} - 1 = \sum_y \frac{p_{xy}^{(t)}}{\pi_y} \cdot \frac{p_{xy}^{(t)} \cdot \pi_x}{\pi_x} - 1$$
$$= \sum_y \frac{p_{xy}^{(t)}}{\pi_y} \cdot \frac{p_{yx}^{(t)} \cdot \pi_y}{\pi_x} - 1 = \frac{1}{\pi_x} \cdot p_{xx}^{(2t)} - 1,$$

wobei die vorletzte Umformung aus der Reversibilität der Markov-Kette folgt (vgl. Aufgabe 1.2).

In der Notation von Lemma 1.4.1 folgt aus $D^{1/2}PD^{-1/2} = U\Lambda U^T$ sofort $P^{2t} = D^{-1/2}U\Lambda^{2t}U^TD^{1/2}$ und somit insbesondere

$$p_{xx}^{(2t)} = \sum_k u_{xk}^2 \lambda_k^{2t}.$$

Verwenden wir nun noch, daß $\lambda_1 = 1$ und $u_1 = D^{1/2}\mathbb{1}$, so folgt

$$\begin{aligned} p_{xx}^{(2t)} &= \pi_x + \sum_{k=2}^N u_{xk}^2 \lambda_k^{2t} \\ &\leq \pi_x + \lambda_{\max}^{2t} \cdot \sum_{k=2}^N u_{xk}^2 \\ &\leq \pi_x + \lambda_{\max}^{2t}, \end{aligned}$$

da mit U natürlich auch U^T orthogonal ist. Dies zeigt die Gültigkeit von (1.4.3).

■

Obere Schranken für die Konvergenzzeit

In Lemma 1.4.2 haben wir eine Beziehung zwischen der Variationsdistanz und $\lambda_{\max}$ aufgestellt. Um die ϵ-Konvergenzzeit durch den Leitwert zu beschränken, benötigen wir zusätzlich noch eine Beziehung zwischen dem Leitwert Φ_{MC} und den Eigenwerten λ_i der Übergangsmatrix P.

Satz 1.4.3. *Für jede ergodische, reversible Markov-Kette mit stationärer Verteilung π gilt*

$$\Phi_{MC}^2 \leq 2(1 - \lambda_2).$$

Beweis. Es sei u ein Eigenvektor von P zum zweitgrößten Eigenwert λ_2. Mit S_+ bezeichnen wir die Menge der positiven Koordinaten von u, also $S_+ = \{i : 1 \leq i \leq N, u_i \geq 0\}$. Da $\mathbb{1}$ Eigenvektor von P zum Eigenwert $\lambda_1 = 1$ ist, gilt $u^T\mathbb{1} = 0$ und daher wegen $\pi > 0$ auch $\pi(S_+) < 1$. Da weiter mit u auch $-u$ ein Eigenvektor von P ist, dürfen wir ohne Einschränkung annehmen, daß $0 < \pi(S_+) \leq 1/2$ gilt.

Weiter nehmen wir ohne Einschränkung an, daß der Zustandsraum $\mathcal{X}$ so geordnet ist, daß gilt

$$u_1 \geq \cdots \geq u_r \geq 0 > u_{r+1} \geq \cdots \geq u_N$$

und definieren einen neuen Vektor ψ, der nur aus den positiven Komponenten von u besteht:

$$\psi = \begin{cases} u_i & \text{für } 1 \leq i \leq r, \\ 0 & \text{sonst.} \end{cases}$$

Wir beweisen Satz 1.4.3 in zwei Schritten. Zunächst zeigen wir

$$\sum_i \pi_i \psi_i^2 \leq 2\Phi_{MC}^{-2} \cdot \sum_{i<j} c_{ij}(\psi_i - \psi_j)^2 \tag{$*$}$$

durch Anwendung der Definition von Φ_{MC} auf ausgewählte Teilmengen von S_+. In einem zweiten Schritt folgern wir dann aus der Tatsache, daß u ein Eigenvektor von P zum Eigenwert λ_2 ist, daß gilt

$$\sum_{i<j} c_{ij}(\psi_i - \psi_j)^2 \leq (1 - \lambda_2) \cdot \sum_i \pi_i \psi_i^2. \tag{$**$}$$

Aus $(*)$ und $(**)$ folgt dann die Behauptung des Satzes durch einige einfache Umformungen, die wir dem Leser überlassen.

Wir beweisen zunächst $(*)$. Nach der Definition von Φ_{MC} gilt $\sum_{i \in S} \pi_i \leq \Phi_{MC}^{-1} \cdot \sum_{i \in S, j \notin S} c_{ij}$ für alle $S \subseteq \mathcal{X}$ mit $0 < \pi(S) \leq 1/2$, insbesondere also für alle nichtleeren Teilmengen von $S_+ = \{1, \ldots, r\}$. Wir wenden diese Ungleichung auf alle Mengen $S_k = \{1, \ldots, k\}$ mit $1 \leq k \leq r$ an:

$$\begin{aligned}
\sum_i \pi_i \psi_i^2 &= \sum_{i=1}^{r} \pi_i \sum_{k=i}^{r} (\psi_k^2 - \psi_{k+1}^2) \\
&= \sum_{k=1}^{r} (\psi_k^2 - \psi_{k+1}^2) \sum_{i \in S_k} \pi_i \\
&\leq \sum_{k=1}^{r} (\psi_k^2 - \psi_{k+1}^2) \Phi_{MC}^{-1} \sum_{\substack{i \in S_k \\ j \notin S_k}} c_{ij} \\
&= \Phi_{MC}^{-1} \cdot \sum_{i<j} c_{ij} \sum_{i \leq k < j} (\psi_k^2 - \psi_{k+1}^2) \\
&= \Phi_{MC}^{-1} \cdot \sum_{i<j} c_{ij} (\psi_i^2 - \psi_j^2) \\
&\leq \Phi_{MC}^{-1} \cdot \sqrt{\sum_{i<j} c_{ij}(\psi_i + \psi_j)^2} \cdot \sqrt{\sum_{i<j} c_{ij}(\psi_i - \psi_j)^2},
\end{aligned}$$

wobei die letzte Ungleichung aus der Cauchy-Schwarzschen Ungleichung folgt. Wegen

$$\sum_{i<j} c_{ij}(\psi_i + \psi_j)^2 \leq 2 \sum_{i<j} c_{ij}(\psi_i^2 + \psi_j^2) \leq \sum_{i,j} c_{ij}(\psi_i^2 + \psi_j^2) = 2 \sum_i \pi_i \psi_i^2$$

folgt daraus sofort $(*)$.

Für den Beweis von $(**)$ setzen wir wieder $D = \mathrm{diag}(\pi_1, \ldots, \pi_N)$ und berechnen den Ausdruck $(D\psi)^T(I-P)u$ auf zwei Arten. Zum einen folgt aus der Wahl von u als Eigenvektor von P zum Eigenwert λ_2 und der Definition von ψ:

$$(D\psi)^T(I-P)u = (1-\lambda_2) \cdot (D\psi)^T u = (1-\lambda_2) \cdot \sum_i \pi_i \psi_i^2.$$

Zum anderen gilt (hier verwenden wir $\psi \geq u$, $\psi_i u_i = \psi_i^2$ für alle i und $c_{ij} = c_{ji}$ für alle i,j):

$$\begin{aligned}
(D\psi)^T(I-P)u &= (D\psi)^T u - \sum_i \pi_i \psi_i (\sum_{j \neq i} p_{ij} u_j + p_{ii} u_i) \\
&\geq \sum_i \pi_i \psi_i u_i - \sum_i \pi_i \psi_i \sum_{j \neq i} p_{ij} \psi_j - \sum_i \pi_i \psi_i (1 - \sum_{j \neq i} p_{ij}) \psi_i \\
&= -\sum_{i \neq j} c_{ij} \psi_i \psi_j + \sum_{i \neq j} c_{ij} \psi_i^2 \\
&= -2 \sum_{i<j} c_{ij} \psi_i \psi_j + \sum_{i<j} c_{ij}(\psi_i^2 + \psi_j^2) \\
&= \sum_{i<j} c_{ij}(\psi_i - \psi_j)^2.
\end{aligned}$$

■

Korollar 1.4.4. *Für die ϵ-Konvergenzzeit einer ergodischen, reversiblen Markovkette mit $\lambda_{\max} = \lambda_2$ gilt*

$$\tau(\epsilon) \leq \frac{2}{\Phi_{MC}^2} \left(\ln \epsilon^{-1} + \ln \pi_{min}^{-1} \right).$$

Beweis. Aus der Definition der ϵ-Konvergenzzeit und der Variationsdistanz folgt sofort

$$\tau(\epsilon) \leq \min\{t : \Delta(t') \leq \epsilon \text{ für alle } t' \geq t\}.$$

Somit gilt nach Lemma 1.4.2 und wegen $1 - x \leq \ln x^{-1}$ für alle $0 < x < 1$:

$$\tau(\epsilon) \leq \frac{\ln \epsilon^{-1} + \ln \pi_{min}^{-1}}{\ln \lambda_{max}^{-1}} \leq \frac{\ln \epsilon^{-1} + \ln \pi_{min}^{-1}}{1 - \lambda_{max}}. \tag{1.4.4}$$

20.6.2001

Falls daher $\lambda_{max} = \lambda_2$ folgt daraus zusammen mit Satz 1.4.3 bereits die behauptete Abschätzung. ■

Beweis von Satz 1.2.1. Im Lichte von Korollar 1.4.4 genügt es zu zeigen, daß für stark aperiodischen Markov-Ketten der zweitgrößte Eigenwert λ_2 immer mit dem betragsmäßig zweitgrößte Eigenwert $\lambda_{\max}$ übereinstimmt.

Um dies zu zeigen, betrachten wir $P' = 2P - I$, wobei P die Übergangsmatrix der stark aperiodischen Markov-Kette $\mathcal{MC}$ sei. Mit P ist auch P' Übergangsmatrix einer reversiblen Markov-Kette $\mathcal{MC}'$, für die ebenfalls π die stationäre Verteilung ist. Die Eigenwerte von P' sind gegeben durch $\lambda'_i = 2\lambda_i - 1$. Da nach Lemma 1.4.1 für den kleinsten Eigenwert λ'_N von P' gilt $\lambda'_N \geq -1$, ergibt sich daraus sofort $\lambda_N \geq 0$ und somit insbesondere $\lambda_{\max} = \lambda_2$. ■

Untere Schranken für die Konvergenzzeit

Im vorangehenden Abschnitt haben wir gezeigt, daß sich die ϵ-Konvergenzzeit einer ergodischen, reversiblen Markov-Kette von oben polynomiell in Φ_{MC}^{-1}, $\ln \epsilon^{-1}$ und $\ln \pi_{\min}^{-1}$ beschränken läßt. In diesem Abschnitt werden wir umgekehrt zeigen, daß sich die ϵ-Konvergenzzeit von unten polynomiell in Φ_{MC}^{-1} und $\frac{1}{\ln \pi_{\min}^{-1}}$ beschränken läßt. Dies impliziert dann natürlich auch, daß eine Markov-Kette für die $\ln \pi_{\min}^{-1}$ polynomiell beschränkt ist, nur dann schnell mischend sein kann, wenn auch Φ_{MC} polynomiell beschränkt ist.

Proposition 1.4.5. *Für die Variationsdistanz $\Delta(t)$ einer Markov-Kette mit stationärer Verteilung π gilt*

$$\Delta(s+t) \leq 2\Delta(s)\Delta(t) \qquad \text{für alle } s, t \in \mathbb{N}.$$

Beweis. Aus der Definition der Variationsdistanz folgt

$$\begin{aligned}
2\Delta(s)\Delta(t) &= 2 \cdot \left(\max_x \frac{1}{2} \sum_y |p_{xy}^{(s)} - \pi_y| \right) \cdot \left(\max_y \frac{1}{2} \sum_z |p_{yz}^{(t)} - \pi_z| \right) \\
&\geq \frac{1}{2} \max_x \sum_y |p_{xy}^{(s)} - \pi_y| \cdot \sum_z |p_{yz}^{(t)} - \pi_z| \\
&\geq \frac{1}{2} \max_x \sum_z |\sum_y p_{xy}^{(s)} p_{yz}^{(t)} - \sum_y \pi_y p_{yz}^{(t)} - \pi_z \sum_y p_{xy}^{(s)} + \pi_z \sum_y \pi_y|.
\end{aligned}$$

Da $\sum_y p_{xy}^{(s)} = 1 = \sum_y \pi_y$ ist und da wegen $P^T \pi = \pi$ auch $(P^t)^T \pi = \pi$ und somit insbesondere $\sum_y \pi_y p_{yz}^{(t)} = \pi_z$ gilt, folgt daraus wie behauptet

$$2\Delta(s)\Delta(t) \geq \frac{1}{2} \max_x \sum_z |p_{xz}^{(s+t)} - \pi_z| = \Delta(s+t).$$

20.6.2001

■

Satz 1.4.6. *Zu jeder Markov-Kette $\mathcal{MC}$ mit stationärer Verteilung π gibt es einen Fluß f mit*

$$\rho \leq 2\tau(\pi_{min}/4) \cdot \lceil \log \pi_{\min}^{-1} + 1 \rceil.$$

Beweis. Schreiben wir abkürzend τ für $\tau(\pi_{\min}/4)$ so gilt nach Proposition 1.4.5 für alle $k \in \mathbb{N}$:

$$\Delta(k\tau) \leq 2^{k-1} \cdot [\Delta(\tau)]^k.$$

Nach der Definition von $\tau = \tau(\pi_{\min}/4)$ gilt $\Delta(\tau) \leq |\mathcal{X}| \cdot \pi_{\min}/4 \leq 1/4$ und somit

$$\Delta(k\tau) \leq 2^{-(k+1)}.$$

Setzen wir nun $t_0 = \tau \cdot \lceil \log \pi_{\min}^{-1} + 1 \rceil$, so folgt sofort

$$\Delta(t_0) \leq 2^{-(\lceil \log \pi_{\min}^{-1} + 1 \rceil + 1)} \leq \frac{1}{4}\pi_{\min}.$$

Wir bezeichnen nun mit $\mathcal{W}'_{xy}(t_0)$ die Menge aller (nicht notwendigerweise einfachen) gerichteten x-y Wege der Länge t_0. Für jeden Weg W in $\mathcal{W}'_{xy}(t_0)$ sei $Prob(W)$ die Wahrscheinlichkeit, daß die in x startende Markov-Kette genau entlang dieses Weges verläuft. Insbesondere gilt also

$$\sum_{W \in \mathcal{W}'_{xy}(t_0)} Prob(W) = p_{xy}^{(t_0)}.$$

Nach den obigen Überlegungen zur Wahl von t_0 wissen wir zudem, daß $\mathcal{W}'_{xy}(t_0)$ nicht leer ist. Wir setzen nun

$$f_{xy}(e) = \sum_{\substack{W \in \mathcal{W}'_{xy}(t_0) \\ e \in W}} \frac{\pi_x \pi_y Prob(W)}{p_{xy}^{(t_0)}}.$$

Dann ist sicherlich $f = \sum f_x y$ ein Fluß. Zu zeigen bleibt noch, daß $\rho(f)$ die behauptete Ungleichung erfüllt.

Für eine beliebige Kante $e = (u, v)$ gilt (hier verwenden wir, daß aus $\Delta(t_0) \leq \frac{1}{4}\pi_{\min}$ nach Definition der Variationsdistanz $\pi_y / p_{xy}^{(t_0)} \leq 2$ folgt):

$$\begin{aligned} f(e) &= \sum_{x,y} \frac{\pi_y}{p_{xy}^{(t_0)}} \sum_{\substack{W \in \mathcal{W}'_{xy}(t_0) \\ e \in W}} \pi_x Prob(W) \\ &\leq 2 \sum_{x,y} \sum_{\substack{W \in \mathcal{W}'_{xy}(t_0) \\ e \in W}} \pi_x Prob(W) \end{aligned}$$

$$\begin{aligned}
&= 2\sum_x \pi_x \sum_{0\le t'\le t_0-1} \sum_y p_{xu}^{(t')} p_{uv} p_{vy}^{(t_0-1-t')} \\
&= 2p_{uv} \sum_{0\le t'\le t_0-1} \underbrace{\sum_x \pi_x p_{xu}^{(t')}}_{=\pi_u} \underbrace{\sum_y p_{vy}^{(t_0-1-t')}}_{=1} \\
&= 2t_0\pi_u p_{uv} \;=\; 2t_0 c(e).
\end{aligned}$$

Für die maximale Kantenauslastung $\rho(f)$ folgt damit sofort

$$\rho(f) = \max_e \frac{f(e)}{c(e)} \le 2t_0.$$

■

Korollar 1.4.7. *Für jede Markov-Kette $\mathcal{MC}$ mit stationärer Verteilung π gilt*

$$\tau(\epsilon) \ge \frac{1}{4\lceil \log \pi_{\min}^{-1} + 1\rceil} \cdot \Phi_{MC}^{-1} \qquad \textit{für alle } \epsilon \le \tfrac{1}{4}\pi_{\min}.$$

Beweis. Wegen der Monotonie der ϵ-Konvergenzzeit genügt es, Korollar 1.4.7 für $\epsilon = \frac{1}{4}\pi_{\min}$ zu zeigen. Aus Satz 1.4.6 folgt

$$\tau(\pi_{\min}/4) \ge \min_f \rho(f) \cdot \frac{1}{2\lceil \log \pi_{\min}^{-1} + 1\rceil},$$

wobei das Minimum über alle Flüsse f gebildet wird. Nach Lemma 1.2.3 gilt andererseits auch $\min_f \rho(f) \ge \frac{1}{2}\Phi_{MC}^{-1}$. ■

Beweis von Satz 1.2.2. Ist Φ_{MC}^{-1} polynomiell beschränkt, so folgt aus Satz 1.2.1 sofort, daß $\mathcal{MC}(I)$ schnell mischend ist. Falls andererseits Φ_{MC}^{-1} nicht polynomiell beschränkt ist, so folgt aus Korollar 1.4.7, daß $\mathcal{MC}(I)$ nicht schnell mischend sein kann. ■

Beweis von Satz 1.2.4. Ist $\mathcal{MC}(I)$ schnell mischend, so folgt aus Satz 1.4.6 die Existenz eines Flusses f mit polynomiell beschränktem $\rho(f)$. Existiert andererseits ein Fluß f für den $\rho(f)$ polynomiell beschränkt ist, so folgt aus Lemma 1.2.3, daß auch Φ_{MC}^{-1} polynomiell beschränkt ist, und somit aus Satz 1.2.2, daß $\mathcal{MC}(I)$ schnell mischend ist. ■

1.5 Übungsaufgaben

1.1. Beweisen Sie die Chapman-Kolmogorow-Gleichungen (1.1.3).

20.6.2001

1.2. Zeigen Sie: ist $\mathcal{MC}$ eine reversible Markov-Kette mit Zustandsraum $\mathcal{X}$ und stationärer Verteilung π, so gilt

$$p_{xy}^{(t)}\pi_x = p_{yx}^{(t)}\pi_y \qquad \text{für alle } x, y \in \mathcal{X} \text{ und } t \in \mathbb{N}.$$

1.3. Zeigen Sie: befindet sich die Markov-Kette $\mathcal{MC}_1(G)$ aus Beispiel 1 in Abschnitt 1.1 zum Zeitpunkt t im stationären Zustand, so gilt

$$Prob(X_{t+1} \notin S \mid X_t \in S) \geq \frac{1}{4} \qquad \text{für jede Menge } S \subseteq \mathcal{X} \text{ mit } |S| \leq |\mathcal{X}|/2.$$

1.4. Beweisen Sie die Aussage (1.2.2).

1.5. Beweisen Sie Proposition 1.3.2 und 1.3.3.
(Tip: Verwenden Sie die Symmetrie der zugrunde liegenden Graphen.) [Dagum, Luby, Mihail, Vazirani (1988)]

1.6. Zeigen Sie, daß es für die Anzahl perfekter Matchings in bipartiten Graphen Graphen $(A \cup B, E)$ mit $|A| = |B| = n$ mit minimalem Grad mindestens $\frac{n}{2}$ ein vollständig polynomielles randomisiertes Approximationsschema (FPAS) gibt.