

7 Markov-Ketten

Um das Matrix-Exponential im nächsten Kapitel einsetzen zu können, sind zunächst einige Vorüberlegungen notwendig.

7.1 Matrix-Normen und Matrix-Exponential

Für den Vektorraum $V = \mathbb{R}^n$ (oder \mathbb{C}^n) mit einer beliebigen Norm $\|\cdot\|$ und einer linearen Abbildung $A: V \rightarrow V$ ist

$$\|A\| := \sup_{x \neq 0} \frac{\|Ax\|}{\|x\|}$$

die zugehörige *Operatornorm*. Dabei gilt

$$\|A\| = \sup_{\|x\|=1} \|Ax\| = \max_{\|x\|=1} \|Ax\|.$$

In der Gleichung ist das Supremum gleich dem Maximum, da die Einheitskugel $\{\|x\| \leq 1\}$ im \mathbb{R}^n (bzw. im \mathbb{C}^n) kompakt ist, und daher ihr Rand, der eine abgeschlossene Teilmenge ist, auch. Dieser Zusammenhang ist nicht trivial und bedeutet, dass der Extremwert für ein x mit $\|x\| = 1$ wirklich angenommen wird. Es gelten weiter folgende Eigenschaften:

$$\|A\| \geq 0, \text{ mit } \|A\| = 0 \Leftrightarrow A = \mathbf{0} \quad (34)$$

$$\|cA\| = |c| \|A\| \text{ für alle } c \in \mathbb{R} \text{ (} c \in \mathbb{C} \text{)} \quad (35)$$

$$\|A + B\| \leq \|A\| + \|B\| \quad (36)$$

$$\|Ax\| \leq \|A\| \|x\| \quad (37)$$

$$\|AB\| \leq \|A\| \|B\| \quad (38)$$

Dabei beschreiben die Eigenschaften (34), (35) und (36) die üblichen Norm-Axiome und (37) ist eine Verträglichkeits-Bedingung. Schließlich ist (38) eine wichtige neue Eigenschaft. Man überlege sich, warum hier i.A. keine Gleichheit gelten kann!

Als Beispiele werden an dieser Stelle die Matrixnormen zu bereits bekannten Normen aufgestellt für eine $n \times n$ -Matrix A mit Elementen in \mathbb{C} (kurz: $A \in \text{Mat}(n, \mathbb{C})$). Die wichtigsten Beispiele sind in Tabelle 2 zusammengefasst. Für eine dieser Normen wird noch ein neuer Begriff benötigt.

Norm $\ \cdot\ $	Operatornorm	$\ \cdot\ $
$\ \cdot\ _1$	Spaltensummennorm:	$\ A\ _1 = \max_{1 \leq j \leq n} \sum_{i=1}^n a_{ij} $
$\ \cdot\ _2$	Spektralnorm:	$\ A\ _2 = \sqrt{\varphi(A^+A)}$, mit $A^+ = \overline{A}^t$
$\ \cdot\ _\infty$	Zeilensummennorm:	$\ A\ _\infty = \max_{1 \leq i \leq n} \sum_{j=1}^n a_{ij} $

Tabelle 2: Einige gängige Matrixnormen

Definition 7.1. Das *Spektrum* einer Matrix $A \in \text{Mat}(n, \mathbb{C})$ ist die Menge ihrer Eigenwerte, also

$$\sigma(A) = \{ \lambda \in \mathbb{C} \mid \lambda \text{ ist EW von } A \}.$$

Dann heißt

$$\varphi(A) = \max_{\lambda \in \sigma(A)} |\lambda|$$

der *Spektralradius* von A und definiert den kleinsten Kreis um 0 in \mathbb{C} , in dem alle Eigenwerte der Matrix A liegen.

Es gibt verschiedene Ansätze, das Matrix-Exponential e^A zu betrachten. Wir beginnen mit der Definition mittels der Potenzreihe der Exponential-Funktion, und erhalten (zunächst formal)

$$e^A = \sum_{n=0}^{\infty} \frac{A^n}{n!}. \quad (39)$$

Wie auch bei der Exponential-Funktion aus Analysis I ist die Konvergenz der Reihe von großer Bedeutung. Daher wird zunächst das Konvergenzverhalten geprüft.

$$\left\| \sum_{m=0}^N \frac{A^m}{m!} \right\| \leq \sum_{m=0}^N \frac{\|A^m\|}{m!} \leq \sum_{m=0}^N \frac{\|A\|^m}{m!} \leq \sum_{m=0}^{\infty} \frac{\|A\|^m}{m!} = e^{\|A\|} < \infty.$$

Daraus folgt die Konvergenz der Reihe für jede gegebene Matrix A , weil wir hier eine Majorante für jedes Element der durch Gl. (39) definierten Matrix berechnet haben (man mache sich die Details als Übung klar).

Beispiel 7.2. Wir berechnen e^{tA} für $A = \begin{pmatrix} 1 & 1 \\ -1 & -1 \end{pmatrix}$. Da $A^2 = 0$, wird die Reihe einfach:

$$e^{tA} = \mathbb{1} + tA + 0 = \begin{pmatrix} 1+t & t \\ -t & 1-t \end{pmatrix}.$$

Für $A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ ist die Rechnung etwas komplizierter. Da $A^2 = \mathbb{1}$, können wir die Summe nach geraden und ungeraden Indices aufteilen, und bekommen (Details: Übung):

$$e^{tA} = \cosh(t) \mathbb{1} + \sinh(t) A = \begin{pmatrix} \cosh(t) & \sinh(t) \\ \sinh(t) & \cosh(t) \end{pmatrix}.$$

Aus unseren obigen Überlegungen folgt auch, dass e^{tA} für jedes $t \in \mathbb{R}$ wohldefiniert ist. Dabei gilt sogar gleichmäßige Konvergenz auf kompakten Intervallen für den Parameter t . Insbesondere gilt $e^{0 \cdot A} = e^0 = \mathbb{1}$.

Nun betrachten wir die Ableitung:

$$\begin{aligned} \frac{d}{dt} e^{tA} &= \frac{d}{dt} \sum_{m \geq 0} \frac{t^m A^m}{m!} = \sum_{m \geq 1} \frac{t^{m-1}}{(m-1)!} A^m \\ &= A \underbrace{\sum_{m \geq 1} \frac{(tA)^{m-1}}{(m-1)!}}_{=e^{tA}} = A e^{tA} = e^{tA} A. \end{aligned}$$

Man beachte, dass der letzte Schritt gilt, weil wir im Schritt davor genauso gut ein A nach hinten hätten herausziehen können. Die Matrizen A und e^{tA} vertauschen also für alle $t \in \mathbb{R}$. Diese Rechnung legt nun außerdem das folgende Resultat nahe.

Satz 7.3. *Sei $A \in \text{Mat}(n, \mathbb{C})$ beliebig, aber fest. Dann besitzt das matrixwertige Anfangswertproblem $\dot{M} = AM$ mit $M(0) = \mathbb{1}$ die eindeutige Lösung $M(t) = e^{tA}$.*

Beweis. Die Lösungseigenschaft ist nach obiger Rechnung klar, die Eindeutigkeit könnte mittels Picard–Lindelöf gezeigt werden. Der Beweis ist auch durch eine konkrete Rechnung möglich. Dazu sei $N(t)$ eine beliebige Lösung. Wir betrachten nun $R(t) = e^{-tA} N(t)$ und werden nachweisen, dass R eine konstante Matrixfunktion ist.

Dazu berechnen wir den Differenzenquotienten (für kleine $h > 0$) wie folgt:

$$\begin{aligned} \frac{1}{h}(R(t+h) - R(t)) &= \frac{1}{h}(e^{-(t+h)A}[N(t+h) - N(t)] + [e^{-(t+h)A} - e^{-tA}]N(t)) \\ &= e^{-(t+h)A} \frac{[N(t+h) - N(t)]}{h} + \frac{[e^{-(t+h)A} - e^{-tA}]}{h} N(t) \\ &\xrightarrow{h \searrow 0} e^{-tA} \underbrace{\dot{N}(t)}_{=A \cdot N(t)} - Ae^{-tA} N(t) \equiv 0 \equiv \dot{R}(t). \end{aligned}$$

Also ist $R(t)$ konstant und es folgt, dass $R(t) = R(0) = \mathbb{1}$ gilt. Es ergibt sich also $N(t) = e^{tA}$ wie behauptet. \square

Eine alternative Herleitung (wie bereits angedeutet) liefert die Lösung einer Differentialgleichung mittels Picard-Iteration. Definiere dafür e^{tA} als Lösung des Anfangswertproblems $\dot{M} = AM$ mit $M(0) = \mathbb{1}$. Daraus ergibt sich

$$M(t) = \underbrace{M(0)}_{=\mathbb{1}} + \int_0^t AM(\tau) d\tau$$

durch Integration. Nun können wir wieder iterieren:

$$M^{(0)}(t) = \mathbb{1} \quad \text{und} \quad M^{(n+1)}(t) := \mathbb{1} + \int_0^t AM^{(n)}(\tau) d\tau \quad \text{für alle } n \in \mathbb{N}_0.$$

Analog zu unserer eindimensionalen Picard-Iteration ergibt sich hier induktiv (nachrechnen!) die Formel

$$M^{(N)}(t) = \sum_{m=0}^N \frac{t^m A^m}{m!} \xrightarrow{N \rightarrow \infty} \sum_{m=0}^{\infty} \frac{t^m A^m}{m!},$$

und wir bekommen wieder die Potenzreihe, mit der wir weiter oben begonnen hatten. Welchen Zugang man bevorzugt, ist Geschmacksache — sie sind äquivalent.

Es gibt noch eine andere wichtige Eigenschaft der Exponentialfunktion, die über eine Funktionalgleichung beschrieben wird.

Satz 7.4. Sei $T: \mathbb{R}_{\geq 0} \rightarrow \text{Mat}(n, \mathbb{C})$ stetig, und erfülle die Funktionalgleichung

$$T(t+s) = T(t)T(s)$$

für alle $t, s \geq 0$ zusammen mit der Anfangsbedingung

$$T(0) = 1.$$

Dann existiert eine Matrix $A \in \text{Mat}(n, \mathbb{C})$, so dass $T(t) = e^{tA}$ für alle $t \geq 0$.

Für einen Beweis verweisen wir auf [8, S. 3 und S. 11]. Der entscheidende Punkt in dieser Aussage ist, dass es eine feste Matrix A gibt, so dass $T(t) = e^{tA}$ gilt, wobei diese Matrix dann aus anderen Daten bestimmt werden muss, die nicht in der Funktionalgleichung vorkommen. Diese Situation tritt etwa in der Phylogenie auf, wenn man Mutations-Generatoren in Stammbäumen bestimmen muss.

Nun kommen wir zu einer Besonderheit in der Berechnung von Matrix-Exponentialen. Sie leitet sich daraus ab, dass der Kommutator zweier Matrizen A und B , also

$$[A, B] := AB - BA$$

i.A. nicht verschwindet, weil die Matrizen nicht miteinander vertauschen. Falls $[A, B] = 0$ gilt, können wir den binomischen Lehrsatz in der üblichen Form benutzen und bekommen

$$\begin{aligned} e^{A+B} &= \sum_{m \geq 0} \frac{(A+B)^m}{m!} = \sum_{m \geq 0} \frac{1}{m!} \sum_{\ell=0}^m \binom{m}{\ell} A^\ell B^{m-\ell} \\ &= \underbrace{\sum_{\ell=0}^{\infty} \frac{A^\ell}{\ell!}}_{=e^A} \underbrace{\sum_{m \geq \ell} \frac{B^{m-\ell}}{(m-\ell)!}}_{=\sum_{k \geq 0} \frac{B^k}{k!} = e^B} = e^A e^B. \end{aligned}$$

Gilt aber $[A, B] \neq 0$, ist der zweite Schritt nicht durchführbar, da AB nicht mit BA vertauscht werden darf:

$$\begin{aligned} (A+B)^2 &= A^2 + AB + BA + B^2 = A^2 + 2AB + B^2 - [A, B] \quad \text{und} \\ (A+B)^3 &= A^3 + 2A^2B + 2AB^2 + B^3 - [A, [A, B]] + [B, [A, B]]. \end{aligned}$$

Entsprechene Formeln gelten für höhere Potenzen.

Satz 7.5. (BCH-Formel nach Baker, Campbell, Hausdorff)

$$\underbrace{\log(e^A e^B)}_{\text{häufig eingesetzte Approximation}} = A + B + \frac{1}{2}[A, B] + \frac{1}{12}([A, [A, B]] - [B, [A, B]]) - \frac{1}{24}[B, [A, [A, B]]] + \dots$$

In Zusammenhang mit der allgemeinen Formel ist auch Dynkin (1947) zu nennen.

Es existieren weitere Formeln, welche zur Approximation eingesetzt werden und verwandt zur BCH-Formel sind:

$$\text{Zassenhaus: } e^{t(A+B)} = \underbrace{e^{tA} e^{tB} e^{-\frac{1}{2}t^2[A,B]} e^{\frac{1}{6}t^3(\dots)}}_{\text{häufige Approximation}} \cdot \dots$$

$$\text{Hadamard: } e^A B e^{-A} = \underbrace{B + [A, B]}_{\text{häufige Approx.}} + \frac{1}{2}[A, [A, B]] + \frac{1}{6}[A, [A, [A, B]]] + \dots$$

Diese und weitere Formeln finden sich auch in den online verfügbaren Lexika.

Um später besser mit speziellen Matrizen (den Markov-Generatoren) umgehen zu können, folgen an dieser Stelle einige allgemeine Rechenregeln und Definitionen.

$$\det(e^A) = e^{\text{tr}(A)} \quad \text{mit der Spur} \quad \text{tr}(A) = \sum_i A_{ii}. \quad (40)$$

Einen Beweis dieser sehr nützlichen Formel kann man rasch für diagonalisierbare Matrizen erbringen, denn dann hat man $A = S^{-1}DS$ mit $D = \text{diag}(\lambda_1, \dots, \lambda_n)$. Also bekommt man

$$S e^A S^{-1} = e^{S A S^{-1}} = e^{\text{diag}(\lambda_1, \dots, \lambda_n)} = \text{diag}(e^{\lambda_1}, \dots, e^{\lambda_n})$$

und folglich auch

$$\det(e^A) = \det(S e^A S^{-1}) = \det(\text{diag}(e^{\lambda_1}, \dots, e^{\lambda_n})) = \prod_{i=1}^n e^{\lambda_i} = e^{\sum_{i=1}^n \lambda_i} = e^{\text{tr}(A)},$$

wobei wir uns daran erinnern, dass die Spur eine Matrix mit der Summe ihrer Eigenwerte (mit Multiplizitäten) übereinstimmt.

Um die Formel (40) in voller Allgemeinheit ableiten zu können, kann man die Jordan'sche Normalform komplexer Matrizen einsetzen. Eine Variante besagt, dass man A stets als $A = B + N$ mit B diagonalisierbar und N nilpotent zerlegen kann, so dass $[B, N] = \mathbf{0}$ gilt. Nilpotent bedeutet, dass $N^m = \mathbf{0}$ für ein $m \in \mathbb{N}$ gilt. Dann folgt

$$e^A = e^{B+N} = e^B e^N,$$

wobei wir schon wissen, wie wir e^B und seine Determinante berechnen. Wegen der Nilpotenz von N ist e^N ein Polynom in N mit Grad kleiner als m , und man zeigt dann noch, dass $\det(e^N) = 1$. Wegen $\det(e^A) = \det(e^B) \det(e^N) = \det(e^B)$ folgt dann unsere Behauptung.

7.2 Markov-Matrizen und Markov-Generatoren

Wir beginnen mit einer kurzen Wiederholung aus der Wahrscheinlichkeitstheorie, wo wir uns schon genauer mit Markov-Ketten in diskreter Zeit befasst hatten. Eine gute allgemeine Referenz für den Hintergrund zu diesem Kapitel ist das Buch von Norris [4].

Definition 7.6. Eine Matrix $M \in \text{Mat}(n, \mathbb{R})$, notiert als $M = (M_{ij})_{1 \leq i, j \leq n}$, heißt *Markov-Matrix*, falls die folgenden beiden Eigenschaften erfüllt sind:

- (1) $M_{ij} \geq 0$ gilt für alle $1 \leq i, j \leq n$;
- (2) $\sum_j M_{ij} = 1$ gilt für alle $1 \leq i \leq n$.

Im weiteren Verlauf wurden absorbierende Zustände diskutiert, aber auch wichtige Kriterien für verschiedene Stufen von Irreduzibilität besprochen (eine Wiederholung des Stoffes ist dringend empfohlen!), nämlich

$$\begin{aligned} \text{irreduzibel:} & \quad \forall i, j: \exists m \in \mathbb{N}: (M^m)_{ij} > 0, \\ \text{primitiv:} & \quad \exists m: \forall i, j: (M^m)_{ij} > 0, \\ \text{aperiodisch:} & \quad \text{ggT}(\text{Zykel-Längen in } \mathcal{G}_M) = 1. \end{aligned}$$

Dabei bezeichnet \mathcal{G}_M den M zugeordneten Graphen, der n Knoten besitzt und eine gerichtete Kante von i nach j genau dann, wenn $M_{ij} > 0$. Man kann die obigen Eigenschaften rasch anhand von \mathcal{G}_M ermitteln. Irreduzibel etwa bedeutet, dass man von jedem Knoten zu jedem Knoten entlang der ‘Einbahnstraßen’ kommen kann. Es sei noch angemerkt, dass die obige, vereinfachte Charakterisierung von ‘aperiodisch’ so i.A. nur für irreduzible Matrizen stimmt, was uns hier aber ausreicht. Die Eigenschaften stehen dabei in folgender Relation:

Satz 7.7. *Eine Markov-Matrix ist primitiv genau dann wenn sie irreduzibel und aperiodisch ist.*

Nun entwickeln wir das Analogon für stetige Zeit.

Definition 7.8. Eine Matrix $A \in \text{Mat}(n, \mathbb{R})$ heißt *Markov-Generator*, falls die folgenden beiden Eigenschaften erfüllt sind:

- (1) $A_{ij} \geq 0$ gilt für alle $i \neq j$;
- (2) $\sum_j A_{ij} = 0$ gilt für alle i .

Ein Beispiel für einen Markov-Generator ist

$$A = \begin{pmatrix} -4 & 1 & 3 \\ 0 & -1 & 1 \\ 1 & 1 & -2 \end{pmatrix}. \quad (41)$$

Um auch einer solchen Matrix einen Graphen zuordnen zu können, vereinbaren wir, dass der Graph \mathcal{G}_A eine gerichtete Kante von i nach j genau dann besitzt, falls $A_{ij} > 0$. Es

kann also niemals eine Kante von i nach i geben, da $A_{ii} \leq 0$, und somit gibt es keine Schleifen. Abb. 18 illustriert \mathcal{G}_A , und wir können hier die Irreduzibilität ablesen, da wir von jedem Knoten starten können und doch entlang der gerichteten Kanten zu jedem Knoten gelangen können. Die Beziehung zwischen irreduzibel und primitiv wird sich hier etwas anders gestalten, wie wir gleich sehen werden.

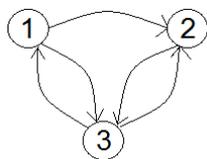


Abbildung 18: Graph \mathcal{G}_A zum Markov-Generator A von Gl. (41).

Satz 7.9. *Ist A ein Markov-Generator, so ist $\{e^{tA} \mid t \geq 0\}$ eine Halbgruppe (sogar Monoid) von Markov-Matrizen. Ist e^{tA} Markov für alle $t \geq 0$, so ist A ein Markov-Generator. Es reicht sogar, dass e^{tA} Markov-Matrix für alle t in einem kleinen Intervall $[0, \varepsilon]$ mit $\varepsilon > 0$ ist.*

Beweis. Für diesen Beweis benötigen wir die Exponentialreihe, also

$$e^{tA} = \mathbb{1} + tA + \frac{t^2}{2}A^2 + \dots,$$

die wir in mehrfacher Weise einsetzen werden.

Sei nun A ein Markov-Generator. Dann gilt für ein beliebiges $1 \leq i \leq n$:

$$\begin{aligned} \sum_j (e^{tA})_{ij} &= 1 + \sum_{m \geq 1} \frac{t^m}{m!} \sum_{j=1}^n (A^m)_{ij} = 1 + \sum_{m \geq 1} \frac{t^m}{m!} \sum_{\ell, j=1}^n (A^{m-1})_{i\ell} A_{\ell j} \\ &= 1 + \sum_{m \geq 1} \frac{t^m}{m!} \sum_{\ell=1}^n (A^{m-1})_{i\ell} \underbrace{\sum_{j=1}^n A_{\ell j}}_{=0} = 1, \end{aligned}$$

woraus sich die Zeilensummeneigenschaft von $M(t) := e^{tA}$ ergibt. Um zu zeigen, dass $M_{ij}(t) \geq 0$ für alle i, j gilt, beachten wir, dass für kleine $t > 0$ gilt:

$$M(t) = e^{tA} = \mathbb{1} + tA + \mathcal{O}(t^2), \quad (42)$$

wobei dies genauer so zu lesen ist, dass es für ein gegebenes $\varepsilon > 0$ zu jedem Indexpaar i, j eine Konstante $c_{ij} > 0$ gibt, so dass das Restglied in $M_{ij}(t) = \delta_{ij} + tA_{ij} + R_{ij}^{(2)}(t)$ die Ungleichung

$$|R_{ij}^{(2)}(t)| \leq c_{ij} t^2$$

erfüllt, und zwar für alle $t \in [0, \varepsilon]$.

Falls nun $A_{ij} > 0$, was nur für $i \neq j$ der Fall sein kann, so wählen wir $t > 0$ so klein, dass $tA_{ij} > c_{ij}t^2$ gilt. Das ist dann für *alle* hinreichend kleinen t wahr, und wir sehen, dass dann $(e^{tA})_{ij} > 0$ für alle hinreichend kleinen $t \geq 0$. Betrachten wir nun den Fall $A_{ij} = 0$ mit $i \neq j$. Hier ist natürlich $M_{ij}(0) = 0$. Wenn nun $M_{ij}(t) < 0$ für $0 < t < \eta$ (und ein sehr kleines $\eta > 0$) wäre, so bekämen wir aus Gl. (42) als Konsequenz aber, dass $A_{ij} < 0$ gelten müsste — was im Widerspruch zur Generator-Eigenschaft von A stünde. Also kann es so ein Intervall nicht geben, und wir haben $M_{ij}(t) \geq 0$ für (mindestens) ein kleines Intervall $[0, \eta']$ mit $\eta' > 0$.

Für die Diagonale, wo wir ja $A_{ii} \leq 0$ haben, können wir Gl. (42) noch vereinfachen zu $M_{ii}(t) = 1 + \mathcal{O}(t)$, und es ist unmittelbar klar, dass sogar $M_{ii}(t) > 0$ für alle hinreichend kleinen t gelten muss. Damit haben wir insgesamt nachgewiesen, dass $M(t)$ für alle hinreichend kleinen t eine Markov-Matrix ist.

Nun bedienen wir uns eines Tricks, und bemerken, dass

$$M(t) = e^{tA} = \left(e^{\frac{t}{n}A}\right)^n = M\left(\frac{t}{n}\right)^n$$

gilt, und zwar für jedes $n \in \mathbb{N}$. Somit können wir für ein gegebenes $t > 0$ immer ein n finden, so dass $M\left(\frac{t}{n}\right)$ nach unserem obigen Argument eine Markov-Matrix ist. Dann muss aber auch $M(t)$ eine Markov-Matrix sein, und die eine Richtung des Beweises ist fertig.

Für die Rückrichtung sei nun $\varepsilon > 0$ beliebig, aber fest, und $M(t) = e^{tA}$ sei Markov für alle $t \in [0, \varepsilon]$. Dann gilt $\sum_j M_{ij}(t) = 1$ für alle $1 \leq i \leq n$ und alle $t \in [0, \varepsilon]$, und wir bekommen durch Differentiation:

$$0 = \frac{d}{dt} \sum_{j=1}^n (e^{tA})_{ij} = \sum_{j=1}^n (e^{tA} A)_{ij} = \sum_{\ell=1}^n (e^{tA})_{i\ell} \sum_{j=1}^n A_{\ell j}.$$

Diese Relation können wir nun bei jedem $t \in [0, \varepsilon]$ einsetzen, insbesondere also auch bei $t = 0$. Da nun $(e^{tA})_{i\ell}|_{t=0} = \delta_{i\ell}$ gilt, bekommen wir damit

$$0 = \sum_{\ell=1}^n \delta_{i\ell} \sum_{j=1}^n A_{\ell j} = \sum_{j=1}^n A_{ij} = 0,$$

womit wir die Zeilensummenbedingung bestätigt haben. Für $i \neq j$ betrachten wir noch einmal die Exponentialreihe, und bemerken, dass wir A_{ij} dann folgendermaßen als Grenzwert bekommen können,

$$A_{ij} = \lim_{h \searrow 0} \frac{1}{h} (e^{hA} - \mathbb{1})_{ij} \stackrel{i \neq j}{=} \lim_{h \searrow 0} \frac{1}{h} (e^{hA})_{ij} \geq 0.$$

Daraus ergibt sich die fehlende Bedingung, und A ist ein Markov-Generator. □

Nun kommen wir zum wichtigen Zusammenhang zwischen ‘irreduzibel’ und ‘primitiv’, der bei Markov-Ketten in stetiger Zeit etwas überraschend ist.

Satz 7.10. *Ist A ein irreduzibler Markov-Generator, so ist e^{tA} für alle $t > 0$ primitiv.*

Beweis. Ist $A_{ij} > 0$, so auch $(e^{tA})_{ij} > 0$ für hinreichend kleine t , wie wir aus dem vorigen Beweis gesehen haben, und damit überhaupt für alle $t > 0$. Da A irreduzibel ist, existiert für alle i, j ein (gerichteter) Weg von i nach j . Nehmen wir an, ein solcher Weg sei $i = i_0 \rightarrow i_1 \rightarrow \dots \rightarrow i_m = j$. Dabei sind dann alle $A_{i_\ell, i_{\ell+1}} > 0$. Daher bekommen wir dann die Ungleichung

$$\begin{aligned} (e^{tA})_{ij} &= ((e^{\frac{t}{m}A})^m)_{ij} = \sum_{j_1=1}^n \cdots \sum_{j_{m-1}=1}^n (e^{\frac{t}{m}A})_{i_0 j_1} (e^{\frac{t}{m}A})_{j_1 j_2} \cdots (e^{\frac{t}{m}A})_{j_{m-1} i_m} \\ &\geq \underbrace{(e^{\frac{t}{m}A})_{i_0 i_1}}_{>0} \underbrace{(e^{\frac{t}{m}A})_{i_1 i_2}}_{>0} \cdots \underbrace{(e^{\frac{t}{m}A})_{i_{m-1} i_m}}_{>0} > 0. \end{aligned}$$

Es existiert also mindestens ein Term, in dem jeder Faktor > 0 ist und damit auch $(e^{tA})_{ij} > 0$. Da dies nun für alle $t > 0$ und alle i, j gilt, ist die Behauptung klar. \square

Aus dem Beweis des vorigen Satzes ergibt sich noch eine recht erstaunliche Folgerung:

Folgerung 7.11. *Ist e^{tA} irreduzibel für irgendein $t > 0$, so ist e^{tA} primitiv für alle $t > 0$. Insbesondere kann in diesem Fall kein Element von e^{tA} bei $t > 0$ verschwinden.*

Beispiel 7.12. Betrachte die Markov-Matrix $M = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, die irreduzibel, aber nicht primitiv ist. Aufgrund der obigen Folgerung kann sie also nie in einer Markov-Halbgruppe vorkommen.

Beispiel 7.13. Nun betrachten wir die Markov-Matrix $M = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$, die primitiv ist, denn $M^2 = \begin{pmatrix} \frac{3}{4} & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$. Dennoch kann auch M nicht in einer Markov-Halbgruppe vorkommen, denn in $M = e^{tA}$ wäre ein Element 0.

Aus diesen Beispielen ergibt sich, dass es eine interessante (und für die Anwendungen wichtige) Frage ist, welche Markov-Matrizen in einer Markov-Halbgruppe vorkommen können. Dies wurde zuerst von Elfving (1936) betrachtet und ist als *Darstellungsproblem* oder *Einbettungsproblem* für Markov-Matrizen bekannt.

Definition 7.14. Eine Markov-Matrix M heisst *einbettbar* (oder *darstellbar*), wenn es einen Markov-Generator Q gibt, so dass $M = e^Q$ gilt.

Man überlegt sich sofort, dass M genau dann einbettbar ist, wenn M in einer Markov-Halbgruppe der Form $\{e^{tQ} \mid t \geq 0\}$ vorkommt. Dies folgt aus der Beobachtung, dass Q genau dann ein Markov-Generator ist, wenn dies auch für tQ mit beliebigen $t > 0$ gilt. Um das Einbettungsproblem besser zu verstehen, benötigen wir auch den Logarithmus einer Matrix, der aber nicht für jede Matrix definiert ist. Dazu beachten wir

$$e^{tA} = \mathbb{1} + \underbrace{(e^{tA} - \mathbb{1})}_{\text{'klein' für } t \searrow 0}$$