Markov Chains Through Semigroup Graph Expansions (A Survey)

John Rhodes and Anne Schilling

Abstract We review the recent approach to Markov chains using the Karnofksy–Rhodes and McCammond expansions in semigroup theory by the authors and illustrate them by two examples.

Keywords Markov chains · Karnofsky–Rhodes expansion · McCammond expansion · Stationary distribution · Mixing time

2010 Mathematics Subject Classification Primary 20M30 · 60J10; Secondary 20M05 · 60B15 · 60C05

1 Introduction

A Markov chain is a model that describes transitions between states in a state space according to certain probabilistic rules. The defining characteristic of a Markov chain is that the transition from one state to another only depends on the current state and the elapsed time, but not how it arrived there. In other words, a Markov chain is "memoryless". Markov chains have many applications, from stock performances, population dynamics to traffic models.

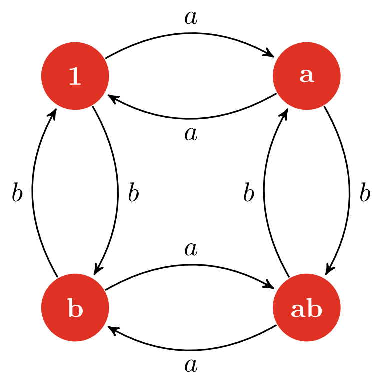
In this paper, we consider Markov chains with a finite state-space Ω. The Markov chain can be described pictorially by its *transition diagram*, where the vertices are the states in Ω and the labeled directed arrows between the vertices indicate the transitions. For example,

Dedicated to K. S. S. Nambooripad.

J. Rhodes
Department of Mathematics, University of California, Berkeley, CA 94720, USA
e-mail: rhodes@math.berkeley.edu; blvdbastille@gmail.com

A. Schilling (✉)
Department of Mathematics, UC Davis, One Shields Ave., Davis, CA 95616-8633, USA
e-mail: anne@math.ucdavis.edu

P. G. Romeo et al. (eds.), *Semigroups, Categories, and Partial Algebras*,
Springer Proceedings in Mathematics & Statistics 345,
https://doi.org/10.1007/978-981-33-4842-4_9



(1.1)

is the transition diagram for a Markov chain with state-space $\Omega = \{\mathbf{1}, \mathbf{a}, \mathbf{b}, \mathbf{ab}\}$. Associating the probability $0 \leqslant x_a \leqslant 1$ to the transition arrows labeled a and $0 \leqslant x_b \leqslant 1$ to the transition arrows labeled b yields a Markov chain assuming that $x_a + x_b = 1$. As the picture indicates, if the Markov chain is in state $\mathbf{1}$, then with probability x_a it transitions to state $\mathbf{a}$ and with probability x_b it transitions to state $\mathbf{b}$, and so on. Note that every vertex has one outgoing arrow labeled a and one outgoing arrow labeled b.

The *transition matrix* T of a Markov chain is a matrix of dimension $|\Omega| \times |\Omega|$. Let A be the set of all edge labels in the transition diagram. Then the entry $T_{s',s}$ in row $s' \in \Omega$ and column $s \in \Omega$ in T is

$$T_{s',s} = \sum_{\substack{a \in A \\ s \xrightarrow{a} s'}} x_a,$$

where the sum is over all $a \in A$ such that $s \xrightarrow{a} s'$ is an edge in the transition diagram. For the transition diagram in (1.1), the transition matrix is given by (ordering the states as $(\mathbf{1}, \mathbf{a}, \mathbf{b}, \mathbf{ab})$)

$$T = \begin{pmatrix} 0 & x_a & x_b & 0 \\ x_a & 0 & 0 & x_b \\ x_b & 0 & 0 & x_a \\ 0 & x_b & x_a & 0 \end{pmatrix}.$$

Note that, since every vertex has precisely one outgoing edge labeled $a \in A$ and since $\sum_{a \in A} x_a = 1$, the column sums of T are equal to one, namely $\sum_{s' \in \Omega} T_{s',s} = 1$. Starting with a distribution ν (where state $s \in \Omega$ occurs with probability ν_s and $\sum_{s \in \Omega} \nu_s = 1$), the distribution of states after t steps in the Markov chain is $T^t \nu$.

Two fundamental questions are to find the *stationary distribution* and the *mixing time* of the Markov chain. Intuitively speaking, the stationary distribution is the distribution of states that the Markov chain will tend to when the chain runs for a long time. For ergodic Markov chains (to be defined later), the stationary distribution is unique and is the right eigenvector of T of eigenvalue one. The mixing time measures how quickly the distribution approaches the stationary distribution.

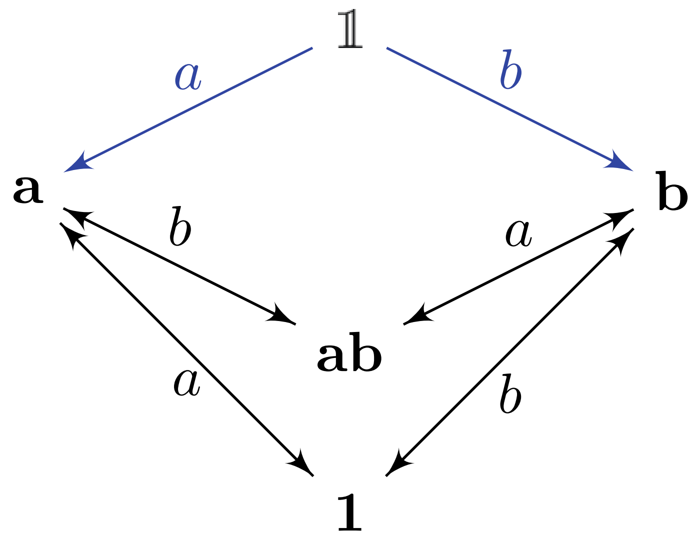


Fig. 1 The right Cayley graph $\mathsf{RCay}(D_2, \{a, b\})$ of the dihedral group with generators $A = \{a, b\}$. Transition edges are indicated in blue. Double edges mean that right multiplication by the label for either vertex yields the other vertex

In this paper, we review the approach of [11, 12] to compute the stationary distribution of a finite Markov chain using expansions of semigroups. For example, the Markov chain with transition diagram (1.1) can be described using the dihedral group

$$D_2 = \langle a, b \mid a^2 = b^2 = 1, (ab)^2 = 1\rangle,$$

generated by two reflections a, b. This group is isomorphic to $Z_2 \times Z_2$. State **1** corresponds to the identity, state **a** corresponds to a, state **b** to b, and state **ab** corresponds to $ab = ba$. The transition from state s to s' is given by left multiplication by one of the generators in $A = \{a, b\}$. In general, it is always possible to describe a finite state Markov chain via a semigroup [1, 8] by the *random letter representation*.

Our focus in this paper is to compute the stationary distribution from the McCammond and Karnofsky–Rhodes expansion of the right Cayley graph of the underlying semigroup S with generators in A. The right Cayley graph $\mathsf{RCay}(S, A)$ of the dihedral group $S = D_2$ with generators $A = \{a, b\}$ is depicted in Fig. 1. In Sect. 2, we will define the right Cayley graph of a semigroup, introduce its Karnofsky–Rhodes and McCammond expansion, and review the main results from [11, 12] on how to compute the stationary distribution of the Markov chain from these. We will illustrate the results in terms of two (running) examples.

2 Stationary Distribution from Semigroup Expansions

2.1 *Markov Chains*

Let $\mathcal{M}$ be a finite Markov chain with state-space Ω and transition matrix T. A Markov chain is *irreducible* if the transition diagram of the Markov chain is strongly connected. It is *aperiodic* if the greatest common divisor of the cycle lengths in the transition diagram of the Markov chain is one. Furthermore, a Markov chain is *ergodic* if it is both irreducible and aperiodic. By the Perron–Frobenius Theorem, an ergodic Markov chain has a unique stationary distribution Ψ and $T^t\nu$ converges to Ψ as $t \to \infty$ for any initial state ν. In fact, the *stationary distribution* is the right eigenvector of eigenvalue one of T

$$T\Psi = \Psi.$$

An important question is how quickly does the Markov chain converges to the stationary distribution. In Markov chain theory, distance is usually the total variation distance. The total variation distance between two probability distributions ν and μ is defined as

$$\|\nu - \mu\| = \max_{A \subseteq \Omega} |\nu(A) - \mu(A)|.$$

For a given small $\epsilon > 0$, the *mixing time* t_{mix} is the smallest t such that

$$\|T^t\nu - \Psi\| \leqslant \epsilon.$$

Brown and Diaconis [2, 3] analyzed Markov chains associated to left regular bands. In particular, they showed [3, Theorem 0] that the total variational distance from stationarity after t steps is bounded above by the probability $\mathsf{Pr}(\tau > t)$, where τ is the first time that the walk hits a certain ideal. The arguments in Brown and Diaconis [2] were generalized to Markov chains for $\mathscr{R}$-trivial semigroups [1]. A unified theory for Markov chains for any finite semigroup was developed in [11, 12].

As explained in [8, Proposition 1.5] and [1, Theorem 2.3], every finite state Markov chain $\mathcal{M}$ has a random letter representation, that is, a representation of a semigroup S acting on the left on the state-space Ω. In this setting, we transition $s \xrightarrow{a} s'$ with probability $0 \leqslant x_a \leqslant 1$, where $s, s' \in \Omega$, $a \in S$ and $s' = a.s$ is the action of a on the state s. It is enough to consider the semigroup S generated by the elements $a \in A$ with $x_a > 0$.

A two-sided *ideal* I (or ideal for short) is a subset $I \subseteq S$ such that $uIv \subseteq I$ for all $u, v \in S^{\mathbb{1}}$, where $S^{\mathbb{1}}$ is the semigroup S with the identity $\mathbb{1}$ added (even if S already contains an identity). If I, J are ideals of S, then $IJ \subseteq I \cap J$, so that $I \cap J \neq \emptyset$. Hence every finite semigroup has a unique *minimal ideal* denoted $K(S)$. An ideal $K(S)$ is *left zero* if $xy = x$ for all $x, y \in K(S)$.

We will determine the stationary distribution from certain expansions of the right Cayley graph $\mathsf{RCay}(S, A)$ of the underlying semigroup S with generators A. The Markov chain itself is a random walk on the minimal ideal $K(S)$ by the left action.

2.2 Right Cayley Graphs

We begin with the definition of a graph.

Definition 2.1 (*Graph*) A *labeled directed graph* Γ (or *graph* for short) consists of a vertex set $V(\Gamma)$, an edge set $E(\Gamma)$, and a labeling set A. An edge $e \in E(\Gamma)$ is a tuple $e = (v, a, w) \in V(\Gamma) \times A \times V(\Gamma)$. We often also write $e \colon v \xrightarrow{a} w$.

A *path* p from vertex v to vertex w in a graph Γ is a sequence of edges

$$p = \left(v = v_0 \xrightarrow{a_1} v_1 \xrightarrow{a_2} \cdots \xrightarrow{a_\ell} v_\ell = w \right),$$

where each tuple $(v_i, a_{i+1}, v_{i+1}) \in E(\Gamma)$ for $0 \leqslant i < \ell$. The initial (resp. terminal) vertex v (resp. w) of p is denoted by $\iota(p)$ (resp. $\tau(p)$). The *length* of p is $\ell(p) := \ell$ and $a_1 \dots a_\ell$ is called the label of the path.

We can define a preorder $\prec$ on $V(\Gamma)$ by $v \prec w$ if there is a path from v to w in Γ. This induces an equivalence relation $\sim$ on $V(\Gamma)$, where $v \sim w$ if $v \prec w$ and $w \prec v$. A *strongly connected component* of Γ is a $\sim$-equivalence class.

Definition 2.2 (*Rooted graph*) A *rooted graph* is a pair (Γ, r), where Γ is a graph and $r \in V(\Gamma)$, such that $r \prec v$ for all $v \in V(\Gamma)$.

A path is called *simple* if it visits no vertex twice. Empty (or trivial) paths are considered simple. For a rooted graph (Γ, r), let $\mathsf{Simple}(\Gamma, r)$ be the set of simple paths of Γ starting at r (including the empty path).

Definition 2.3 (*Right Cayley graph*) Let (S, A) be a finite semigroup S together with a set of generators A. The *right Cayley graph* $\mathsf{RCay}(S, A)$ of S with respect to A is the rooted graph with vertex set $V(\mathsf{RCay}(S, A)) = S^{\mathbb{1}}$, root $r = \mathbb{1} \in S^{\mathbb{1}}$, and edges $s \xrightarrow{a} s'$ for all $(s, a, s') \in S^{\mathbb{1}} \times A \times S^{\mathbb{1}}$, such that $s' = sa$ in $S^{\mathbb{1}}$.

An example of a right Cayley graph is given in Fig. 1.

For a semigroup S, two elements $s, s' \in S$ are in the same $\mathscr{R}$-class if the corresponding right ideals are equal, that is, $sS^{\mathbb{1}} = s'S^{\mathbb{1}}$. The strongly connected components of $\mathsf{RCay}(S, A)$ are precisely the $\mathscr{R}$-classes of $S^{\mathbb{1}}$. In other words, the vertices of a strongly connected component are exactly the vertices that represent the elements in an $\mathscr{R}$-class of $S^{\mathbb{1}}$. Edges that go between distinct strongly connected components will turn out to play an important role in the Karnofksy–Rhodes expansion.

Definition 2.4 (*Transition edges*) Let Γ be a graph. Then $e = (v, a, w) \in E(\Gamma)$ with $v, w \in V(\Gamma)$ and $a \in A$ is a *transition edge* if $v \nsim w$. In other words, there is no path from w to v in Γ.

In Fig. 1, the transition edges are indicated in blue. Note that the edges leaving $\mathbb{1}$ in the right Cayley graph are always transitional. Other edges might or might not be transitional. In this example $K(S)$ consists of all vertices in $\mathsf{RCay}(S, A)$ except the root $\mathbb{1}$.

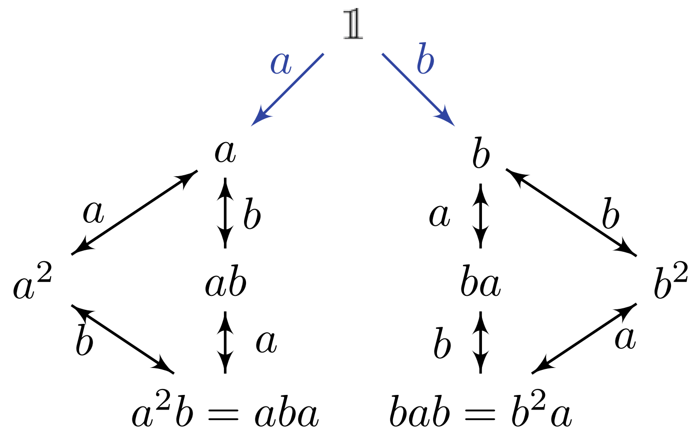


Fig. 2 The Karnofsky–Rhodes expansion $\mathsf{KR}(S, A)$ of the right Cayley graph of Fig. 1

2.3 *The Karnofsky–Rhodes Expansion*

To compute explicit expressions for the stationary distributions of Markov chains on finite semigroups, we need the *Karnofsky–Rhodes expansion* [7] of the right Cayley graph $\mathsf{RCay}(S, A)$. See also [9, Definition 4.15] and [10, Sect. 3.4].

Denote by (A^+, A), the free semigroup with generators in A. In other words, A^+ is the set of all words $a_1 \dots a_\ell$ of length $\ell \geqslant 1$ over A with multiplication given by concatenation. Furthermore, let $A^\star = A^+ \cup \{1\}$, so that $A^\star$ is A^+ with the identity added; it is the free monoid generated by A.

Definition 2.5 (*Karnofksy–Rhodes expansion*) The *Karnofsky–Rhodes expansion* $\mathsf{KR}(S, A)$ is obtained as follows. Start with the right Cayley graph $\mathsf{RCay}(A^+, A)$. Identify the endpoints of two paths in $\mathsf{RCay}(A^+, A)$

$$p := \left(\mathbb{1} \xrightarrow{a_1} v_1 \xrightarrow{a_2} \cdots \xrightarrow{a_\ell} v_\ell\right) \quad \text{and} \quad p' := \left(\mathbb{1} \xrightarrow{a'_1} v'_1 \xrightarrow{a'_2} \cdots \xrightarrow{a'_{\ell'}} v'_{\ell'}\right)$$

in $\mathsf{KR}(S, A)$ if and only if the corresponding paths in $\mathsf{RCay}(S, A)$

$$[p]_S := \left(\mathbb{1} \xrightarrow{a_1} [w_1]_S \xrightarrow{a_2} \cdots \xrightarrow{a_\ell} [w_\ell]_S\right) \quad \text{and} \quad [p']_S := \left(\mathbb{1} \xrightarrow{a'_1} [w'_1]_S \xrightarrow{a'_2} \cdots \xrightarrow{a'_{\ell'}} [w'_{\ell'}]_S\right),$$

where $w_i = a_1 a_2 \dots a_i$ and $w'_i = a'_1 a'_2 \dots a'_i$ end at the same vertex $[w_\ell]_S = [w'_{\ell'}]_S$ and in addition the set of transition edges of $[p]_S$ and $[p']_S$ in $\mathsf{RCay}(S, A)$ is equal.

An example for $\mathsf{KR}(S, A)$ is given in Fig. 2. In this figure, the paths a^2b and aba are equal because they end in the same vertex when projected onto S and they share the same transition edge, which is the first a. On the other hand, the paths ab and ba are distinct even though $ab = ba$ in D_2 because for the first path the transition edge is the first a and for the second path the transition edge is the first b.

Proposition 2.6 ([12, Proposition 2.15]) $\mathsf{KR}(S, A)$ *is the right Cayley graph of a semigroup, also denoted by* $\mathsf{KR}(S, A)$.

2.4 The McCammond Expansion

The McCammond expansion [9] of a rooted graph is intimately related to the unique simple path property.

Definition 2.7 (*Unique simple path property*) A rooted graph (Γ, r) has the *unique simple path property* if for each vertex $v \in V(\Gamma)$ there is a unique simple path from the root r to v.

As proven in [9, Proposition 2.32], the unique simple path property is equivalent to (Γ, r) admitting a unique directed spanning tree T. Note that the unique simple path property not only depends on the graph Γ, but also on the chosen root r. In this paper, we always choose $r = \mathbb{1}$. It was established in [9, Sect. 2.7] that every rooted graph (Γ, r) has a universal simple cover, which has the unique simple path property.

If p and q are paths, $\ell(q) = k \leqslant \ell(p)$, and the first $k+1$ vertices and k edges of p and q agree, we say that q is an *initial segment* of p, written $q \subseteq p$.

Definition 2.8 (*McCammond expansion*) For a rooted graph (Γ, r), define its *McCammond expansion* $(\Gamma^{\mathsf{Mc}}, r)$ as the graph with

$$\begin{aligned} V(\Gamma^{\mathsf{Mc}}) &= \mathsf{Simple}(\Gamma, r), \\ E(\Gamma^{\mathsf{Mc}}) &= \{(p, a, q) \in V(\Gamma^{\mathsf{Mc}}) \times A \times V(\Gamma^{\mathsf{Mc}}) \mid (\tau(p), a, \tau(q)) \in E(\Gamma), \\ &\qquad \ell(q) = \ell(p) + 1 \text{ or } (q \subseteq p \text{ and } \ell(q) \leqslant \ell(p))\}. \end{aligned}$$

Note that by definition there are two types of edges $(p, a, q) \in E(\Gamma^{\mathsf{Mc}})$: either $\ell(q) = \ell(p) + 1$ or $\ell(q) \leqslant \ell(p)$ as paths in $\mathsf{Simple}(\Gamma, r)$. The spanning tree T has vertex set $V(\Gamma^{\mathsf{Mc}})$ and only those edges $(p, a, q) \in E(\Gamma^{\mathsf{Mc}})$ such that $\ell(q) = \ell(p) + 1$.

From now on choose $r = \mathbb{1}$. The simple path

$$\mathbb{1} \xrightarrow{a_1} v_1 \xrightarrow{a_2} \cdots \xrightarrow{a_\ell} v_\ell$$

in $\mathsf{Simple}(\Gamma, \mathbb{1})$ is naturally indexed by the word $a_1 a_2 \ldots a_\ell$. We will use this labeling for the McCammond expansion of $\mathsf{KR}(S, A)$. In particular, if $a_1 a_2 \ldots a_\ell \in \mathsf{Simple}(\Gamma, \mathbb{1})$ and $a_1 a_2 \ldots a_\ell a \in \mathsf{Simple}(\Gamma, \mathbb{1})$, then the edge $a_1 a_2 \ldots a_\ell \xrightarrow{a} a_1 a_2 \ldots a_\ell a$ is in the spanning tree T. Otherwise, we have $a_1 a_2 \ldots a_\ell \xrightarrow{a} a_1 a_2 \ldots a_k$ for some unique $1 \leqslant k < \ell$. Thus under the right action of $a \in A$ on $a_1 a_2 \ldots a_\ell$, we either move forward in the spanning tree or fall backwards somewhere on the unique geodesic from $\mathbb{1}$ to $a_1 a_2 \ldots a_\ell$, but staying in the same $\mathscr{R}$-class. An example of a McCammond expansion of a Karnofsky–Rhodes graph is given in Fig. 3.

For a non-simple path in $(\Gamma^{\mathsf{Mc}}, \mathbb{1})$, we can remove loops; it does not matter in which order these loops are removed. This is also known as the Church–Rosser property [4] or a Knuth–Bendix rewriting system. This is proved in [9].

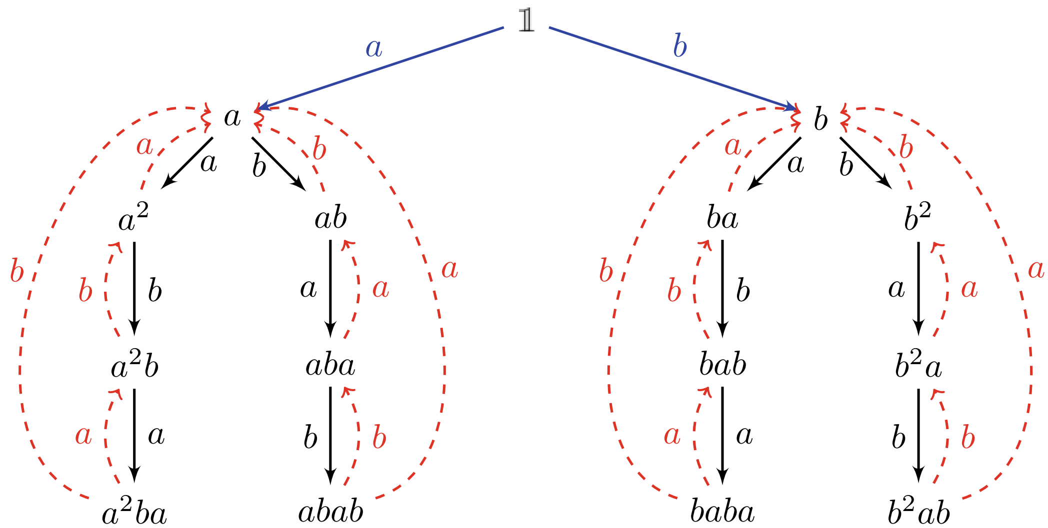


Fig. 3 The McCammond expansion $\mathsf{Mc} \circ \mathsf{KR}(D_2, \{a, b\})$ of Fig. 2

We denote the McCammond expansion of a semigroup (S, A) with generators in A by $\mathsf{Mc}(S, A)$, which is the McCammond expansion of its right Cayley graph.

2.5 Stationary Distribution

Denote by $\mathcal{M}(S, A)$, the Markov chain associated with the semigroup (S, A). As mentioned before, this is the random walk on $K(S)$ by the left action. The probability x_a is associated with generator $a \in A$. The stationary distribution for $\mathcal{M}(S, A)$ was computed in [12] using $\mathsf{Mc} \circ \mathsf{KR}(S, A)$. The treatment depends on whether the minimal ideal $K(S)$ is left zero or not. We first start with the former case, stated in [12, Corollaries 2.23 & 2.28].

Theorem 2.9 *If $K(S)$ is left zero, the stationary distribution of the Markov chain $\mathcal{M}(S, A)$ labeled by $w \in K(S)$ is given by*

$$\Psi_w^{\mathcal{M}(S,A)} = \sum_p \prod_{a \in p} x_a,$$

where the sum is over all paths p in $\mathsf{Mc} \circ \mathsf{KR}(S, A)$ starting at $\mathbb{1}$ and ending in s such that $[s]_S = w$.

The case when $K(S)$ is not left zero was treated in [12, Sect. 2.9] by adding a zero element $\Box$ to the semigroup S and the generators A. This new generator $\Box$ has its own probability $x_\Box$. The minimal ideal of the semigroup $(S \cup \{\Box\}, A \cup \{\Box\})$ is left zero and by taking the limit $x_\Box \to 0$, the stationary distribution of the original Markov chain (S, A) is retrieved as stated in [12, Corollary 2.33].

Theorem 2.10 *If $K(S)$ is not left zero, the stationary distribution of the Markov chain $\mathcal{M}(S, A)$ labeled by $w \in K(S)$ is given by*

$$\Psi_w^{\mathcal{M}(S,A)} = \lim_{x_\Box \to 0} \Big(\sum_p \prod_{a \in p} x_a \Big), \tag{2.1}$$

where the sum is over all paths p in $\mathsf{Mc} \circ \mathsf{KR}(S \cup \{\Box\}, A \cup \{\Box\})$ starting at $\mathbb{1}$ and ending in s such that $[s]_{S\cup\{\Box\}} = w\Box$.

In [11], we developed a strategy using *loop graphs* to compute the expressions in Theorems 2.9 and 2.10 as rational functions in the probabilities x_a for $a \in A$. This is done in two steps:

(1) Using McCammond's Pict map, we map the McCammond expansion together with a simple path from $\mathbb{1}$ to the ideal to a loop graph. See Sect. 2.6.
(2) The set of all paths from $\mathbb{1}$ to the element in the ideal in a loop graph can be written as a Kleene expression. The Kleene expression immediately yields a rational expression for the stationary distribution. See Sect. 2.7.

2.6 Loop Graphs

A *loop* of size ℓ is a connected directed graph with ℓ vertices such that each vertex has exactly one incoming and one outgoing edge. In other words, a loop is a directed circle of ℓ vertices. A *loop graph* can be defined recursively. Start with a directed straight line path. Recursively, attach a loop of an arbitrary finite size to any existing chosen vertex. Repeat or stop. The edges of the loop graph can be labeled. An example of a loop graph is given in Fig. 4.

Recall that an important property of the McCammond expansion is the unique simple path property (see Definition 2.7). We now define the map Pict from the set of tuples (Γ, p), where Γ is a graph with the unique simple path property and p is a simple path in Γ starting at $\mathbb{1}$, to the set of loop graphs. The straight line, which the loop graph is based on, corresponds to the chosen simple path p. We follow [11, Sect. 3.2].

Definition 2.11 (*McCammond* [11, Definition 3.5]) Let Γ be a graph with the unique simple path property and p a simple path in Γ starting at $\mathbb{1}$. Then $\mathsf{Pict}(\Gamma, p)$ is defined by the principle of induction.

Induction basis: Set $P = p$ and start at vertex $v_0 = \mathbb{1}$.

Induction step: Suppose one is at vertex $v_0 \neq \tau(p)$ on path p. Take the edge e from v_0 to v_1 in p.

(1) If there is no edge in Γ coming into v_1 besides e, continue with the unique next vertex in p, now denoted v_1 (with the current vertex v_1 relabeled v_0), unless $v_1 = \tau(p)$. If $v_1 = \tau(p)$, then output $\mathsf{Pict}(\Gamma, p) = P$.

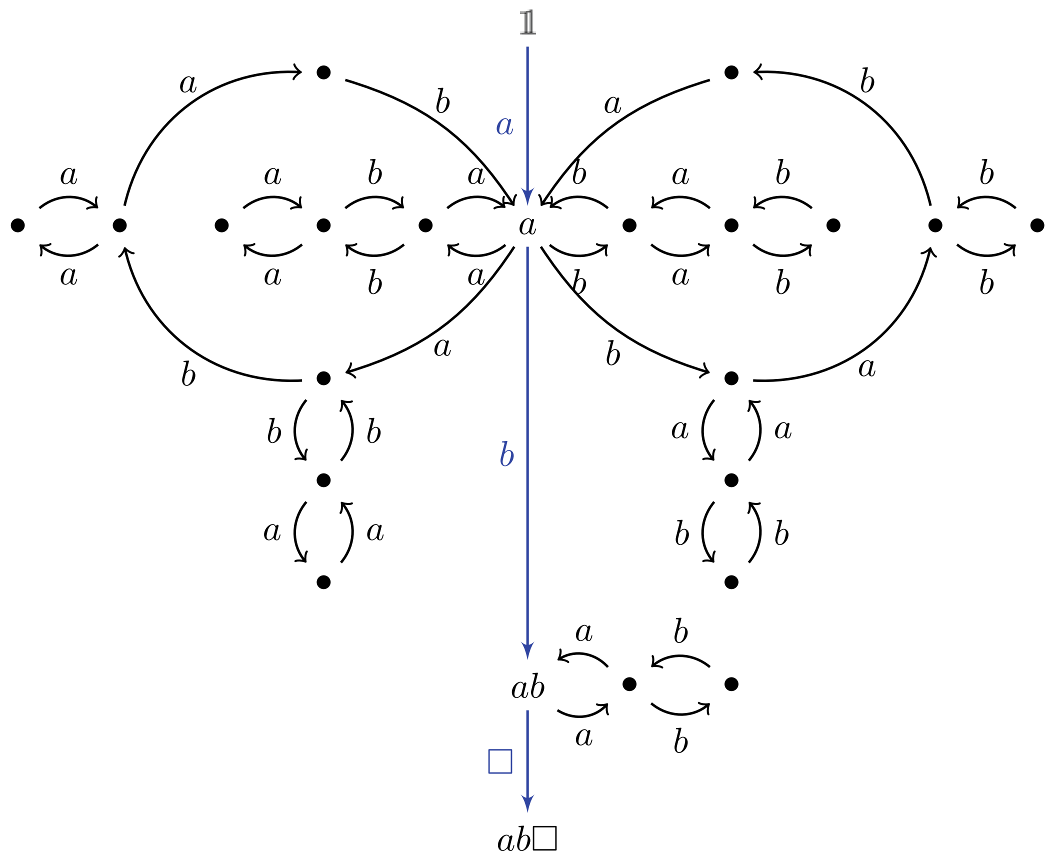


Fig. 4 Loop graph for $(\mathsf{Mc} \circ \mathsf{KR}(D_2 \cup \{\square\}, \{a, b, \square\}), ab\square)$

(2) Otherwise there is at least one edge $e' \neq e$ in Γ going into v_1, given by $e' = \left(v' \xrightarrow{a} v_1\right)$ for some $a \in A$. Since Γ has the unique simple path property by assumption, there must be a unique simple path starting at $\mathbb{1}$ going to v_0 along the path p followed by the path p' starting at v_0, going along e to v_1, and ending at v'.

  (a) Run the induction on p' in a subgraph Γ' of Γ, consisting of all edges and vertices on circuits containing a vertex of p'. Note that p' is simple in Γ'. The output is $P' = \mathsf{Pict}(\Gamma', p')$.
  (b) Modify P by attaching P' disjointly except at v_1 and adding edge e' from v' in P' back to v_1.

(3) Repeat step (2) for each edge $e' \neq e$ at vertex v_1.
(4) Continue with the induction step unless $v_1 = \tau(p)$. If $v_1 = \tau(p)$, then output $\mathsf{Pict}(\Gamma, p) = P$.

Remark 2.12 The map Pict has the property that the set of all paths in Γ from $\mathbb{1}$ to $\tau(p)$ is in bijection with the set of all paths in $\mathsf{Pict}(\Gamma, p)$ from $\mathbb{1}$ to $\tau(p)$ such that the labels of the paths are preserved.

Example 2.13 Let us compute the example $\mathsf{Pict}(\Gamma, ab\square)$, where $\Gamma = \mathsf{Mc} \circ \mathsf{KR}(D_2 \cup \{\square\}, \{a, b, \square\})$. The McCammond expansion is given in Fig. 3 if we attach to each vertex an edge labeled $\square$ to the ideal $\{\square\}$. The straight path is $ab\square$. The vertex labeled a in Fig. 3 has four dashed edges coming in. Two are labeled a and two are labeled b. By the algorithm described in Definition 2.11, the long dashed arrows labeled a and b give rise to loops of length 4. The other two dashed arrows give rise to loops of length 2. Repeating the process yields the loop graph in Fig. 4.

2.7 Kleene Expressions

Denote the set of all paths in a loop graph G starting at $\mathbb{1}$ and ending at $\tau(p)$, where p is the straight line the loop graph is based on, by $\mathcal{P}_G$. We represent a path $q \in \mathcal{P}_G$ by

$$\mathbb{1} \xrightarrow{a_1} v_1 \xrightarrow{a_2} \cdots \xrightarrow{a_k} v_k = \tau(p),$$

where v_i are vertices in G and $a_i \in A$ are the labels on the edges.

There is a simple inductive way to describe $\mathcal{P}_G$ using *Kleene expressions* (see [11, Sect. 1.3]). Given a set L, define $L^0 = \{\varepsilon\}$ given by the empty string, $L^1 = L$, and recursively $L^{i+1} = \{wa \mid w \in L^i, a \in L\}$ for each integer $i > 0$. Then the *Kleene star* is

$$L^\star = \bigcup_{i \geqslant 0} L^i.$$

A Kleene expression only involves letters in A, concatenation, unions, and $\star$. To obtain a Kleene expression for $\mathcal{P}_G$, perform the following doubly recursive procedure:

Algorithm 1. Assume that the straight line path corresponding to the loop graph G is indexed as

$$\mathbb{1} \longrightarrow 1 \longrightarrow 2 \longrightarrow \quad \cdots \quad \longrightarrow \tau(p)$$

Induction basis: Start at vertex $\mathbb{1}$ and with the empty expression E.

Induction step: Suppose one is at vertex $i \neq \tau(p)$ (or $\mathbb{1}$) on the straight line path underlying G.

(1) Continue to the next vertex $i + 1$ (or 1) on the straight line path underlying G and append the label a on the edge from $i \xrightarrow{a} i + 1$ (or $\mathbb{1} \xrightarrow{a} 1$) to E.

(2) If there are loops $\ell_1, \ell_2, \ldots, \ell_k$ at vertex $i + 1$ (or 1), append the formal expression

$$\{\ell_1, \ell_2, \ldots, \ell_k\}^\star$$

to E. The loops $\ell_1, \ell_2, \ldots, \ell_k$ are in one-to-one correspondence with the edges coming into vertex $i + 1$.

(3) If $i + 1 \neq \tau(p)$, continue with the next induction step. Else stop and output E.

Algorithm 2. For each symbol ℓ_i in the expression for E, do the following:

(1) Consider the loop $\ell_i = \left(v_0 \xrightarrow{a_1} v_1 \xrightarrow{a_2} \cdots \xrightarrow{a_k} v_k = v_0\right)$ from vertex v_0 to v_0 in G. Consider the subgraph of G with straight line $v_1 \xrightarrow{a_2} \cdots \xrightarrow{a_k} v_k$ and all further loops that are attached to any of the vertices v_i in G. Attach $\mathbb{1}$ to v_1. The resulting graph $G^{(i)}$ is a new loop graph. Perform Algorithm 1 on $G^{(i)}$ to obtain a Kleene expression $E^{(i)}$. Replace the symbol ℓ_i in E by $E^{(i)}$.

(2) Continue this process until E does not contain any further expressions ℓ_i for some loop ℓ_i, that is, E only contains unions, $\star$ and elements in the alphabet A. Then the Kleene expression for $\mathcal{P}_G$ is E.

The resulting expressions can be made into unionless expressions by using *Zimin words*

$$\{a\}^\star = a^\star \qquad \text{and} \qquad \{a,b\}^\star = (a^\star b)^\star a^\star \qquad \text{for a,b} \in \text{A}. \tag{2.2}$$

Expressions for larger unions can be obtained by induction using (2.2).

Example 2.14 Let us continue Example 2.13 and compute the Kleene expression for $\mathcal{P}_{\mathsf{Pict}(\Gamma, ab\square)}$ for $\Gamma = \mathsf{Mc} \circ \mathsf{KR}(D_2 \cup \{\square\}, \{a, b, \square\})$. By Algorithm 1, we obtain the expression

$$E = a\{\ell_1, \ell_2, \ell_3, \ell_4\}^\star b \ell_5^\star \square.$$

Using Algorithm 2 repeatedly for $\ell_1, \ldots, \ell_5$, we obtain

$$\begin{aligned}
\ell_1 &= a(b(aa)^\star b)^\star b(aa)^\star ab,\\
\ell_2 &= a(b(aa)^\star b)^\star a,\\
\ell_3 &= b(a(bb)^\star a)^\star a(bb)^\star ba,\\
\ell_4 &= b(a(bb)^\star a)^\star b,\\
\ell_5 &= a(bb)^\star a.
\end{aligned}$$

2.8 *From Kleene Expressions to Rational Functions*

Our aim is to evaluate the expressions for $\Psi_w^{\mathcal{M}(S,A)}$ in Theorems 2.9 and 2.10. Let G be a loop graph with straight line path p. Define

$$\Psi_G(x_1, \ldots, x_n) = \sum_q \prod_{a \in q} x_a, \tag{2.3}$$

where the sum is over all paths q from $\mathbb{1}$ to $\tau(p)$ in G. In [11, Definition 1.3] this is also called the *normal distribution* of the loop graph G. Note that $\Psi_w^{\mathcal{M}(S,A)}$ is the sum of $\Psi_G(x_1, \ldots, x_n)$ for various loop graphs with straight line paths p such that $\tau(p) = w$ (see also [11, Theorem 1.4]). The Kleene expressions from Sect. 2.7 give us an expression for the set of relevant paths q in G. Now we discuss how to get from the Kleene expressions to rational functions.

The main idea is that concatenation in Kleene expressions corresponds to products, unions corresponds to sums, and $\star$ corresponds to the geometric series. More concretely, for a path $p = a_1 \cdots a_k$ we obtain

$$\prod_{a \in p} x_a = x_{a_1} x_{a_2} \cdots x_{a_k}.$$

For $\star$-expressions with a single letter a, we obtain

$$\sum_{s \in a^\star} \prod_{i \in s} x_i = \sum_{\ell=0}^{\infty} x_a^\ell = \frac{1}{1 - x_a}.$$

Similarly

$$\sum_{s \in \{a,b\}^\star} \prod_{i \in s} x_i = \sum_{s \in a^\star (b a^\star)^\star} \prod_{i \in s} x_i = \frac{1}{1 - x_a} \cdot \frac{1}{1 - \frac{x_b}{1 - x_a}} = \frac{1}{1 - x_a - x_b}.$$

In general, using the recursion (2.2) we derive by induction

$$\sum_{s \in \{a_1, a_2, \ldots, a_n\}^\star} \prod_{i \in s} x_i = \frac{1}{1 - x_{a_1} - x_{a_2} - \cdots - x_{a_n}}. \tag{2.4}$$

Example 2.15 Let us now compute

$$\Psi_{ab\square} = \sum_{p \in E} \prod_{a \in p} x_a$$

for the Kleene expression E of Example 2.14. We find (see also [11, Example 3.8])

$$\begin{aligned}
\Psi_{ab\square} &= \frac{x_a x_b x_\square}{\left(1 - \frac{x_a^2 x_b^2}{\left(1 - \frac{x_b^2}{1 - x_a^2}\right)(1 - x_a^2)} - \frac{x_a^2}{1 - \frac{x_b^2}{1 - x_a^2}} - \frac{x_a^2 x_b^2}{\left(1 - \frac{x_a^2}{1 - x_b^2}\right)(1 - x_b^2)} - \frac{x_b^2}{1 - \frac{x_a^2}{1 - x_b^2}}\right)\left(1 - \frac{x_a^2}{1 - x_b^2}\right)} \\
&= \frac{x_a x_b x_\square (1 - x_b^2)}{\left(1 - \frac{2 x_a^2 x_b^2}{1 - x_a^2 - x_b^2} - \frac{x_a^2 (1 - x_a^2)}{1 - x_a^2 - x_b^2} - \frac{x_b^2 (1 - x_b^2)}{1 - x_a^2 - x_b^2}\right)(1 - x_a^2 - x_b^2)} \\
&= \frac{x_a x_b x_\square (1 - x_b^2)}{1 - 2x_a^2 - 2x_b^2 + (x_a^2 - x_b^2)^2}.
\end{aligned}$$

In the limit as $x_\square \to 0$, we obtain

$$\lim_{x_\square \to 0} \Psi_{ab\square} = \frac{1-x_b^2}{8}.$$

In [11, Example 3.8], the remaining stationary distributions were computed using that $x_a + x_b + x_\square = 1$ and by taking the limit $x_\square \to 0$

$$\begin{aligned}
\Psi_\square &= x_\square & &\xrightarrow{x_\square \to 0} & & 0\\
\Psi_{a\square} &= \frac{x_a(1-x_a^2-x_b^2)x_\square}{1-2x_a^2-2x_b^2+(x_a^2-x_b^2)^2} & &\xrightarrow{x_\square \to 0} & & \frac{x_a}{4}\\
\Psi_{ab\square} &= \frac{x_a x_b x_\square (1-x_b^2)}{1-2x_a^2-2x_b^2+(x_a^2-x_b^2)^2} & &\xrightarrow{x_\square \to 0} & & \frac{1-x_b^2}{8}\\
\Psi_{aba\square} &= \frac{x_a^2 x_b x_\square}{1-2x_a^2-2x_b^2+(x_a^2-x_b^2)^2} & &\xrightarrow{x_\square \to 0} & & \frac{x_a}{8}\\
\Psi_{abab\square} &= \frac{x_a^2 x_b^2 x_\square}{1-2x_a^2-2x_b^2+(x_a^2-x_b^2)^2} & &\xrightarrow{x_\square \to 0} & & \frac{x_a x_b}{8}\\
\Psi_{a^2\square} &= \frac{x_a^2(1-x_a^2)x_\square}{1-2x_a^2-2x_b^2+(x_a^2-x_b^2)^2} & &\xrightarrow{x_\square \to 0} & & \frac{x_a(1+x_a)}{8}\\
\Psi_{a^2b\square} &= \frac{x_a^2 x_b x_\square}{1-2x_a^2-2x_b^2+(x_a^2-x_b^2)^2} & &\xrightarrow{x_\square \to 0} & & \frac{x_a}{8}\\
\Psi_{a^2ba\square} &= \frac{x_a^3 x_b x_\square}{1-2x_a^2-2x_b^2+(x_a^2-x_b^2)^2} & &\xrightarrow{x_\square \to 0} & & \frac{x_a^2}{8}
\end{aligned}$$

and similarly for the cases with a and b interchanged by symmetry.

For a word w in $\{a,b\}$, denote by $[w]$ the corresponding element in D_2. For example, in D_2 we have $[a]=[bab]=[b^2a]$. Note that by Theorem 2.10

$$\Psi_s^{\mathcal{M}(D_2,\{a,b\})} = \frac{1}{4} \qquad \text{for all s} \in \mathrm{D}_2$$

by summing the appropriate results for $\lim_{x_\square\to 0}\Psi_{w\square}$ as above. For example,

$$\Psi_{[ab]}^{\mathcal{M}(D_2,\{a,b\})} = \lim_{x_\square \to 0}\left(\Psi_{ab\square}+\Psi_{ba\square}+\Psi_{a^2ba\square}+\Psi_{b^2ab\square}\right) = \frac{1-x_b^2}{8}+\frac{1-x_a^2}{8}+\frac{x_a^2}{8}+\frac{x_b^2}{8}=\frac{1}{4}.$$

This shows that the stationary distribution is uniform.

Remark 2.16 Note that the Markov chain in (1.1) is not ergodic since the greatest common divisor of the cycle length is 2 and not 1. We can make the Markov chain ergodic by introducing a new generator c, which acts as the identity. In other words, this would introduce loops at each vertex in (1.1) labeled c. In turn, this would introduce loops labeled c at each vertex in the McCammond expansion in Fig. 3 and the loop graph in Fig. 4. This would change the Kleene expressions in Example 2.14 to

$$E = a\{\ell_1, \ell_2, \ell_3, \ell_4, c\}^\star b\{\ell_5, c\}^\star \square$$

with

$$\begin{aligned}
\ell_1 &= a\{b\{ac^\star a, c\}^\star b, c\}^\star b\{ac^\star a, c\}^\star ac^\star b,\\
\ell_2 &= a\{b\{ac^\star a, c\}^\star b, c\}^\star a,\\
\ell_3 &= b\{a\{bc^\star b, c\}^\star a, c\}^\star a\{bc^\star b, c\}^\star bc^\star a,\\
\ell_4 &= b\{a\{bc^\star b, c\}^\star a, c\}^\star b,\\
\ell_5 &= a\{bc^\star b, c\}^\star a.
\end{aligned}$$

In this setting, we find

$$\begin{aligned}
\Psi_{ab\square} &= \frac{x_a x_b x_\square}{\left(1 - \frac{x_a^2 x_b^2}{\left(1-\frac{x_b^2}{1-\frac{x_a^2}{1-x_c}-x_c}-x_c\right)\left(1-\frac{x_a^2}{1-x_c}-x_c\right)(1-x_c)} - \frac{x_a^2}{1-\frac{x_b^2}{1-\frac{x_a^2}{1-x_c}-x_c}-x_c} - \frac{x_a^2 x_b^2}{\left(1-\frac{x_a^2}{1-\frac{x_b^2}{1-x_c}-x_c}-x_c\right)\left(1-\frac{x_b^2}{1-x_c}-x_c\right)(1-x_c)} - \frac{x_b^2}{1-\frac{x_a^2}{1-\frac{x_b^2}{1-x_c}-x_c}-x_c} - x_c\right)\left(1 - \frac{x_a^2}{1-\frac{x_b^2}{1-x_c}-x_c} - x_c\right)}\\
&= \frac{x_a x_b x_\square (1 + x_b - x_c)(1 - x_b - x_c)}{(1 - x_a - x_b - x_c)(1 + x_a + x_b - x_c)(1 - x_a + x_b - x_c)(1 + x_a - x_b - x_c)}.
\end{aligned}$$

Using $x_a + x_b + x_c + x_\square = 1$, we see that the term $(1 - x_a - x_b - x_c)$ in the denominator cancels with $x_\square$ in the numerator. Hence in the limit $x_\square \to 0$, we obtain

$$\lim_{x_\square \to 0} \Psi_{ab\square} = \frac{(1 + x_b - x_c)(1 - x_b - x_c)}{8(x_a + x_b)}.$$

As $x_c \to 0$, we recover the result from Example 2.15.

2.9 Mixing Time

Recall from Sect. 2.1, that for a given small $\epsilon > 0$, the *mixing time* t_{mix} is the smallest t such that

$$\|T^t \nu - \Psi\| \leqslant \epsilon.$$

Many references about mixing time can be found in [5, 8].

Let τ be the first time that the Markov chain hits the ideal (when starting at $\mathbb{1}$ in $\mathsf{RCay}(S, A)$ or $\mathsf{Mc} \circ \mathsf{KR}(S, A)$). Denote by $\mathsf{Pr}(\tau > t)$ the probability that τ is bigger than a given t. In [1], it was shown that $\mathsf{Pr}(\tau > t)$ gives a bound on the mixing time.

Theorem 2.17 ([1]) *Let S be a finite semigroup whose minimal ideal $K(S)$ is a left zero semigroup and let T be the transition matrix of the associated Markov chain. Then*

$$\|T^t \nu - \Psi\| \leqslant \mathsf{Pr}(\tau > t).$$

In [13], we provide a way to compute $\mathsf{Pr}(\tau > t)$ from particular rational expressions for the stationary distribution. Let G be a loop graph and recall $\Psi_G(x_1, \dots, x_n)$ from (2.3), which is a rational function in $x_1, \dots, x_n$. Let $\mathsf{Pr}_G(\tau \geqslant t)$ be the probability that the length of the paths in the loop graph G from $\mathbb{1}$ to s in the ideal is weakly bigger than t. Let $\Psi_G^{\geqslant t}(x_1, \dots, x_n)$ be the truncation of the formal power series associated to the rational function $\Psi_G(x_1, \dots, x_n)$ to terms of degree weakly bigger than t and let $\Psi_G^{<t}(x_1, \dots, x_n)$ be the truncation of the formal power series associated to the rational function $\Psi_G(x_1, \dots, x_n)$ to terms of degree strictly smaller than t. Note that

$$\Psi_G(x_1, \dots, x_n) = \Psi_G^{<t}(x_1, \dots, x_n) + \Psi_G^{\geqslant t}(x_1, \dots, x_n).$$

Theorem 2.18 ([13]) *Suppose the Markov chain satisfies the conditions of Theorem 2.17. If $\Psi_G(x_1, \dots, x_n)$ is represented by a rational function such that each term of degree ℓ in its formal power sum expansion corresponds to a path in G of length ℓ, we have*

$$\mathsf{Pr}_G(\tau \geqslant t) = \frac{\Psi_G^{\geqslant t}(x_1, \dots, x_n)}{\Psi_G(x_1, \dots, x_n)} = 1 - \frac{\Psi_G^{<t}(x_1, \dots, x_n)}{\Psi_G(x_1, \dots, x_n)}.$$

By Markov's inequality (see for example [6, 8]), we have

$$\mathsf{Pr}(\tau > t) \leqslant \frac{E[\tau]}{t+1}, \tag{2.5}$$

where $E[\tau]$ is the expected value for τ, the first time the walk hits the ideal. Hence knowing $E[\tau]$ gives an upper bound on the mixing time. In [13], we find a way to compute $E[\tau]$ from certain representations of the stationary distribution.

Theorem 2.19 ([13]) *Suppose the Markov chain satisfies the conditions of Theorem 2.17. If $\Psi_G(x_1, \dots, x_n)$ is represented by a rational function such that each term of degree ℓ in its formal power sum expansion corresponds to a path in G of length ℓ, we have*

$$E_G[\tau] = \left(\sum_{i=1}^n x_i \frac{\partial}{\partial x_i} \right) \ln \Psi_G(x_1, \dots, x_n). \tag{2.6}$$

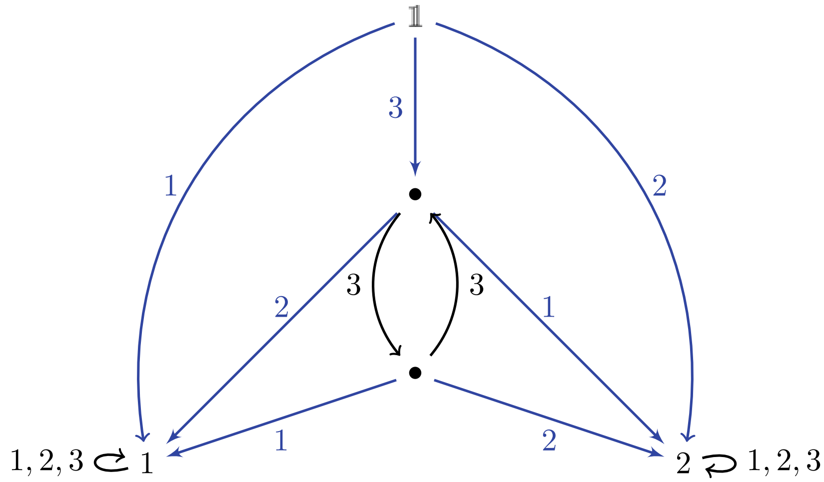
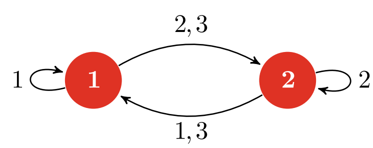


Fig. 5 The right Cayley graph $\mathsf{RCay}(S, A)$ of the semigroup that gives the Markov chain in Sect. 2.10

2.10 Another Example

Let us illustrate the concepts and algorithms in terms of another example. Consider the Markov chain with state-space $\Omega = \{\mathbf{1}, \mathbf{2}\}$ given by the transition diagram:



(2.7)

The transition matrix of this Markov chain is given by

$$T = \begin{pmatrix} x_1 & x_1 + x_3 \\ x_2 + x_3 & x_2 \end{pmatrix}.$$

Pick as generators of the semigroup $A = \{1, 2, 3\}$. Then the right Cayley graph of the semigroup that gives the above Markov chain by left multiplication is depicted in Fig. 5, where $K(S) = \{1, 2\}$. Indeed, the left action $11 = 1$, $21 = 2$, and $31 = 2$, which gives all the edges out of **1** in (2.7). Similarly, the left action $12 = 1$, $22 = 2$, and $32 = 1$, which gives all the edges out of **2** in (2.7).

The McCammond and Karnofsky–Rhodes expansion of the right Cayley graph is given in Fig. 6.

The Kleene expression for all paths from $\mathbb{1}$ to $32 \in K(S)$ is $3(33)^{\star}2$ and similarly for paths with other endpoints. From this we easily compute

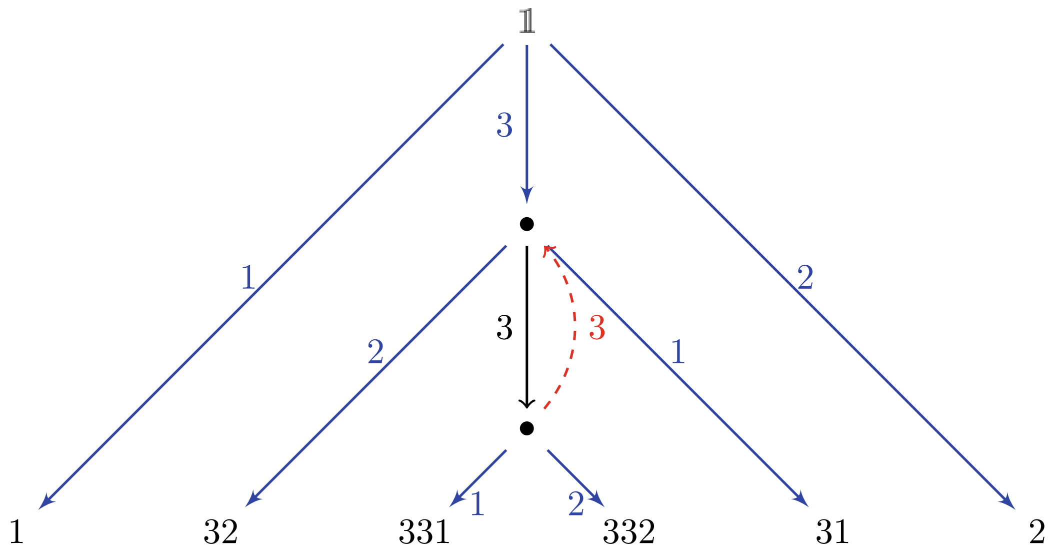


Fig. 6 $\mathsf{Mc} \circ \mathsf{KR}(S, A)$ of $\mathsf{RCay}(S, A)$ in Fig. 5 with loops on the ideal omitted

$$\Psi_1 = x_1 \qquad \Psi_2 = x_2,$$
$$\Psi_{32} = \frac{x_2 x_3}{1 - x_3^2} \qquad \Psi_{31} = \frac{x_1 x_3}{1 - x_3^2},$$
$$\Psi_{331} = \frac{x_1 x_3^2}{1 - x_3^2} \qquad \Psi_{332} = \frac{x_2 x_3^2}{1 - x_3^2}.$$

Furthermore,

$$\Psi_1^{\mathcal{M}(S,A)} = \Psi_1 + \Psi_{32} + \Psi_{331} = \frac{x_1 + x_2 x_3}{1 - x_3^2},$$
$$\Psi_2^{\mathcal{M}(S,A)} = \Psi_2 + \Psi_{31} + \Psi_{332} = \frac{x_2 + x_1 x_3}{1 - x_3^2}.$$

Using that $x_1 + x_2 + x_3 = 1$, we find indeed that $\Psi_1^{\mathcal{M}(S,A)} + \Psi_2^{\mathcal{M}(S,A)} = 1$.

Since the expressions for $\Psi_1^{\mathcal{M}(S,A)}$ and $\Psi_2^{\mathcal{M}(S,A)}$ were computed directly from $\mathsf{Mc} \circ \mathsf{KR}(S, A)$ (or the corresponding loop graphs) without using that $x_1 + x_2 + x_3 = 1$, each term of degree ℓ in the expansion of the rational function corresponds to a path in the graph. Hence we may use Theorem 2.19 to give an upper bound on the mixing time

$$E_1[\tau] = \left(x_1 \frac{\partial}{\partial x_1} + x_2 \frac{\partial}{\partial x_2} + x_3 \frac{\partial}{\partial x_3} \right) \ln \Psi_1^{\mathcal{M}(S,A)} = \frac{x_1}{x_1 + x_2 x_3} + \frac{2 x_2 x_3}{x_1 + x_2 x_3} + \frac{2 x_3^2}{1 - x_3^2}.$$

Inserting $x_1 = x_2 = x_3 = \frac{1}{3}$ yields $E_1[\tau] = E_2[\tau] = \frac{3}{2}$, so that $t_{\mathsf{mix}} \leqslant 3$ if $\epsilon = \frac{1}{2}$.

Acknowledgements The last author was partially supported by NSF grants DMS–1760329 and DMS–1764153. This material is based upon work supported by the Swedish Research Council under grant no. 2016-06596 while the last author was in residence at Institute Mittag–Leffler in

Djursholm, Sweden during Spring 2020. The authors thank the organizers of the International conference on semigroups and applications held at Cochin University of Science and Technology in India December 9–12, 2019, where this work was presented. We would like to thank Igor Pak for discussions.

References

1. Ayyer, A., Schilling, A., Steinberg, B., Thiéry, N.M.: Markov chains, $\mathscr{R}$-trivial monoids and representation theory. Internat. J. Algebra Comput. **25**(1–2), 169–231 (2015)
2. Brown, K.S., Diaconis, P.: Random walks and hyperplane arrangements. Ann. Probab. **26**(4), 1813–1854 (1998)
3. Brown, K.S.: Semigroups, rings, and Markov chains. J. Theoret. Probab. **13**(3), 871–938 (2000)
4. Church, A., Rosser, J.B.: Some properties of conversion. Trans. Amer. Math. Soc. **39**(3), 472–482 (1936)
5. Diaconis, P.: The mathematics of mixing things up. J. Stat. Phys. **144**(3), 445–458 (2011)
6. Devroye, L., Lugosi, G.: Combinatorial Methods in Density Estimation. Springer Series in Statistics. Springer, New York (2001)
7. Elston, G.Z.: Semigroup expansions using the derived category, kernel, and Malcev products. J. Pure Appl. Algebra **136**(3), 231–265 (1999)
8. Levin, D.A., Peres, Y.: Markov chains and mixing times. American Mathematical Society, Providence (2017). Second edition of [MR2466937], With contributions by Elizabeth L. Wilmer, With a chapter on “Coupling from the past” by James G. Propp and David B. Wilson
9. McCammond, J., Rhodes, J., Steinberg, B.: Geometric semigroup theory (2011). arXiv:1104.2301
10. Margolis, S., Saliola, F., Steinberg, B.: Combinatorial topology and the global dimension of algebras arising in combinatorics. J. Eur. Math. Soc. (JEMS) **17**(12), 3037–3080 (2015)
11. Rhodes, J., Schilling, A.: Normal distributions of finite Markov chains. Internat. J. Algebra Comput. **29**(8), 1431–1449 (2019)
12. Rhodes, J., Schilling, A.: Unified theory for finite Markov chains. Adv. Math. **347**, 739–779 (2019)
13. Rhodes, J., Schilling, A.: Bounds on mixing time of Markov chains via loop graphs (2020). Preprint arXiv:2010.08879