Strong stationary duality for Möbius monotone Markov chains

Paweł Lorek · Ryszard Szekli

Abstract For Markov chains with a finite, partially ordered state space, we show strong stationary duality under the condition of Möbius monotonicity of the chain. We give examples of dual chains in this context which have no downwards transitions. We illustrate general theory by an analysis of nonsymmetric random walks on the cube with an interpretation for unreliable networks of queues.

1 Introduction

The motivation of this paper stems from a study on the speed of convergence to stationarity for unreliable queueing networks, as in Lorek and Szekli [13]. The problem of bounding the speed of convergence for networks is a rather complex one, and is related to transient analysis of Markov processes, spectral analysis, coupling or duality constructions, drift properties, monotonicity properties, among others (see for more details Dieker and Warren [9], Aldous [1], Lorek and Szekli [13]). In order to give bounds on the speed of convergence for some unreliable queueing networks, it is necessary to study the availability vector of unreliable network processes. This vector is a Markov chain with the state space representing sets of stations with *down* or *up* status via the power set of the set of nodes (typical state is the set of broken nodes).

P. Lorek · R. Szekli (✉)
Mathematical Institute, University of Wrocław, pl. Grunwaldzki 2/4, 50-384 Wrocław, Poland
e-mail: Ryszard.Szekli@math.uni.wroc.pl

P. Lorek
e-mail: Pawel.Lorek@math.uni.wroc.pl

Such a chain represents at the same time a random walk on the vertices of the finite dimensional cube. We are concerned in this paper with walks on the vertices of the finite dimensional cube which are up–down in the natural (inclusion) ordering on the power set. This study is a special case of a general duality construction for monotone Markov chains.

To be more precise, we shall study strong stationary duality (SSD) which is a probabilistic approach to the problem of speed of convergence to stationarity for Markov chains. SSD was introduced by Diaconis and Fill [7]. This approach involves strong stationary times (SST) introduced earlier by Aldous and Diaconis [2, 3] who gave a number of examples showing useful bounds on the total variation distance for convergence to stationarity in cases where other techniques utilizing eigenvalues or coupling were not easily applicable. A *strong stationary time* for a Markov chain (X_n) is a stopping time T for this chain for which X_T has the stationary distribution π and is independent of T. Diaconis and Fill [7] constructed an absorbing *dual* Markov chain with its absorption time equal to the strong stationary time T for (X_n). In general, there is no recipe for constructing particular dual chains. However, a few cases are known and tractable. One of the most basic and interesting ones is given by Diaconis and Fill [7] (Theorem 4.6) when the state space is linearly ordered. In this case, under the assumption of stochastic monotonicity for the time reversed chain, and under the condition that for the initial distribution ν, $\nu \leq_{mlr} \pi$ (that is, for any $k_1 > k_2$, $\frac{\nu(k_1)}{\pi(k_1)} \leq \frac{\nu(k_2)}{\pi(k_2)}$) it is possible to construct a dual chain on the same state space. A special case is a stochastically monotone birth-and-death process for which the strong stationary time has the same distribution as the time to absorption in the dual chain, which turns out to be again a birth-and-death process on the same state space. Times to absorption are usually more tractable objects in a direct analysis than times to stationarity. In particular, a well-known theorem, usually attributed to Keilson, states that, for an irreducible continuous-time birth-and-death chain on $\mathbb{E} = \{0, \ldots, M\}$, the passage time from state 0 to state M is distributed as a sum of M independent exponential random variables. Fill [11] uses the theory of strong stationary duality to give a stochastic proof of an analogous result for discrete-time birth-and-death chains and geometric random variables. He shows a link for the parameters of the distributions to eigenvalue information about the chain. The obtained dual is a pure birth chain. Similar structure holds for more general chains. An (upward) skip-free Markov chain with the set of nonnegative integers as a state space is a chain for which upward jumps may be only of unit size; there is no restriction on downward jumps. Brown and Shao [5] determined, for an irreducible continuous-time skip-free chain and any M, the passage time distribution from state 0 to state M. When the eigenvalues of the generator are all real, their result states that the passage time is distributed as the sum of M independent exponential random variables with rates equal to the eigenvalues. Fill [12] gives another proof of this theorem. In the case of birth-and-death chains, this proof leads to an explicit representation of the passage time as a sum of independent exponential random variables. Diaconis and Miclo [8] recently obtained such a representation, using an involved duality construction; for some recent references related to duality and stationarity, see this paper.

Our main result is an SSD construction which generalizes the above mentioned construction of Diaconis and Fill [7]. We consider a partially ordered state space

instead of a linearly ordered one and utilize Möbius monotonicity instead of the usual stochastic monotonicity. This construction opens new ways to study particular Markov chains by a dual approach and is of independent interest. It has a special feature that the dual state space is again the same state space as for the original chain, similarly as in SSD for birth-and-death processes. Moreover, we show that the dual chain can have an upwards drift in the sense that it has no downwards transitions. We formulate the main result in Sect. 3, explaining the needed notation and definitions in detail in Sect. 2. We elaborate on the topic of Möbius monotonicity because it is almost not present in the literature. The only papers we are aware of are the following two: Massey [14] recalls Möbius monotonicity as considered earlier by Adrianus Kester in his PhD thesis, and proves that Möbius monotonicity implies a weak stochastic monotonicity. The second paper is by Falin [10] where a similar result to the one by Massey can be found. We introduce two versions of Möbius monotonicity, and we define a new notion of Möbius monotone functions which appear in a natural way in our main result on SSD. We characterize Möbius monotonicity by an invariance property on the set of Möbius monotone functions. Utilization of Möbius monotonicity involves a general problem of inverting a sum ranging over a partially ordered set, which appears in many combinatorial contexts; see, for example, Rota [15]. The inversion can be carried out by defining an analog of the difference operator relative to a given partial ordering. Such an operator is the Möbius function, and the analog of the fundamental theorem of calculus obtained in this context is the Möbius inversion formula on a partially ordered set, which we recall in Sect. 2.

In Sect. 3 we present our main result on SSD with a proof and give some corollaries which show other possible duals, including an alternative dual for linearly ordered state spaces (Corollary 3.2). In Sect. 4 we show an SSD result for nonsymmetric nearest neighbor walks on the finite dimensional cube. It gives an additional insight into the structure of eigenvalues of this chain. It is interesting that the dual (absorbing) chain here is a chain which jumps only upwards to neighboring states or stays at the same state. This structure of the dual chain allows us to read all eigenvalues for the transition matrix $\mathbf{P}$ and its dual $\mathbf{P}^*$ from the diagonal of $\mathbf{P}^*$ since $\mathbf{P}^*$ is upper-triangular. The symmetric walk was considered by Diaconis and Fill [7]. They used the symmetry to reduce the problem of the speed of convergence to a birth-and-death chain setting. The problem of the speed of convergence to stationarity for the non-symmetric case was studied by Brown [4], were the eigenvalues were identified by a different method. Finally, it is worth mentioning that Möbius monotonicity of non-symmetric nearest neighbor walks is a stronger property than the usual stochastic monotonicity for this chain.

2 SSD, Möbius monotonicity

2.1 Time to stationarity and strong stationary duality

Let $\mathbf{P}$ be an irreducible aperiodic transition matrix on a finite, partially ordered state space $(\mathbb{E}, \preceq)$. We enumerate the states using natural numbers $\mathbb{N}$ in such a way that for the partial order $\preceq$, for all $i, j \in \mathbb{N}$, $\mathbf{e}_i \preceq \mathbf{e}_j$ implies $i < j$. Each distribution ν on $\mathbb{E}$ we regard as a row vector, and $\nu\mathbf{P}$ denotes the usual vector times matrix multiplication.

Consider a Markov chain $\mathbf{X} = (X_n)_{n \geq 0}$ with transition matrix $\mathbf{P}$, initial distribution ν, and (unique) stationary distribution π. One possibility of measuring distance to stationarity is to use the separation distance (see Aldous and Diaconis [3]), given by $s(\nu \mathbf{P}^n, \pi) = \max_{\mathbf{e} \in \mathbb{E}} (1 - \nu \mathbf{P}^n(\mathbf{e})/\pi(\mathbf{e}))$. Separation distance s provides an upper bound on the total variation distance: $s(\nu \mathbf{P}^n, \pi) \geq d(\nu \mathbf{P}^n, \pi) := \max_{B \subset \mathbb{E}} |\nu \mathbf{P}^n(B) - \pi(B)|$.

A random variable T is a *Strong Stationary Time* (SST) if it is a randomized stopping time for $\mathbf{X} = (X_n)_{n \geq 0}$ such that T and X_T are independent, and X_T has distribution π. SST was introduced by Aldous and Diaconis in [2, 3]. In [3], they prove that $s(\nu \mathbf{P}^n, \pi) \leq P(T > n)$ (T implicitly depends on ν). Diaconis [6] gives some examples of bounds on the rates of convergence to stationarity via an SST. However, the method to find an SST is specific to each example.

Diaconis and Fill [7] introduced the so-called *Strong Stationary Dual* (SSD) chains. Such chains have a special feature, namely for them the SST for the original process has the same distribution as the time to absorption in the SSD one.

To be more specific, let $\mathbf{X}^*$ be a Markov chain with transition matrix $\mathbf{P}^*$, initial distribution ν^* on a state space $\mathbb{E}^*$. Assume that $\mathbf{e}_a^*$ is an absorbing state for $\mathbf{X}^*$. Let $\Lambda \equiv \Lambda(\mathbf{e}^*, \mathbf{e})$, $\mathbf{e}^* \in \mathbb{E}^*$, $\mathbf{e} \in \mathbb{E}$, be a kernel, called a *link*, such that $\Lambda(\mathbf{e}_a^*, \cdot) = \pi$ for $\mathbf{e}_a^* \in \mathbb{E}^*$. Diaconis and Fill [7] prove that if $(\nu^*, \mathbf{P}^*)$ is an SSD of $(\nu, \mathbf{P})$ with respect to Λ in the sense that

$$\nu = \nu^* \Lambda \quad \text{and} \quad \Lambda \mathbf{P} = \mathbf{P}^* \Lambda, \tag{2.1}$$

then there exists a bivariate Markov chain $(\mathbf{X}, \mathbf{X}^*)$ with the following marginal properties:

$\mathbf{X}$ is Markov with the initial distribution ν and the transition matrix $\mathbf{P}$,
$\mathbf{X}^*$ is Markov with the initial distribution ν^* and the transition matrix $\mathbf{P}^*$,
the absorption time T^* of $\mathbf{X}^*$ is an SST for $\mathbf{X}$.

Recall that $\overleftarrow{\mathbf{X}} = (\overleftarrow{X}_n)_{n \geq 0}$ is the time reversed process if its transition matrix is given by

$$\overleftarrow{\mathbf{P}} = \big(\mathbf{diag}(\pi)\big)^{-1} \mathbf{P}^T \big(\mathbf{diag}(\pi)\big),$$

where $\mathbf{diag}(\pi)$ denotes the matrix which is diagonal with stationary vector π on the diagonal.

The following theorem (Diaconis and Fill [7], Theorem 4.6) gives an SSD chain for linearly ordered state spaces under some stochastic monotonicity assumption. In the formulation below, we set $g(M + 1) = 0$, $\overleftarrow{P}(M + 1, \{1, \ldots, i\}) = 0$, for all $i \in \mathbb{E}$.

Theorem 1 *Let $\mathbf{X} \sim (\nu, \mathbf{P})$ be an ergodic Markov chain on a finite state space $\mathbb{E} = \{1, \ldots, M\}$, linearly ordered by $\leq$, having initial distribution ν, and stationary distribution π. Assume that*

(i) $g(i) = \frac{\nu(i)}{\pi(i)}$ *is non-increasing,*
(ii) $\overleftarrow{\mathbf{X}}$ *is stochastically monotone.*

Then there exists a Strong Stationary Dual chain $\mathbf{X}^* \sim (\nu^*, \mathbf{P}^*)$ *on* $\mathbb{E}^* = \mathbb{E}$ *with the following link kernel*

$$\Lambda(j, i) = \mathbb{I}(i \le j) \frac{\pi(i)}{H(j)},$$

where $H(j) = \sum_{k:k \le j} \pi(k)$. *Moreover, the SSD chain is uniquely determined by*

$$\nu^*(i) = H(i)\big(g(i) - g(i+1)\big), \quad i \in \mathbb{E},$$

$$\mathbf{P}^*(i, j) = \frac{H(j)}{H(i)}\big(\overleftarrow{P}\big(j, \{1, \ldots, i\}\big) - \overleftarrow{P}\big(j+1, \{1, \ldots, i\}\big)\big), \quad i, j \in \mathbb{E}.$$

Theorem 2 is our main result on SSD chains. It is an extension of Theorem 1 to Markov chains on partially ordered state spaces by replacing monotonicity in condition (i) and stochastic monotonicity in condition (ii) with Möbius monotonicity. We state this theorem in Sect. 3, after introducing required definitions and background material. Theorem 2 reveals the role of Möbius functions in finding SSD chains. Consequently, it is possible to reformulate Theorem 1 in terms of the corresponding Möbius function (in a similar way as in Corollary 3.2).

2.2 Möbius monotonicities

Consider a finite, partially ordered set $\mathbb{E} = \{\mathbf{e}_1, \ldots, \mathbf{e}_M\}$, and denote a partial order on $\mathbb{E}$ by $\preceq$. We select the above enumeration of $\mathbb{E}$ to be consistent with the partial order, i.e., $\mathbf{e}_i \preceq \mathbf{e}_j$ implies $i < j$.

Let $\mathbf{X} = (X_n)_{n \ge 0} \sim (\nu, \mathbf{P})$ be a time homogeneous Markov chain with an initial distribution ν and transition function $\mathbf{P}$ on the state space $\mathbb{E}$. We identify the transition function with the corresponding matrix written for the fixed enumeration of the state space. Suppose that $\mathbf{X}$ is ergodic with the stationary distribution π.

We shall use $\wedge$ for the meet (greatest lower bound) and $\vee$ for the join (least upper bound) in $\mathbb{E}$. If $\mathbb{E}$ is a lattice, it has unique minimal and maximal elements, denoted by $\mathbf{e}_1 := \hat{\mathbf{0}}$ and $\mathbf{e}_M := \hat{\mathbf{1}}$, respectively.

Recall that the *zeta function* ζ of the partially ordered set $\mathbb{E}$ is defined by: $\zeta(\mathbf{e}_i, \mathbf{e}_j) = 1$ if $\mathbf{e}_i \preceq \mathbf{e}_j$ and $\zeta(\mathbf{e}_i, \mathbf{e}_j) = 0$ otherwise. If the states are enumerated in such a way that $\mathbf{e}_i \preceq \mathbf{e}_j$ implies $i < j$ (assumed in this paper), then ζ can be represented by an upper-triangular, 0–1 valued matrix $\mathbf{C}$, which is invertible. It is well known that ζ is an element of *the incidence algebra* (see Rota [15], p. 344), which is invertible in this algebra, and the inverse to ζ, denoted by μ, is called the *Möbius function*. Using the enumeration which defines $\mathbf{C}$, the corresponding matrix describing μ is given by the usual matrix inverse $\mathbf{C}^{-1}$.

Throughout the paper, μ will denote the Möbius function of the corresponding ordering.

For the state space $\mathbb{E} = \{\mathbf{e}_1, \ldots, \mathbf{e}_M\}$ with the partial ordering $\preceq$, we define the following operators acting on all functions $f : \mathbb{E} \to \mathbb{R}$

$$S_\downarrow f(\mathbf{e}_i) = \sum_{\mathbf{e} \in \mathbb{E}} f(\mathbf{e})\zeta(\mathbf{e}, \mathbf{e}_i) = \sum_{\mathbf{e}:\mathbf{e} \preceq \mathbf{e}_i} f(\mathbf{e}) =: F(\mathbf{e}_i) \tag{2.2}$$

and

$$S_\uparrow f(\mathbf{e}_i) = \sum_{\mathbf{e} \in \mathbb{E}} \zeta(\mathbf{e}_i, \mathbf{e}) f(\mathbf{e}) = \sum_{\mathbf{e}:\mathbf{e} \succeq \mathbf{e}_i} f(\mathbf{e}) =: \bar{F}(\mathbf{e}_i). \qquad (2.3)$$

In the matrix notation, we shall use the corresponding bold letters for functions, and we have $\mathbf{F} = \mathbf{f}\mathbf{C}$, $\bar{\mathbf{F}} = \mathbf{f}\mathbf{C}^T$, where $\mathbf{f} = (f(\mathbf{e}_1), \ldots, f(\mathbf{e}_M))$, $\mathbf{F} = (F(\mathbf{e}_1), \ldots, F(\mathbf{e}_M))$, and $\bar{\mathbf{F}} = (\bar{F}(\mathbf{e}_1), \ldots, \bar{F}(\mathbf{e}_M))$.

The following difference operators $D_\downarrow$ and $D_\uparrow$ are the inverse operators to the summation operators $S_\downarrow$ and $S_\uparrow$, respectively,

$$D_\downarrow f(\mathbf{e}_i) = \sum_{\mathbf{e} \in \mathbb{E}} f(\mathbf{e})\mu(\mathbf{e}, \mathbf{e}_i) = \sum_{\mathbf{e}:\mathbf{e} \preceq \mathbf{e}_i} f(\mathbf{e})\mu(\mathbf{e}, \mathbf{e}_i) =: g(\mathbf{e}_i),$$

and

$$D_\uparrow f(\mathbf{e}_i) = \sum_{\mathbf{e} \in \mathbb{E}} \mu(\mathbf{e}_i, \mathbf{e}) f(\mathbf{e}) = \sum_{\mathbf{e}:\mathbf{e} \succeq \mathbf{e}_i} \mu(\mathbf{e}_i, \mathbf{e}) f(\mathbf{e}) =: h(\mathbf{e}_i).$$

In the matrix notation, we have $\mathbf{g} = \mathbf{f}\mathbf{C}^{-1}$ and $\mathbf{h} = \mathbf{f}(\mathbf{C}^T)^{-1}$.

If, for example, the relations (2.2) and (2.3) hold then

$$f(\mathbf{e}_i) = \sum_{\mathbf{e}:\mathbf{e} \preceq \mathbf{e}_i} F(\mathbf{e})\mu(\mathbf{e}, \mathbf{e}_i) = D_\downarrow\big(S_\downarrow f(\mathbf{e}_i)\big)$$

and

$$f(\mathbf{e}_i) = \sum_{\mathbf{e}:\mathbf{e} \succeq \mathbf{e}_i} \mu(\mathbf{e}_i, \mathbf{e})\bar{F}(\mathbf{e}) = D_\uparrow\big(S_\uparrow f(\mathbf{e}_i)\big), \qquad (2.4)$$

respectively.

Definition 2.1 For a Markov chain $\mathbf{X}$ with the transition function $\mathbf{P}$, we say that $\mathbf{P}$ (or alternatively that $\mathbf{X}$) is

$\downarrow$-Möbius monotone if

$$\mathbf{C}^{-1}\mathbf{P}\mathbf{C} \geq 0,$$

$\uparrow$-Möbius monotone if

$$(\mathbf{C}^T)^{-1}\mathbf{P}\mathbf{C}^T \geq 0,$$

where $\mathbf{P}$ is the matrix of the transition probabilities written using the enumeration which defines $\mathbf{C}$, and ≥ 0 means that each entry of a matrix is nonnegative.

Definition 2.2 A function $f : \mathbb{E} \to \mathbb{R}$ is

$\downarrow$-Möbius monotone if $\mathbf{f}(\mathbf{C}^T)^{-1} \geq 0$,
$\uparrow$-Möbius monotone if $\mathbf{f}\mathbf{C}^{-1} \geq 0$.

For example, in terms of the Möbius function μ and the transition probabilities, $\downarrow$- Möbius monotonicity of $\mathbf{P}$ means that for all $(\mathbf{e}_i, \mathbf{e}_j \in \mathbb{E})$

$$\sum_{\mathbf{e}:\mathbf{e} \succeq \mathbf{e}_i} \mu(\mathbf{e}_i, \mathbf{e}) P\big(\mathbf{e}, \{\mathbf{e}_j\}^{\downarrow}\big) \geq 0,$$

where $P(\cdot, \cdot)$ denotes the corresponding transition kernel, i.e., $P(\mathbf{e}_i, \{\mathbf{e}_j\}^{\downarrow}) = \sum_{\mathbf{e}:\mathbf{e} \preceq \mathbf{e}_j} \mathbf{P}(\mathbf{e}_i, \mathbf{e})$, and $\{\mathbf{e}_j\}^{\downarrow} = \{\mathbf{e} : \mathbf{e} \preceq \mathbf{e}_j\}$. In order to check such a condition, an explicit formula for μ is needed. Note that the above definition for monotonicity can be rewritten as follows, f is $\downarrow$-Möbius monotone if for some nonnegative vector $\mathbf{m} \geq 0$, $\mathbf{f} = \mathbf{m}\mathbf{C}^T$ holds, and f is $\uparrow$-Möbius monotone if $\mathbf{f} = \mathbf{m}\mathbf{C}$. The last equality means that f is a nonnegative linear combination of the rows of matrix $\mathbf{C}$. This monotonicity implies that f is non-decreasing in the usual sense (f non-decreasing means: $\mathbf{e}_i \preceq \mathbf{e}_j$ implies $f(\mathbf{e}_i) \leq f(\mathbf{e}_j)$).

Proposition 2.1 $\mathbf{P}$ *is* $\uparrow$*-Möbius monotone iff*

f *is* $\uparrow$*-Möbius monotone implies that* $\mathbf{P}\mathbf{f}^T$ *is* $\uparrow$*-Möbius monotone.*

Proof Suppose that $\mathbf{P}$ is $\uparrow$-Möbius monotone, that is, $(\mathbf{C}^T)^{-1}\mathbf{P}\mathbf{C}^T \geq 0$. Take arbitrary f which is $\uparrow$-Möbius monotone, i.e., take $\mathbf{f} = \mathbf{m}\mathbf{C}$ for some arbitrary $\mathbf{m} \geq 0$. Then $(\mathbf{C}^T)^{-1}\mathbf{P}\mathbf{C}^T\mathbf{m}^T \geq 0$, which is (using transposition) equivalent to $\mathbf{f}\mathbf{P}^T\mathbf{C}^{-1} \geq 0$, which, in turn, gives (by definition) that $\mathbf{P}\mathbf{f}^T$ is $\uparrow$-Möbius monotone. Conversely, for all $\mathbf{f} = \mathbf{m}\mathbf{C}$, where $\mathbf{m} \geq 0$, we have $\mathbf{f}\mathbf{P}^T\mathbf{C}^{-1} \geq 0$ since $\mathbf{P}\mathbf{f}^T$ is $\uparrow$-Möbius monotone. This implies that $(\mathbf{C}^T)^{-1}\mathbf{P}\mathbf{C}^T\mathbf{m}^T \geq 0$ and $(\mathbf{C}^T)^{-1}\mathbf{P}\mathbf{C}^T \geq 0$.   □

Many examples can be produced using the fact that the set of Möbius monotone matrices is a convex subset of the set of transition matrices. We shall give some basic examples in Sect. 4. These examples can be used to build up a large class of Möbius monotone matrices.

Proposition 2.2

(i) *If* $\mathbf{P}_1$ *and* $\mathbf{P}_2$ *are* $\uparrow$*-Möbius monotone* ($\downarrow$*-Möbius monotone) then* $\mathbf{P}_1\mathbf{P}_2$ *is* $\uparrow$*-Möbius monotone* ($\downarrow$*-Möbius monotone).*

(ii) *If* $\mathbf{P}$ *is* $\uparrow$*-Möbius monotone* ($\downarrow$*-Möbius monotone) then* $(\mathbf{P})^k$ *is* $\uparrow$*-Möbius monotone* ($\downarrow$*-Möbius monotone) for each* $k \in \mathbb{N}$.

(iii) *If* $\mathbf{P}_1$ *is* $\uparrow$*-Möbius monotone* ($\downarrow$*-Möbius monotone) and* $\mathbf{P}_2$ *is* $\uparrow$*-Möbius monotone* ($\downarrow$*-Möbius monotone) then*

$$p\mathbf{P}_1 + (1 - p)\mathbf{P}_2$$

is $\uparrow$*-Möbius monotone* ($\downarrow$*-Möbius monotone) for all* $p \in (0, 1)$.

Proof (i) Since $(\mathbf{C}^T)^{-1}\mathbf{P}_1\mathbf{C}^T \geq 0$ and $(\mathbf{C}^T)^{-1}\mathbf{P}_2\mathbf{C}^T \geq 0$, one has

$$\big(\mathbf{C}^T\big)^{-1}\mathbf{P}_1\mathbf{P}_2\mathbf{C}^T = \big(\big(\mathbf{C}^T\big)^{-1}\mathbf{P}_1 C^T\big)\big(\big(\mathbf{C}^T\big)^{-1}\mathbf{P}_2\mathbf{C}^T\big) \geq 0.$$

The statements (ii), (iii) are immediate by definition.   □

3 Main result: SSD for Möbius monotone chains

Now we are prepared to state our main result on SSD.

Theorem 2 *Let* $\mathbf{X} \sim (v, \mathbf{P})$ *be an ergodic Markov chain on a finite state space* $\mathbb{E} = \{\mathbf{e}_1, \dots, \mathbf{e}_M\}$*, partially ordered by* $\preceq$*, with a unique maximal state* $\mathbf{e}_M$*, and with stationary distribution* π*. Assume that*

(i) $g(\mathbf{e}) = \frac{v(\mathbf{e})}{\pi(\mathbf{e})}$ *is* $\downarrow$*-Möbius monotone,*

(ii) $\overleftarrow{\mathbf{X}}$ *is* $\downarrow$*-Möbius monotone.*

Then there exists a Strong Stationary Dual chain $\mathbf{X}^* \sim (v^*, \mathbf{P}^*)$ *on* $\mathbb{E}^* = \mathbb{E}$ *with the following link kernel*

$$\Lambda(\mathbf{e}_j, \mathbf{e}_i) = \mathbb{I}(\mathbf{e}_i \preceq \mathbf{e}_j) \frac{\pi(\mathbf{e}_i)}{H(\mathbf{e}_j)},$$

where $H(\mathbf{e}_j) = S_\downarrow \pi(\mathbf{e}_j) = \sum_{\mathbf{e}: \mathbf{e} \preceq \mathbf{e}_j} \pi(\mathbf{e})$ $(\mathbf{H} = \pi \mathbf{C})$*. Moreover, the SSD chain is uniquely determined by*

$$v^*(\mathbf{e}_i) = H(\mathbf{e}_i) \sum_{\mathbf{e}: \mathbf{e} \succeq \mathbf{e}_i} \mu(\mathbf{e}_i, \mathbf{e}) g(\mathbf{e}) = S_\downarrow \pi(\mathbf{e}_i) D^\uparrow g(\mathbf{e}_i),$$

$$\mathbf{P}^*(\mathbf{e}_i, \mathbf{e}_j) = \frac{H(\mathbf{e}_j)}{H(\mathbf{e}_i)} \sum_{\mathbf{e}: \mathbf{e} \succeq \mathbf{e}_j} \mu(\mathbf{e}_j, \mathbf{e}) \overleftarrow{P}(\mathbf{e}, \{\mathbf{e}_i\}^\downarrow) = \frac{S_\downarrow \pi(\mathbf{e}_j)}{S_\downarrow \pi(\mathbf{e}_i)} D^\uparrow \overleftarrow{P}(\mathbf{e}_j, \{\mathbf{e}_i\}^\downarrow).$$

The corresponding matrix formulas are given by

$$v^* = \mathbf{g}(\mathbf{C}^T)^{-1} \mathbf{diag}(\pi \mathbf{C}),$$

$$\mathbf{P}^* = (\mathbf{diag}(\mathbf{H}) \, \mathbf{C}^{-1} \overleftarrow{\mathbf{P}} \, \mathbf{C} \, \mathbf{diag}(\mathbf{H})^{-1})^T$$

$$= \mathbf{diag}(\pi \mathbf{C})^{-1} (\mathbf{C}^T \mathbf{diag}(\pi)) \mathbf{P} (\mathbf{C}^T \mathbf{diag}(\pi))^{-1} \mathbf{diag}(\pi \mathbf{C}),$$

where $\mathbf{g} = (g(\mathbf{e}_1), \dots, g(\mathbf{e}_M))$ *(row vector).*

Proof of Theorem 2 We have to check the conditions (2.1). The first condition given in (2.1), $v = v^* \Lambda$, reads for arbitrary $\mathbf{e}_i \in \mathbb{E}$ as

$$v(\mathbf{e}_i) = \sum_{\mathbf{e} \succeq \mathbf{e}_i} v^*(\mathbf{e}) \frac{\pi(\mathbf{e}_i)}{H(\mathbf{e})},$$

which is equivalent to

$$\frac{v(\mathbf{e}_i)}{\pi(\mathbf{e}_i)} = \sum_{\mathbf{e}: \mathbf{e} \succeq \mathbf{e}_i} \frac{v^*(\mathbf{e})}{H(\mathbf{e})}.$$

From the Möbius inversion formula (2.4), we get

$$\frac{v^*(\mathbf{e}_i)}{H(\mathbf{e}_i)} = \sum_{\mathbf{e}: \mathbf{e} \succeq \mathbf{e}_i} \mu(\mathbf{e}_i, \mathbf{e}) \frac{v(\mathbf{e})}{\pi(\mathbf{e})},$$

which gives the required formula. From the assumption that $g = \frac{\nu}{\pi}$ is $\downarrow$-Möbius monotone, it follows that $\nu^* \geq 0$. Moreover, since $\nu = \nu^* \Lambda$ and Λ is a transition kernel, it is clear that ν^* is a probability vector.

The second condition given in (2.1), $\Lambda \mathbf{P} = \mathbf{P}^* \Lambda$, means that for all $\mathbf{e}_i, \mathbf{e}_j \in \mathbb{E}$

$$\sum_{\mathbf{e} \in \mathbb{E}} \Lambda(\mathbf{e}_i, \mathbf{e}) \mathbf{P}(\mathbf{e}, \mathbf{e}_j) = \sum_{\mathbf{e} \in \mathbb{E}} \mathbf{P}^*(\mathbf{e}_i, \mathbf{e}) \Lambda(\mathbf{e}, \mathbf{e}_j).$$

Taking the proposed Λ, we have to check that

$$\sum_{\mathbf{e}:\mathbf{e} \preceq \mathbf{e}_i} \frac{\pi(\mathbf{e})}{H(\mathbf{e}_i)} \mathbf{P}(\mathbf{e}, \mathbf{e}_j) = \sum_{\mathbf{e}:\mathbf{e} \succeq \mathbf{e}_j} \frac{\pi(\mathbf{e}_j)}{H(\mathbf{e})} \mathbf{P}^*(\mathbf{e}_i, \mathbf{e}),$$

that is,

$$\frac{1}{H(\mathbf{e}_i)} \sum_{\mathbf{e}:\mathbf{e} \preceq \mathbf{e}_i} \frac{\pi(\mathbf{e})}{\pi(\mathbf{e}_j)} \mathbf{P}(\mathbf{e}, \mathbf{e}_j) = \sum_{\mathbf{e}:\mathbf{e} \succeq \mathbf{e}_j} \frac{\mathbf{P}^*(\mathbf{e}_i, \mathbf{e})}{H(\mathbf{e})}.$$

Using $\frac{\pi(\mathbf{e})}{\pi(\mathbf{e}_j)} \mathbf{P}(\mathbf{e}, \mathbf{e}_j) = \overleftarrow{\mathbf{P}}(\mathbf{e}_j, \mathbf{e})$, we have

$$\frac{1}{H(\mathbf{e}_i)} \overleftarrow{P}\left(\mathbf{e}_j, \{\mathbf{e}_i\}^{\downarrow}\right) = \sum_{\mathbf{e}:\mathbf{e} \succeq \mathbf{e}_j} \frac{\mathbf{P}^*(\mathbf{e}_i, \mathbf{e})}{H(\mathbf{e})}. \tag{3.1}$$

For each fixed $\mathbf{e}_i$ we treat $\frac{1}{H(\mathbf{e}_i)} \overleftarrow{P}(\mathbf{e}_j, \{\mathbf{e}_i\}^{\downarrow})$ as a function of $\mathbf{e}_j$ and again use the Möbius inversion formula (2.4) to get from (3.1)

$$\frac{\mathbf{P}^*(\mathbf{e}_i, \mathbf{e}_j)}{H(\mathbf{e}_j)} = \sum_{\mathbf{e}:\mathbf{e} \succeq \mathbf{e}_j} \mu(\mathbf{e}_j, \mathbf{e}) \frac{\overleftarrow{P}(\mathbf{e}, \{\mathbf{e}_i\}^{\downarrow})}{H(\mathbf{e}_i)}.$$

In the matrix notation, we have

$$\mathbf{C}^{-1} \mathbf{P} \mathbf{C}(\mathbf{e}_i, \mathbf{e}_j) = \sum_{\mathbf{e}:\mathbf{e} \succeq \mathbf{e}_i} \mu(\mathbf{e}_i, \mathbf{e}) P(\mathbf{e}, \{\mathbf{e}_j\}^{\downarrow}),$$

therefore,

$$\mathbf{P}^* = \left(\mathbf{diag}(\mathbf{H}) \, \mathbf{C}^{-1} \overleftarrow{\mathbf{P}} \, \mathbf{C} \, \mathbf{diag}(\mathbf{H})^{-1} \right)^T.$$

Since, from our assumption, $\mathbf{C}^{-1} \overleftarrow{\mathbf{P}} \mathbf{C} \geq 0$, we have $\mathbf{P}^* \geq 0$. Now $\Lambda \mathbf{P} = \mathbf{P}^* \Lambda$ implies that $\mathbf{P}^*$ is a transition matrix. $\qquad \square$

Note that in the context of Theorem 2, if the original chain starts with probability 1 in the minimal state, i.e., $\nu = \delta_{\mathbf{e}_1}$, then $\nu^* = \delta_{\mathbf{e}_1}$.

For $\mathbb{E} = \{1, \dots, M\}$, with linear ordering $\leq$, the Möbius function is given by $\mu(k, k) = 1, \mu(k-1, k) = -1$, and μ equals 0 otherwise. In this case, the link is given by $\Lambda(j, i) = \mathbb{I}(i \leq j) \frac{\pi(i)}{H(j)}$, and we obtain from Theorem 2 (as a special case) Theorem 1, which is a reformulation of Theorem 4.6 from Diaconis and Fill [7].

In a similar way, we construct an analog SSD chain for $^\uparrow$-Möbius monotone **P**. We skip the corresponding matrix formulation and a proof. This analog SSD chain will be used in Corollary 3.2 to give an alternative SSD chain to the one given in Theorem 1.

Corollary 3.1 *Let* $\mathbf{X} \sim (\nu, \mathbf{P})$ *be an ergodic Markov chain on a finite state space* $\mathbb{E} = \{\mathbf{e}_1, \ldots, \mathbf{e}_M\}$, *partially ordered by* $\preceq$, *with a unique minimal state* $\mathbf{e}_1$, *and with the stationary distribution* π. *Assume that*

(i) $g(\mathbf{e}) = \frac{\nu(\mathbf{e})}{\pi(\mathbf{e})}$ *is* $^\uparrow$-*Möbius monotone,*

(ii) $\overleftarrow{\mathbf{X}}$ *is* $^\uparrow$-*Möbius monotone.*

Then there exists a Strong Stationary Dual chain $\mathbf{X}^\bullet \sim (\nu^\bullet, \mathbf{P}^\bullet)$ *on* $\mathbb{E}^\bullet = \mathbb{E}$ *with the following link*

$$\Lambda^\bullet(\mathbf{e}_j, \mathbf{e}_i) = \mathbb{I}(\mathbf{e}_i \succeq \mathbf{e}_j) \frac{\pi(\mathbf{e}_i)}{\bar{H}(\mathbf{e}_j)},$$

where $\bar{H}(\mathbf{e}_j) = S_\uparrow \pi(\mathbf{e}_j)$. *Moreover, the SSD is uniquely determined by*

$$\nu^\bullet(\mathbf{e}_i) = \bar{H}(\mathbf{e}_i) \sum_{\mathbf{e}:\mathbf{e} \preceq \mathbf{e}_i} g(\mathbf{e})\mu(\mathbf{e}, \mathbf{e}_i) = S_\uparrow \pi(\mathbf{e}_i) D^\downarrow g(\mathbf{e}_i),$$

$$\mathbf{P}^\bullet(\mathbf{e}_i, \mathbf{e}_j) = \frac{\bar{H}(\mathbf{e}_j)}{\bar{H}(\mathbf{e}_i)} \sum_{\mathbf{e}:\mathbf{e} \preceq \mathbf{e}_j} \overleftarrow{P}\big(\mathbf{e}, \{\mathbf{e}_i\}^\uparrow\big)\mu(\mathbf{e}, \mathbf{e}_j) = \frac{S_\uparrow \pi(\mathbf{e}_j)}{S_\uparrow \pi(\mathbf{e}_i)} D^\downarrow \overleftarrow{P}\big(\mathbf{e}_j, \{\mathbf{e}_i\}^\uparrow\big).$$

Note that in the setting of Corollary 3.1, if the original chain starts with probability 1 in the maximal state, i.e., $\nu = \delta_{\mathbf{e}_M}$, then $\nu^\bullet = \delta_{\mathbf{e}_M}$.

From Corollary 3.1 we obtain an alternative dual result for linearly ordered spaces assuming that $\frac{\nu(i)}{\mu(i)}$ is non-decreasing. Roughly speaking, Theorem 1 and Corollary 3.2 describe two complementary situations, namely when an initial distribution for the original chain is in a sense (mlr ordering) smaller or bigger than the stationary distribution, then one can create (and use) different (alternative) dual chains as described in these statements.

Corollary 3.2 *Let* $\mathbf{X} \sim (\nu, \mathbf{P})$ *be an ergodic Markov chain on a finite state space* $\mathbb{E} = \{1, \ldots, M\}$, *linearly ordered by* $\leq$, *with stationary distribution* π. *Assume that*

(i) $g(i) = \frac{\nu(i)}{\pi(i)}$ *is non-decreasing,*

(ii) $\overleftarrow{\mathbf{X}}$ *is stochastically monotone.*

Then there exists a Strong Stationary Dual chain $\mathbf{X}^\bullet \sim (\nu^\bullet, \mathbf{P}^\bullet)$ *on* $\mathbb{E}^\bullet = \mathbb{E}$ *with the following link kernel*

$$\Lambda^\bullet(j, i) = \mathbb{I}(i \geq j) \frac{\pi(i)}{\bar{H}(j)},$$

where $\bar{H}(j) = \sum_{k:k \geq j} \pi(k)$. *Moreover, the SSD is uniquely determined by*

$$v^{\bullet}(i) = \bar{H}(i) \sum_{k:k \leq i} \mu(k,i)g(k) = \bar{H}(i)\big(g(i) - g(i-1)\big), \quad i \in \mathbb{E},$$

$$\mathbf{P}^{\bullet}(i,j) = \frac{\bar{H}(j)}{\bar{H}(i)} \sum_{k:k \leq j} \mu(k,j) \overleftarrow{P}\big(k, \{i, \dots, M\}\big)$$

$$= \frac{\bar{H}(j)}{\bar{H}(i)}\big(\overleftarrow{P}\big(j, \{i, \dots, M\}\big) - \overleftarrow{P}\big(j-1, \{i, \dots, M\}\big)\big), \quad i, j \in \mathbb{E}.$$

4 Nearest neighbor Möbius monotone walks on a cube

Consider the discrete time Markov chain $\mathbf{X} = \{X_n, n \geq 0\}$, with the state space $\mathbb{E} = \{0, 1\}^d$, and transition matrix $\mathbf{P}$ given by

$$\mathbf{P}(\mathbf{e}, \mathbf{e} + \mathbf{s}_i) = \alpha_i \mathbb{I}_{\{e_i = 0\}},$$
$$\mathbf{P}(\mathbf{e}, \mathbf{e} - \mathbf{s}_i) = \beta_i \mathbb{I}_{\{e_i = 1\}}, \tag{4.1}$$
$$\mathbf{P}(\mathbf{e}, \mathbf{e}) = 1 - \sum_{i:e_i = 0} \alpha_i - \sum_{i:e_i = 1} \beta_i,$$

where $\mathbf{e} = (e_1, \dots, e_d) \in \mathbb{E}$, $e_i \in \{0, 1\}$, and $\mathbf{s}_i = (0, \dots, 0, 1, 0, \dots, 0)$ with 1 at the ith coordinate.

We use the following partial order:

$$\mathbf{e} = (e_1, e_2, \dots, e_d) \preceq \mathbf{e}' = \big(e_1', e_2', \dots, e_d'\big) \quad \text{if } e_i \leq e_i', \text{ for all } i = 1, \dots, d.$$

To make our presentation simpler, we assume that $v = \delta_{(0,\dots,0)}$ (this assumption can be waived).

For example, such a Markov chain is a model for a set of working unreliable servers where the repairs and breakdowns of servers are independent for different servers and only one server can be broken or repaired at a transition time.

Assume that all α_i and β_i are positive and that there exists at least one state $\mathbf{e}$ such that $\mathbf{P}(\mathbf{e}, \mathbf{e}) > 0$. Then the chain is ergodic.

Theorem 3 *For the Markov chain* $\mathbf{X} = \{X_n, n \geq 0\}$, $\mathbb{E} = \{0, 1\}^d$ *with the transition matrix* $\mathbf{P}$ *given by* (4.1), *and* $v = \delta_{(0,\dots,0)}$, *assume that* $\sum_{i=1}^{d}(\alpha_i + \beta_i) \leq 1$. *Then there exists a dual chain on* $\mathbb{E}^* = \mathbb{E}$ *given by* $v^* = v$, *and*

$$\mathbf{P}^*(\mathbf{e}, \mathbf{e} + \mathbf{s}_i) = \alpha_i + \beta_i \quad \textit{if } e_i = 0,$$
$$\mathbf{P}^*(\mathbf{e}, \mathbf{e}) = 1 - \sum_{i:e_i = 0}(\alpha_i + \beta_i),$$
$$\mathbf{P}^*(\mathbf{e}, \mathbf{e}') = 0 \quad \textit{otherwise.}$$

Proof A direct check shows that **X** is time-reversible with stationary distribution

$$\pi(x) = \prod_{i:x_i=1} \frac{\alpha_i}{\alpha_i + \beta_i} \prod_{i:x_i=0} \frac{\beta_i}{\alpha_i + \beta_i}.$$

Let $|\mathbf{e}| = \sum_{i=1}^d e_i$. Note that $\mathbb{E}$ with $\preceq$ is a Boolean lattice, and the corresponding Möbius function is given by

$$\mu(\mathbf{e}, \mathbf{e}') = \begin{cases} (-1)^{|\mathbf{e}'|-|\mathbf{e}|} & \text{if } \mathbf{e} \preceq \mathbf{e}', \\ 0 & \text{otherwise.} \end{cases}$$

The assumption $\sum_{i=1}^d (\alpha_i + \beta_i) \le 1$ implies $\downarrow$-Möbius monotonicity. Indeed, calculating

$$\mathbf{P}^*(\mathbf{e}_i, \mathbf{e}_j) = \frac{H(\mathbf{e}_j)}{H(\mathbf{e}_i)} \sum_{\mathbf{e}:\mathbf{e}\succeq\mathbf{e}_j} \mu(\mathbf{e}_j, \mathbf{e}) \overleftarrow{P}(\mathbf{e}, \{\mathbf{e}_i\}^\downarrow),$$

we find conditions for its nonnegativity, which implies $\downarrow$-Möbius monotonicity of the chain and its time reversed chain:

$$\mathbf{P}^*\big((0, \dots, 0), (0, \dots, 0)\big)$$
$$= \sum_{\mathbf{e}\succeq(0,\dots,0)} \mu\big((0, \dots, 0), \mathbf{e}\big) \overleftarrow{P}\big(\mathbf{e}, \big\{(0, \dots, 0)\big\}^\downarrow\big)$$
$$= 1 - (\alpha_1 + \dots + \alpha_d) - \beta_1 - \dots - \beta_d = 1 - \sum_{i=1}^d \alpha_i - \sum_{i=1}^d \beta_i,$$

which is nonnegative (because we assumed that $\sum_{i=1}^d (\alpha_i + \beta_i) \le 1$).

Fix $\mathbf{s}_i = (0, \dots, 0, 1, 0, \dots, 0)$ with 1 in position i. Then

$$\mathbf{P}^*(\mathbf{s}_i, \mathbf{s}_i) = \sum_{\mathbf{e}:\mathbf{e}\succeq\mathbf{s}_i} \mu(\mathbf{s}_i, \mathbf{e}) \overleftarrow{P}(\mathbf{e}, \{\mathbf{s}_i\}^\downarrow) = 1 - \sum_{k=1}^d \alpha_k + \alpha_i - \sum_{k=1}^d \beta_k + \beta_i.$$

For each state of the form $\mathbf{e}^i = (e_1, \dots, e_{i-1}, 0, e_{i+1}, \dots, e_d)$,

$$\mathbf{P}^*\big(\mathbf{e}^i, \mathbf{e}^i + \mathbf{s}_i\big) = \frac{H\big(\mathbf{e}^i + \mathbf{s}_i\big)}{H\big(\mathbf{e}^i\big)} \sum_{\mathbf{e}:\mathbf{e}\succeq\mathbf{e}^i+\mathbf{s}_i} \mu\big(\mathbf{e}^i + \mathbf{s}_i, \mathbf{e}\big) \overleftarrow{P}\big(\mathbf{e}, \{\mathbf{e}^i\}^\downarrow\big)$$
$$= \frac{H\big(\mathbf{e}^i + \mathbf{s}_i\big)}{H\big(\mathbf{e}^i\big)} \big(\mu\big(\mathbf{e}^i + \mathbf{s}_i, \mathbf{e}^i + \mathbf{s}_i\big) \overleftarrow{P}\big(\mathbf{e}^i + \mathbf{s}_i, \{\mathbf{e}^i\}^\downarrow\big)\big) = \frac{H\big(\mathbf{e}^i + \mathbf{s}_i\big)}{H\big(\mathbf{e}^i\big)} \beta_i.$$

Denote by $z(\mathbf{e}) = \{k : e_k = 0\}$ the index set of zero coordinates. We shall compute $\frac{H(\mathbf{e}^i + \mathbf{s}_i)}{H(\mathbf{e}^i)}$. Let $G = \prod_{j=1}^d (\alpha_j + \beta_j)$.

$$H\big(\mathbf{e}^i\big) = \sum_{\mathbf{e}:\mathbf{e}\preceq\mathbf{e}^i} \pi(\mathbf{e}) = \frac{1}{G} \sum_{\mathbf{e}:\mathbf{e}\preceq\mathbf{e}^i} \prod_{k:e_k=1} \alpha_k \prod_{k:e_k=0} \beta_k$$

$$= \frac{1}{G} \prod_{k:e_k^i=0} \beta_k \left(\sum_{A \subseteq \{1,\ldots,d\}\backslash z(\mathbf{e}^i)} \prod_{j \in A} \alpha_j \prod_{j \in A^C} \beta_j \right),$$

$$H\left(\mathbf{e}^i + \mathbf{s}_i\right) = \sum_{\mathbf{e} \preceq \mathbf{e}^i + \mathbf{s}_i} \pi(\mathbf{e}) = H\left(\mathbf{e}^i\right) + \sum_{\substack{\mathbf{e} \preceq \mathbf{e}^i + \mathbf{s}_i \\ e_i = 1}} \pi(\mathbf{e})$$

$$= H\left(\mathbf{e}^i\right) + \frac{1}{G} \prod_{k:e_k=0} \beta_k \alpha_i \left(\sum_{A \subseteq \{1,\ldots,d\}\backslash z(\mathbf{e}^i)} \prod_{j \in A} \alpha_j \prod_{j \in A^C} \beta_j \right)$$

$$= H\left(\mathbf{e}^i\right) + \frac{1}{G} \prod_{k:e_k^i=0} \beta_k \frac{\alpha_i}{\beta_i} \left(\sum_{A \subseteq \{1,\ldots,d\}\backslash z(\mathbf{e}^i)} \prod_{j \in A} \alpha_j \prod_{j \in A^C} \beta_j \right).$$

And thus

$$\frac{H\left(\mathbf{e}^i + \mathbf{s}_i\right)}{H\left(\mathbf{e}^i\right)} = 1 + \frac{\alpha_i}{\beta_i} = \frac{\alpha_i + \beta_i}{\beta_i}$$

and

$$\mathbf{P}^*\left(\mathbf{e}^i, \mathbf{e}^i + \mathbf{s}_i\right) = \alpha_i + \beta_i.$$

Now, fix some ${}^i\mathbf{e} = \{e_1, \ldots, e_{i-1}, 1, e_{i+1}, \ldots, e_d\}$.

$$\mathbf{P}^*\left({}^i\mathbf{e}, {}^i\mathbf{e} - \mathbf{s}_i\right) = \frac{H\left({}^i\mathbf{e} - \mathbf{s}_i\right)}{H\left({}^i\mathbf{e}\right)} \sum_{\mathbf{e}:\mathbf{e} \succeq {}^i\mathbf{e} - \mathbf{s}_i} \mu\left({}^i\mathbf{e} - \mathbf{s}_i, \mathbf{e}\right) \overleftarrow{P}\left(\mathbf{e}, \{{}^i\mathbf{e}\}^\downarrow\right).$$

Fix $j \in \{1, \ldots, d\} \setminus z({}^i\mathbf{e})$. The following cases are possible:

$$\mathbf{e} = {}^i\mathbf{e} - \mathbf{s}_i : \qquad \mu\left({}^i\mathbf{e} - \mathbf{s}_i, {}^i\mathbf{e} - \mathbf{s}_i\right) \overleftarrow{P}\left({}^i\mathbf{e} - \mathbf{s}_i, \{{}^i\mathbf{e}\}^\downarrow\right) = 1 - \sum_{k \in z({}^i\mathbf{e})} \alpha_k;$$

$$\mathbf{e} = {}^i\mathbf{e} : \qquad \mu\left({}^i\mathbf{e} - \mathbf{s}_i, {}^i\mathbf{e}\right) \overleftarrow{P}\left({}^i\mathbf{e}, \{{}^i\mathbf{e}\}^\downarrow\right) = -\left(1 - \sum_{k \in z({}^i\mathbf{e})} \alpha_k\right);$$

$$\mathbf{e} = {}^i\mathbf{e} - \mathbf{s}_i + \mathbf{s}_j : \quad \mu\left({}^i\mathbf{e} - \mathbf{s}_i, {}^i\mathbf{e} - \mathbf{s}_i + \mathbf{s}_j\right) \overleftarrow{P}\left({}^i\mathbf{e} - \mathbf{s}_i + \mathbf{s}_j, \{{}^i\mathbf{e}\}^\downarrow\right) = -\beta_j;$$

$$\mathbf{e} = {}^i\mathbf{e} + \mathbf{s}_j : \qquad \mu\left({}^i\mathbf{e} - \mathbf{s}_i, {}^i\mathbf{e} + \mathbf{s}_j\right) \overleftarrow{P}\left({}^i\mathbf{e} + \mathbf{s}_j, \{{}^i\mathbf{e}\}^\downarrow\right) = \beta_j.$$

Summing up all possibilities, we get

$$\mathbf{P}^*\left({}^i\mathbf{e}, {}^i\mathbf{e} - \mathbf{s}_i\right) = 0.$$

For each $\mathbf{e}$, we have

$$\mathbf{P}^*(\mathbf{e}, \mathbf{e}) = \sum_{\mathbf{e}':\mathbf{e}' \succeq \mathbf{e}} \mu(\mathbf{e}, \mathbf{e}') \overleftarrow{P}\left(\mathbf{e}', \{\mathbf{e}\}^\downarrow\right);$$

$$\mathbf{e}' = \mathbf{e}: \qquad \mu(\mathbf{e}, \mathbf{e}) \overleftarrow{P}\left(\mathbf{e}, \{\mathbf{e}\}^{\downarrow}\right) = 1 - \sum_{i \in z(\mathbf{e})} \alpha_i;$$

$$\mathbf{e}' = \mathbf{e} + \mathbf{s}_i: \quad \mu(\mathbf{e}, \mathbf{e} + \mathbf{s}_i) \overleftarrow{P}\left(\mathbf{e} + \mathbf{s}_i, \{\mathbf{e}\}^{\downarrow}\right) = -1 \cdot \beta_i.$$

Therefore, we get

$$\mathbf{P}^*(\mathbf{e}, \mathbf{e}) = \mu(\mathbf{e}, \mathbf{e}) \overleftarrow{P}\left(\mathbf{e}, \{\mathbf{e}\}^{\downarrow}\right) + \sum_{i \in z(\mathbf{e})} \mu(\mathbf{e}, \mathbf{e} + s_i) \overleftarrow{P}\left(\mathbf{e} + \mathbf{s}_i, \{\mathbf{e}\}^{\downarrow}\right)$$

$$= 1 - \sum_{i \in z(\mathbf{e})} (\alpha_i + \beta_i). \qquad \qquad \Box$$

It is worth mentioning that the condition for $\downarrow$-Möbius monotonicity (i.e., $\sum_{i=1}^{d} (\alpha_i + \beta_i) \leq 1$) is equivalent to the condition that all eigenvalues of $\mathbf{P}$ are non-negative.

The time to absorption of the above defined dual chain has the following "balls and bins" interpretation. Consider n multinomial trials with cell probabilities $p_i = \alpha_i + \beta_i, i = 1, \ldots, d$ and $p_{d+1} = 1 - \sum_{i=1}^{d} (\alpha_i + \beta_i)$. Then, the time to absorption of the dual chain $\mathbf{P}^*$ is equal (in distribution) to the waiting time until all cells are occupied. To be more specific, let T be the waiting time until all cells $1, \ldots, d$ are occupied and let A_n be the event that at least one cell is empty. Then, since T is some SST for $\mathbf{P}$, we have

$$s\left(\delta_{(0,\ldots,0)}\mathbf{P}^n, \pi\right) \leq P(T > n) = P(A_n).$$

In particular, for $\alpha_i = \beta_i = \frac{1}{2d}$, we have $\mathbf{P}(\mathbf{e}, \mathbf{e}) = 1/2$, and $\mathbf{P}^*(\mathbf{e}, \mathbf{e}) = \frac{|\mathbf{e}|}{d}$. Moreover, T is equal in distribution to $\sum_{i=1}^{d} N_i$, where (N_i) are independent, N_i has geometric distribution with parameter $\frac{i}{d}$. In the "balls and bins" scheme, $p_i = \frac{1}{d}$, $i = 1, \ldots, d$, and $p_{d+1} = 0$. From the coupon collector's problem solution, we recover a well known bound

$$s\left(\delta_{(0,\ldots,0)}\mathbf{P}^{d\log d + cn}, \pi\right) \leq P(T > d\log d + cn) \leq e^{-c}, \quad c > 0.$$

4.1 Further research

The problem of finding SSD chains for walks on the cube, which are not nearest neighbor walks, is open and seems to be a difficult one. Moreover, a more difficult task is to find SSD chains which have an upper triangular form (potentially useful for finding bounds on times to absorption). We have some observations for three dimensional cubes which might be of some interest. Consider the random walk on the three-dimensional cube, $\mathbb{E} = \{0, 1\}^3$, which is a special case of the random walk given in (4.1). We define on $\mathbb{E}$ the partial ordering: for all $\mathbf{e} = (e_1, e_2, e_3) \in \mathbb{E}$, $\mathbf{e}' = (e_1', e_2', e_3') \in \mathbb{E}$, $\mathbf{e} \preceq \mathbf{e}'$ iff $e_1 \leq e_1', e_2 \leq e_2', e_3 \leq e_3'$.

We consider the transition matrix $\mathbf{P}$ under the state space enumeration: $\mathbf{e}_1 = (0, 0, 0)$, $\mathbf{e}_2 = (1, 0, 0)$, $\mathbf{e}_3 = (0, 1, 0)$, $\mathbf{e}_4 = (0, 0, 1)$, $\mathbf{e}_5 = (1, 1, 0)$, $\mathbf{e}_6 = (1, 0, 1)$,

$\mathbf{e}_7 = (0, 1, 1)$, $\mathbf{e}_8 = (1, 1, 1)$ of the form

$$
\begin{bmatrix}
1-3\alpha & \alpha & \alpha & \alpha & 0 & 0 & 0 & 0 \\
\beta & 1-\beta-2\alpha & 0 & 0 & \alpha & \alpha & 0 & 0 \\
\beta & 0 & 1-\beta-2\alpha & 0 & \alpha & 0 & \alpha & 0 \\
\beta & 0 & 0 & 1-\beta-2\alpha & 0 & \alpha & \alpha & 0 \\
0 & \beta & \beta & 0 & 1-2\beta-\alpha & 0 & 0 & \alpha \\
0 & \beta & 0 & \beta & 0 & 1-2\beta-\alpha & 0 & \alpha \\
0 & 0 & \beta & \beta & 0 & 0 & 1-2\beta-\alpha & \alpha \\
0 & 0 & 0 & 0 & \beta & \beta & \beta & 1-3\beta
\end{bmatrix}
$$

with the dual $\mathbf{P}^*$

$$
\begin{bmatrix}
1-3\alpha-3\beta & \beta+\alpha & \beta+\alpha & \beta+\alpha & 0 & 0 & 0 & 0 \\
0 & 1-2\alpha-2\beta & 0 & 0 & \beta+\alpha & \beta+\alpha & 0 & 0 \\
0 & 0 & 1-2\alpha-2\beta & 0 & \beta+\alpha & 0 & \beta+\alpha & 0 \\
0 & 0 & 0 & 1-2\alpha-2\beta & 0 & \beta+\alpha & \beta+\alpha & 0 \\
0 & 0 & 0 & 0 & 1-\beta-\alpha & 0 & 0 & \beta+\alpha \\
0 & 0 & 0 & 0 & 0 & 1-\beta-\alpha & 0 & \beta+\alpha \\
0 & 0 & 0 & 0 & 0 & 0 & 1-\beta-\alpha & \beta+\alpha \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{bmatrix}
$$

$$(4.2)$$

One possibility to extend the model to allow up–down jumps not only to neighboring states is to take powers of the nearest neighbor transitions matrix $\mathbf{P}$, that is, to look at two step chain. It turns out that the matrix $\mathbf{P}^2$ is again Möbius monotone, and has a dual with an upper-triangular form if $\alpha = \beta$.

Another way to modify the nearest neighbor walk is to transform some rows of $\mathbf{P}$ to get distributions bigger in the supermodular ordering. To be more precise, recall that we say that two random elements X, Y of $\mathbb{E}$ are supermodular stochastically ordered (and write $X \prec_{sm} Y$ or $Y \succ_{sm} X$) if $Ef(X) \leq Ef(Y)$ for all supermodular functions, i.e., functions such that, for all $x, y \in \mathbb{E}$,

$$f(x \wedge y) + f(x \vee y) \geq f(x) + f(y).$$

A simple sufficient criterion for $\prec_{sm}$ order when $\mathbb{E}$ is a discrete (countable) lattice is given as follows.

Lemma 4.1 *Let P_1 be a probability measure on a discrete lattice ordered space $\mathbb{E}$ and assume that for not comparable points $x \neq y \in \mathbb{E}$ we have $P_1(x) \geq \kappa$ and $P_1(y) \geq \kappa$ for some $\kappa > 0$. Define a new probability measure P_2 on $\mathbb{E}$ by*

$$
\begin{aligned}
P_2(x) &= P_1(x) - \kappa, & P_2(x \vee y) &= P_1(x \vee y) + \kappa, \\
P_2(y) &= P_1(y) - \kappa, & P_2(x \wedge y) &= P_1(x \wedge y) + \kappa, \\
P_2(z) &= P_1(z) \quad \textit{otherwise.}
\end{aligned}
$$

$$(4.3)$$

Then $P_1 \prec_{sm} P_2$.

If in Lemma 4.1 the state space $\mathbb{E}$ is the set of all subsets of a finite set (i.e., the cube) then the transformation described in (4.3) is called as in Li and Xu [16] a *pairwise g^+ transform*, and Lemma 4.1 specializes then to their Proposition 5.5.

If we modify rows numbered 1, 3, 6, 8 by such a transformation (notice that $\mathbf{e}_1, \mathbf{e}_3, \mathbf{e}_6, \mathbf{e}_8$ lie on a symmetry axis), that is, we consider an up–down walk which allows jumps *not only* to the nearest neighbors, then it can be checked that it is Möbius monotone, and the dual matrix again has an upper-triangular form, for an appropriate selection of α and κ.

An upper-triangular form of dual matrices gives us the corresponding eigenvalues since they are equal to the diagonal elements. They can be used to find bounds on the speed of convergence to stationarity via

$$d\left(\nu\mathbf{P}^n, \pi\right) \leq s\left(\nu\mathbf{P}^n, \pi\right) \leq P\left(T^* > n\right).$$

If $\nu = \delta_{\mathbf{e}_1}$, $\mathbf{P}^*$ has an upper-triangular form, and in addition it has positive values only directly above the diagonal then $T^* = \sum_{i=1}^{M-1} N_i$, where N_i are independent geometric random variables with parameters $1 - \lambda_i$, $i = 1, \ldots, M-1$, where $\lambda_1, \ldots, \lambda_M = 1$ denote the diagonal entries of $\mathbf{P}^*$. This case corresponds to a skip-free structure to the right as described for example by Fill [12]. In other cases, it is possible to bound $P(T^* > n)$ by $P(T' > n)$, where T' is the time to absorption in a reduced chain representing the stochastically maximal passage time from $\mathbf{e}_1$ to $\mathbf{e}_M$. We will skip details of such a possibility giving, however, an example which illustrates this idea. Consider $\mathbf{P}^*$ given above in (4.2). Analyzing all possible paths from $\mathbf{e}_1$ to $\mathbf{e}_M$, we see that T^* is stochastically bounded by T', which is the time to absorption in a chain with the following transition matrix

$$\mathbf{P}' = \begin{bmatrix} 1 - 3(\alpha + \beta) & 3(\alpha + \beta) & 0 & 0 \\ 0 & 1 - 2(\alpha + \beta) & 2(\alpha + \beta) & 0 \\ 0 & 0 & 1 - (\alpha + \beta) & \alpha + \beta \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

We have $T' = \sum_{i=1}^{3} N_i$, where N_i are independent geometric random variables with parameters $i(\alpha + \beta)$, $i = 1, 2, 3$. The expected time to absorption is $ET' = \frac{11}{6}(\alpha + \beta)$, and using Markov inequality we have (for any $c > 0$)

$$d\left(\nu\mathbf{P}^n, \pi\right) \leq s\left(\nu\mathbf{P}^n, \pi\right) \leq P\left(T^* > n\right) \leq P\left(T' > n\right) \leq \frac{ET}{n} = \frac{1}{c}$$

for $n = c \cdot \frac{11}{6}(\alpha + \beta)$.

There are several other examples of chains which are Möbius monotone on some other state spaces. We shall study this topic in a subsequent paper.

References

1. Aldous, D.J.: Finite-time implications of relaxation times for stochastically monotone processes. Probab. Theory Relat. Fields **77**, 137–145 (1988)
2. Aldous, D.J., Diaconis, P.: Shuffling cards and stopping times. Am. Math. Mon. **93**, 333–348 (1986)
3. Aldous, D.J., Diaconis, P.: Strong uniform times and finite random walks. Adv. Appl. Math. **8**, 69–97 (1987)
4. Brown, M.: Consequences of monotonicity for Markov transition functions. Technical report, City college, CUNY (1990)
5. Brown, M., Shao, Y.S.: Identifying coefficients in the spectral representation for first passage time distributions. Probab. Eng. Inf. Sci. **1**, 69–74 (1987)
6. Diaconis, P.: Group Representations in Probability and Statistics. IMS, Hayward (1988)
7. Diaconis, P., Fill, J.A.: Strong stationary times via a new form of duality. Ann. Probab. **18**, 1483–1522 (1990)
8. Diaconis, P., Miclo, L.: On times to quasi-stationarity for birth and death processes. J. Theor. Probab. **22**, 558–586 (2009)
9. Dieker, A.B., Warren, J.: Series Jackson networks and non-crossing probabilities. Math. Oper. Res. **35**, 257–266 (2010)
10. Falin, G.I.: Monotonicity of random walks in partially ordered sets. Russ. Math. Surv. **43**, 167–168 (1988)
11. Fill, J.A.: The passage time distribution for a birth-and-death chain: Strong stationary duality gives a first stochastic proof. J. Theor. Probab. **22**, 543–557 (2009)
12. Fill, J.A.: On hitting times and fastest strong stationary times for skip-free and more general chains. J. Theor. Probab. **22**, 587–600 (2009)
13. Lorek, P., Szekli, R.: On the speed of convergence to stationarity via spectral gap: queueing networks with breakdowns and repairs (submitted to JAP). arXiv:1101.0332 [math.PR]
14. Massey, W.A.: Stochastic ordering for Markov processes on partially ordered spaces. Math. Oper. Res. **12**, 350–367 (1987)
15. Rota, G.C.: On the foundations of combinatorial theory I. Theory of Möbius functions. Z. Wahrscheinlichkeitstheor. **2**, 340–368 (1964)
16. Xu, S.H., Li, H.: Majorization of weighted trees: A new tool to study correlated stochastic systems. Math. Oper. Res. **25**, 298–323 (2000)