ESAIM: Probability and Statistics Will be set by the publisher
URL: `http://www.emath.fr/ps/`

ANTIDUALITY AND MÖBIUS MONOTONICITY: GENERALIZED COUPON COLLECTOR PROBLEM *

Paweł Lorek [1]

Abstract. For a given absorbing Markov chain X^* on finite state space, the chain X is a sharp antidual of X^* if Fastest Strong Stationary time of X is equal, in distribution, to the absorption time of X^*. In this paper we show a systematic way of finding such antidual based on some partial ordering of the state space. We use the theory of Strong Stationary Duality developed recently for Möbius monotone Markov chains. We give several sharp antidual chains for Markov chain corresponding to Generalized Coupon Collector Problem. As a consequence - utilizing known results on limiting distribution of absorption time - we indicate separation cutoff (with its window size) in several chains.

1991 Mathematics Subject Classification. 60J10, 60G40, 06A06.

November 29, 2016.

1. Introduction

Strong Stationary Times (SST) are a probabilistic tool for bounding the rate of convergence to stationarity for Markov chains. Aldous and Diaconis [1], [2] gave several examples of chains where SST were found *ad hoc*. Later in [4] the authors introduced a more systematic way of finding SSTs. For a given general ergodic chain they showed that one can construct a so-called Strong Stationary Dual (SSD) chain, a chain whose absorption time is equal in distribution to some SST of the original chain. Moreover, they proved that there always exists *sharp* SSD, in the sense that its corresponding SST is stochastically the smallest, in which case it is called the *Fastest Strong Stationary Time* (FSST).

Their construction for general chains is purely theoretical (it involves the knowledge of the distribution of the chain at each step). However, they give a detailed recipe on how to construct such SSD assuming that the time reversed chain is stochastically monotone w.r.t. linear ordering. In particular, they consider birth and death chain, for which SST has the same distribution as absorption time in a dual chain, which turns out to be an absorbing birth and death chain. They also show that assuming that time reversed chain is stochastically monotone one can always construct set-valued SSD (see their Section 3.4 “greedy construction of a set-valued dual”). However, in this paper we actually start with some absorbing chain (which is SSD of some other one), calculating antidual chain in set-valued settings would be much more complicated. Recently, in [21] authors provided the recipe for constructiong SSD on the same state space for chains, whose time reversal is Möbius monotone w.r.t to some partial ordering of the state space. This significantly enlarges the class of chains for

Keywords and phrases: Markov chains; Strong Stationary Duality; Antiduality; Absorption times; Fastest Strong Stationary Times; Generalized Coupon Collector Problem; Double Dixie Cup Problem; separation cutoff; partial ordering; perfect simulation

* *Work supported by NCN Research Grant DEC-2013/10/E/ST1/00359*

[1] Mathematical Institute, University of Wrocław, pl. Grunwaldzki 2/4, 50-384 Wrocław, Poland. Email: Pawel.Lorek@math.uni.wroc.pl

which SSD can be found. In many chains there is usually some natural underlying ordering of the state space which is only partial. Moreover, the method yields the sharp SSD which is crucial for our applications.

Studying the rate of convergence of a chain to its stationary distribution, one is often interested in so-called *mixing time* (i.e., the time until the chain is "close" to its stationary distribution). However, sometimes we can say much more than just a mixing time by showing that so-called **cutoff** phenomenon occurs. Roughly speaking, this phenomenon describes a sharp transition in the convergence of the chain to its stationary distribution over a negligible period of time (cutoff window). There are two most commonly studied phenomena: separation cutoff and total variation cutoff, which differ in a distance used to measure the convergence (separation vs. total variation distance).

The total variation cutoff was first shown for the *random transposition* card shuffling in [6]. The name comes from [1], where authors showed that *top-to-random* card shuffling exhibits total variation cutoff. Separation cutoff has recently been studied in few contexts. For example: In [5] authors gave if and only if conditions for existence of separation cutoff for birth and death chains; necessary and sufficient condition for skip-free chains was given in [22]; some other specific chains were considered in [3]; in [10] author gives a formula for separation distance for *Tsetlin library* chain specifying weights for which there is and there is no separation cutoff.

Although many chains are believed to exhibit cutoff, proving that one occurs is often extremely challenging task. That is why there are relatively few examples for which cutoff has been rigorously proven, and even fewer with given window size. As mentioned before, FSST is equal in distribution to absorption time of the sharp SSD chain. Thus, there is a close relation between sharp SSD and separation cutoff. Roughly speaking, this cutoff can be studied by studying the limiting distribution of absorption time of the SSD. This can be extremely difficult task. However, since examples of chains with proven separation cutoff are always welcome, we can reverse the procedure: starting with some already absorbing chain we can try to find an ergodic **sharp antidual** chain (or even a class of such antidual chains). (Such approach was considered in [12] in context of birth and death chains only). Using this approach we will indicate separation cutoff time and window size in several examples of chains utilizing (nontrivial) results for the limiting distribution of the absorption time in some generalizations of the classical Coupon Collector Problem. That is why we need a recipe for sharp antidual chains, what will be given based on results from [21]. Most of the examples that follow deal with some product-type chains. It is however worth noting that taking a product of chains where each chain exhibits a cutoff does not have to yield a chain (on product space) exhibiting a cutoff. Such example was recently given in [17].

The absorption time of many absorbing chains is distributed as a mixture of sums of geometric random variables with parameters being the eigenvalues of transition matrix. E.g, absorption time of discrete time birth and death chain starting at minimal state with maximal one being absorbing is distributed as a sum of geometric random variables with such parameters, provided the chain is stochastically monotone. The result is usually attributed to Karlin and McGregor [15] or Keilson [16]. Fill [12] gave a first stochastic proof of this result using also the theory of SSD, later it was extended to skip-free Markov chains in Fill [11]. Miclo [23] showed that for large class of absorbing chains on finite state space, the absorption time is distributed as mixture of sums of geometric random variables. A natural question arises: *Given a mixture of sums of geometric random variables and some distribution π can we find an ergodic chain whose stationary distribution is π and whose FSST is equal in distribution to this mixture?*. Or, a special case of the question, *Given some distribution π can be construct an ergodic chain whose stationary distribution is π having deterministic FSST*? We provide positive answers to both questions (some assumptions on distributions are needed). In particular, we present two ergodic chains on completely different state spaces having the same FSST.

The main goals of the paper are: i) we give a systematic way (based on partial ordering of the state space and Möbius monotonicity) for finding a class of sharp antidual chains; ii) we present nontrivial antidual chains related to some generalizations of Coupon Collector Problem and, as a consequence, we show cutoff phenomena in some cases; iii) we present a construction of a chain with prescribed FSST and prescribed stationary distribution.

There is yet another potential application which served as a motivation for the paper (however, not exploited here): Given a probability distribution π on $\mathbb{E}$, how to simulate a sample from this distribution? Markov Chain Monte Carlo methods come with the answer: construct a chain with stationary distribution π and run it *long enough*. The most common algorithms for such constructions are Metropolis-Hastings algorithm and Gibbs sampler. Unfortunately, none says anything about the rate of convergence (each example must be studied separately). This paper suggests an alternative approach: given π on $\mathbb{E}$ find some absorbing chain on $\mathbb{E}$ and then calculate sharp antidual chain having this π as stationary distribution. Knowing, e.g., expectation and variance of absorption time, one can quite precisely determine the number of steps needed for simulation. Moreover, having a sharp SSD actually can allow for a *perfect simulation* from distribution π. One can construct an appropriate coupling of the absorbing chain and its antidual, so that stopping antidual chain when its SSD is absorbed yields an unbiased sample from π. The reader is referred for details to [4] (Section 2.4) or [13] (Section 1.1). We want to emphasize that utilizing this was not the purpose of this paper, and the stationary distributions which appear in most of the examples are of product form, which means we can easily simulate them coordinate by coordinate.

The paper is organized as follows. In Section 2 we introduce preliminaries on Strong Stationary Duality and separation cutoff. In Section 3 we recall the notion of Möbius monotonicity and give a matrix-form proof of the result from [21]. In Section 4 we present our main results. Firstly, in Section 4.1 in Theorem 4.1 we give a systematic way for finding a class of sharp antidual chains. Secondly, in Section 4.2 we introduce in details the chain corresponding to the Generalized Coupon Collector Problem and present sharp antidual chains in Theorems 4.4 and 4.6. Then, in Section 4.3, we proceed with presenting separation cutoff results for some cases. In Section 4.4 we present our results concerning construction of ergodic chain with prescribed stationary distribution and with prescribed FSST. Section 5 includes main proofs. Section 5.1 contains proofs of Theorems 4.4 and 4.6, whereas Section 5.2 contains the proof of Theorem 4.16.

2. Preliminaries

2.1. Strong Stationary Duality

Consider ergodic Markov chain $X \sim (\nu, \mathbf{P})$ on finite state space $\mathbb{E} = \{\mathbf{e}_1, \ldots, \mathbf{e}_M\}$ with initial distribution ν, stationary distribution π and transition matrix $\mathbf{P}$. Let $\mathbb{E}^* = \{\mathbf{e}^*_1, \ldots, \mathbf{e}^*_N\}$ be a state space of absorbing Markov chain $X^* \sim (\nu^*, \mathbf{P}^*)$, whose unique absorbing state is denoted by $\mathbf{e}^*_N$. Define Λ, a matrix of size $N \times M$, to be a *link* if it is a stochastic matrix with the property: $\Lambda(\mathbf{e}^*_N, \mathbf{e}) = \pi(\mathbf{e})$ for all $\mathbf{e} \in \mathbb{E}$. We say that X^* is a *Strong Stationary Dual* (SSD) of X with link Λ if

$$\nu = \nu^* \Lambda \quad \text{and} \quad \Lambda \mathbf{P} = \mathbf{P}^* \Lambda. \tag{1}$$

Diaconis and Fill [4] prove that then, the absorption time T^* of X^* is so called *Strong Stationary Time* (SST) for X. This is such a random variable T that X_T has distribution π and T is independent from X_T. The main application is in studying the rate of convergence of an ergodic chain to its stationary distribution, since for such random variable we always have: $d_{TV}(\nu\mathbf{P}^k, \pi) \leq sep(\nu\mathbf{P}^k, \pi) := \max_{\mathbf{e}\in\mathbb{E}} \left(1 - \nu\mathbf{P}^k(\mathbf{e})/\pi(\mathbf{e})\right) \leq P(T > k)$, where d_{TV} stands for *total variation distance*, and *sep* stands for *separation "distance"*. The corresponding T^* is **sharp** if $sep(\nu\mathbf{P}^k, \pi) = P(T^* > k)$. In such the case, T^* is called *the Fastest Strong Stationary Time* for X, which we denote by T_{FSST}. For more details on this duality consult [4]. Moreover, duality relation (1) allows for stochastic constructions, see, e.g., [12], where stochastic proof for passage time distribution for birth and death chain was given.

Note that once we fix $\mathbb{E}^*$ and link Λ, and if there exists right-inverse of Λ, i.e., Λ^{-1} we can simply calculate from (1):

$$\mathbf{P}^* = \Lambda \mathbf{P} \Lambda^{-1} \text{ and } \nu^* = \nu \Lambda^{-1}.$$

If the resulting $\mathbf{P}^*$ is a stochastic matrix and ν^* is a probability distribution, then (it will always correspond to absorbing chain) we have found a SSD. However, we can start with some already absorbing chain $\mathbf{P}^*$, then find some $\mathbb{E}$ and *some* probability distribution π on $\mathbb{E}$, and a link Λ, so that

$$\mathbf{P} = \Lambda^{-1}\mathbf{P}^*\Lambda \text{ and } \nu = \nu^*\Lambda.$$

If the resulting $\mathbf{P}$ is a stochastic matrix, then $X \sim (\nu, \mathbf{P})$ is an ergodic chain with stationary distribution π, and T^* (time to absorption of X^*) is a SST for X. In such the case, X is called **antidual** of X^*. Moreover, if we somehow know, that for some class of links relation (1) implies that T^* is sharp (see Corollary 3.6), then we can possibly find many different antiduals, which all have the same Fastest Strong Stationary Time T^*, which has a phase-type distribution. In such the case X is called **sharp antidual** of X^*.

2.2. **Separation cutoff**

The forthcoming Theorem 4.1 indeed gives a recipe on how to construct a sharp antidual chain X with specified stationary distribution π given absorbing chain X^*, both on the same state space. It means, that we have

$$sep(\nu\mathbf{P}^k, \pi) = P(T_{FSST} > k) = P(T^* > k). \tag{2}$$

Thus, studying the distribution of T_{FSST} is equivalent to study the distribution of T^*. Furthermore, the so-called separation cutoff can be studied by studying the properties of T^*. In what follows, we introduce the notion of separation cutoff. Since the dfntn involves increasing state space, we add a subscript (d) to transition matrices, distributions, state space and absorption time. Suppose we have a sequence of ergodic Markov chains $X_{(d)} \sim (\nu_{(d)}, \mathbf{P}_{(d)})$ indexed by $d = 1, 2, \ldots$ Denote by $\pi_{(d)}$ the stationary distribution of $X_{(d)}$. We say that this sequence exhibits a **separation cutoff at time** t_d with a **window size** $w_d = o(t_d)$ if

$$\begin{aligned} \lim_{c\to\infty} \ \limsup_{d\to\infty} \ & sep(\nu_{(d)}\mathbf{P}_{(d)}^{t_d+cw_d}, \pi_{(d)}) = 0, \\ \lim_{c\to\infty} \ \liminf_{d\to\infty} \ & sep(\nu_{(d)}\mathbf{P}_{(d)}^{t_d-cw_d}, \pi_{(d)}) = 1. \end{aligned} \tag{3}$$

If the convergence to stationarity is measured in total variation distance, we say about **total variation cutoff**.

3. Möbius monotonicity and duality

In general, there is no recipe on how to find SSD, i.e., a triplet $\mathbb{E}^*, \mathbf{P}^*, \Lambda$. In [4] authors give a recipe for dual on the same state space $\mathbb{E}^* = \mathbb{E}$ provided that time reversed chain $\overleftarrow{X}$ is stochastically monotone with respect to total ordering. In [21] we give an extension of this result to state spaces which are only partially ordered by $\preceq$. Then, provided time reversed chain $\overleftarrow{X}$ is *Möbius* monotone (plus some conditions on initial distribution), we give a formula for sharp SSD on the same state space $\mathbb{E}^* = \mathbb{E}$.

The Möbius monotonicity seems to be a natural one for extension of main result from [4] to partially ordered state spaces. In [19] we show that it is equivalent to the existence of Siegmund dual of a chain with given partial ordering. For linearly ordered state space, stochastic monotonicity of a chain is required for existence of Siegmund dual (see [27]), and stochastic monotonicity of time reversal is required for existence of SSD with link being a truncated stationary distribution (see [4]). Both results fail for non-linear orderings, since both require Möbius monotonicity, which, in general, is different than the stochastic one. The monotonicities are equivalent for linear ordering. For more relations between these (and not only) monotonicities consult [20], and for applications Siegmund duality to some generalizatoins of gambler's ruin problem consult [18].

We will introduce this monotonicity by trying to solve (1) with some given link Λ.

For function $f : \mathbb{E} \to \mathbf{R}$, by lower-case bold symbol $\boldsymbol{f}$ we denote the row vector $\boldsymbol{f} = (f(\mathbf{e}_1), \ldots, f(\mathbf{e}_M))$.

The idea is to find a SSD $\mathbf{P}^*$ on the same state space $\mathbb{E}^* = \mathbb{E}$ with link, whose row corresponding to $\mathbf{e}$ is a stationary distribution of X truncated to $\{\mathbf{e}\}^\downarrow := \{\mathbf{e}' : \mathbf{e}' \preceq \mathbf{e}\}$, i.e.,

$$\Lambda(\mathbf{e}_i, \mathbf{e}_j) = \frac{\pi(\mathbf{e}_j)}{\sum_{\mathbf{e}':\mathbf{e}'\preceq\mathbf{e}_i} \pi(\mathbf{e}')}\mathbf{1}(\mathbf{e}_j \preceq \mathbf{e}_i). \tag{4}$$

Note that $\forall(\mathbf{e}\in\mathbb{E})$ we have $\Lambda(\mathbf{e}_M,\mathbf{e})=\pi(\mathbf{e})$, as required. For a given ordering let $\mathbf{C}(\mathbf{e}_i,\mathbf{e}_j)=\mathbf{1}(\mathbf{e}_i\preceq\mathbf{e}_j)$. For the partial ordering we require only that the state which is absorbing for X^*, denoted throughout the paper by $\mathbf{e}_M$, is the unique maximal one (i.e., $\mathbf{C}(\mathbf{e}_M,\mathbf{e}_j)=\mathbf{1}(\mathbf{e}_j=\mathbf{e}_M)$). We always identify ordering $\preceq$ with matrix $\mathbf{C}$, keeping in mind, that enumeration of states in $\mathbf{C}$ and $\mathbf{P}$ must be the same. Then, the link can be written in matrix form:

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

where $\mathbf{diag}(\boldsymbol{g})$ is a diagonal matrix with entries $g(\mathbf{e}_1),\ldots,g(\mathbf{e}_M)$. The states can always be rearranged in such a way that $\mathbf{C}(\mathbf{e}_i,\mathbf{e}_j)=1$ implies $i\leq j$, what means that $\mathbf{C}$, and thus Λ, are invertible. Often, $\mu\equiv\mathbf{C}^{-1}$ is called a *Möbius function* of partial order $\preceq$. Solving (1) for $\mathbf{P}^*$ yields (recall that the transitions of time reversed chains are given by $\overleftarrow{\mathbf{P}}=(\mathbf{diag}(\boldsymbol{\pi}))^{-1}\mathbf{P}^T(\mathbf{diag}(\boldsymbol{\pi})))$

$$\begin{aligned}\mathbf{P}^*=\Lambda\mathbf{P}\Lambda^{-1}&=(\mathbf{diag}(\boldsymbol{\pi}\mathbf{C}))^{-1}\mathbf{C}^T\mathbf{diag}(\boldsymbol{\pi})\mathbf{P}\mathbf{diag}(\boldsymbol{\pi})^{-1}(\mathbf{C}^{-1})^T(\mathbf{diag}(\boldsymbol{\pi}\mathbf{C}))\\&=(\mathbf{diag}(\boldsymbol{\pi}\mathbf{C})(\mathbf{C}^{-1}\overleftarrow{\mathbf{P}}\mathbf{C})(\mathbf{diag}(\boldsymbol{\pi}\mathbf{C}))^{-1})^T,\end{aligned}$$

which is a stochastic matrix if and only if $(\mathbf{C}^{-1}\overleftarrow{\mathbf{P}}\mathbf{C})\geq 0$ (each entry nonnegative), in other words, we say that $\overleftarrow{\mathbf{P}}$ is Möbius monotone. This way we proved the main part of Theorem 2 of [21]. We include it here, since this is a little bit different (matrix-form) proof. We will restate the thrm for completeness, introducing formal dfntns of monotonicities first.

Definition 3.1. X is Möbius monotone if $\mathbf{C}^{-1}\mathbf{P}\mathbf{C}\geq 0$ (each entry nonnegative). In terms of transition probabilities, it means that

$$\forall(\mathbf{e}_i,\mathbf{e}_j\in\mathbb{E})\quad\sum_{\mathbf{e}\succeq\mathbf{e}_i}\mu(\mathbf{e}_i,\mathbf{e})\mathbf{P}(\mathbf{e},\{\mathbf{e}_j\}^{\downarrow})\geq 0,$$

where $\mathbf{P}(\mathbf{e},\{\mathbf{e}_j\}^{\downarrow})=\sum_{\mathbf{e}':\mathbf{e}'\preceq\mathbf{e}_j}\mathbf{P}(\mathbf{e},\mathbf{e}')$.

Recall that for Möbius function we always have $\mu(\mathbf{e}_i,\mathbf{e})=0$ whenever $\mathbf{e}\not\preceq\mathbf{e}_i$.

Definition 3.2. A function $f:\mathbb{E}\to\mathbf{R}$ is Möbius monotone if $\mathbf{f}(\mathbf{C}^T)^{-1}\geq 0$ (each entry nonnegative). It means that

$$\forall(\mathbf{e}_i\in\mathbb{E})\quad\sum_{\mathbf{e}:\mathbf{e}\succeq\mathbf{e}_i}\mu(\mathbf{e}_i,\mathbf{e})f(\mathbf{e})\geq 0.$$

Remark 3.3. In Lorek, Szekli [21] this Möbius monotonicity (of both, function and chain) was called $^{\downarrow}$*-Möbius monotonicity* (see Definitions 2.1 and 2.2 therein).

Definition 3.4. X is $^{\uparrow}$-Möbius monotone if $(\mathbf{C}^T)^{-1}\mathbf{P}\mathbf{C}^T\geq 0$ (each entry nonnegative).

Theorem 3.5 (Theorem 2 of [21])**.** *Let $X\sim(\nu,\mathbf{P})$ 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 maximal state $\mathbf{e}_M$, and with stationary distribution π. Assume that*

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

(ii) *time reversed chain $\overleftarrow{X}$ is Möbius monotone.*

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

$$\Lambda=(\mathbf{diag}(\pi\mathbf{C}))^{-1}\mathbf{C}^T\mathbf{diag}(\pi).\tag{5}$$

Let $H(\mathbf{e})=\sum_{\mathbf{e}'\preceq\mathbf{e}}\pi(\mathbf{e}')$. SSD chain is uniquely determined by

$$\begin{aligned}\nu^*&=\nu\Lambda^{-1}\quad &i.e.,\quad \nu^*(\mathbf{e}_i)&=H(\mathbf{e}_i)\sum_{\mathbf{e}:\mathbf{e}\succeq\mathbf{e}_i}\mu(\mathbf{e}_i,\mathbf{e})g(\mathbf{e}),\\ \mathbf{P}^*&=\Lambda\mathbf{P}\Lambda^{-1},\quad &i.e.,\quad \mathbf{P}^*(\mathbf{e}_i,\mathbf{e}_j)&=\frac{H(\mathbf{e}_j)}{H(\mathbf{e}_i)}\sum_{\mathbf{e}:\mathbf{e}\succeq\mathbf{e}_j}\mu(\mathbf{e}_j,\mathbf{e})\overleftarrow{\mathbf{P}}(\mathbf{e},\{\mathbf{e}_i\}^{\downarrow}).\end{aligned}$$

The following Corollary will play a crucial role:

Corollary 3.6. *The SSD constructed in Theorem 3.5 is* ***sharp****.*

Proof. The link given in (5) is lower-triangular, thus, by Remark 2.39 in [4], the resulting SSD is sharp. □

4. Main results

4.1. General procedure for sharp anti-dual chains

The main contribution is a systematic way of finding sharp antidual (on the same state space $\mathbb{E} = \mathbb{E}^*$) chain of some given already absorbing chain $X^* \sim (\nu^*, \mathbf{P}^*)$ with unique absorbing state $\mathbf{e}_M$. The idea is clear from the previous section: introduce some partial ordering and some distribution π on $\mathbb{E}$. Then solve $\Lambda\mathbf{P} = \mathbf{P}^*\Lambda$ for $\mathbf{P}$ with link given in (5). If the resulting matrix is nonnegative, it will be a stochastic matrix of ergodic Markov chain X with stationary distribution π. Moreover, changing π and/or ordering usually will yield a different sharp antidual. It means we can have a class of chains, all having the same Fastest Strong Stationary Time T_{FFST}.

Fix some partial ordering $\preceq$ on $\mathbb{E}^*$ (expressed by $\mathbf{C}$) having unique maximal state $\mathbf{e}_M$ and some distribution π on $\mathbb{E}$. For given $\mathbf{P}^*$ define

$$\widehat{\mathbf{P}^*} = \mathbf{diag}(\boldsymbol{\pi}\mathbf{C})\mathbf{P}^*(\mathbf{diag}(\boldsymbol{\pi}\mathbf{C}))^{-1}.$$

With slight abuse of notation we will assume that $\widehat{\mathbf{P}^*}$ is $^{\uparrow}$-Möbius monotone meaning that $(\mathbf{C}^T)^{-1}\widehat{\mathbf{P}^*}\mathbf{C}^T \geq 0$. Definition 3.4 was stated for Markov chain X with transition matrix $\mathbf{P}$, note however that $\widehat{\mathbf{P}^*}$ does not have to be a stochastic matrix.

Theorem 4.1. *Let $\mathbf{X}^* \sim (\nu^*, \mathbf{P}^*)$ be an absorbing Markov chain on $\mathbb{E}^* = \{\mathbf{e}_1, \ldots, \mathbf{e}_M\}$ with unique absorbing state $\mathbf{e}_M$. Let $\mathcal{C}$ be the class of all partial orderings on $\mathbb{E}^*$ with $\mathbf{e}_M$ being unique maximal state. Consider the class of pairs of distributions and partial orderings:*

$$\mathcal{P}(\mathbf{P}^*) = \left\{(\pi, \mathbf{C}) : \mathbf{C} \in \mathcal{C}, \widehat{\mathbf{P}^*} \text{ is } ^{\uparrow}\text{-Möbius monotone}\right\}.$$

Then, for any $(\pi, \mathbf{C}) \in \mathcal{P}(\mathbf{P}^*)$ the chain $X \sim (\nu, \mathbf{P})$ with the link Λ defined in (5) and with

$$\nu = \nu^*\Lambda, \quad \mathbf{P} = (\mathbf{diag}(\boldsymbol{\pi}))^{-1}(\mathbf{C}^T)^{-1}\widehat{\mathbf{P}^*}\mathbf{C}^T\mathbf{diag}(\boldsymbol{\pi})$$

is a sharp antidual for $\mathbf{P}^*$, i.e., $\mathbf{P}^*$ is a sharp SSD for $\mathbf{P}$. Equivalently, $\mathbf{P} = \Lambda^{-1}\mathbf{P}^*\Lambda$, where, for given π and $\mathbf{C}$, the link is defined in (5).

Proof. Since ν^* is a distribution on $\mathbb{E}$ and Λ is a stochastic matrix, ν is a distribution on $\mathbb{E}$. By assumption that $\widehat{\mathbf{P}^*}$ is $^{\uparrow}$-Möbius monotone, the matrix $\mathbf{P}$ is nonnegative. We will show that π is its stationary distribution. Let $\boldsymbol{\eta} = (0, \ldots, 0, 1)$. Last row of Λ is equal to $\boldsymbol{\pi}$ what can be expressed as $\boldsymbol{\eta}\Lambda = \boldsymbol{\pi}$, thus $\boldsymbol{\eta} = \boldsymbol{\pi}\Lambda^{-1}$. We have

$$\boldsymbol{\pi}\mathbf{P} = \boldsymbol{\pi}\Lambda^{-1}\mathbf{P}^*\Lambda = \boldsymbol{\eta}\mathbf{P}^*\Lambda = \boldsymbol{\eta}\Lambda = \boldsymbol{\pi}.$$

Now we will show that $\mathbf{P}(1, \ldots, 1)^T = (1, \ldots, 1)^T$. We have

$$\begin{aligned}
\mathbf{P}(1, \ldots, 1)^T &= (\mathbf{diag}\boldsymbol{\pi})^{-1}(\mathbf{C}^T)^{-1}\widehat{\mathbf{P}^*}\mathbf{C}^T\mathbf{diag}(\boldsymbol{\pi})(1, \ldots, 1)^T \\
&= (\mathbf{diag}\boldsymbol{\pi})^{-1}(\mathbf{C}^T)^{-1}\widehat{\mathbf{P}^*}\mathbf{C}^T\boldsymbol{\pi}^T = (\mathbf{diag}\boldsymbol{\pi})^T(\mathbf{C}^T)^{-1}\widehat{\mathbf{P}^*}(\boldsymbol{\pi}\mathbf{C})^T \\
&= (\mathbf{diag}\boldsymbol{\pi})^{-1}(\mathbf{C}^T)^{-1}\mathbf{diag}(\boldsymbol{\pi}\mathbf{C})\mathbf{P}^*(\mathbf{diag}(\boldsymbol{\pi}\mathbf{C}))^{-1}(\boldsymbol{\pi}\mathbf{C})^T \\
&= (\mathbf{diag}\boldsymbol{\pi})^{-1}(\mathbf{C}^T)^{-1}\mathbf{diag}(\boldsymbol{\pi}\mathbf{C})\mathbf{P}^*(1, \ldots, 1)^T \\
&= (\mathbf{diag}\boldsymbol{\pi})^{-1}(\mathbf{C}^T)^{-1}\mathbf{diag}(\boldsymbol{\pi}\mathbf{C})(1, \ldots, 1)^T = (1, \ldots, 1)^T
\end{aligned}$$

Thus, $\mathbf{P}$ is a stochastic matrix and thus $X \sim (\nu, \mathbf{P})$ is a Markov chain with stationary distribution π. Since (1) holds, X^* is a SSD for X. The chain X^* is constructed as in Theorem 3.5, by Corollary 3.6 it is sharp. $\square$

Remark 4.2. If, in addition, within ordering $\preceq$ we have a unique minimal state, say $\mathbf{e}_1$, and X^* starts from this state (i.e., $\nu^* = \delta_{\mathbf{e}_1}$), then the antidual chain also starts from this state, i.e. $\nu = \delta_{\mathbf{e}_1}$. This is the case in all examples that follow.

Remark 4.3. The condition that $\widehat{\mathbf{P}^*}$ is $(\pi, \mathbf{C})$-Möbius monotone is equivalent to nonnegativity of resulting antidual $\mathbf{P}$. In examples, it is often more convenient to calculate Λ and Λ^{-1} directly.

4.2. Antidual chains for Generalized Coupon Collector Problem

Consider d different types of coupons. These are sampled independently with replacement. Sampled types are recorded. For $1 \leq k \leq d$ let $p_k > 0$ be the probability that the coupon of type k is sampled, with $\sum_{k=1}^d p_k \leq 1$. With remnant probability, i.e., with probability $1 - \sum_{k=1}^d p_k$ no new type is sampled. We start with no coupons of any type. Let T^* be the number of steps it takes to collect N_j coupons of type j, $j = 1, \ldots, d$ for some fixed integers $N_1, \ldots, N_d$. The distribution of T^* is the time to absorption in the state $(N_1, \ldots, N_d)$ of the chain $X^* \sim (\nu^*, \mathbf{P}^*)$ on the state space $\mathbb{E}^* = \{(i_1, \ldots, i_d) : 0 \leq i_j \leq N_j, 1 \leq j \leq d\}$ with initial distribution $\nu^* = \delta_{(0,\ldots,0)}$ and the following transition matrix:

$$\mathbf{P}^*((i_1, \ldots, i_d), (i'_1, \ldots, i'_d)) = \begin{cases} p_j & \text{if} \quad i'_j = i_j + 1, i'_k = i_k, k \neq j, \\ 1 - \displaystyle\sum_{\substack{k=1,\ldots,d \\ i_k \neq N_k}} p_k & \text{if} \quad i'_j = i_j, 1 \leq j \leq d. \end{cases} \tag{6}$$

We refer to $\mathbf{P}^*$ as to *Generalized Coupon Collector chain*. The case $N_j = 1, j = 1, \ldots, d$ and $p_k = 1/d$ is *the classic Coupon Collector Problem*, which has a long history, see for example [9]. The term *generalized* is not unique. It is used when sequence $\{p_k\}$ is general but $N_1 = \ldots = N_d = 1$ (e.g., [24]) or when $p_k = 1/d$ but we are to collect more coupons of each type (see, e.g., [25], [7]). Although the chain $\mathbf{P}^*$ given in (6) includes both mentioned generalizations, we consider two antidual chains for two different cases separately:

a) for general $N_j \geq 1$ and $p_j, j = 1, \ldots, d$ with uniform stationary distribution of antidual chain;
b) for general p_j but $N_j = 1, j = 1, \ldots, d$ with more general stationary distribution of antidual chain (including uniform one as special case).

The proofs are postponed to Section 5.1.

For convenience denote $\mathbf{i} = (i_1, \ldots, i_d)$ and $\mathbf{i}^{(k)} = (i_1^{(k)}, \ldots, i_d^{(k)})$. Define $\mathbf{s}_k := (0, \ldots, 1, \ldots, 0)$ (with 1 on position k).

Case: general $N_j \geq 1$ and $p_j, j = 1, \ldots, d$ and uniform stationary distribution of antidual

Theorem 4.4. *Let $X^* \sim (\nu^*, \mathbf{P}^*)$ be a Generalized Coupon Collector chain with matrix given in (6) with fixed integers $N_j \geq 1, j = 1, \ldots, d$. Moreover, assume that*

$$\sum_{j=1}^d \left(\frac{2N_j + 1}{N_j + 1}\right) p_j \leq 1. \tag{7}$$

Then the chain $X \sim (\nu, \mathbf{P})$ with $\nu = \delta_{(0,\ldots,0)}$ and with transition matrix

$$\mathbf{P}(\mathbf{i}^{(1)},\mathbf{i}^{(2)}) =$$

$$\begin{cases} \dfrac{i_k^{(1)}+1}{i_k^{(1)}+2}p_k & \textit{if} \quad \mathbf{i}^{(2)} = \mathbf{i}^{(1)} + \mathbf{s}_k, \\ \left(\dfrac{\mathbf{1}(i_k^{(1)} < N_k)}{(i_k^{(1)}+1)(i_k^{(1)}+2)} + \dfrac{\mathbf{1}(i_k^{(1)} = N_k)}{N_k+1}\right) p_k & \textit{if} \quad \mathbf{i}^{(2)} = \mathbf{i}^{(1)} - m\cdot \mathbf{s}_k \\ & \quad \textit{with} \ \ 1 \le m \le i_k, \\ 1 - \sum\limits_{j: i_j^{(1)} < N_j} \left(1 - \dfrac{1}{(i_j^{(1)}+1)(i_j^{(1)}+2)}\right) p_j - \sum\limits_{j: i_j^{(1)} = N_j} \dfrac{N_j}{N_j+1} p_j & \textit{if} \quad \mathbf{i}^{(2)} = \mathbf{i}^{(1)} \end{cases} \tag{8}$$

is an ergodic Markov chain with uniform distribution on $\mathbb{E} = \mathbb{E}^*$ which is a sharp antidual for $\mathbf{P}^*$.

Remark 4.5. The condition (7) is sufficient but not necessary. In particular, for $N_1 = \ldots = N_j = 1$ it is enough to have $\sum_{j=1}^d p_j \le 1$.

Roughly speaking, the antidual has the following transitions. Being in state $(i_1,\ldots,i_d)$ it can increase each coordinate by one (if feasible), it can stay in this state or it can change one of the coordinates to anything smaller. I.e., in one step it can only go to states $(i_1,\ldots,i_j+1,i_{j+1},\ldots,i_d)$, $(i_1,\ldots,i_j-m,i_{j+1},\ldots,i_d)$, $1\le m\le i_j$, or it can stay in $(i_1,\ldots,i_d)$. Changing some coordinate depends only on the value of this coordinate, and decreasing coordinate, say from i_j to $i_j - m$ is constant for all $1 \le m < i_j$ (the probability depends only on i_j and the formula itself is different on the border, i.e., when $i_j = N_j$, than on all other states).

Case: general p_j and $N_j = 1, j = 1,\ldots,d$ and non-uniform distribution of antidual.

Theorem 4.6. *Let $X^* \sim (\nu^*, \mathbf{P}^*)$ be a Generalized Coupon Collector chain with matrix given in (6). Assume that $N_1 = \ldots = N_d = 1$. Let $a_k \in (0,1)$ for $k = 1,\ldots,d$. Then, the chain $X \sim (\nu, \mathbf{P})$ on the same state space $\mathbb{E} = \mathbb{E}^* = \{0,1\}^d$ with initial distribution $\nu = \nu^* = \delta_{(0,\ldots,0)}$ and transition matrix:*

$$\mathbf{P}(\mathbf{i}^{(1)},\mathbf{i}^{(2)}) = \begin{cases} a_k p_k & \text{if} \quad \mathbf{i}^{(2)} = \mathbf{i}^{(1)} + \mathbf{s}_k, \\ 1 - \sum\limits_{j: i_j^{(1)}=0} a_j p_j - \sum\limits_{j: i_j^{(1)}=1} (1-a_j) p_j & \text{if} \quad \mathbf{i}^{(2)} = \mathbf{i}^{(1)}, \\ (1-a_k) p_k & \text{if} \quad \mathbf{i}^{(2)} = \mathbf{i}^{(1)} - \mathbf{s}_k, \end{cases} \tag{9}$$

is an ergodic Markov chain which is sharp antidual for $\mathbf{P}^*$. The stationary distribution is following:

$$\pi(\mathbf{e}) \quad = \quad \prod_{j=1}^d [a_j \mathbf{1}(i_j = 1) + (1-a_j)\mathbf{1}(i_j = 0)]. \tag{10}$$

Taking the following concrete sequences of a_k: $a_k = \frac{b}{a+b}$ or $a_k = \frac{1}{2}, j = 1,\ldots,d$ we obtain the following special cases:

Corollary 4.7. *The chains $X^{(i)} \sim (\nu, \mathbf{P}_i), i = 1,2$ with common initial distribution $\nu = \delta_{(0,\ldots,0)}$ and transition matrices*

$$\mathbf{P}_1(\mathbf{i}^{(1)},\mathbf{i}^{(2)}) = \begin{cases} \frac{1}{2}p_k & \text{if} \quad \mathbf{i}^{(2)} = \mathbf{i}^{(1)} + \mathbf{s}_k, \\ 1 - \frac{1}{2}\sum_{j=1}^{d} p_j & \text{if} \quad \mathbf{i}^{(2)} = \mathbf{i}^{(1)}, \\ \frac{1}{2}p_k & \text{if} \quad \mathbf{i}^{(2)} = \mathbf{i}^{(1)} - \mathbf{s}_k. \end{cases}$$

$$\mathbf{P}_2(\mathbf{i}^{(1)},\mathbf{i}^{(2)}) = \begin{cases} \frac{b}{a+b}p_k & \text{if} \quad \mathbf{i}^{(2)} = \mathbf{i}^{(1)} + \mathbf{s}_k, \\ 1 - \frac{b}{a+b}\sum_{j:i_j^{(1)}=0} p_r - \frac{a}{a+b}\sum_{j:i_j^{(1)}=1} p_r & \text{if} \quad \mathbf{i}^{(2)} = \mathbf{i}^{(1)}, \\ \frac{a}{a+b}p_k & \text{if} \quad \mathbf{i}^{(2)} = \mathbf{i}^{(1)} - \mathbf{s}_k. \end{cases}$$

and with the respective stationary distributions

$$\pi_1(\mathbf{i}) = \frac{1}{2^d}, \quad \pi_2(\mathbf{i}) = \frac{a^{d-|\mathbf{i}|}b^{|\mathbf{i}|}}{(a+b)^d}$$

are sharp antidual chains for $\mathbf{P}^*$ *given in (6), where* $|\mathbf{i}| = \sum_{j=1}^d i_j$*, called the level of* $\mathbf{i}$.

Remark 4.8. The antidual $X \sim (\nu, \mathbf{P})$ has transitions consistent with partial ordering, i.e., at each step it can stay or it can either change one coordinate from 0 to 1 or vice-versa. This is not the case for any distribution π (eg., for some π it can happen that two coordinates change at a time).

Remark 4.9. In [21] we considered chain on $\mathbb{E} = \{0,1\}^d$ with transition matrix $\mathbf{P}_3$ given by

$$\mathbf{P}_3(\mathbf{i}^{(1)},\mathbf{i}^{(2)}) = \begin{cases} \alpha_k & \text{if} \quad \mathbf{i}^{(2)} = \mathbf{i}^{(1)} + \mathbf{s}_k, \\ 1 - \sum_{j:i_j^{(1)}=0} \alpha_j - \sum_{j:i_j^{(1)}=1} \beta_j & \text{if} \quad \mathbf{i}^{(2)} = \mathbf{i}^{(1)}, \\ \beta_k & \text{if} \quad \mathbf{i}^{(2)} = \mathbf{i}^{(1)} - \mathbf{s}_k. \end{cases}$$

The chain is reversible with product form stationary distribution:

$$\pi_3(\mathbf{i}) = \prod_{j:i_j=0} \frac{\beta_j}{\alpha_j+\beta_j} \prod_{j:i_j=1} \frac{\alpha_j}{\alpha_j+\beta_j}. \tag{11}$$

We showed that the chain is Möbius monotone if and only if $\sum_{j=1}^d(\alpha_j+\beta_j) \leq 1$. As partial ordering, coordinate-wise was used. Then we obtained the following dual chain:

$$\mathbf{P}^*(\mathbf{i}^{(1)},\mathbf{i}^{(2)}) = \begin{cases} \alpha_k + \beta_k & \text{if} \quad \mathbf{i}^{(2)} = \mathbf{i}^{(1)} + \mathbf{s}_k, \\ 1 - \sum_{j:i_j^{(1)}=0} (\alpha_j+\beta_j) & \text{if} \quad \mathbf{i}^{(2)} = \mathbf{i}^{(1)}, \end{cases}$$

what is our absorbing dual (6) we started with, with $p_j = \alpha_j + \beta_j$ and $N_j = 1, j = 1, \ldots, d$. Note that $\mathbf{P}_3$ is a special case of $\mathbf{P}$ given in (9) with $a_j = \frac{\alpha_j}{\alpha_j+\beta_j}$.

Corollary 4.10. *The matrices* $\mathbf{P}$ *given in (8) and in (9) have eigenvalues of the form:*

$$\lambda_A = 1 - \sum_{k\in A} p_k, \qquad \textit{for}\ \ A \subseteq \{1,\ldots,d\}$$

(the multiplicity of which depends on the case).

Proof. We can order the states of X^* in such a way that $\mathbf{P}^*$ given in (6) is upper triangular, thus eigenvalues are the entries on the diagonal. If the link Λ is invertible (which is the case), then the transition matrices $\mathbf{P}$ and $\mathbf{P}^*$ of SSD have the same set of eigenvalues, what is a direct consequence of relation (1). □

Remark 4.11. Fix d and $N_j = N, j = 1,\ldots,d$. One can ask the following question: For what sequence $\{p_k\}$ is the associated T_{FSST} stochastically the smallest? Conjecture 2 in [7] suggests that this is in the case of equal probabilities $p_k = 1/d$.

4.3. Results on separation cutoff

Since obtained antidual chains are sharp (i.e., (2) holds), we can present a series of results on separation cutoff utilizing existing results on limiting distribution of T^*.

We start with the simplest chain corresponding to the classical Coupon Collector Problem.

Corollary 4.12. *Consider a sequence of Markov chains* $X_{(d)}$ *indexed by* $d = 1,2,\ldots$ *on* $\mathbb{E}_{(d)} = \{0,1\}^d$ *with initial distribution* $\nu_{(d)} = \delta_{(0,\ldots,0)}$ *and transition matrix* $\mathbf{P}_{(d)}$ *given in (9) with* $p_k = \frac{1}{d}$ *and any* $a_k \in (0,1)$ *for* $k = 1,\ldots,d$*. The stationary distribution* $\pi_{(d)}$ *is given in (10). The sequence exhibits a separation cutoff at time* $d\log d$ *with window size* d*.*

Proof. Denote the FSST of the chain by T^*_d. It is known that $ET^*_d = d\sum_{i=1}^d \frac{1}{i} \approx d\log d$. Moreover, $\frac{1}{d}(T^*_d - d\log d)$ converges in distribution (as $d\to\infty$) to a standard Gumbel random variable Z (with c.d.f $P(Z\le c) = e^{-e^{-c}}$), see [14].
Taking $t_d = d\log d$ and $w_d = d$ we have

$$\begin{aligned}
sep(\nu_{(d)}\mathbf{P}_d^{d\log d+cd}, \pi_d) &= P(T^*_d > d\log d + cd) = 1 - P\left(\tfrac{1}{d}(T^*_d - d\log d) \le c\right),\\
sep(\nu_{(d)}\mathbf{P}_d^{d\log d-cd}, \pi_d) &= P(T^*_d > d\log d - cd) = 1 - P\left(\tfrac{1}{d}(T^*_d - d\log d)) \le -c\right).
\end{aligned}$$

Taking the limits as $d\to\infty$ we have

$$\begin{aligned}
\limsup_{d\to\infty}\ sep(\nu_{(d)}\mathbf{P}_d^{d\log d+cd}, \pi_d) &= 1 - e^{-e^{-c}},\\
\liminf_{d\to\infty}\ sep(\nu_{(d)}\mathbf{P}_d^{d\log d-cd}, \pi_d) &= 1 - e^{-e^{c}}.
\end{aligned}$$

Taking the limit as $c\to\infty$ finishes the proof. □

Results on limiting distribution of T^*_d from [24] let us indicate separation cutoffs for cases with non-constant probabilities p_k. For example we can have the following Corollary.

Corollary 4.13. *Consider piecewise constant probability density function on* $[0,1]$*:*

$$f(y) = \lambda_j, \quad n_{j-1} < x \le n_j, \quad 1 \le j \le k,$$

where $\lambda_1,\ldots,\lambda_k > 0$ *and* $0 = n_0 < n_1 < \cdots n_k = 1$*. Without loss of generality assume that* $\lambda_1 < \lambda_2 < \ldots < \lambda_k$*. Consider a sequence of Markov chains* $X_{(d)}$ *indexed by* $d = 1,2,\ldots$ *on* $\mathbb{E}_{(d)} = \{0,1\}^d$ *with initial distribution*

$\nu_{(d)} = \delta_{(0,\dots,0)}$ and transition matrix $\mathbf{P}_{(d)}$ given in (9) with

$$p_k = \int_{(k-1)/d}^{k/d} f(y)dy, \quad k = 1, \dots, d$$

and any $a_k \in (0,1)$ for $k = 1, \dots, d$. The stationary distribution $\pi_{(d)}$ is given in (10). The sequence exhibits a separation cutoff at time $t_d = \frac{d}{\lambda_1}(\log d - \log(n_1))$ with window size $w_d = \frac{d}{\lambda_1}$.

Proof. We have

$$\begin{aligned} sep(\nu_{(d)}\mathbf{P}_d^{t_d+cd}, \pi_d) &= P\left(T_d^* > \tfrac{d}{\lambda_1}(\log d - \log(n_1)) + c\tfrac{d}{\lambda_1}\right) \\ &= 1 - P\left(\tfrac{1}{d}(T_d^* - \tfrac{1}{\lambda_1} d\log d) \le \tfrac{\log(n_1)}{\lambda_1} + \tfrac{c}{\lambda}\right). \end{aligned}$$

Lemma 3.1 in [24] implies that $\frac{1}{d}(T_d^* - \frac{1}{\lambda_1}d\log d)$ converges in distribution to random variable Z with c.d.f $P(Z \le c) = e^{-n_1 e^{-\lambda_1 c}}$. Thus, we have

$$\limsup_{d\to\infty} sep(\nu_{(d)}\mathbf{P}_d^{t_d+cd}, \pi_d) = 1 - e^{-n_1 e^{-\lambda_1\left(\frac{\log(n_1)}{\lambda_1} + \frac{c}{\lambda}\right)}} = 1 - e^{-e^{-c}}.$$

Similarly

$$sep(\nu_{(d)}\mathbf{P}_d^{t_d-cd}, \pi_d) = 1 - P\left(\tfrac{1}{d}(T_d^* - \tfrac{1}{\lambda_1}d\log d) \le \tfrac{\log(n_1)}{\lambda_1} - \tfrac{c}{\lambda}\right)$$

and

$$\liminf_{d\to\infty} sep(\nu_{(d)}\mathbf{P}_d^{t_d-cd}, \pi_d) = 1 - e^{-n_1 e^{-\lambda_1\left(\frac{\log(n_1)}{\lambda_1} - \frac{c}{\lambda}\right)}} = 1 - e^{-e^{c}}.$$

Taking limits as $c \to \infty$ finishes the proof. □

Next corollaries utilize results on time until some set of coupons is collected.

Corollary 4.14. *Consider a sequence of Markov chains $X_{(d)}$ indexed by $d = 1, 2, \dots$ on $\mathbb{E}_{(d)} = \{0, 1, \dots, N\}^d$ with initial distribution $\nu_{(d)} = \delta_{(0,\dots,0)}$ and transition matrix $\mathbf{P}_{(d)}$ given in (8) with $p_k = \frac{1}{d}$ and $N_1 = \dots = N_d = N \ge 2$ (so that (7) holds). The stationary distribution $\pi_{(d)}$ is uniform. The sequence of chains exhibits a separation cutoff at time $d\log d + (N-1)d\log\log d - d[\gamma - \log(N-1)!]$ with window size d, where $\gamma = 0.57721\dots$ is the Euler-Mascheroni constant.*

Proof. In [8] authors derived limiting distribution of T_d^* showing that

$$\frac{1}{d}(T_d^* - d\log d - (N-1)d\log\log d + d[\gamma - \log(N-1)!])$$

converges in distribution to a standard Gumbel random variable. Similar calculations as in Corollary 4.12 finish the proof. □

Recently authors in [7] extended the result of [8] obtaining the limiting distribution of T_d^* for $N_1 = \dots = N_d = N$ and for quite general choices of probabilities p_k. Let us indicate here one example (which actually includes result of Corollary 4.14 as a special case).

Corollary 4.15. *Consider a sequence of Markov chains $X_{(d)}$ indexed by $d = 1, 2, \dots$ on $\mathbb{E}_{(d)} = \{0, 1, \dots, N\}^d$ with initial distribution $\nu_{(d)} = \delta_{(0,\dots,0)}$ and transition matrix $\mathbf{P}_{(d)}$ given in (8) with*

$$p_k = \frac{1}{(\log k)^p}\frac{1}{K_d}, \quad K_d = \sum_{k=1}^{d} \frac{1}{(\log k)^p}, \quad p \in (0,1), \quad k = 1, \dots, d$$

and $N_1 = \ldots = N_d = N \geq 2$ (so that (7) holds). The stationary distribution $\pi_{(d)}$ is uniform. The sequence of chains exhibits a separation cutoff at time $d\log d + (N-1)d\log\log d - d[\gamma + p - \log(p+1) - \log(N-1)!]$ with window size d.

Proof. In [7] authors prove that

$$\frac{1}{d}(T_d^* - d\log d - (N-1)d\log\log d + d[\gamma + p - \ln(p+1) - \log(N-1)!])$$

converges in distribution to a standard Gumbel random variable. Again, similar calculations as in Corollary 4.12 finish the proof. □

4.4. Constructing ergodic chain with prespecified FSST and arbitrary stationary distribution

Let us ask the following question (which was one of the main motivations for the paper):

How to construct a Markov chain on a state space of size M with arbitrary stationary distribution π whose FSST T is deterministic, $P(T = M-1) = 1$?

The recipe is clear from previous sections: Start with some absorbing chain X^* for which $P(T^* = M-1) = 1$, where T^* is absorption time. Probably the simplest one is the following: take $\mathbb{E} = \{1, \ldots, M\}$ with transitions $\mathbf{P}_0^*(k, k+1) = 1$ for $k < N$ and $\mathbf{P}_0^*(N, N) = 1$ and start it at state 1. Then of course we have desired absorption time and thus the antidual would have desired stationary distribution and FSST.

The above example will be a special case of a more general result. Many absorbing chains have absorption time T^* distributed as a mixture of sums of independent geometric random variables with parameters being the eigenvalues. E.g., for stochastically monotone discrete time birth and death chain starting at 1 with $d > 1$ being absorbing state, time to absorption is distributed as a sum of geometric random variables with parameters being the eigenvalues. This result follows from Karlin and McGregor [15] or Keilson [16]. Fill [12] gave a first stochastic proof of this result (using dualities). This was extended to skip-free Markov chains in Fill [11]. Miclo [23] showed that for any absorbing chain on $\mathbb{E} = \{\mathbf{e}_1, \ldots, \mathbf{e}_M\}$ with positive eigenvalues and some reversibility condition (involving substochastic kernel corresponding to transition matrix with row and column corresponding to absorbing state removed) there exists a measure $a = (a_1, \ldots, a_M)$ such that the time do absorption T^* has distribution

$$T^* \sim \sum_{i=1}^{M-1} a_i \mathcal{G}(\lambda_i, \lambda_{i+1}, \ldots, \lambda_{M-1}),$$

where λ_i are the eigenvalues of transition matrix sorted in nonincreasing order and $\mathcal{G}(p_1, \ldots, p_k)$ denotes the distribution of $\sum_{j=1}^k X_j$, where $X_j \sim Geo(p_j)$.

For convenience denote $H(k) := \sum_{j=1}^k \pi(j)$. Our result is following.

Theorem 4.16. *Let $\mathbb{E} = \{1, \ldots, M\}$ and $p_k \in (0,1], k = 1, \ldots, M-1$. Let $a_k, \pi(k), k = 1, \ldots, M$ be two probability distributions on $\mathbb{E}$ such that $a_k \geq 0, \pi(k) > 0$ for all $k \in \mathbb{E}$. Define the matrix*

$$\mathbf{P}(k,s) = \begin{cases} \dfrac{\pi(1) + \pi(2)(1-p_2)}{\pi(1)+\pi(2)} & \textit{if} \quad k = s = 1, \\ \dfrac{p_2\pi(2)}{\pi(1)+\pi(2)} & \textit{if} \quad k = 1, s = 2, \\ \dfrac{\pi(s)}{\pi(k)}\left[p_{k-1}\left(1 - \dfrac{H(k-1)}{H(k)}\right) - p_k\left(1 - \dfrac{H(k)}{H(k+1)}\right)\right] & \textit{if} \quad 1 < k < M, s < k, \\ (1-p_k)\dfrac{1}{H(k)} + p_k\dfrac{H(k)}{H(k+1)} - p_{k-1}\dfrac{H(k-1)}{H(k)} & \textit{if} \quad 1 < k < M, s = k, \\ p_k\dfrac{H(k)}{H(k+1)}\dfrac{\pi(k+1)}{\pi(k)} & \textit{if} \quad 1 < k < M, s = k+1, \\ p_{M-1}\pi(s) & \textit{if} \quad k = M, s \leq M-1, \\ 1 - p_{M-1} + p_{M-1}\pi(M) & \textit{if} \quad k = M, s = M. \end{cases}$$

Assume that π and sequence $\{p_k\}_{k=1,\ldots,M}$ are such that that matrix $\mathbf{P}$ is nonnegative. Then Markov chain X with transition matrix $\mathbf{P}$ with initial distribution $\nu = (\nu(1), \ldots, \nu(M))$ given by

$$\nu(k) = \pi(k)\sum_{i=k}^{M}\frac{a_i}{H(i)}$$

has FSST T distributed as

$$\sum_{i=1}^{M-1} a_i\mathcal{G}(p_i, p_{i+1}, \ldots, p_{M-1}) \tag{12}$$

and π is its stationary distribution. Moreover, $\{1-p_1, \ldots, 1-p_{M-1}, 1\}$ are the eigenvalues of $\mathbf{P}$.

Note that X is a skip-free chain: for given k the only nonzero entries of $\mathbf{P}$ are $\mathbf{P}(k,s)$ for $s \leq k+1$. The proof of the Theorem is postponed to Section 5.2.
We can relatively easy have some corollaries being interesting special cases of Theorem 4.16. Applying the Theorem with $p_k = 1, k = 1, \ldots, M-1, p_M = 0$ and $a_1 = 1, a_k = 0, k = 2, \ldots, M$ we obtain the following Corollary.

Corollary 4.17. *Consider a distribution π on $\mathbb{E} = \{1, \ldots, M\}$ such that $\pi(k) > 0$ for all $k \in \mathbb{E}$. The Markov chain X on $\mathbb{E}$ with transition matrix*

$$\mathbf{P}_0(k,r) = \begin{cases} \dfrac{\pi(r)}{\pi(1)+\pi(2)} & \textit{for} \quad r = 1, 2, k = 1, \\ \dfrac{\pi(r)}{\pi(k)}\left[\dfrac{H(k)}{H(k+1)} - \dfrac{H(k-1)}{H(k)}\right] & \textit{for} \quad 1 < k < N, r \leq k, \\ \dfrac{\pi(k+1)}{\pi(k)}\dfrac{H(k)}{H(k+1)} & \textit{for} \quad 1 < k < N, r = k+1, \\ \pi(r) & \textit{for} \quad r \leq k = N. \end{cases}$$

is ergodic with stationary distribution π. Assume the initial distribution is $\nu = \delta_1$ (i.e., $P(X_0 = 1) = 1$). Then the chain has deterministic Fastest Strong Stationary Time T such that $P(T = M-1) = 1$.

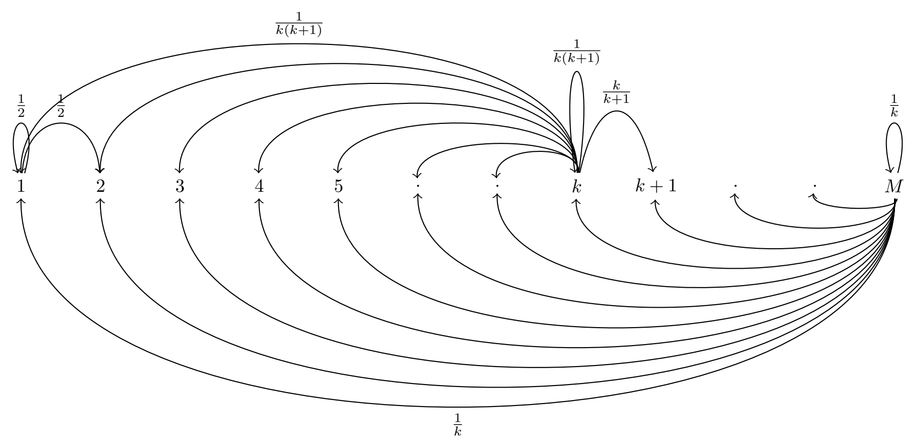

FIGURE 1. Chain X on $\mathbb{E} = \{1, \dots, M\}$ with uniform stationary distribution with deterministic FSST $T : P(T = M) = 1$

Note that for this chain we have

$$sep(\nu \mathbf{P}^k, \pi) = P(T > k) = \begin{cases} 1 & \text{if } k \leq M-2, \\ 0 & \text{if } k \geq M-1. \end{cases}$$

Thus this is an extreme example for separation cutoff: For any $k \leq M - 2$ the chain is completely not mixed (the separation distance between stationary distribution and distribution at step k is 1) and the chain mixes completely exactly at step $k = M - 1$ (the distance is 0).
Simplifying the chain further by taking additionally uniform distribution $\pi(k) = \frac{1}{M}$ in Corollary 4.17 we obtain

$$\mathbf{P}_0(k, r) = \begin{cases} \dfrac{1}{2} & \text{for} \quad r = 1, 2, k = 1, \\[2ex] \dfrac{1}{k(k+1)} & \text{for} \quad 1 < k < N, r \leq k, \\[2ex] \dfrac{k}{k+1} & \text{for} \quad 1 < k < N, r = k+1, \\[2ex] \dfrac{1}{k} & \text{for} \quad r \leq k = N. \end{cases}$$

The chain is sketched in Fig. 1

Two Markov chains on essentially different state spaces with the same FSST
So far in this Section we considered chains on totally ordered state space $\mathbb{E} = \{1, \dots, M\}$. We can also consider another state spaces. We will consider chain on $\mathbb{E}^{(2)} = \{0, 1\}^d$. We will not present full generality one can have, instead we will present two chains, one on $\mathbb{E}^{(1)} = \{1, \dots, d\}$ and the other on $\mathbb{E}^{(2)}$ both with uniform distributions and the same FSST distributed as $\sum_{k=1}^{d-1} X_k$, where $X_k \sim Geo(k \cdot p)$ for some fixed $p \leq \frac{1}{d}$. Note that in particular the sizes of the state spaces are completely different, 2^d versus d

Corollary 4.18. *Fix some integer $d > 1$ and $0 < p \le \frac{1}{d}$. Let $X^{(1)}$ be a Markov chain on $\mathbb{E}^{(1)} = \{1, \ldots, d\}$ with initial distribution $\nu^{(1)} = (1, 0, \ldots, 0)$ and transitions*

$$\mathbf{P}^{(1)}(k,s) = \begin{cases} 1-(2-d)p & \text{if} \quad k=s=1, \\ (d-2)p & \text{if} \quad k=1, s=2, \\ \dfrac{d+1}{k(k+1)} & \text{if} \quad 1<k<M, s<k, \\ (1-(d-k)p)\dfrac{M}{k} + (d-k)p\dfrac{k}{k+1} - (d-k+1)\dfrac{k-1}{k} & \text{if} \quad 1<k<M, s=k, \\ (d-k)p\dfrac{k}{k+1} & \text{if} \quad 1<k<M, s=k+1, \\ \dfrac{p}{d} & \text{if} \quad k=M, s\le M-1, \\ \dfrac{p}{d}+(1-p) & \text{if} \quad k=M, s=M. \end{cases}$$

Let $X^{(2)}$ be a Markov chain on $\mathbb{E}^{(2)} = \{0,1\}^d$ with initial distribution $\nu^{(2)}((0,\ldots,0)) = \nu^{(2)}((1,0,\ldots,0)) = 1/2$ and with transitions

$$\mathbf{P}^{(2)}(\mathbf{i},\mathbf{i}') = \begin{cases} \frac{1}{2}p & \text{if} \quad \mathbf{i}'' = \mathbf{i} \pm \mathbf{s}_k, \\ 1-\frac{1}{2}dp & \text{if} \quad \mathbf{i}' = \mathbf{i}. \end{cases}$$

(Recall that $|\mathbf{i}| = \sum_{j=1}^{d} i_j$ and was called a level of $\mathbf{i}$).

Then, the FSSTs $T^{(1)}$ and $T^{(2)}$ of both chains have the same distribution:

$$T^{(1)} \overset{(d)}{=} T^{(2)} = \sum_{k=1}^{d-1} X_k, \ \ \textit{where} \ \ X_k \sim Geo(k\cdot p).$$

Both chains have uniform stationary distribution on respective state spaces.

Proof. We will show that chains $X^{(1)}$ and $X^{(2)}$ are sharp antidual chains of different chains $X^{*(1)}$ and $X^{*(2)}$, whose absorption times are equal to the statement.

- Chain $X^{(1)}$
  This is a special case of the chain given in Theorem 4.16 with $p_k = (d-k)p$ and uniform stationary distribution π. Taking $a_1 = 1, a_k = 0, k = 2, \ldots, M$ we have that the initial distribution $v = (1, 0, \ldots, 0)$ and that FSST $T^{(1)}$ is distributed as $\sum_{k=1}^{d-1} X_k, X_k \sim Geo(p_k)$ with $p_k = (d-k)p$. The distribution of $T^{(1)}$ is equal to $\sum_{k=1}^{d-1} Y_k$ with $Y_k \sim Geo(k \cdot p)$
- Chain $X^{(2)}$
  This is a special case of chain $\mathbf{P}_1$ given in Corollary 4.7 with $p_k = p$. Thus, its sharp dual chain is given in (6). Recall this is the case $N_j = 1, j = 1, \ldots, d$, let us explicitly write the transitions of this $\mathbf{P}^*$ using notation from this Section:

$$\mathbf{P}^*(\mathbf{i},\mathbf{i}') = \begin{cases} p & \text{if} \quad \mathbf{i}' = \mathbf{i} + \mathbf{s}_k, \\ 1-(d-|\mathbf{i}|)p & \text{if} \quad \mathbf{i}' = \mathbf{i} \end{cases}$$

  Roughly speaking, this is the following random walk on hypercube $\{0,1\}^d$. Being at some state $\mathbf{i} = (i_1, \ldots, i_d), i_k \in \{0,1\}$ either we change one coordinate from 0 to 1 with probability p or with the remaining probability we do nothing. State $(1,\dots,1)$ is an absorbing state. Since the probability of changing 0 into 1 does not depend on actual state, the time to increase the current level depends only on the level. Being at any state on level $|\mathbf{i}|=l$ time to reach next level has distribution $Geo((d-l)p)$ (since there are $(d-l)$ of zeros, each of which can be changed into 1 with probability p). Thus, if the chain starts somewhere on level 1, say $\nu^*((1,0,\dots,0))=1$ then absorption time is equal in distribution to $\sum_{k=1}^{d-1}X_k$, where $X_k\sim Geo(k\cdot p)$. What remains to show is that $\nu=\nu^*\Lambda$ yields $\nu^{(2)}((0,\dots,0))=\nu^{(2)}((1,0,\dots,0))=1/2$. All the proofs of Theorems 4.4 and 4.6 are based on coordinate-wise ordering, i.e.

$$\mathbf{i}\preceq\mathbf{i}' \text{ if } i_j\le i'_j, j=1,\dots,d.$$

Recall the link Λ (it is given in (4))

$$\Lambda(\mathbf{i},\mathbf{i}')=\frac{\pi(\mathbf{i}')}{\sum_{\mathbf{i}_0:\mathbf{i}_0\preceq\mathbf{i}}\pi(\mathbf{i}_0)}\mathbf{1}(\mathbf{i}'\preceq\mathbf{i}).$$

We have

$$\begin{aligned}
\nu(0,\dots,0) &= \sum_{\mathbf{i}}\nu^*(\mathbf{i})\Lambda(\mathbf{i},(0,\dots,0))=\Lambda((1,0,\dots,0),(0,\dots,0))\\
&= =\frac{\pi((0,\dots,0))}{\pi((0,\dots,0))+\pi((1,0,\dots,0))}=\frac{1}{2},\\
\nu(1,0,\dots,0) &= \sum_{\mathbf{i}}\nu^*(\mathbf{i})\Lambda(\mathbf{i},(1,0,\dots,0))=\Lambda((1,0,\dots,0),(1,0,\dots,0))\\
&= =\frac{\pi((1,0,\dots,0))}{\pi((0,\dots,0))+\pi((1,0,\dots,0))}=\frac{1}{2},
\end{aligned}$$

what finishes the proof.

□

5. Proofs

5.1. Proofs of Theorems 4.4 and 4.6

In both proofs we use the coordinate-wise ordering

$$\mathbf{i}\preceq\mathbf{i}' \text{ iff } i_j\le i'_j, j=1,\dots,d.$$

In this ordering $\mathbf{i}_{min}=(0,\dots,0)$ is a unique minimal and $\mathbf{i}_{max}=(N_1,\dots,N_d)$ is a unique maximal one.

Proof of Theorem 4.4. For ordering under consideration, directly from Proposition 5 in Rota [26], we find the corresponding Möbius function

$$\mu((i_1,\dots,i_d),(i_1+r_1,\dots,i_d+r_d))=\begin{cases}(-1)^{\sum_{k=1}^d r_k} & r_j\in\{0,1\},\ i_j+r_j\le N_j,\ k=1,\dots,d\\ 0 & \text{otherwise.}\end{cases}$$

Let

$$\rho(\mathbf{i})=\prod_{j=1}^d(i_j+1).$$

We will apply Theorem 4.1 with above ordering and uniform distribution π on $\mathbb{E}^*$, i.e., $\pi(\mathbf{i})=\frac{1}{\rho(\mathbf{i}_{max})}$. Since X^* starts at minimal state, so does - by Remark 4.2 - the antidual chain.

The link $\Lambda(\mathbf{i},\mathbf{i}')$ is a uniform distribution truncated to $\{\mathbf{i}'\preceq\mathbf{i}\}$, from (**??**) we have $\Lambda=(\mathbf{diag}(\pi\mathbf{C}))^{-1}\mathbf{C}^T\mathbf{diag}(\pi)$, thus

$$\Lambda(\mathbf{i},\mathbf{i}')=\frac{\mathbf{C}(\mathbf{i}',\mathbf{i})\frac{1}{\rho(\mathbf{i}_{max})}}{\displaystyle\sum_{\mathbf{i}^{(2)}}\frac{1}{\rho(\mathbf{i}_{max})}\mathbf{C}(\mathbf{i}^{(2)},\mathbf{i})}=\frac{\mathbf{1}(\mathbf{i}'\preceq\mathbf{i})}{\rho(\mathbf{i})}.$$

The inverse is given by $\Lambda^{-1}=(\mathbf{diag}(\pi))^{-1}(\mathbf{C}^{-1})^T\mathbf{diag}(\pi\mathbf{C})$, thus

$$\Lambda^{-1}(\mathbf{i}^{(1)},\mathbf{i}^{(2)})=\frac{1}{\rho(\mathbf{i}_{max})}\mathbf{C}^{-1}(\mathbf{i}^{(2)},\mathbf{i}^{(1)})\frac{\rho(\mathbf{i}^{(2)})}{\rho(\mathbf{i}_{max})}=\rho(\mathbf{i}^{(2)})\mathbf{C}^{-1}(\mathbf{i}^{(2)},\mathbf{i}^{(1)}).$$

Instead of calculating $\widehat{\mathbf{P}^*}$, we will calculate Λ^{-1} and then directly antidual from $\mathbf{P}=\Lambda^{-1}\mathbf{P}^*\Lambda$ (the conditions on $(\pi,\mathbf{C})$-Möbius monotonicity will be read from resulting antidual, see Remark 4.3). We have to calculate

$$\mathbf{P}(\mathbf{i}^{(1)},\mathbf{i}^{(2)})=(\Lambda^{-1}\mathbf{P}^*\Lambda)(\mathbf{i}^{(1)},\mathbf{i}^{(2)})=\sum_{\mathbf{i}}\Lambda^{-1}(\mathbf{i}^{(1)},\mathbf{i})(\mathbf{P}^*\Lambda)(\mathbf{i},\mathbf{i}^{(2)}).$$

Because of the form of Λ^{-1}, we need only to consider states which differ from $\mathbf{i}^{(1)}$ at most by 1 on each coordinate.

$$\begin{aligned}\mathbf{P}(\mathbf{i}^{(1)},\mathbf{i}^{(2)}) &= \sum_{\mathbf{r}=(r_1,\ldots,r_d)\in\{0,1\}^d:\ \mathbf{i}^{(1)}-\mathbf{r}\in\mathbb{E}^*}\Lambda^{-1}(\mathbf{i}^{(1)},\mathbf{i}^{(1)}-\mathbf{r})(\mathbf{P}^*\Lambda)(\mathbf{i}^{(1)}-\mathbf{r},\mathbf{i}^{(2)})\\ &= \sum_{\mathbf{r}=(r_1,\ldots,r_d)\in\{0,1\}^d:\ \mathbf{i}^{(1)}-\mathbf{r}\in\mathbb{E}^*}(-1)^{|\mathbf{r}|}\rho(\mathbf{i}^{(1)}-\mathbf{r})(\mathbf{P}^*\Lambda)(\mathbf{i}^{(1)}-\mathbf{r},\mathbf{i}^{(2)})\end{aligned}$$

We need to calculate

$$(\mathbf{P}^*\Lambda)(\mathbf{i}^{(1)}-\mathbf{r},\mathbf{i}^{(2)})=\sum_{\mathbf{i}}\mathbf{P}^*(\mathbf{i}^{(1)}-\mathbf{r},\mathbf{i})\Lambda(\mathbf{i},\mathbf{i}^{(2)})=\sum_{\mathbf{i}}\mathbf{P}^*(\mathbf{i}^{(1)}-\mathbf{r},\mathbf{i})\frac{\mathbf{1}(\mathbf{i}^{(2)}\preceq\mathbf{i})}{\rho(\mathbf{i})}$$

Note that for given $\mathbf{i}^{(1)}-\mathbf{r}\in\mathbb{E}^*$ the only nonzero entries of $\mathbf{P}^*(\mathbf{i}^{(1)}-\mathbf{r},\mathbf{i})$ are for $\mathbf{i}=\mathbf{i}^{(1)}-\mathbf{r}$ or $\mathbf{i}=\mathbf{i}^{(1)}-\mathbf{r}+\mathbf{s}_j$ (if $\mathbf{i}\in\mathbb{E}^*$), where $\mathbf{s}_j=(0,\ldots,0,1,0,\ldots,0)$ (with 1 at position j). Thus, we have

$$\begin{aligned}&(\mathbf{P}^*\Lambda)(\mathbf{i}^{(1)}-\mathbf{r},\mathbf{i}^{(2)})=\\ &\mathbf{P}^*(\mathbf{i}^{(1)}-\mathbf{r},\mathbf{i}^{(1)}-\mathbf{r})\frac{\mathbf{1}(\mathbf{i}^{(2)}\preceq\mathbf{i}^{(1)}-\mathbf{r})}{\rho(\mathbf{i}^{(1)}-\mathbf{r})}+\sum_{j:i^{(1)}_j-r_j<N_j}\mathbf{P}^*(\mathbf{i}^{(1)}-\mathbf{r},\mathbf{i}^{(1)}-\mathbf{r}+\mathbf{s}_j)\frac{\mathbf{1}(\mathbf{i}^{(2)}\preceq\mathbf{i}^{(1)}-\mathbf{r}+\mathbf{s}_j)}{\rho(\mathbf{i}^{(1)}-\mathbf{r}+\mathbf{s}_j)}\\ &=\left(1-\sum_{j:i^{(1)}_j-r_j<N_j}p_j\right)\frac{\mathbf{1}(\mathbf{i}^{(2)}\preceq\mathbf{i}^{(1)}-\mathbf{r})}{\rho(\mathbf{i}^{(1)}-\mathbf{r})}+\sum_{j:i^{(1)}_j-r_j<N_j}\frac{\mathbf{1}(\mathbf{i}^{(2)}\preceq\mathbf{i}^{(1)}-\mathbf{r}+\mathbf{s}_j)}{\rho(\mathbf{i}^{(1)}-\mathbf{r}+\mathbf{s}_j)}p_j\end{aligned}$$

We have

$$\begin{aligned}\mathbf{P}(\mathbf{i}^{(1)},\mathbf{i}^{(2)})=&\sum_{\mathbf{r}=(r_1,\ldots,r_d)\in\{0,1\}^d:\ \mathbf{i}^{(1)}-\mathbf{r}\in\mathbb{E}^*}(-1)^{|\mathbf{r}|}\rho(\mathbf{i}^{(1)}-\mathbf{r})\\ &\times\left[\left(1-\sum_{j:i^{(1)}_j-r_j<N_j}p_j\right)\frac{\mathbf{1}(\mathbf{i}^{(2)}\preceq\mathbf{i}^{(1)}-\mathbf{r})}{\rho(\mathbf{i}^{(1)}-\mathbf{r})}+\sum_{j:i^{(1)}_j-r_j<N_j}\frac{\mathbf{1}(\mathbf{i}^{(2)}\preceq\mathbf{i}^{(1)}-\mathbf{r}+\mathbf{s}_j)}{\rho(\mathbf{i}^{(1)}-\mathbf{r}+\mathbf{s}_j)}p_j\right]\end{aligned}\qquad(13)$$

We will consider cases:

- Let $\mathbf{i}^{(2)} = \mathbf{i}^{(1)} + \mathbf{s}_k$. Then note that for any $\mathbf{r} \in \{0,1\}^d$ all indicators $\mathbf{1}(\mathbf{i}^{(2)} \preceq \mathbf{i}^{(1)} - \mathbf{r})$ are equal to 0. Considering indicators $\mathbf{1}(\mathbf{i}^{(2)} \preceq \mathbf{i}^{(1)} - \mathbf{r} + \mathbf{s}_j)$ the only nonzero is for $j = k$ and $\mathbf{r} = (0, \ldots, 0)$. Then we have

$$\mathbf{P}(\mathbf{i}^{(1)}, \mathbf{i}^{(1)} + \mathbf{s}_k) = (-1)^0 \rho(\mathbf{i}^{(1)}) \frac{p_k}{\rho(\mathbf{i}^{(1)} + \mathbf{s}_k)} = \frac{i_k^{(1)} + 1}{i_k^{(1)} + 2} p_k$$

- Let $\mathbf{i}^{(2)} = \mathbf{i}^{(1)} - m \cdot \mathbf{s}_k$ with $1 \leq m \leq i_k$. Then, both indicators are nonzero only when $\mathbf{r} = (0, \ldots, 0)$ or when $\mathbf{r} = \mathbf{s}_k$.

$$\mathbf{P}(\mathbf{i}^{(1)}, \mathbf{i}^{(1)} - m \cdot \mathbf{s}_k) =$$

$$\begin{aligned}
= \quad & \rho(\mathbf{i}^{(1)}) \quad \times \Bigg[\Bigg(1 - \sum_{j: i_j^{(1)} - 0 < N_j} p_j \Bigg) \frac{1}{\rho(\mathbf{i}^{(1)})} \mathbf{1}(\mathbf{i}^{(1)} - m \cdot \mathbf{s}_k \preceq \mathbf{i}^{(1)}) \\
& \qquad + \sum_{j: i_j^{(1)} - 0 < N_j} \frac{p_j}{\rho(\mathbf{i}^{(1)} + \mathbf{s}_j)} \mathbf{1}(\mathbf{i}^{(1)} - m \cdot \mathbf{s}_k \preceq \mathbf{i}^{(1)} + \mathbf{s}_j) \Bigg] \\
& -\rho(\mathbf{i}^{(1)} - s_k) \quad \times \Bigg[\Bigg(1 - \sum_{j: i_j^{(1)} - \mathbf{1}(j=k) < N_j} p_j \Bigg) \frac{1}{\rho(\mathbf{i}^{(1)} - \mathbf{s}_k)} \mathbf{1}(\mathbf{i}^{(1)} - m \cdot \mathbf{s}_k \preceq \mathbf{i}^{(1)} - \mathbf{s}_k) + \\
& \qquad \sum_{j: i_j^{(1)} - \mathbf{1}(j=k) < N_j} \frac{p_j}{\rho(\mathbf{i}^{(1)} - \mathbf{s}_k + \mathbf{s}_j)} \mathbf{1}(\mathbf{i}^{(1)} - m \cdot \mathbf{s}_k \preceq \mathbf{i}^{(1)} - \mathbf{s}_k + \mathbf{s}_j) \Bigg]
\end{aligned}$$

$$\begin{aligned}
= \quad & \Bigg(1 - \sum_{j: i_j^{(1)} - 0 < N_j} p_j \Bigg) + \sum_{j: i_j^{(1)} - 0 < N_j} \frac{\rho(\mathbf{i}^{(1)})}{\rho(\mathbf{i}^{(1)} + \mathbf{s}_j)} p_j \\
- \quad & \Bigg(1 - \sum_{j: i_j^{(1)} - \mathbf{1}(j=k) < N_j} p_j \Bigg) - \sum_{j: i_j^{(1)} - \mathbf{1}(j=k) < N_j} \frac{\rho(\mathbf{i}^{(1)} - \mathbf{s}_k)}{\rho(\mathbf{i}^{(1)} - \mathbf{s}_k + \mathbf{s}_j)} p_j \\
= \quad & p_k - \mathbf{1}(i_k^{(1)} < N_k) p_k + \sum_{\substack{j: i_j^{(1)} < N_j \\ j \neq k}} \left(\frac{\rho(\mathbf{i}^{(1)})}{\rho(\mathbf{i}^{(1)} + \mathbf{s}_j)} - \frac{\rho(\mathbf{i}^{(1)} - \mathbf{s}_k)}{\rho(\mathbf{i}^{(1)} - \mathbf{s}_k + \mathbf{s}_j)} \right) p_j \\
+ \quad & \frac{\rho(\mathbf{i}^{(1)})}{\rho(\mathbf{i}^{(1)} + \mathbf{s}_k)} p_k \mathbf{1}(i_k^{(1)} < N_k) - \frac{\rho(\mathbf{i}^{(1)} - \mathbf{s}_k)}{\rho(\mathbf{i}^{(1)})} p_k \\
= \quad & \left[1 - \frac{\rho(\mathbf{i}^{(1)} - \mathbf{s}_k)}{\rho(\mathbf{i}^{(1)})} - \mathbf{1}(i_k^{(1)} < N_k) \left(1 - \frac{\rho(\mathbf{i}^{(1)})}{\rho(\mathbf{i}^{(1)} + \mathbf{s}_k)} \right) \right] p_k \\
= \quad & \left[\frac{1}{i_k^{(1)} + 1} - \mathbf{1}(i_k^{(1)} < N_k) \left(\frac{1}{i_k^{(1)} + 2} \right) \right] p_k.
\end{aligned}$$

Finally, we have

$$\mathbf{P}(\mathbf{i}^{(1)},\mathbf{i}^{(1)}-m\cdot\mathbf{s}_k)=\begin{cases}\dfrac{1}{(i_k^{(1)}+1)(i_k^{(1)}+2)}p_k & \text{if } i_k^{(1)}<N_k,\\ \dfrac{1}{N_k+1}p_k & \text{if } i_k^{(1)}=N_k.\end{cases}$$

- Let $\mathbf{i}^{(2)}=\mathbf{i}^{(1)}$. Then, the indicator $\mathbf{1}(\mathbf{i}^{(2)}\preceq\mathbf{i}^{(1)}-\mathbf{r})$ is nonzero only when $\mathbf{r}=(0,\ldots,0)$, whereas the indicator $\mathbf{1}(\mathbf{i}^{(2)}\preceq\mathbf{i}^{(1)}-\mathbf{r}+\mathbf{s}_j)$ is nonzero when $\mathbf{r}=(0,\ldots,0)$ and any $j=1,\ldots,d$ or when $\mathbf{r}=\mathbf{s}_j$. We have

$$\mathbf{P}(\mathbf{i}^{(1)},\mathbf{i}^{(1)})=(-1)^0\rho(\mathbf{i}^{(1)})\quad\left[\left(1-\sum_{j:i_j^{(1)}-0<N_j}p_j\right)\frac{1}{\rho(\mathbf{i}^{(1)})}+\sum_{j:i_j^{(1)}-0<N_j}\frac{p_j}{\rho(\mathbf{i}^{(1)}+\mathbf{s}_j)}\right]$$
$$-\sum_{k:i_k^{(1)}-1\geq 0}\rho(\mathbf{i}^{(1)}-\mathbf{s}_k)\qquad\left[\sum_{j:i_j^{(1)}-\mathbf{1}(j=k)<N_j}\frac{p_j}{\rho(\mathbf{i}^{(1)}-\mathbf{s}_k+\mathbf{s}_j)}\mathbf{1}(\mathbf{i}^{(1)}\preceq\mathbf{i}^{(1)}-\mathbf{s}_k+\mathbf{s}_j)\right]$$

$$\begin{aligned}
&= 1-\sum_{j:i_j^{(1)}<N_j}p_j+\sum_{j:i_j^{(1)}<N_j}\frac{\rho(\mathbf{i}^{(1)})}{\rho(\mathbf{i}^{(1)})+\mathbf{s}_j)}p_j-\sum_{k:i_k^{(1)}\geq 1}\frac{\rho(\mathbf{i}^{(1)}-\mathbf{s}_k)}{\rho(\mathbf{i}^{(1)})}p_k\\
&= 1-\sum_{j:i_j^{(1)}<N_j}\left(1-\frac{i_j^{(1)}+1}{i_j^{(1)}+2}\right)p_j-\sum_{k=1}^{d}\frac{i_k^{(1)}}{i_k^{(1)}+1}p_k\\
&= 1-\sum_{j:i_j^{(1)}<N_j}\left(\frac{1}{i_j^{(1)}+2}\right)p_j-\sum_{j:i_j^{(1)}<N_j}\frac{i_j^{(1)}}{i_j^{(1)}+1}p_j-\sum_{j:i_j^{(1)}=N_j}\frac{i_j^{(1)}}{i_j^{(1)}+1}p_j\\
&= 1-\sum_{j:i_j^{(1)}<N_j}\left(1-\frac{1}{(i_j^{(1)}+1)(i_j^{(1)}+2)}\right)p_j-\sum_{j:i_j^{(1)}=N_j}\frac{N_j}{N_j+1}p_j.
\end{aligned}$$

The assumption (7) implies that $\mathbf{P}(\mathbf{i}^{(1)},\mathbf{i}^{(1)})\geq 0$. All other transitions have probability 0. For example, let $\mathbf{i}^{(2)}=\mathbf{i}^{(1)}+m\cdot\mathbf{s}_k$ with $2\leq m\leq N_k-i_k$. Then both indicators in (13) are equal to 0. Similarly, if two or more coordinates of $\mathbf{i}^{(2)}$ are strictly larger then appropriate coordinates of $\mathbf{i}^{(1)}$, the indicators are also equal to 0. However, instead of considering all other cases, it is enough to show that all already calculated transitions sum up to 1 (see proof of Theorem 4.1). We have (with convention $\sum_{m=1}^{0}f(m)\equiv 0$)

$$\sum_{\mathbf{i}^{(2)}\in\mathbb{E}^*}\mathbf{P}(\mathbf{i}^{(1)},\mathbf{i}^{(2)})=$$

$$\begin{aligned}
&\mathbf{P}(\mathbf{i}^{(1)},\mathbf{i}^{(1)}) + \sum_{j:i_j^{(1)}<N_j} \mathbf{P}(\mathbf{i}^{(1)},\mathbf{i}^{(1)}+\mathbf{s}_j) + \sum_{j=1}^{d}\sum_{m=1}^{i_j^{(1)}} \mathbf{P}((\mathbf{i}^{(1)},\mathbf{i}^{(1)}-m\cdot\mathbf{s}_j) \\
= \;& 1 - \sum_{j:i_j^{(1)}<N_j}\left(1-\frac{1}{(i_j^{(1)}+1)(i_j^{(1)}+2)}\right)p_j - \sum_{j:i_j^{(1)}=N_j}\frac{N_j}{N_j+1}p_j \\
&+ \sum_{j:i_j^{(1)}<N_j}\frac{i^{(1)}+1}{i^{(1)}+2}p_j + \sum_{j:i_j^{(1)}<N_j}\sum_{m=1}^{i_j^{(1)}}\mathbf{P}((\mathbf{i}^{(1)},\mathbf{i}^{(1)}-m\cdot\mathbf{s}_j) + \sum_{j:i_j^{(1)}=N_j}\sum_{m=1}^{i_j^{(1)}}\mathbf{P}((\mathbf{i}^{(1)},\mathbf{i}^{(1)}-m\cdot\mathbf{s}_j) \\
= \;& 1 - \sum_{j:i_j^{(1)}<N_j}\left(1-\frac{1}{(i_j^{(1)}+1)(i_j^{(1)}+2)} - \frac{i^{(1)}+1}{i^{(1)}+2}\right)p_j - \sum_{j:i_j^{(1)}=N_j}\frac{N_j}{N_j+1}p_j \\
&+ \sum_{j:i_j^{(1)}<N_j}\frac{i_j^{(1)}}{(i_j^{(1)}+1)(i_j^{(1)}+2)} + \sum_{j:i_j^{(1)}=N_j}\frac{N_j}{N_j+1}p_j \\
= \;& 1 - \sum_{j:i_j^{(1)}<N_j}\left(1-\frac{1}{(i_j^{(1)}+1)(i_j^{(1)}+2)} - \frac{i^{(1)}+1}{i^{(1)}+2} - \frac{i_j^{(1)}}{(i_j^{(1)}+1)(i_j^{(1)}+2)}\right)p_j = 1
\end{aligned}$$

□

Proof of Theorem 4.6. We start with general distribution π on $\mathbb{E}$. Note that $(0,\ldots,0)$ is a minimal state, and X^* starts at this state $\nu^* = \delta_{(0,\ldots,0)}$, thus - by Remark 4.2 - this is also the initial distribution of antidual chain, i.e., $\nu = \nu^*$, regardless the stationary distribution π.

Let us start with $\widehat{\mathbf{P}^*}$. For convenience, define

$$f(\mathbf{i},k) = \frac{\sum_{\mathbf{i}'\preceq\mathbf{i}}\pi(\mathbf{i}')}{\sum_{\mathbf{i}''\preceq\mathbf{i}+\mathbf{s}_k}\pi(\mathbf{i}'')} \quad \text{for} \quad \mathbf{i}: i_k = 0.$$

We have

$$\widehat{\mathbf{P}^*}(\mathbf{i}^{(2)},\mathbf{i}^{(1)}) = \frac{(\boldsymbol{\pi}\mathbf{C})(\mathbf{i}^{(2)})}{(\boldsymbol{\pi}\mathbf{C})(\mathbf{i}^{(1)})}\mathbf{P}^*(\mathbf{i}^{(2)},\mathbf{i}^{(1)}) = \begin{cases} f(\mathbf{i}^{(2)},k)p_k & \text{if} \quad \mathbf{i}^{(1)} = \mathbf{i}^{(2)}+\mathbf{s}_k, \\ 1 - \sum\limits_{j:i_j^{(2)}=0} p_j & \text{if} \quad \mathbf{i}^{(1)} = \mathbf{i}^{(2)}. \end{cases}$$

Note that $\widehat{\mathbf{P}^*}$ is not a stochastic matrix, since we have

$$\sum_{\mathbf{i}}\widehat{\mathbf{P}^*}(\mathbf{i}^{(2)},\mathbf{i}) = \sum_{j:i_j^{(2)}=0} f(\mathbf{i}^{(2)},j)p_j + 1 - \sum_{j:i_j^{(2)}=0} p_j = 1 - \sum_{j:i_j^{(2)}=0}\Big(1-f(\mathbf{i}^{(2)},j)\Big)p_j. \tag{14}$$

Now, calculating antidual from Theorem 4.1, we have

$$\mathbf{P}(\mathbf{i}^{(1)},\mathbf{i}^{(2)}) = \frac{\pi(\mathbf{i}^{(2)})}{\pi(\mathbf{i}^{(1)})}((\mathbf{C}^T)^{-1}\widehat{\mathbf{P}^*}\mathbf{C}^T)(\mathbf{i}^{(1)},\mathbf{i}^{(2)}) = \frac{\pi(\mathbf{i}^{(2)})}{\pi(\mathbf{i}^{(1)})}(\mathbf{C}(\widehat{\mathbf{P}^*})^T\mathbf{C}^{-1})(\mathbf{i}^{(2)},\mathbf{i}^{(1)})$$

$$= \frac{\pi(\mathbf{i}^{(2)})}{\pi(\mathbf{i}^{(1)})}\sum_{\mathbf{i}\preceq\mathbf{i}^{(1)}}\widehat{\mathbf{P}^*}(\mathbf{i},\{\mathbf{i}^{(2)}\}^{\uparrow})(-1)^{|\mathbf{i}^{(1)}-\mathbf{i}|}, \tag{15}$$

where we have used the fact, that for this ordering the Möbius function fulfills $\mathbf{C}^{-1}(\mathbf{i},\mathbf{i}^{(1)}) = (-1)^{|\mathbf{i}^{(1)}-\mathbf{i}|}\mathbf{1}(\mathbf{i}\preceq\mathbf{i}^{(1)})$ (see, e.g., Corollary on p. 345 of [26]).

We proceed with (15) by considering cases:

- **Case 1:** $\mathbf{i}^{(2)} = \mathbf{i}^{(1)} + \mathbf{s}_k$ for some $k : i_k^{(1)} = 0$.

  Then, the sum in (15) is following $\sum_{\mathbf{i}\preceq\mathbf{i}^{(1)}} \widehat{\mathbf{P}^*}(\mathbf{i}, \{\mathbf{i}^{(1)}+\mathbf{s}_k\}^{\uparrow})(-1)^{|\mathbf{i}^{(1)}-\mathbf{i}|}$, the only non-zero term is for $\mathbf{i} = \mathbf{i}^{(1)}$, thus

$$\mathbf{P}(\mathbf{i}^{(1)},\mathbf{i}^{(1)}+\mathbf{s}_k) = \frac{\pi(\mathbf{i}^{(1)}+\mathbf{s}_k)}{\pi(\mathbf{i}^{(1)})}\widehat{\mathbf{P}^*}(\mathbf{i}^{(1)},\{\mathbf{i}^{(1)}+\mathbf{s}_k\}^{\uparrow}) = \frac{\pi(\mathbf{i}^{(1)}+\mathbf{s}_k)}{\pi(\mathbf{i}^{(1)})} f(\mathbf{i}^{(1)},k)p_k. \tag{16}$$

- **Case 2:** $\mathbf{i}^{(2)} = \mathbf{i}^{(1)} + \mathbf{s}_{k_1} + \ldots \mathbf{s}_{k_M}$ for some $k_1,\ldots,k_M : i_{k_1}^{(1)} = \ldots = i_{k_M}^{(1)} = 0$ and $M \geq 2$.

  Then, for any $\mathbf{i}\preceq\mathbf{i}^{(1)}$ we have that $\widehat{\mathbf{P}^*}(\mathbf{i},\{\mathbf{i}^{(1)}+\mathbf{s}_{k_1}+\ldots\mathbf{s}_{k_M}\}^{\uparrow}) = 0$, thus $\mathbf{P}(\mathbf{i}^{(1)},\mathbf{i}^{(1)}+\mathbf{s}_{k_1}+\ldots\mathbf{s}_{k_M}) = 0$.
- **Case 3:** $\mathbf{i}^{(2)} = \mathbf{i}^{(1)}$. Then we have

$$\mathbf{P}(\mathbf{i}^{(1)},\mathbf{i}^{(1)}) = \sum_{\mathbf{i}\preceq\mathbf{i}^{(1)}} \widehat{\mathbf{P}^*}(\mathbf{i},\{\mathbf{i}^{(1)}\}^{\uparrow})(-1)^{|\mathbf{i}^{(1)}-\mathbf{i}|}$$

$$= \widehat{\mathbf{P}^*}(\mathbf{i}^{(1)},\{\mathbf{i}^{(1)}\}^{\uparrow}) - \sum_{j:\mathbf{i}_j^{(1)}=1} \widehat{\mathbf{P}^*}(\mathbf{i}^{(1)}-\mathbf{s}_j,\{\mathbf{i}^{(1)}\}^{\uparrow}).$$

  First term is equal to $\sum_{\mathbf{i}} \widehat{\mathbf{P}^*}(\mathbf{i}^{(1)},\mathbf{i})$, in the latter, the only possibility is to change j-th coordinate of $\mathbf{i}^{(1)}-\mathbf{s}_j$ to one:

$$\mathbf{P}(\mathbf{i}^{(1)},\mathbf{i}^{(1)}) = 1 - \sum_{j:i_j^{(1)}=0} p_j(1-f(\mathbf{i}^{(1)},j)) - \sum_{j:i_j^{(1)}=1} f(\mathbf{i}^{(1)}-\mathbf{s}_j,j)p_j. \tag{17}$$

- **Case 4:** $\mathbf{i}^{(2)} = \mathbf{i}^{(1)} - \mathbf{s}_k$ for some $k : i_k^{(1)} = 1$. Consider the sum from (15)

$$\sum_{\mathbf{i}\preceq\mathbf{i}^{(1)}} \widehat{\mathbf{P}^*}(\mathbf{i},\{\mathbf{i}^{(1)}-\mathbf{s}_k\}^{\uparrow})(-1)^{|\mathbf{i}^{(1)}-\mathbf{i}|} = \widehat{\mathbf{P}^*}(\mathbf{i}^{(1)},\{\mathbf{i}^{(1)}-\mathbf{s}_k\}^{\uparrow})$$

$$+\widehat{\mathbf{P}^*}(\mathbf{i}^{(1)}-\mathbf{s}_k,\{\mathbf{i}^{(1)}-\mathbf{s}_k\}^{\uparrow})(-1) + \sum_{j:i_j^{(1)}=0} \widehat{\mathbf{P}^*}(\mathbf{i}^{(1)}-\mathbf{s}_j,\{\mathbf{i}^{(1)}-\mathbf{s}_k\}^{\uparrow})(-1)$$

$$+ \sum_{\substack{j:i_j^{(1)}=1\\ j\neq k}} \widehat{\mathbf{P}^*}(\mathbf{i}^{(1)}-\mathbf{s}_j-\mathbf{s}_k,\{\mathbf{i}^{(1)}-\mathbf{s}_k\}^{\uparrow})(-1)^2 =: S_1 - S_2 - S_3 + S_4.$$

  S_1 is for $\mathbf{i} = \mathbf{i}^{(1)}$, S_2 and S_3 are for all $\mathbf{i}\preceq\mathbf{i}^{(1)}$ which are on level $|\mathbf{i}^{(1)}|-1$, and S_4 for states on level $|\mathbf{i}|-2$.

  Concerning S_1 and S_2. All the "transitions" (recall, $\widehat{\mathbf{P}^*}$ is not a stochastic matrix) are only upward w.r.t. $\preceq$, thus, by (14) we have

$$\begin{aligned} S_1 &= 1 - \sum_{j:i_j^{(1)}=0} p_j(1-f(\mathbf{i}^{(1)},j)),\\ S_2 &= 1 - \sum_{j:i_j^{(1)}=0} p_j(1-f(\mathbf{i}^{(1)}-\mathbf{s}_k,j)) - p_k(1-f(\mathbf{i}^{(1)}-\mathbf{s}_k,k)). \end{aligned}$$

Concerning S_3: Note that the only possibility is to change j-th coordinate of $\mathbf{i}^{(1)} - \mathbf{s}_j$:

$$S_3 := \sum_{\substack{j: i_j^{(1)}=1 \\ j\neq k}} \widehat{\mathbf{P}^*}(\mathbf{i}^{(1)} - \mathbf{s}_j, \mathbf{i}^{(1)}) = \sum_{\substack{j: i_j^{(1)}=1 \\ j\neq k}} f(\mathbf{i}^{(1)} - \mathbf{s}_j, j) p_j.$$

Concerning S_4: Similarly, here we can only change j-th coordinate:

$$S_4 := \sum_{\substack{j: i_j^{(1)}=1 \\ j\neq k}} \widehat{\mathbf{P}^*}(\mathbf{i}^{(1)} - \mathbf{s}_j - \mathbf{s}_k, \mathbf{i}^{(1)} - \mathbf{s}_k) = \sum_{\substack{j: i_j^{(1)}=1 \\ j\neq k}} f(\mathbf{i}^{(1)} - \mathbf{s}_j - \mathbf{s}_k, j).$$

Summarizing, $S_1 + S_2 + S_3 + S_4 = p_k(1 - f(\mathbf{i}^{(1)} - \mathbf{s}_k, k))$

$$- \sum_{j: i_j^{(1)}=0} p_j \Big(f(\mathbf{i}^{(1)}, j) - f(\mathbf{i}^{(1)} - \mathbf{s}_k, j) \Big) - \sum_{\substack{j: i_j^{(1)}=1 \\ j\neq k}} p_j \Big(f(\mathbf{i}^{(1)} - \mathbf{s}_j - \mathbf{s}_k, j) - f(\mathbf{i}^{(1)} - \mathbf{s}_j, j) \Big),$$

thus

$$\mathbf{P}(\mathbf{i}^{(1)}, \mathbf{i}^{(1)} - \mathbf{s}_k) = \frac{\pi(\mathbf{i}^{(1)} - \mathbf{s}_k)}{\pi(\mathbf{i}^{(1)})} \Big(p_k (1 - f(\mathbf{i}^{(1)} - \mathbf{s}_k, k))$$

$$+ \sum_{j: i_j^{(1)}=0} \Big(f(\mathbf{i}^{(1)}, j) - f(\mathbf{i}^{(1)} - \mathbf{s}_k, j) \Big) + \sum_{\substack{j: i_j^{(1)}=1 \\ j\neq k}} \Big(f(\mathbf{i}^{(1)} - \mathbf{s}_j - \mathbf{s}_k, j) - f(\mathbf{i}^{(1)} - \mathbf{s}_j, j) \Big) \Big). \tag{18}$$

For general stationary distribution there are more cases to consider than the four considered above. However, for the specific stationary distribution under consideration, these are all the cases.

For our stationary distribution π given in (10) we have

$$\frac{\pi(\mathbf{i}^{(1)} + \mathbf{s}_k)}{\pi(\mathbf{i}^{(1)})} = \frac{a_k}{1 - a_k}, \quad \frac{\pi(\mathbf{i}^{(1)} - \mathbf{s}_k)}{\pi(\mathbf{i}^{(1)})} = \frac{1 - a_k}{a_k},$$

$$f(\mathbf{i}, k) = \frac{\displaystyle\sum_{\mathbf{i}' \preceq \mathbf{i}} \prod_{j=1}^{d} [a_j \mathbf{1}(i_j = 1) + (1 - a_j)\mathbf{1}(i'_j = 0)]}{\displaystyle\sum_{\mathbf{i}'' \preceq \mathbf{i} + \mathbf{s}_k} \prod_{j=1}^{d} [a_j \mathbf{1}(i''_j = 1) + (1 - a_j)\mathbf{1}(i''_j = 0)]}.$$

Denote

$$\xi(\mathbf{i}, k) = \prod_{\substack{j=1 \\ j\neq k}}^{d} [a_j \mathbf{1}(i_j = 1) + (1 - a_j)\mathbf{1}(i_j = 0)].$$

The sum in denominator of $f(\mathbf{i}, k)$ can be split into two sums: for $\mathbf{i}'' : i''_k = 0$ and $\mathbf{i}'' : i''_k = 1$. We have

$$f(\mathbf{i}, k) = \frac{\displaystyle\sum_{\mathbf{i}' \preceq \mathbf{i}} \xi(\mathbf{i}', k)(1 - a_k)}{\displaystyle\sum_{\substack{\mathbf{i}'' \preceq \mathbf{i} + \mathbf{s}_k \\ i''_k = 0}} \xi(\mathbf{i}'', k)(1 - a_k) + \sum_{\substack{\mathbf{i}'' \preceq \mathbf{i} + \mathbf{s}_k \\ i''_k = 1}} \xi(\mathbf{i}'', k) a_k} = 1 - a_k.$$

From (16), (17) and (18) we obtain transitions given by matrix $\mathbf{P}$ given in (9). □

5.2. **Proof of Theorem 4.16**

Let $\mathbf{X}^*$ be an absorbing chain on $\mathbb{E} = \{1, \ldots, M\}$ with transition matrix:

$$\mathbf{P}^*(k,s) = \begin{cases} p_k & \text{if } s = k+1, \\ \\ 1 - p_k & \text{if } s = k, \end{cases}$$

where, for convenience, we set $p_M = 0$. Let $\nu^* = (a_1, \ldots, a_M)$ be its initial distribution. This is a pure birth chain, thus its absorption time T^* is distributed as (12). We will show that $\mathbf{P}$ is its sharp antidual chain.
We consider total ordering $\preceq := \leq$. Then the link given in (4) reads

$$\Lambda(k,s) = \frac{\pi(s)\mathbf{1}(s \leq k)}{H(k)}.$$

The inverse Λ^{-1} can be easily derived:

$$\Lambda^{-1}(k,s) = \frac{H(k)}{\pi(k)}\mathbf{1}(s=k) - \frac{H(k-1)}{\pi(k)}\mathbf{1}(s=k-1) = \begin{cases} \dfrac{H(K)}{\pi(k)} & \text{if} \quad s = k, \\ -\dfrac{H(k-1)}{\pi(k)} & \text{if} \quad s = k-1. \end{cases}$$

Let us calculate

$$\begin{aligned} \mathbf{P}^*\Lambda(k,s) &= \sum_r \mathbf{P}^*(k,r)\Lambda(r,s) \\ &= \mathbf{1}(k < M)\left(\mathbf{P}^*(k,k+1)\Lambda(k+1,s)\right) + \mathbf{1}(k = M)\left(\mathbf{P}^*(M,M)\Lambda(M,s)\right) \\ &= \mathbf{1}(k < M)\frac{\pi(s)\mathbf{1}(s \leq k+1)}{H(k+1)} + \mathbf{1}(k = M)\pi(s). \end{aligned}$$

Calculating transitions of antidual chain:

$$\begin{aligned} \mathbf{P}(k,r) &= \Lambda^{-1}\mathbf{P}^*_0\Lambda(k,r) = \sum_s \Lambda^{-1}(k,s)\mathbf{P}^*_0\Lambda(s,r) \\ &= \frac{H(k)}{\pi(k)}\mathbf{P}^*_0\Lambda(k,r) - \mathbf{1}(k>1)\frac{H(k-1)}{\pi(k)}\mathbf{P}^*_0\Lambda(k-1,r) \end{aligned}$$

Consider separately the cases

- $k = 1$. Then $\mathbf{P}(1,s) = \frac{H(1)}{\pi(1)}\mathbf{P}^*\Lambda(1,s) = \mathbf{P}^*\Lambda(1,s)$. This is nonzero only if $s = 1$ or $s = 2$.

$$\begin{aligned} \mathbf{P}(1,1) &= (1-p_1)\frac{\pi(1)}{H(1)} + p_1\frac{\pi(1)}{H(2)} = \frac{(1-p_1)(\pi(1)+\pi(2)) + p_1\pi(1)}{\pi(1)+\pi(2)} = \frac{\pi(1)+\pi(2)(1-p_2)}{\pi(1)+\pi(2)}, \\ \mathbf{P}(1,2) &= p_1\frac{\pi(2)}{H(2)} = \frac{p_1\pi(2)}{\pi(1)+\pi(2)}. \end{aligned}$$

- $k = M$. We have

$$\begin{aligned}
\mathbf{P}^*\Lambda(M,s) &= \pi(s)\\
\mathbf{P}^*\Lambda(M-1,s) &= (1-p_{M-1})\frac{\pi(s)\mathbf{1}(s\le M-1)}{H(M-1)} + \mathbf{1}(M-1<M)\frac{\pi(s)\mathbf{1}(s\le M)}{H(M)}p_{M-1}\\
&= (1-p_{M-1})\frac{\pi(s)\mathbf{1}(s\le M-1)}{H(M-1)} + \pi(s)p_{M-1}.
\end{aligned}$$

Thus

$$\begin{aligned}
\mathbf{P}(M,s) &= \frac{H(M)}{\pi(M)}\mathbf{P}^*\Lambda(M,s) - \mathbf{1}(M>1)\frac{H(M-1)}{\pi(M)}\mathbf{P}^*\Lambda(M-1,s)\\
&= \frac{H(M)}{\pi(M)}\pi(s) - \frac{H(M-1)}{\pi(M)}\left((1-p_{M-1})\frac{\pi(s)\mathbf{1}(s\le M-1)}{H(M-1)} + \pi(s)p_{M-1}\right)\\
&= \frac{\pi(s)}{\pi(M)} - \frac{\pi(s)}{\pi(M)}(1-p_{M-1})\mathbf{1}(s\le M-1) - \frac{\pi(s)}{\pi(M)}H(M-1)p_{M-1}\\
&= \frac{\pi(s)}{\pi(M)}\left[1-\pi_{M-1} + (1-p_{M-1})\mathbf{1}(s\le M-1) + \pi(M)p_{M-1}\right]\\
&= \begin{cases} p_{M-1}\pi(s) & \text{if} \quad s\le M-1,\\ 1-p_{M-1}+p_{M-1}\pi(M) & \text{if} \quad s=M.\end{cases}
\end{aligned}$$

- $1 < k < M$. We have

$$\mathbf{P}^*\Lambda(k-1,s) = (1-p_{k-1})\frac{\pi(s)\mathbf{1}(s\le k-1)}{H(k-1)} + p_{k-1}\frac{\pi(s)\mathbf{1}(s\le k)}{H(k)}.$$

Thus

$$\begin{aligned}
\mathbf{P}(k,s) &= \frac{H(k)}{\pi(k)}\mathbf{P}^*\Lambda(k,s) - \frac{H(k-1)}{\pi(k)}\mathbf{P}^*\Lambda(k-1,s)\\
&= \frac{H(k)}{\pi(k)}\left[(1-p_k)\frac{\pi(s)\mathbf{1}(s\le k)}{H(k)} + p_k\frac{\pi(s)\mathbf{1}(s\le k+1)}{H(k+1)})\right]\\
&- \frac{H(k-1)}{\pi(k)}\left[(1-p_{k-1})\frac{\pi(s)\mathbf{1}(s\le k-1)}{H(k-1)} + p_{k-1}\frac{\pi(s)\mathbf{1}(s\le k)}{H(k)}\right].
\end{aligned}$$

Consider three subcases:

◇ $s = k+1$. Then we have

$$\mathbf{P}(k,k+1) = p_k\frac{H(k)}{H(k+1)}\frac{\pi(k+1)}{\pi(k)}.$$

◇ $s = k$. Then we have

$$\begin{aligned}
\mathbf{P}(k,k) &= \frac{H(k)}{\pi(k)}\left[(1-p_k)\frac{\pi(k)}{H(k)} + p_k\frac{\pi(k)}{H(k+1)}\right] - \frac{H(k-1)}{\pi(k)}\left[p_{k-1}\frac{\pi(k)}{H(k)}\right]\\
&= (1-p_k)\frac{1}{H(k)} + p_k\frac{H(k)}{H(k+1)} - p_{k-1}\frac{H(k-1)}{H(k)}.
\end{aligned}$$

$\diamond$ $s < k$. Then we have

$$\begin{aligned}\mathbf{P}(k,s) &= \frac{H(k)}{\pi(k)}\left[(1-p_k)\frac{\pi(s)}{H(k)} + p_k\frac{\pi(s)}{H(k+1)}\right] \\ &- \frac{H(k-1)}{\pi(k)}\left[(1-p_{k-1})\frac{\pi(s)}{H(k-1)} + p_{k-1}\frac{\pi(s)}{H(k)}\right] \\ &= (1-p_k)\frac{\pi(s)}{\pi(k)} + p_k\frac{\pi(s)}{\pi(k)}\frac{H(k)}{H(k+1)} - (1-p_{k-1})\frac{\pi(s)}{\pi(k)} - p_{k-1}\frac{\pi(s)}{\pi(k)}\frac{H(k-1)}{H(k)} \\ &= \frac{\pi(s)}{\pi(k)}\left[p_{k-1}\left(1-\frac{H(k-1)}{H(k)}\right) - p_k\left(1-\frac{H(k)}{H(k+1)}\right)\right].\end{aligned}$$

This way we considered all the cases. The only thing left to calculate is the initial distribution of antidual chain. Using relation (1) we have

$$\nu(k) = \sum_{i=1}^{M}\nu^*(i)\Lambda(i,k) = \pi(k)\sum_{i=1}^{M}\frac{a_i\mathbf{1}(k\le i)}{H(i)} = \pi(k)\sum_{i=k}^{M}\frac{a_i}{H(i)}.$$

The matrix $\mathbf{P}^*$ is upper-triangular, thus $\{1-p_1,\ldots,1-p_{M-1},1\}$ are its eigenvalues. Because of the relation (1) these are also the eigenvalues of $\mathbf{P}$.

Acknowledgements

Author thanks Zbigniew Palmowski and Rafał Kulik for helpful suggestions and discussions.

References

[1] D. Aldous and P. Diaconis. Shuffling cards and stopping times. *American Mathematical Monthly*, 93(5):333–348, 1986.
[2] D. Aldous and P. Diaconis. Strong Uniform Times and Finite Random Walks. *Advances in Applied Mathematics*, 97:69–97, 1987.
[3] S. B. Connor. Separation and coupling cutoffs for tuples of independent Markov processes. *Latin American Journal of Probability and Mathematical Statistics*, 7(3):65–77, 2010.
[4] P. Diaconis and J. A. Fill. Strong stationary times via a new form of duality. *The Annals of Probability*, 18(4):1483–1522, 1990.
[5] P. Diaconis and L. Saloff-Coste. Separation cut-offs for birth and death chains. *The Annals of Applied Probability*, 16(4):2098–2122, 2006.
[6] P. Diaconis and M. Shahshahani. Generating a random permutation with random transpositions. *Zeitschrift fur Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 57(2):159–179, 1981.
[7] A. V. Doumas and V. G. Papanicolaou. The Coupon Collector's Problem Revisited: Generalizing the Double Dixie Cup Problem of Newman and Shepp. *ESAIM: Probability and Statistics*, 20:367—-399, 2016.
[8] P. L. Erdős and A. Rényi. On a classical problem of probability theory. *Publ. Math. Inst. Hung. Acad. Sci.*, Ser. A(6):215–220, 1961.
[9] W. Feller. *An introduction to probability theory and its applications.* 1971.
[10] J. A. Fill. An exact formula for the move-to-front rule for self-organizing lists. *Journal of Theoretical Probability*, 9(1):113–160, Jan. 1996.
[11] J. A. Fill. On hitting times and fastest strong stationary times for skip-free and more general chains. *Journal of Theoretical Probability*, 22(3):587–600, June 2009.
[12] J. A. Fill. The Passage Time Distribution for a Birth-and-Death Chain: Strong Stationary Duality Gives a First Stochastic Proof. *Journal of Theoretical Probability*, 22(3):543–557, 2009.
[13] J. A. Fill and V. Lyzinski. Hitting times and interlacing eigenvalues: a stochastic approach using intertwinings. *Journal of Theoretical Probability*, 27(3):954—-981, 2014.
[14] L. Holst. Extreme Value Distributions for Random Coupon Collector and Birthday Problems. *Extremes*, 4(2):129–145, 2001.
[15] S. Karlin and J. McGregor. Coincidence properties of birth and death processes. *Pacific Journal of Mathematics*, 9(4):1109–1140, 1959.

[16] J. Keilson. Log-Concavity and Log-Convexity in Passage Time Densities of Diffusion and Birth-Death Processes on JSTOR. *Journal of applied probability*, 8(2):391—-398, 1971.
[17] H. Lacoin. The Cutoff profile for the Simple-Exclusion process on the circle. *Ann. Probab.*, 44(5):3399–3430, 2016.
[18] P. Lorek. Generalized Gambler's Ruin Problem: Explicit Formulas via Siegmund Duality. *Methodology and Computing in Applied Probability*, pages 1–11, 2016.
[19] P. Lorek. Siegmund duality for Markov chains on partially ordered state spaces. *Submitted*, pages 1–14, 2016.
[20] P. Lorek and P. Markowski. Monotonicity requirements for exact sampling with Markov chains. *Submitted*, 2016.
[21] P. Lorek and R. Szekli. Strong stationary duality for Möbius monotone Markov chains. *Queueing Systems*, 71(1-2):79–95, Mar. 2012.
[22] Y.-h. Mao, C. Zhang, and Y.-h. Zhang. Separation cutoff for upward skip-free chains. *to appear in Journal of Applied Probability*, pages 1–15, 2015.
[23] L. Miclo. On absorption times and Dirichlet Eigenvalues. *ESAIM: Probability and Statistics*, 14:117—-150, 2010.
[24] P. Neal. The generalised coupon collector problem. *Journal of Applied Probability*, 45(3):621–629, Sept. 2008.
[25] D. J. Newman. The double dixie cup problem. *American Mathematical Monthly*, 67(1):58–61, 1960.
[26] G.-C. Rota. On the foundations of combinatorial theory I. Theory of Möbius functions. *Probability theory and related fields*, 368:340–368, 1964.
[27] D. Siegmund. The Equivalence of Absorbing and Reflecting Barrier Problems for Stochastically Monotone Markov Processes. *The Annals of Probability*, 4(6):914–924, 1976.