ESAIM: PS 23 (2019) 739–769
https://doi.org/10.1051/ps/2019004

ESAIM: Probability and Statistics
www.esaim-ps.org

ANTIDUALITY AND MÖBIUS MONOTONICITY: GENERALIZED COUPON COLLECTOR PROBLEM☆

Paweł Lorek*

Abstract. For a given absorbing Markov chain X^* on a finite state space, a chain X is a sharp antidual of X^* if the fastest strong stationary time (FSST) of X is equal, in distribution, to the absorption time of X^*. In this paper, we show a systematic way of finding such an antidual based on some partial ordering of the state space. We use a theory of strong stationary duality developed recently for Möbius monotone Markov chains. We give several sharp antidual chains for Markov chain corresponding to a generalized coupon collector problem. As a consequence – utilizing known results on the limiting distribution of the absorption time – we indicate separation cutoffs (with their window sizes) in several chains. We also present a chain which (under some conditions) has a prescribed stationary distribution and its FSST is distributed as a prescribed mixture of sums of geometric random variables.

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

Received November 28, 2016. Accepted March 5, 2019.

1. Introduction

Strong stationary times (SSTs) are a probabilistic tool for bounding a rate of convergence to stationarity for Markov chains. Aldous and Diaconis [1, 2] gave several examples of chains where SST was found *ad hoc*. Later in [8] 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. Moreover, they provide a probabilistic construction, *i.e.*, the coupling of the ergodic chain with its SSD is presented (their construction works for any initial distribution). The authors also proved that there always exists a *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 an SSD assuming that the time reversed chain is stochastically monotone w.r.t. the linear ordering. In particular, they consider a 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 a set-valued SSD (see their Sect. 3.4 "greedy construction of a set-valued

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

Keywords and phrases: Markov chains, strong stationary duality, antiduality, absorption times, fastest strong stationary times, Möbius monotonicity, generalized coupon collector problem, Double Dixie cup problem, separation cutoff, partial ordering, perfect simulation.

Mathematical Institute, University of Wrocław, pl. Grunwaldzki 2/4, 50-384 Wrocław, Poland.

* Corresponding author: pawel.lorek@math.uni.wroc.pl

Article published by EDP Sciences

dual"). In this paper, we actually start with some absorbing chain and show that it is a sharp SSD of a class (which we indicate) of ergodic chains. We exploit the results from [30], where the authors provided the recipe for constructing 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 which an 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 a so-called *mixing time* (*i.e.*, the time when the chain is "close" to its stationary distribution). However, sometimes we can say much more than just a mixing time by showing that a 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: the separation cutoff and the total variation cutoff, which differ in the way used to measure the convergence (the separation *vs.* the total variation distance).

The total variation cutoff was first shown for a *random transposition* card shuffling in [12]. The name comes from [1], where the authors showed that a *top-to-random* card shuffling exhibits a total variation cutoff. The separation cutoff has been studied in various contexts. For example, in [11] authors gave if and only if conditions for the existence of a separation cutoff for birth and death chains (they use duality theory to convert convergence rates to hitting times and Keilson's representation of first hitting times) – they show that there is a cutoff if and only if the product of a spectral gap and a mixing time tends to infinity; this was somehow extended – in [4] authors show that there is a cutoff measured in L^p-norm ($1 < p \leq \infty$) if and only if the spectral gap and max-L^p mixing time tend to infinity; computation of a cutoff time and a window size in a variety of birth and death chains is given in [5]; the separation cutoff for skip-free chains was given in [32]; some other specific chains were considered in [7]; in [17] author gives a formula for the separation for *Tsetlin library* chain specifying weights for which there is and there is no separation cutoff. Several examples of both separation and total variation cutoffs are given in [26], some characterization of the total variation cutoff for lazy (*i.e.*, with the probability of staying $\geq 1/2$) chains was given in [3]. In [6] authors give sufficient condition for skip-free chains to have real eigenvalues, they use Siegmund duality – actually antiduality – a type of transitions of their (anti)dual resembles some chains we obtain for a coupon collector problem. It is worth mentioning that although a sequence of birth and death chains exhibits the total variation cutoff if and only if it exhibits the separation cutoff [11, 13], it is not the case (in general) for other chains, as shown in [21].

As mentioned before, FSST is equal in distribution to the absorption time of the sharp SSD chain. Thus, there is a close relation between a sharp SSD and the separation cutoff. Roughly speaking, this cutoff can be studied by studying the limiting distribution of the 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 an approach was considered in [19] in a context of birth and death chains only. A connection between the separation cutoff and a coupon collector problem (including some generalizations, *e.g.*, sampling $k > 1$ different coupons at a time) was given in [36].

Using this approach we will indicate the separation cutoff time and a 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 [30]. 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 a product space) exhibiting a cutoff. Such an example was given in [25].

The absorption time of many absorbing chains is distributed as a mixture of sums of geometric random variables with parameters being the eigenvalues of the transition matrix. For example, the absorption time of a discrete time birth and death chain starting at the minimal state with the 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 [23] or Keilson [24]. Fill [19] gave a stochastic proof of this result using also the theory of SSD (the result was simultaneously obtained in [9]), later it was extended to skip-free Markov chains in Fill [18]. Miclo [33] showed that for a large class of absorbing chains on a finite state space, the absorption time is distributed as a 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 we 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 a 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; and (iii) we present a construction of a chain with a prescribed FSST and a 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 the stationary distribution π and run it *long enough*. The most common algorithms for such constructions are Metropolis–Hastings algorithm and Gibbs sampler (for studies on rate of convergence for Metropolis–Hasting algorithms see, *e.g.*, [10], the cutoff for Gibbs sampler for Ising model on the lattice was studied on [31]). This paper suggests an alternative approach: given π on $\mathbb{E}$ finds some absorbing chain on $\mathbb{E}$ and then calculates sharp antidual chain having this π as stationary distribution. Knowing, *e.g.*, the expectation and the variance of the absorption time, one can quite precisely determine the number of steps needed for simulations. Moreover, having a sharp SSD actually can allow for a *perfect simulation* from the 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 [8] (Sect. 2.4), [20] (Sect. 1.1), or [29] (Sect. 2.3, Algorithm 4). 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 a strong stationary duality and the separation cutoff. In Section 3, we recall a notion of Möbius monotonicity and give a matrix-form proof of the result from [30]. 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 an ergodic chain with a prescribed stationary distribution and with a 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 an ergodic (*i.e.*, irreducible and aperiodic) Markov chain $X \sim (\nu, \mathbf{P})$ on a finite state space $\mathbb{E} = \{\mathbf{e}_1, \ldots, \mathbf{e}_M\}$ with an initial distribution ν, a stationary distribution π and a transition matrix $\mathbf{P}$. Let $\mathbb{E}^* = \{\mathbf{e}_1^*, \ldots, \mathbf{e}_N^*\}$ be a state space of an absorbing Markov chain $X^* \sim (\nu^*, \mathbf{P}^*)$, whose unique absorbing state and unique irreducible class are 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 an *SSD* of X with link Λ if

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

Diaconis and Fill [8] prove that then the absorption time T^* of X^* is a so-called *SST* for X. This is a finite random variable T such 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 a random variable we always have: $d_{TV}(\nu\mathbf{P}^k,\pi) \leq \mathrm{sep}(\nu\mathbf{P}^k,\pi) := \max_{\mathbf{e}\in\mathbb{E}}\left(1-\nu\mathbf{P}^k(\mathbf{e})/\pi(\mathbf{e})\right) \leq P(T>k)$ for any $k \in \mathbb{Z}_+$, where d_{TV} stands for *the total variation distance*, and sep stands for *the separation*. Note that sep is not symmetric and thus is not a distance between probability measures. The corresponding T^* is *sharp* if $\mathrm{sep}(\nu\mathbf{P}^k,\pi) = P(T^*>k)$ for any $k\in\mathbb{Z}_+$. In such a case T^* is called *the FSST* for X, which we denote by T_{FSST}. For more details on this duality consult [8]. Moreover, duality relation (2.1) allows for stochastic constructions, see, *e.g.*, [19], where stochastic proof for passage time distribution for a birth and death chain was given.

Note that once we fix $\mathbb{E}^*$ and a link Λ, and if there exists a right-inverse of Λ, *i.e.*, Λ^{-1} we can simply calculate from (2.1):

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

If the resulting $\mathbf{P}^*$ is a stochastic and aperiodic matrix, which is irreducible outside the absorbing point which can be attained from other points, and if ν^* is a probability distribution, then (it will always correspond to an absorbing chain) we have found an 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 (again, Λ^{-1} is its right-inverse), 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 an SST for X. In such a case X is called an *sharp antidual* of X^*. Moreover, if we somehow know, that for some class of links relation (2.1) implies that T^* is sharp (see Cor. 3.3), then we can possibly find many different antiduals, which all have the same FSST T^*, which has a phase-type distribution. In such a case X is called a *sharp antidual* of X^*.

2.2. The separation cutoff

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

$$\mathrm{sep}(\nu\mathbf{P}^k,\pi) = P(T_{\mathrm{FSST}}>k) = P(T^*>k). \tag{2.2}$$

Thus, studying the distribution of T_{FSST} is equivalent to studying the distribution of T^*. Furthermore, the separation cutoff can be studied by studying the properties of T^*. In what follows, we introduce the notion of the separation cutoff. Since the definition of the cutoff involves a sequence of state spaces, 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 the *separation cutoff* at time t_d with the *window size* $w_d = o(t_d)$ if

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

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

3. Möbius monotonicity and duality

In general, there is no recipe on how to find an SSD, *i.e.*, a triplet $\mathbb{E}^*, \mathbf{P}^*, \Lambda$. In [8] authors give a recipe for a dual on the same state space $\mathbb{E}^* = \mathbb{E}$ provided that a time reversed chain $\overleftarrow{X}$ is stochastically monotone with respect to total ordering. In [30], we give an extension of this result to state spaces which are only partially ordered by $\preceq$. Then, provided that the time reversed chain $\overleftarrow{X}$ is *Möbius* monotone (plus some conditions on the initial distribution), we give a formula for a sharp SSD on the same state space $\mathbb{E}^* = \mathbb{E}$.

The Möbius monotonicity seems to be a natural one for extension of the main result from [8] to partially ordered state spaces. In [28] we show that it is equivalent to the existence of a Siegmund dual of a chain with given partial ordering. For a linearly ordered state space, stochastic monotonicity of a chain is required for the existence of a Siegmund dual (see [38]), and stochastic monotonicity of a time reversal is required for the existence of an SSD with a link being a truncated stationary distribution (see [8]). 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 [29], and for applications of a Siegmund duality to some generalizations of a gambler's ruin problem consult [27], we will introduce this monotonicity by trying to solve (2.1) with some given link Λ.

We consider a finite state space $\mathbb{E} = \{\mathbf{e}_1, \ldots, \mathbf{e}_M\}$ partially ordered by $\preceq$ such that $\mathbf{e}_M$ is the unique maximal state. For a 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 an SSD X^* with a transition matrix $\mathbf{P}^*$ on the same state space $\mathbb{E}^* = \mathbb{E}$ with a 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.*,

$$\Lambda(\mathbf{e}_i, \mathbf{e}_j) = \frac{\pi(\mathbf{e}_j)}{\sum_{\mathbf{e}':\mathbf{e}' \preceq \mathbf{e}_i} \pi(\mathbf{e}')} \mathbf{1}(\mathbf{e}_j \preceq \mathbf{e}_i). \tag{3.1}$$

Note that for all $\mathbf{e} \in \mathbb{E}$ 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.*, $\mathbf{C}(\mathbf{e}_M, \mathbf{e}_j) = \mathbf{1}(\mathbf{e}_j = \mathbf{e}_M)$ for all j and there is no $\mathbf{e}_{M_2} \neq \mathbf{e}_M$ such that $\mathbf{C}(\mathbf{e}_{M_2}, \mathbf{e}_j) = \mathbf{1}(\mathbf{e}_j = \mathbf{e}_{M_2})$ for all j). We identify the ordering $\preceq$ with the matrix $\mathbf{C}$. Then the link can be written in a matrix form:

$$\Lambda = (\mathbf{diag}(\boldsymbol{\pi}\mathbf{C}))^{-1}\mathbf{C}^T\mathbf{diag}(\boldsymbol{\pi}), \tag{3.2}$$

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 $i \leq j$, which means that $\mathbf{C}$, and thus Λ, is invertible. Often, $\mu \equiv \mathbf{C}^{-1}$ is called the *Möbius function* or the *Möbius matrix* of the partial order $\preceq$. Solving (2.1) for $\mathbf{P}^*$ yields (recall that the transitions of time reversed chains are given by $\overleftarrow{\mathbf{P}} = (\mathbf{diag}(\boldsymbol{\pi}))^{-1}\mathbf{P}^T(\mathbf{diag}(\boldsymbol{\pi})))$

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

which is a stochastic matrix if and only if each entry of $\mathbf{C}^{-1}\overleftarrow{\mathbf{P}}\mathbf{C}$ is non-negative, in other words we say that $\overleftarrow{\mathbf{P}}$ is Möbius monotone. This way we proved the main part of Theorem 2 of [30]. We include it here, since this is a little bit different (matrix-form) proof. We will restate the theorem for completeness, introducing formal definitions of monotonicities first. For given partial ordering $\preceq$ and any matrix $\mathbf{P}$ (not necessarily stochastic) we define $\mathbf{P}(\mathbf{e}, \{\mathbf{e}_j\}^\downarrow) = \sum_{\mathbf{e}':\mathbf{e}' \preceq \mathbf{e}_j} \mathbf{P}(\mathbf{e}, \mathbf{e}')$ and similarly $\mathbf{P}(\mathbf{e}, \{\mathbf{e}_j\}^\uparrow) = \sum_{\mathbf{e}':\mathbf{e}' \succeq \mathbf{e}_j} \mathbf{P}(\mathbf{e}, \mathbf{e}')$.

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

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

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

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 non-negative). 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 3.1. In Lorek and Szekli [30], this Möbius monotonicity (of both, function and chain) was called ${}^{\downarrow}$-*Möbius monotonicity* (see Defs. 2.1 and 2.2 therein).

Definition 3.3. X is ${}^{\uparrow}$-Möbius monotone if $(\mathbf{C}^T)^{-1}\mathbf{P}\mathbf{C}^T \geq 0$ (each entry non-negative).

Theorem 3.2 ([30], Thm. 2)**.** *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 a 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 an SSD chain $X^* \sim (\nu^*, \mathbf{P}^*)$ on $\mathbb{E}^* = \mathbb{E}$ with the following link

$$\Lambda = (\mathbf{diag}(\pi\mathbf{C}))^{-1}\mathbf{C}^T\mathbf{diag}(\pi). \tag{3.3}$$

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

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

The following corollary will play a crucial role:

Corollary 3.3. *The SSD constructed in Theorem 3.2 is sharp.*

Proof. The link given in (3.3) is lower-triangular, thus, by Remark 2.39 in [8], 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 a sharp antidual (on the same state space $\mathbb{E} = \mathbb{E}^*$) chain of some given already absorbing chain $X^* \sim (\nu^*, \mathbf{P}^*)$ with the unique absorbing state $\mathbf{e}_M$. The idea is clear from the previous section: introduce some partial ordering and some distribution π on $\mathbb{E}$ such that $\pi(\mathbf{e}) > 0$ for all $\mathbf{e} \in \mathbb{E}$. Then solve $\Lambda\mathbf{P} = \mathbf{P}^*\Lambda$ for $\mathbf{P}$ with the link given in (3.3). If the resulting matrix is non-negative, it will be a stochastic matrix. Since $\mathbf{P}$ and $\mathbf{P}^*$ have the same spectrum, the eigenvalue $\lambda_1 = 1$ for $\mathbf{P}$ has multiplicity 1, this implies that $\mathbf{P}$ has a unique recurrence class, which is the whole set $\mathbb{E}$ (since $\pi(\mathbf{e}) > 0$ for all $\mathbf{e} \in \mathbb{E}$). Thus, $\mathbf{P}$ is a transition matrix of an ergodic Markov chain X with the 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 FSST T_{FSST}.

Fix some partial ordering $\preceq$ on $\mathbb{E}^*$ (expressed by $\mathbf{C}$) having the unique maximal state $\mathbf{e}_M$ and some distribution π on $\mathbb{E}$, such that $\pi(\mathbf{e}) > 0$ for all $\mathbf{e} \in \mathbb{E}$. 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 3.3 was stated for a Markov chain X with a 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 $\mathbb{E}^* = \{\mathbf{e}_1, \dots, \mathbf{e}_M\}$ with the 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 such that $\widehat{\mathbf{P}^*}$ is $^{\uparrow}$-Möbius monotone:*

$$\mathcal{P}(\mathbf{P}^*) = \left\{(\pi, \mathbf{C}) : \mathbf{C} \in \mathcal{C}, \forall(\mathbf{e} \in \mathbb{E})\pi(\mathbf{e}) > 0, \widehat{\mathbf{P}^*} \text{ is } ^{\uparrow}\text{-Möbius monotone}\right\}.$$

Then for any $(\pi, \mathbf{C}) \in \mathcal{P}(\mathbf{P}^*)$ the chain $X \sim (\nu, \mathbf{P})$ with the link Λ defined in (3.3) 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 (3.3).

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 non-negative. We will show that π is its stationary distribution. Let $\boldsymbol{\eta} = (0, \dots, 0, 1)$. 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 the rows of $\mathbf{P}$ sum up to 1, *i.e.*, that $\mathbf{P}(1, \dots, 1)^T = (1, \dots, 1)^T$. We have

$$\begin{aligned}
\mathbf{P}(1, \dots, 1)^T &= (\mathbf{diag}\boldsymbol{\pi})^{-1}(\mathbf{C}^T)^{-1}\widehat{\mathbf{P}^*}\mathbf{C}^T\mathbf{diag}(\boldsymbol{\pi})(1, \dots, 1)^T \\
&= (\mathbf{diag}\boldsymbol{\pi})^{-1}(\mathbf{C}^T)^{-1}\widehat{\mathbf{P}^*}\mathbf{C}^T\boldsymbol{\pi}^T = (\mathbf{diag}\boldsymbol{\pi})^{-1}(\mathbf{C}^T)^{-1}\widehat{\mathbf{P}^*}(\boldsymbol{\pi}\mathbf{C})^T \\
&= (\mathbf{diag}\boldsymbol{\pi})^{-1}(\mathbf{C}^T)^{-1}\mathbf{diag}(\boldsymbol{\pi}\mathbf{C})\mathbf{P}^*(\mathbf{diag}(\boldsymbol{\pi}\mathbf{C}))^{-1}(\boldsymbol{\pi}\mathbf{C})^T \\
&= (\mathbf{diag}\boldsymbol{\pi})^{-1}(\mathbf{C}^T)^{-1}\mathbf{diag}(\boldsymbol{\pi}\mathbf{C})\mathbf{P}^*(1, \dots, 1)^T \\
&= (\mathbf{diag}\boldsymbol{\pi})^{-1}(\mathbf{C}^T)^{-1}\mathbf{diag}(\boldsymbol{\pi}\mathbf{C})(1, \dots, 1)^T \stackrel{(*)}{=} (1, \dots, 1)^T. \qquad (4.1)
\end{aligned}$$

To show $(*)$ we need to show that $\sum_{\mathbf{e}' \in \mathbb{E}^*}((\mathbf{diag}\boldsymbol{\pi})^{-1}(\mathbf{C}^T)^{-1}\mathbf{diag}(\boldsymbol{\pi}\mathbf{C}))(\mathbf{e}, \mathbf{e}') = 1$ for any $\mathbf{e} \in \mathbb{E}^*$. For diagonal matrices $\mathbf{D}_1$, $\mathbf{D}_2$ and a square matrix $\mathbf{A}$ (all of the same sizes) we have $\mathbf{D}_1\mathbf{A}\mathbf{D}_2(\mathbf{e}, \mathbf{e}') =$

$\mathbf{D}_1(\mathbf{e},\mathbf{e})\mathbf{A}(\mathbf{e},\mathbf{e}')\mathbf{D}_2(\mathbf{e}',\mathbf{e}')$, thus

$$\sum_{\mathbf{e}'\in\mathbb{E}^*}((\mathbf{diag}\boldsymbol{\pi})^{-1}(\mathbf{C}^T)^{-1}\mathbf{diag}(\boldsymbol{\pi}\mathbf{C}))(\mathbf{e},\mathbf{e}') = \frac{1}{\pi(\mathbf{e})}\sum_{\mathbf{e}'\in\mathbb{E}^*}\mathbf{C}^{-1}(\mathbf{e}',\mathbf{e})\mathbf{diag}(\pi\mathbf{C})(\mathbf{e}',\mathbf{e}')$$
$$= \frac{1}{\pi(\mathbf{e})}(\boldsymbol{\pi}\mathbf{C}\mathbf{C}^{-1})(\mathbf{e}) = \frac{1}{\pi(\mathbf{e})}\pi(\mathbf{e}) = 1.$$

Thus, $\mathbf{P}$ is a stochastic matrix and thus $X\sim(\nu,\mathbf{P})$ is a Markov chain with the stationary distribution π. Since (2.1) holds, X^* is an SSD for X. Theorem 3.2 and Corollary 3.3 imply that X^* is a sharp SSD of X. □

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 Möbius monotone (w.r.t. π and $\mathbf{C}$) is equivalent to non-negativity of the resulting matrix $\mathbf{P}$. In examples, it is often more convenient to calculate Λ and Λ^{-1} directly.

4.2. Antidual chains for a generalized coupon collector problem

Consider d different types of coupons. These are sampled independently with replacement. Sampled types are recorded. For $1\le k\le d$ let $p_k>0$ be the probability that the coupon of type k is sampled, with $\sum_{k=1}^d p_k\le 1$. With the remaining probability, *i.e.*, with probability $1-\sum_{k=1}^d p_k$, no coupon 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 $N_1,\ldots,N_d$. Let $(i_1,\ldots,i_d)$ denote that a coupon of type j was sampled i_j times. If $i_j=N_j$ and coupon of type j is sampled, the chain does not move. 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 $\mathbb{E}^*=\{(i_1,\ldots,i_d): 0\le i_j\le N_j, 1\le j\le d\}$ with the initial distribution $\nu^*=\delta_{(0,\ldots,0)}$ and the following transition matrix:

$$\mathbf{P}^*((i_1,\ldots,i_d),(i_1',\ldots,i_d')) = \begin{cases} p_j & \text{if}\quad i_j'=i_j+1, i_k'=i_k, k\neq j,\\ 1-\sum_{k=1}^d p_k+\sum_{k:i_k=N_k} p_k & \text{if}\quad i_j'=i_j, 1\le j\le d.\end{cases} \tag{4.2}$$

We refer to $\mathbf{P}^*$ as to a *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 [16]. The term *generalized* is not unique. It is used when sequence $\{p_k\}$ is general but $N_1=\ldots=N_d=1$ (*e.g.*, [34]) or when $p_k=1/d$ but we are to collect more coupons of each type (see, *e.g.*, [14, 35]). Although the chain $\mathbf{P}^*$ given in (4.2) includes both mentioned generalizations, we consider two antidual chains for two different cases separately:

(a) for general $N_j\ge 1$ and $p_j, j=1,\ldots,d$ with the uniform stationary distribution of an antidual chain;
(b) for general p_j but $N_j=1, j=1,\ldots,d$ with more general stationary distribution of an 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 the position k).

Case: general $N_j \geq 1$ and $p_j, j = 1, \dots, d$ and a uniform stationary distribution of an antidual chain.

Theorem 4.4. *Let $X^* \sim (\nu^*, \mathbf{P}^*)$ be a generalized coupon collector chain with the transition matrix given in (4.2) with fixed integers $N_j \geq 1, j = 1, \dots, d$. Moreover, assume that*

$$\sum_{j=1}^{d} \left(1 - \frac{1}{N_j(N_j+1)}\right) p_j \leq 1. \tag{4.3}$$

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

$$\mathbf{P}(\mathbf{i}^{(1)}, \mathbf{i}^{(2)}) = \begin{cases} \dfrac{i_k^{(1)}+1}{i_k^{(1)}+2} p_k & \text{if} \quad \mathbf{i}^{(2)} = \mathbf{i}^{(1)} + \mathbf{s}_k, \\ \left(\dfrac{\mathbf{1}(i_k^{(1)} < N_k)}{(i_k^{(1)}+1)(i_k^{(1)}+2)} + \dfrac{\mathbf{1}(i_k^{(1)} = N_k)}{N_k+1}\right) p_k & \text{if} \quad \mathbf{i}^{(2)} = \mathbf{i}^{(1)} - m \cdot \mathbf{s}_k \\ & \quad \text{with} \quad 1 \leq m \leq i_k, \\ 1 - \sum\limits_{j: i_j^{(1)} < N_j} \left(1 - \dfrac{1}{(i_j^{(1)}+1)(i_j^{(1)}+2)}\right) p_j - \sum\limits_{j: i_j^{(1)} = N_j} \dfrac{N_j}{N_j+1} p_j & \text{if} \quad \mathbf{i}^{(2)} = \mathbf{i}^{(1)} \end{cases} \tag{4.4}$$

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

Remark 4.5. Note that for example for $N_1 = \ldots = N_j = 1$, the condition (4.3) if always fulfilled.

Roughly speaking, the antidual has the following transitions. Being in state $(i_1, \dots, i_d)$ 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. 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 $1 \leq m < i_j$ (the probability depends only on i_j and the formula itself is different on the border, *i.e.*, when $i_j = N_j$).

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

Theorem 4.6. *Let $X^* \sim (\nu^*, \mathbf{P}^*)$ be a generalized coupon collector chain with the transition matrix given in (4.2). Assume that $N_1 = \ldots = N_d = 1$. Let $a_k \in (0,1)$ for $k = 1, \dots, d$. Then the chain $X \sim (\nu, \mathbf{P})$ on the same state space $\mathbb{E} = \mathbb{E}^* = \{0,1\}^d$ with the initial distribution $\nu = \nu^* = \delta_{(0,\dots,0)}$ and the transition matrix*

$$\mathbf{P}(\mathbf{i}^{(1)}, \mathbf{i}^{(2)}) = \begin{cases} a_k p_k & if \quad \mathbf{i}^{(2)} = \mathbf{i}^{(1)} + \mathbf{s}_k, \\ 1 - \sum\limits_{j: i_j^{(1)} = 0} a_j p_j - \sum\limits_{j: i_j^{(1)} = 1} (1 - a_j) p_j & if \quad \mathbf{i}^{(2)} = \mathbf{i}^{(1)}, \\ (1 - a_k) p_k & if \quad \mathbf{i}^{(2)} = \mathbf{i}^{(1)} - \mathbf{s}_k, \end{cases} \tag{4.5}$$

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

$$\pi(\mathbf{i}) \ = \ \prod_{j=1}^{d} [a_j \mathbf{1}(i_j = 1) + (1 - a_j) \mathbf{1}(i_j = 0)]. \tag{4.6}$$

Remark 4.7. The proof of Theorem 4.6 implies that the antidual chain $X \sim (\nu, \mathbf{P})$ has transitions consistent with the 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. It can happen, that for some π two coordinates change at a time or antidual does not exist (since some entries of $\mathbf{P}$ can be negative). This is further commented after a proof in Remark 5.1.

Taking the following sequences: a_k: $a_k = \frac{b}{a+b}, j = 1, \ldots, d$ we obtain the following special case.

Corollary 4.8. *The chains $X^{(1)} \sim (\nu, \mathbf{P}_1)$ with an initial distribution $\nu = \delta_{(0,\ldots,0)}$, the transition matrix*

$$\mathbf{P}_1(\mathbf{i}^{(1)}, \mathbf{i}^{(2)}) \;=\; \begin{cases} \dfrac{b}{a+b} p_k & if \quad \mathbf{i}^{(2)} = \mathbf{i}^{(1)} + \mathbf{s}_k, \\ 1 - \dfrac{b}{a+b} \displaystyle\sum_{j: i_j^{(1)} = 0} p_r - \dfrac{a}{a+b} \displaystyle\sum_{j: i_j^{(1)} = 1} p_r & if \quad \mathbf{i}^{(2)} = \mathbf{i}^{(1)}, \\ \dfrac{a}{a+b} p_k & if \quad \mathbf{i}^{(2)} = \mathbf{i}^{(1)} - \mathbf{s}_k, \end{cases}$$

and with the stationary distribution

$$\pi_1(\mathbf{i}) = \frac{a^{d-|\mathbf{i}|} b^{|\mathbf{i}|}}{(a+b)^d}$$

(where $|\mathbf{i}| = \sum_{j=1}^d i_j$, called a level of $\mathbf{i}$) is a sharp antidual chain for $\mathbf{P}^*$ given in (4.2).

Note that for $a_k = \frac{1}{2}, j = 1, \ldots, d$ the stationary distribution of $X^{(1)}$ is uniform.

Remark 4.9. In [30] we considered the chain on $\mathbb{E} = \{0,1\}^d$ (*i.e.*, $N_1 = N_2 = \ldots = N_d = 1$) with a transition matrix $\mathbf{P}_2$ given by

$$\mathbf{P}_2(\mathbf{i}^{(1)}, \mathbf{i}^{(2)}) = \begin{cases} \alpha_k & \text{if} \quad \mathbf{i}^{(2)} = \mathbf{i}^{(1)} + \mathbf{s}_k, \\ 1 - \displaystyle\sum_{j: i_j^{(1)} = 0} \alpha_j - \displaystyle\sum_{j: i_j^{(1)} = 1} \beta_j & \text{if} \quad \mathbf{i}^{(2)} = \mathbf{i}^{(1)}, \\ \beta_k & \text{if} \quad \mathbf{i}^{(2)} = \mathbf{i}^{(1)} - \mathbf{s}_k. \end{cases}$$

The chain is reversible with product form stationary distribution:

$$\pi_2(\mathbf{i}) = \prod_{j: i_j = 0} \frac{\beta_j}{\alpha_j + \beta_j} \prod_{j: i_j = 1} \frac{\alpha_j}{\alpha_j + \beta_j}.$$

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

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

what is our absorbing dual (4.2) we started with, with $p_j = \alpha_j + \beta_j$ and $N_j = 1, j = 1, \ldots, d$. Note that $\mathbf{P}_2$ is a special case of $\mathbf{P}$ given in (4.5) with $a_j = \frac{\alpha_j}{\alpha_j + \beta_j}$.

Corollary 4.10. *The matrices* $\mathbf{P}$ *given in (4.4) and in (4.5) have eigenvalues of the form:*

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

(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 (4.2) is upper triangular, thus the 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 an SSD have the same set of eigenvalues, what is a direct consequence of relation (2.1). □

Remark 4.11. Fix d and $N_j = N, j = 1,\dots,d$. One can ask the following question: for what sequence $\{p_k\}$ is the associated T_{FSST} stochastically the smallest? Conjecture 2 in [14] suggests that this is in the case of equal probabilities $p_k = 1/d$.

4.3. Results on the separation cutoff

Since obtained antidual chains are sharp (*i.e.*, (2.2) holds), we can present a series of results on the separation cutoff utilizing existing results on the 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* $d = 1, 2,\dots$ *on* $\mathbb{E}_{(d)} = \{0,1\}^d$ *with an initial distribution* $\nu_{(d)} = \delta_{(0,\dots,0)}$ *and the transition matrix* $\mathbf{P}_{(d)}$ *given in (4.5) with* $p_k = \frac{1}{d}$ *and any* $a_k \in (0,1)$ *for* $k = 1,\dots,d$*. The stationary distribution* $\pi_{(d)}$ *is given in (4.6). The sequence exhibits the separation cutoff at time* $d\log d$ *with the 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 [22].

Taking $t_d = d\log d$ and $w_d = d$ we have

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

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

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

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

Results on the limiting distribution of T^*_d from [34] allow us to 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 $\lambda_1,\dots,\lambda_k > 0$ *and* $0 = n_0 < n_1 < \dots < n_k = 1$*. Assume that* $\lambda_1 < \lambda_2 < \ldots < \lambda_k$*. Consider a sequence of Markov chains* $X_{(d)}$ *indexed by* $d = 1,2,\dots$ *on* $\mathbb{E}_{(d)} = \{0,1\}^d$ *with an initial distribution* $\nu_{(d)} = \delta_{(0,\dots,0)}$ *and* *the transition matrix* $\mathbf{P}_{(d)}$ *given in (4.5) with*

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

and any $a_k \in (0,1)$ *for* $k = 1, \ldots, d$. *The stationary distribution* $\pi_{(d)}$ *is given in (4.6). The sequence exhibits the separation cutoff at time* $t_d = \frac{d}{\lambda_1}(\log d - \log(n_1))$ *with the window size* $w_d = \frac{d}{\lambda_1}$.

Proof. Denote the FSST of the chain by T_d^* (which is equal, in distribution, to collecting d coupons). We have

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

Lemma 3.1 in [34] implies that $\frac{1}{d}(T_d^* - \frac{1}{\lambda_1}d\log d)$ converges in distribution to a random variable Z with c.d.f. $P(Z \leq c) = e^{-n_1 e^{-\lambda_1 c}}$. Thus, we have

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

Similarly

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

and

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

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* $d = 1, 2, \ldots$ *on* $\mathbb{E}_{(d)} = \{0, 1, \ldots, N\}^d$ *with an initial distribution* $\nu_{(d)} = \delta_{(0,\ldots,0)}$ *and the transition matrix* $\mathbf{P}_{(d)}$ *given in (4.4) with* $p_k = \frac{1}{d}$ *and* $N_1 = \ldots = N_d = N \geq 2$ *(so that (4.3) holds). The stationary distribution* $\pi_{(d)}$ *is uniform. The sequence of chains exhibits the separation cutoff at time* $d\log d + (N-1)d\log\log d$ *with the window size* d.

Proof. In [15] authors derived the limiting distribution of T_d^* showing that

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

(where $\gamma = 0.57721\ldots$ is the Euler–Mascheroni constant) converges in distribution to a standard Gumbel random variable. Similar calculations as in Corollary 4.12 finish the proof.

□

Authors in [14] extended the result of [15] obtaining the limiting distribution of T_d^* for $N_1 = \ldots = N_d = N$ and for quite general choices of probabilities p_k. Let us indicate here one example (which actually includes the result of Cor. 4.14 as a special case).

Corollary 4.15. *Consider a sequence of Markov chains $X_{(d)}$ indexed by $d = 1, 2, \ldots$ on $\mathbb{E}_{(d)} = \{0, 1, \ldots, N\}^d$ with an initial distribution $\nu_{(d)} = \delta_{(0,\ldots,0)}$ and the transition matrix $\mathbf{P}_{(d)}$ given in (4.4) with*

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

and $N_1 = \ldots = N_d = N \geq 2$ (so that (4.3) holds). The stationary distribution $\pi_{(d)}$ is uniform. The sequence of chains exhibits the separation cutoff at time $d \log d + (N-1) d \log \log d$ with the window size d.

Proof. In [14] 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 an ergodic chain with a prespecified FSST and an 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 the 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 $\mathbf{P}_0^*(M, M) = 1$ and start it at state 1. Then of course we have a desired absorption time and thus the antidual would have a desired stationary distribution and FSST.

The above example will be a special case of a more general result. Many absorbing chains have the absorption time T^* distributed as a mixture of sums of independent geometric random variables with parameters being the eigenvalues of the transition matrix. For example, for a stochastically monotone discrete time birth and death chain starting at 1 with $d > 1$ being the absorbing state, the time to absorption is distributed as a sum of geometric random variables with parameters being the eigenvalues of the transition matrix (which are positive in this case). This result follows from Karlin and McGregor [23] or Keilson [24]. Fill [19] gave a first stochastic proof of this result using dualities (the result was simultaneously obtained in [9]). This was extended to skip-free Markov chains in Fill [18]. Miclo [33] 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 the transition matrix with row and column corresponding to the absorbing state removed) there exists a measure $a = (a_1, \ldots, a_M)$ such that the time to absorption T^* has a 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 the transition matrix sorted in a non-increasing 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 a 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*

$$\mathbf{P}(k,s) = \begin{cases} 1 - \dfrac{\pi(2)}{\pi(1)+\pi(2)} p_1 & \text{if} \quad k = s = 1, \\ \dfrac{\pi(2)}{\pi(1)+\pi(2)} p_1 & \text{if} \quad k = 1, s = 2, \\ \dfrac{\pi(s)}{\pi(k)} \left[p_{k-1} \left(1 - \dfrac{H(k-1)}{H(k)} \right) - p_k \left(1 - \dfrac{H(k)}{H(k+1)} \right) \right] & \text{if} \quad 1 < k < M, s < k, \\ 1 - p_k \left(1 - \dfrac{H(k)}{H(k+1)} \right) - p_{k-1} \dfrac{H(k-1)}{H(k)} & \text{if} \quad 1 < k < M, s = k, \\ p_k \dfrac{H(k)}{H(k+1)} \dfrac{\pi(k+1)}{\pi(k)} & \text{if} \quad 1 < k < M, s = k+1, \\ p_{M-1} \pi(s) & \text{if} \quad k = M, s \leq M-1, \\ 1 - p_{M-1} + p_{M-1} \pi(M) & \text{if} \quad k = M, s = M. \end{cases}$$

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

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

has the FSST T distributed as

$$\sum_{i=1}^{M-1} a_i \mathcal{G}(p_i, p_{i+1}, \ldots, p_{M-1}) \tag{4.7}$$

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 4.16 with $p_k = 1, k = 1, \ldots, M-1$ 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 the transition matrix*

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

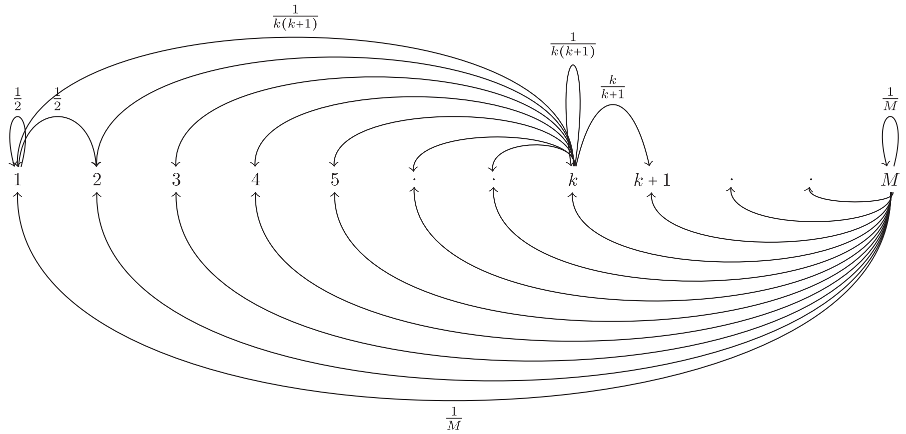

FIGURE 1. The chain X on $\mathbb{E} = \{1, \ldots, M\}$ with the uniform stationary distribution and with the deterministic FSST $T : P(T = M - 1) = 1$ (if X starts at 1).

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

Note that for this chain we have

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

Thus, this is an extreme example of the separation cutoff: for any $k \leq M - 2$ the chain is completely not mixed (the separation 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 the uniform distribution $\pi(k) = \frac{1}{M}$ in Corollary 4.17 we obtain

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

The chain is sketched in Figure 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 $\mathbb{E} = \{1, \ldots, M\}$. 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 $\mathbb{E}^{(1)} = \{1, \ldots, d\}$ 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 an initial distribution $\nu^{(1)}=(1,0,\ldots,0)$ and transitions*

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

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

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

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 the 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 the 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 the chain $\mathbf{P}_1$ given in Corollary 4.8 with $p_k=p$. Thus, its sharp dual chain is given in (4.2). 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 (recall that $|\mathbf{i}| = \sum_{j=1}^{d} i_j$):

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

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 $(1,\ldots,1)$ is absorbing. Since the probability of changing 0 into 1 does not depend on the actual state, the time to increase the current level depends only on the level. Being at any state on a level $|\mathbf{i}| = l$ the time to reach a next level has distribution $Geo((d-l)p)$ (since there are $(d-l)$ zeros, each of which can be changed into 1 with probability p). Thus, if the chain starts somewhere on level 1, say $\nu^*((1,0,\ldots,0)) = 1$, then the 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 $\nu^{(2)}((0,\ldots,0)) = \nu^{(2)}((1,0,\ldots,0)) = 1/2$. All the proofs of Theorems 4.4 and 4.6 are based on the coordinate-wise ordering, *i.e.*,

$$\mathbf{i} \preceq \mathbf{i}' \text{ if } i_j \le i'_j, j = 1,\ldots,d. \tag{4.8}$$

Recall the link Λ (it is given in (3.1))

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

We have

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

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 (defined in (4.8)) for which $\mathbf{i}_{\min} = (0,\ldots,0)$ is the minimum and $\mathbf{i}_{\max} = (N_1,\ldots,N_d)$ is the maximum.

Proof of Theorem 4.4. For the ordering under consideration, directly from Proposition 5 in Rota [37], we find the corresponding Möbius function

$$\mu((i_1,\ldots,i_d),(i_1+r_1,\ldots,i_d+r_d)) = \begin{cases} (-1)^{\sum_{k=1}^d r_k} & r_j\in\{0,1\},\ i_j+r_j\le N_j,\ k=1,\ldots,d \\ 0 & \text{otherwise.}\end{cases} \tag{5.1}$$

Let

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

We will apply Theorem 4.1 with the above ordering and the uniform distribution π on $\mathbb{E}^*$, *i.e.*, $\pi(\mathbf{i}) = \frac{1}{\rho(\mathbf{i}_{\max})}$. Since X^* starts at the minimal state, so does – by Remark 4.2 – the antidual chain. The link $\Lambda(\mathbf{i},\mathbf{i}')$ is the uniform distribution truncated to $\{\mathbf{i}'\preceq\mathbf{i}\}$, from (3.2) we have $\Lambda = (\mathbf{diag}(\pi\mathbf{C}))^{-1}\mathbf{C}^T\mathbf{diag}(\pi)$, thus

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

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

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

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

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

thus

$$
\begin{aligned}
\mathbf{P}(\mathbf{i}^{(1)}, \mathbf{i}^{(2)}) = &\sum_{\mathbf{r} = (r_1, \ldots, r_d) \in \{0,1\}^d:\ \mathbf{i}^{(1)} - \mathbf{r} \in \mathbb{E}^*} (-1)^{|\mathbf{r}|} \rho(\mathbf{i}^{(1)} - \mathbf{r}) \\
&\times \left[\left(1 - \sum_{j: i_j^{(1)} - r_j < N_j} p_j\right) \frac{\mathbf{1}(\mathbf{i}^{(2)} \preceq \mathbf{i}^{(1)} - \mathbf{r})}{\rho(\mathbf{i}^{(1)} - \mathbf{r})} + \sum_{j: i_j^{(1)} - r_j < N_j} \frac{\mathbf{1}(\mathbf{i}^{(2)} \preceq \mathbf{i}^{(1)} - \mathbf{r} + \mathbf{s}_j)}{\rho(\mathbf{i}^{(1)} - \mathbf{r} + \mathbf{s}_j)} p_j\right].
\end{aligned} \tag{5.2}
$$

For convenience, define

$$
\begin{aligned}
H_1(\mathbf{i}^{(1)}, \mathbf{i}^{(2)}, \mathbf{r}) &:= \left(1 - \sum_{j: i_j^{(1)} - r_j < N_j} p_j\right) \frac{\mathbf{1}(\mathbf{i}^{(2)} \preceq \mathbf{i}^{(1)} - \mathbf{r})}{\rho(\mathbf{i}^{(1)} - \mathbf{r})}, \\
H_2(\mathbf{i}^{(1)}, \mathbf{i}^{(2)}, \mathbf{r}) &:= \sum_{j: i_j^{(1)} - r_j < N_j} \frac{\mathbf{1}(\mathbf{i}^{(2)} \preceq \mathbf{i}^{(1)} - \mathbf{r} + \mathbf{s}_j)}{\rho(\mathbf{i}^{(1)} - \mathbf{r} + \mathbf{s}_j)} p_j.
\end{aligned}
$$

Consider cases:

- **Case 1**. *Increasing some coordinates*: $\mathbf{i}^{(2)} = \mathbf{i}^{(1)} + m_1 \mathbf{s}_{k_1} + \ldots + m_t \mathbf{s}_{k_t}$, where $1 \le k_i \le d, i = 1, \ldots, t$ are $t \ge 1$ distinct integers and $m_i \ge 1, i = 1, \ldots, t$. When $t \ge 2$, then indicators in both, H_1 and H_2 are equal to 0. When $t = 1$, then the indicator in H_1 is equal to 0, whereas the indicator in H_2 can be nonzero only in case $m_1 = 1$, $j = k_1$ and $\mathbf{r} = (0, \ldots, 0)$. Then we have

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

- **Case 2**. *Increasing two or more coordinates and decreasing any number of coordinates*: because of the same reasons as in previous case, indicators in both, H_1 and H_2 are equal to 0.

- **Case 3**. *Decreasing some coordinates*: $\mathbf{i}^{(2)} = \mathbf{i}^{(1)} - m_1 \mathbf{s}_{k_1} - \ldots - m_t \mathbf{s}_{k_t}$, where $1 \le k_i \le d, i = 1, \ldots, t$ are $t \ge 1$ distinct integers and $1 \le m_i \le i_{k_i}, i = 1, \ldots, t$.

Let $\boldsymbol{\kappa} = (\kappa_1, \ldots, \kappa_d)$, where $\kappa_{k_i} = 1, i = 1, \ldots, t$ and $\kappa_j = 0$ for $j \notin \{k_1, \ldots, k_t\}$. In (5.2) we sum over all $\mathbf{r} \in \{0,1\}^d$ such that $\mathbf{i}^{(1)} - \mathbf{r} \in \mathbb{E}^*$. Let us split this sum into two sums over disjoint sets I_1 and I_2, where

$$I_1 := \{\mathbf{e} \in \{0,1\}^d : \mathbf{e} \preceq \boldsymbol{\kappa}, \mathbf{i}^{(1)} - \mathbf{e} \in \mathbb{E}^*\}, \qquad I_2 := \{\mathbf{e} \in \{0,1\}^d : \mathbf{e} \not\preceq \boldsymbol{\kappa}, \mathbf{i}^{(1)} - \mathbf{e} \in \mathbb{E}^*\}.$$

Consider $\mathbf{r}' = (r'_1, \ldots, r'_d) \in I_2$. Since it is incomparable with $\boldsymbol{\kappa}$ it means that for some $q \geq 1$ we have $a_1, \ldots, a_q$ such that $\{a_1, \ldots, a_q\} \cap \{k_1, \ldots, k_t\} = \emptyset$ and $r'_{a_i} = 1, i = 1, \ldots, q$. Then the indicator in H_1 is equal to 0. The second indicator can be nonzero only when $q = 1$ and $j = a_1$. In this case for any $\mathbf{r} \in I_2$ we have $\mathbf{r} - s_n \in I_1$, for all $1 \leq n \leq d$ such that $n \neq k_i, i = 1, \ldots, t$. We have

$$\sum_{\mathbf{r} \not\preceq \boldsymbol{\kappa} : \mathbf{i}^{(1)} - \mathbf{r} \in \mathbb{E}^*} (-1)^{|\mathbf{r}|} \rho(\mathbf{i}^{(1)} - \mathbf{r}) \left(H_1(\mathbf{i}^{(1)}, \mathbf{i}^{(2)}, \mathbf{r}) + H_2(\mathbf{i}^{(1)}, \mathbf{i}^{(2)}, \mathbf{r}) \right)$$

$$= \sum_{\substack{n: 1 \leq n \leq d \\ n \neq \kappa_i, i=1,\ldots,t}} \sum_{\mathbf{r} \preceq \boldsymbol{\kappa} : \mathbf{i}^{(1)} - \mathbf{r} \in \mathbb{E}^*} (-1)^{|\mathbf{r} + \mathbf{s}_n|} \rho(\mathbf{i}^{(1)} - \mathbf{r} - \mathbf{s}_n) \left(\sum_{j : i_j^{(1)} - r_j < N_j} \frac{\mathbf{1}(\mathbf{i}^{(1)} - m_1 \mathbf{s}_{k_1} - \ldots - m_t \mathbf{s}_{k_t} \preceq \mathbf{i}^{(1)} - \mathbf{r} - \mathbf{s}_n + \mathbf{s}_j)}{\rho(\mathbf{i}^{(1)} - \mathbf{r} - \mathbf{s}_n + \mathbf{s}_j)} p_j \right).$$

The indicator is nonzero only when $j = n$, and $r_n = 0$ for $n \notin \{k_1, \ldots, k_t\}$, thus

$$\begin{aligned} &= \sum_{\mathbf{r} \preceq \boldsymbol{\kappa} : \mathbf{i}^{(1)} - \mathbf{r} \in \mathbb{E}^*} (-1)^{|\mathbf{r}|+1} \sum_{\substack{n: 1 \leq n \leq d \\ n \neq \kappa_i, i=1,\ldots,t}} \left(\mathbf{1}(i_n^{(1)} - r_n < N_n) \frac{\rho(\mathbf{i}^{(1)} - \mathbf{r} - \mathbf{s}_n)}{\rho(\mathbf{i}^{(1)} - \mathbf{r})} p_j \right) \\ &= - \sum_{\mathbf{r} \preceq \boldsymbol{\kappa} : \mathbf{i}^{(1)} - \mathbf{r} \in \mathbb{E}^*} (-1)^{|\mathbf{r}|} \sum_{\substack{n: 1 \leq n \leq d \\ n \neq \kappa_i, i=1,\ldots,t}} \mathbf{1}(i_n^{(1)} - r_n < N_n) \frac{i_n^{(1)}}{i_n^{(1)} + 1} p_j = 0, \end{aligned}$$

since there is at least one coordinate of $\mathbf{i}^{(1)}$ which is not zero.

Consider $\mathbf{r} \in I_1$. Then indicators in both H_1 and H_2 are nonzero, we have

$$\begin{aligned} S_1 &:= \sum_{\mathbf{r} \preceq \boldsymbol{\kappa}} (-1)^{|\mathbf{r}|} \rho(\mathbf{i}^{(1)} - \mathbf{r}) \left(H_1(\mathbf{i}^{(1)}, \mathbf{i}^{(2)}, \mathbf{r}) + H_2(\mathbf{i}^{(1)}, \mathbf{i}^{(2)}, \mathbf{r}) \right) \\ &= \sum_{\mathbf{r} \preceq \boldsymbol{\kappa}} (-1)^{|\mathbf{r}|} \left[\left(1 - \sum_{j : i_j^{(1)} - r_j < N_j} p_j \right) \frac{\rho(\mathbf{i}^{(1)} - \mathbf{r})}{\rho(\mathbf{i}^{(1)} - \mathbf{r})} + \sum_{j : i_j^{(1)} - r_j < N_j} \frac{\rho(\mathbf{i}^{(1)} - \mathbf{r})}{\rho(\mathbf{i}^{(1)} - \mathbf{r} + \mathbf{s}_j)} p_j \right] \\ &= \sum_{\mathbf{r} \preceq \boldsymbol{\kappa}} (-1)^{|\mathbf{r}|} \left[1 - \sum_{j : i_j^{(1)} - r_j < N_j} p_j + \sum_{j : i_j^{(1)} - r_j < N_j} \left\{ \frac{i_j^{(1)} + 1}{i_j^{(1)} + 2} \mathbf{1}(r_j = 0) + \frac{i_j^{(1)}}{i_j^{(1)} + 1} \mathbf{1}(r_j = 1) \right\} p_j \right]. \end{aligned}$$

Consider cases:

(a) $t = 1$, *i.e.*, we decrease only one coordinate. In this case $\boldsymbol{\kappa} = (0, \dots, 0, 1, 0, \dots, 0)$ with only one 1 at position k. Thus, there are only two $\mathbf{r}$ such that $\mathbf{r} \preceq \boldsymbol{\kappa}$, namely $\mathbf{r} = (0, \dots, 0)$ or $\mathbf{r} = \boldsymbol{\kappa}$. We have

$$\begin{aligned} S_1 \;=\; & \left[1 - \sum_{j: i_j^{(1)} - 0 < N_j} p_j + \sum_{j: i_j^{(1)} - 0 < N_j} \left\{ \frac{i_j^{(1)} + 1}{i_j^{(1)} + 2} \right\} p_j \right] \\ & - \left[1 - \sum_{j: i_j^{(1)} - \mathbf{1}(j=k) < N_j} p_j + \sum_{j: i_j^{(1)} - \mathbf{1}(j=k) < N_j} \left\{ \frac{i_j^{(1)} + 1}{i_j^{(1)} + 2} \mathbf{1}(j \neq k) + \frac{i_j^{(1)}}{i_j^{(1)} + 1} \mathbf{1}(j = k) \right\} p_j \right]. \end{aligned}$$

Note that for $j \neq k$ all the corresponding terms (for $\mathbf{r} = (0, \dots, 0)$ and $\mathbf{r} = \boldsymbol{\kappa}$) are the same, thus they sum up to 0. The remaining terms:

$$\begin{aligned} S_1 \;&=\; \left(p_k - \frac{i_k^{(1)}}{i_k^{(1)} + 1} p_k \right) \mathbf{1}(i_k^{(1)} - 1 < N_k) + \left(-p_k + \frac{i_k^{(1)} + 1}{i_k^{(1)} + 2} p_k \right) \mathbf{1}(i_k^{(1)} < N_k) \\ &=\; \frac{1}{i_k^{(1)} + 1} p_k - \frac{1}{i_k^{(1)} + 2} p_k \mathbf{1}(i_k^{(1)} < N_k). \end{aligned}$$

Finally, we have

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

(b) $t \geq 0$. Things are different in this case. Consider $\mathbf{r} = (r_1, \dots, r_d) \preceq \boldsymbol{\kappa}$ and fixed r_j, where $j \in \{k_1, \dots, k_t\}$. Then there are 2^{t-1} different $\mathbf{r}$ in S_1, from which exactly 2^{t-2} gives $(-1)^{|\mathbf{r}|} = 1$ and exactly 2^{t-2} gives $(-1)^{|\mathbf{r}|} = -1$, resulting in vanishing the terms $\frac{i_j^{(1)}+1}{i_j^{(1)}+2}$ or $\frac{i_j^{(1)}}{i_j^{(1)}+1}$ (depending on the value of r_n). This implies that $S_1 = 0$. For example, for $t = 2$ and, for simplicity, for $d = 2$, we have (consecutive lines correspond to $\mathbf{r} = (0, 0), \mathbf{r} = (0, 1), \mathbf{r} = (1, 0)$ and $\mathbf{r} = (1, 1)$, respectively)

$$\begin{aligned} S_1 = {} & 1 - \mathbf{1}(i_1^{(1)} - 0 < N_1) p_1 - \mathbf{1}(i_2^{(1)} - 0 < N_2) p_2 + \mathbf{1}(i_1^{(1)} - 0 < N_1) \frac{i_1^{(1)} + 1}{i_1^{(1)} + 2} p_1 + \mathbf{1}(i_2^{(1)} - 0 < N_2) \frac{i_2^{(1)} + 1}{i_2^{(1)} + 2} p_2 \\ & - \left[1 - \mathbf{1}(i_1^{(1)} - 0 < N_1) p_1 - \mathbf{1}(i_2^{(1)} - 1 < N_2) p_2 + \mathbf{1}(i_1^{(1)} - 0 < N_1) \frac{i_1^{(1)} + 1}{i_1^{(1)} + 2} p_1 + \mathbf{1}(i_2^{(1)} - 1 < N_2) \frac{i_2^{(1)}}{i_2^{(1)} + 1} p_2 \right] \\ & - \left[1 - \mathbf{1}(i_1^{(1)} - 1 < N_1) p_1 - \mathbf{1}(i_2^{(1)} - 0 < N_2) p_2 + \mathbf{1}(i_1^{(1)} - 1 < N_1) \frac{i_1^{(1)}}{i_1^{(1)} + 1} p_1 + \mathbf{1}(i_2^{(1)} - 0 < N_2) \frac{i_2^{(1)} + 1}{i_2^{(1)} + 2} p_2 \right] \\ & + 1 - \mathbf{1}(i_1^{(1)} - 1 < N_1) p_1 - \mathbf{1}(i_2^{(1)} - 1 < N_2) p_2 + \mathbf{1}(i_1^{(1)} - 1 < N_1) \frac{i_1^{(1)}}{i_1^{(1)} + 1} p_1 + \mathbf{1}(i_2^{(1)} - 1 < N_2) \frac{i_2^{(1)}}{i_2^{(1)} + 1} p_2 \end{aligned}$$

what sums up to 0.

Remark: in case $t = 1$ for fixed r_{k_1} there was no corresponding $j \neq k_1$ which could make the terms vanish.

– **Case 4**. *Increasing one, decreasing another coordinate:* $\mathbf{i}^{(2)} = \mathbf{i}^{(1)} - m_1\mathbf{s}_{k_1} + m_2\mathbf{s}_{k_2}$. We have shown that increasing/decreasing $t \geq 2$ coordinates has probability 0, thus there is no need to consider the case where we increase and decrease any number of coordinates in one step.
In this case, the indicator in H_1 is zero. Concerning H_2. Let, $\boldsymbol{\kappa} = (0, \ldots, 0, 1, 0, \ldots, 0)$ with one 1 at position k_1. Note that for $\mathbf{r} \not\preceq \boldsymbol{\kappa}$, the indicator in H_2 is also 0. Thus, the only nonzero terms are for either $\mathbf{r} = (0, \ldots, 0)$ or $\mathbf{r} = \boldsymbol{\kappa}$ (and then $j = k_2$) and for $m_2 = 1$:

$$\mathbf{r} = (0,\ldots,0): \qquad H_2(\mathbf{i}^{(1)}, \mathbf{i}^{(1)} - m_1\mathbf{s}_{k_1} + \mathbf{s}_{k_2}, \mathbf{r}) = \mathbf{1}(i^{(1)}_{k_2} - 0 < N_{k_2}) \frac{1}{\rho(\mathbf{i}^{(1)} + s_{k_2})} p_{k_2} =: A$$

$$\mathbf{r} = \boldsymbol{\kappa}: \qquad H_2(\mathbf{i}^{(1)}, \mathbf{i}^{(1)} - m_1\mathbf{s}_{k_1} + \mathbf{s}_{k_2}, \mathbf{r}) = \mathbf{1}(i^{(1)}_{k_2} - 0 < N_{k_2}) \frac{1}{\rho(\mathbf{i}^{(1)} - s_{k_1} + s_{k_2})} p_{k_2} =: B$$

and we have

$$\begin{aligned}
\mathbf{P}(\mathbf{i}^{(1)}, \mathbf{i}^{(1)} - m_1\mathbf{s}_{k_1} + \mathbf{s}_{k_2}) &= (-1)^0 \rho(\mathbf{i}^{(1)}) \cdot A + (-1)^1 \rho(\mathbf{i}^{(1)} - s_{k_1}) \cdot B \\
&= \mathbf{1}(i^{(1)}_{k_2} - 0 < N_{k_2}) \left[\frac{\rho(\mathbf{i}^{(1)})}{\rho(\mathbf{i}^{(1)} + s_{k_2})} - \frac{\rho(\mathbf{i}^{(1)} - s_{k_1})}{\rho(\mathbf{i}^{(1)} - s_{k_1} + s_{k_2})} \right] p_{k_2} \\
&= \mathbf{1}(i^{(1)}_{k_2} - 0 < N_{k_2}) \left[\frac{i^{(1)}_{k_2} + 1}{i^{(1)}_{k_2} + 2} - \frac{i^{(1)}_{k_2} + 1}{i^{(1)}_{k_2} + 2} \right] p_{k_2} = 0.
\end{aligned}$$

– **Case 5**. *Staying at the same state:* $\mathbf{i}^{(2)} = \mathbf{i}^{(1)}$. 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

$$\begin{aligned}
\mathbf{P}(\mathbf{i}^{(1)}, \mathbf{i}^{(1)}) &= (-1)^0 \rho(\mathbf{i}^{(1)}) \left[\left(1 - \sum_{j: i^{(1)}_j - 0 < N_j} p_j \right) \frac{1}{\rho(\mathbf{i}^{(1)})} + \sum_{j: i^{(1)}_j - 0 < N_j} \frac{p_j}{\rho(\mathbf{i}^{(1)} + \mathbf{s}_j)} \right] \\
&\quad - \sum_{k: i^{(1)}_k - 1 \geq 0} \rho(\mathbf{i}^{(1)} - \mathbf{s}_k) \left[\sum_{j: i^{(1)}_j - \mathbf{1}(j=k) < N_j} \frac{p_j}{\rho(\mathbf{i}^{(1)} - \mathbf{s}_k + \mathbf{s}_j)} \mathbf{1}(\mathbf{i}^{(1)} \preceq \mathbf{i}^{(1)} - \mathbf{s}_k + \mathbf{s}_j) \right] \\
&= 1 - \sum_{j: i^{(1)}_j < N_j} p_j + \sum_{j: i^{(1)}_j < N_j} \frac{\rho(\mathbf{i}^{(1)})}{\rho(\mathbf{i}^{(1)}) + \mathbf{s}_j)} p_j - \sum_{k: i^{(1)}_k \geq 1} \frac{\rho(\mathbf{i}^{(1)} - \mathbf{s}_k)}{\rho(\mathbf{i}^{(1)})} p_k \\
&= 1 - \sum_{j: i^{(1)}_j < N_j} \left(1 - \frac{i^{(1)}_j + 1}{i^{(1)}_j + 2} \right) p_j - \sum_{k=1}^{d} \frac{i^{(1)}_k}{i^{(1)}_k + 1} p_k \\
&= 1 - \sum_{j: i^{(1)}_j < N_j} \left(\frac{1}{i^{(1)}_j + 2} \right) p_j - \sum_{j: i^{(1)}_j < N_j} \frac{i^{(1)}_j}{i^{(1)}_j + 1} p_j - \sum_{j: i^{(1)}_j = N_j} \frac{i^{(1)}_j}{i^{(1)}_j + 1} p_j \\
&= 1 - \sum_{j: i^{(1)}_j < N_j} \left(1 - \frac{1}{(i^{(1)}_j + 1)(i^{(1)}_j + 2)} \right) p_j - \sum_{j: i^{(1)}_j = N_j} \frac{N_j}{N_j + 1} p_j.
\end{aligned}$$

The assumption (4.3) implies that $\mathbf{P}(\mathbf{i}^{(1)},\mathbf{i}^{(1)}) \geq 0$. We have considered all the transitions. Checking that $\sum_{\mathbf{i}^{(2)}\in\mathbb{E}^*}\mathbf{P}(\mathbf{i}^{(1)},\mathbf{i}^{(2)})$ is left to the reader. □

Proof of Theorem 4.6. Note that $(0,\ldots,0)$ is the 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 the antidual chain, *i.e.*, $\nu = \nu^*$.

For convenience, define

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

For the stationary distribution π given in (4.6) we have

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

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

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 the 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

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

Let us proceed with $\widehat{\mathbf{P}^*}$.

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

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

$$\sum_{\mathbf{i}}\widehat{\mathbf{P}^*}(\mathbf{i}^{(2)},\mathbf{i}) = \sum_{j:i^{(2)}_j=0}f(\mathbf{i}^{(2)},j)p_j + 1 - \sum_{j:i^{(2)}_j=0}p_j = 1-\sum_{j:i^{(2)}_j=0}\left(1-f(\mathbf{i}^{(2)},j)\right)p_j = 1-\sum_{j:i^{(2)}_j=0}a_jp_j < 1.$$

Now, calculating the antidual chain from Theorem 4.1, we have

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

$$= \frac{\pi(\mathbf{i}^{(2)})}{\pi(\mathbf{i}^{(1)})}\sum_{\mathbf{i}\preceq\mathbf{i}^{(1)}}\widehat{\mathbf{P}^*}(\mathbf{i},\{\mathbf{i}^{(2)}\}^{\uparrow})(-1)^{|\mathbf{i}^{(1)}-\mathbf{i}|}, \tag{5.3}$$

where we applied the Möbius function for this ordering: $\mathbf{C}^{-1}(\mathbf{i},\mathbf{i}^{(1)}) = (-1)^{|\mathbf{i}^{(1)}-\mathbf{i}|}\mathbf{1}(\mathbf{i}\preceq\mathbf{i}^{(1)})$ (a consequence of (5.1)). We proceed with (5.3) by considering cases:

– **Case 1.** *Increasing some coordinates*: $\mathbf{i}^{(2)} = \mathbf{i}^{(1)} + \mathbf{s}_{k_1} + \ldots \mathbf{s}_{k_t}$ for some distinct $t \geq 1$ integers $1 \leq k_i \leq d, i = 1,\ldots,d$.

  First note that if $t \geq 2$, than, for any $\mathbf{i} \preceq \mathbf{i}^{(1)}$ we have $\widehat{\mathbf{P}^*}(\mathbf{i},\{\mathbf{i}^{(1)} + \mathbf{s}_{k_1} + \ldots \mathbf{s}_{k_t}\}^{\uparrow}) = 0$, thus $\mathbf{P}(\mathbf{i}^{(1)},\mathbf{i}^{(1)} + \mathbf{s}_{k_1} + \ldots \mathbf{s}_{k_M}) = 0$.

  For $t = 1$ the sum in (5.3) 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 nonzero term is for $\mathbf{i} = \mathbf{i}^{(1)}$, thus

$$\begin{aligned}\mathbf{P}(\mathbf{i}^{(1)},\mathbf{i}^{(1)}+\mathbf{s}_k) &= \frac{\pi(\mathbf{i}^{(1)}+\mathbf{s}_k)}{\pi(\mathbf{i}^{(1)})}\widehat{\mathbf{P}^*}(\mathbf{i}^{(1)},\{\mathbf{i}^{(1)}+\mathbf{s}_k\}^{\uparrow}) = \frac{\pi(\mathbf{i}^{(1)}+\mathbf{s}_k)}{\pi(\mathbf{i}^{(1)})}f(\mathbf{i}^{(1)},k)p_k \\ &= \frac{a_k}{1-a_k}(1-a_k)p_k = a_kp_k.\end{aligned}$$

– **Case 2.** *Increasing two or more coordinates and decreasing any number of coordinates*: because of the same reasons as in previous case (we would have to increase at least two coordinates in one step) such transition has probability 0.

– **Case 3.** $\mathbf{i}^{(2)} = \mathbf{i}^{(1)} - \mathbf{s}_{k_1} - \ldots - \mathbf{s}_{k_t}, t \geq 1$. Let us split $\{\mathbf{e} \preceq \mathbf{i}^{(1)}\}$ into five disjoint sets:

$$I_1 = \{\mathbf{i}^{(1)}\},\quad I_2 = \{\mathbf{i}^{(2)}\},\quad I_3 = \{\mathbf{e} : \mathbf{e} \prec \mathbf{i}^{(2)}\},\quad I_4 = \{\mathbf{e} : \mathbf{i}^{(2)} \prec \mathbf{e} \prec \mathbf{i}^{(1)}\},\quad I_5 = \{\mathbf{e} : \mathbf{e} \not\preceq \mathbf{i}^{(2)}\},$$

  where $\mathbf{e} \prec \mathbf{e}'$ means that $\mathbf{e} \preceq \mathbf{e}'$ and $\mathbf{e} \neq \mathbf{e}'$, and $\mathbf{e} \not\preceq \mathbf{e}'$ means that $\mathbf{e}$ and $\mathbf{e}'$ are incomparable. Define also

$$S_m := \sum_{\mathbf{i}\in I_m}\widehat{\mathbf{P}^*}(\mathbf{i},\{\mathbf{i}^{(2)}\}^{\uparrow})(-1)^{|\mathbf{i}^{(1)}-\mathbf{i}|},\quad m = 1,2,3,4,5.$$

We have

$$\begin{aligned}
S_1 &= \widehat{\mathbf{P}^*}(\mathbf{i}^{(1)}, \{\mathbf{i}^{(2)}\}^{\uparrow}) = 1 - \sum_{j:i_j^{(1)}=0} a_j p_j. \\
S_2 &= \widehat{\mathbf{P}^*}(\mathbf{i}^{(2)}, \{\mathbf{i}^{(2)}\}^{\uparrow})(-1)^{|\mathbf{i}^{(1)}-\mathbf{i}^{(2)}|} = \left(1 - \sum_{j:i_j^{(1)}=0} a_j p_j - \sum_{j\in\{k_1,\ldots,k_t\}} a_j p_j\right)(-1)^{|\mathbf{i}^{(1)}-\mathbf{i}^{(2)}|}. \\
S_3 &= \sum_{\substack{j:i_j^{(1)}=1 \\ j\notin\{k_1,\ldots,k_t\}}} \widehat{\mathbf{P}^*}(\mathbf{i}^{(2)} - \mathbf{s}_j, \{\mathbf{i}^{(2)}\}^{\uparrow})(-1)^{|\mathbf{i}^{(1)}-\mathbf{i}|} = (-1)^{|\mathbf{i}^{(1)}-\mathbf{i}^{(2)}-1|} \sum_{\substack{j:i_j^{(1)}=1 \\ j\notin\{k_1,\ldots,k_t\}}} (1-a_j)p_j. \\
S_4 &= \sum_{\mathbf{i}\in I_4}\left(1 - \sum_{j:i_j=0} a_j p_j\right)(-1)^{|\mathbf{i}^{(1)}-\mathbf{i}|} = \sum_{\mathbf{i}\in I_4}\left(1 - \sum_{j:i_j^{(1)}=0} a_j p_j - \sum_{\substack{j\in\{k_1,\ldots,k_j\} \\ i_j=0}} a_j p_j\right)(-1)^{|\mathbf{i}^{(1)}-\mathbf{i}|}. \\
S_5 &= \sum_{\emptyset\neq\{b_1,\ldots,b_z\}\subseteq\{k_1,\ldots,k_t\}} \sum_{\substack{j:i_j^{(1)}=1 \\ j\notin\{k_1,\ldots,k_t\}}} \widehat{\mathbf{P}^*}(\mathbf{i}^{(1)} - \mathbf{s}_{b_1} - \ldots, \mathbf{s}_{b_z} - \mathbf{s}_j, \{\mathbf{i}^{(2)}\}^{\uparrow})(-1)^{t-z+1} \\
&= \sum_{\emptyset\neq\{b_1,\ldots,b_z\}\subseteq\{k_1,\ldots,k_t\}} \sum_{\substack{j:i_j^{(1)}=1 \\ j\notin\{k_1,\ldots,k_t\}}} (1-a_j)p_j(-1)^{t-z+1} = (-1)^t \sum_{\substack{j:i_j^{(1)}=1 \\ j\notin\{k_1,\ldots,k_t\}}} (1-a_j)p_j.
\end{aligned}$$

Let us consider cases $t = 1$ and $t \geq 2$ separately.

(a) $t = 1$, *i.e.*, $\mathbf{i}^{(2)} = \mathbf{i}^{(1)} - \mathbf{s}_k$ ($k_1 \equiv k$). Note that then $I_4 = \emptyset$. We have

$$\begin{aligned}
S_1 &= 1 - \sum_{j:i_j^{(1)}=0} a_j p_j. \\
S_2 &= -\left(1 - \sum_{j:i_j^{(1)}=0} a_j p_j - a_k p_k\right). \\
S_3 &= \sum_{\substack{j:i_j^{(1)}=1 \\ j\neq k}} (1-a_j)p_j. \\
S_5 &= -\sum_{j:i_j^{(1)}=1, j\neq k} \widehat{\mathbf{P}^*}(\mathbf{i}^{(1)} - \mathbf{s}_j, \{\mathbf{i}^{(2)}\}^{\uparrow}) = -\sum_{j:i_j^{(1)}=1, j\neq k} (1-a_j)p_j.
\end{aligned}$$

We have $S_1 + S_2 + S_3 + S_4 + S_5 = a_k p_k$ and finally

$$\mathbf{P}(\mathbf{i}^{(1)}, \mathbf{i}^{(1)} - \mathbf{s}_k) = \frac{\pi(\mathbf{i}^{(1)} - \mathbf{s}_k)}{\pi(\mathbf{i}^{(1)})}(S_1 + S_2 + S_3 + S_4) = \frac{1 - a_k}{a_k} a_k p_k = (1 - a_k)p_k.$$

(b) $t \geq 2$. Consider first $t = 2$. Assume thus that $\mathbf{i}^{(2)} = \mathbf{i}^{(1)} - \mathbf{s}_{k_1} - \mathbf{s}_{k_2}$. We have

$$
\begin{aligned}
S_1 &= 1 - \sum_{j: i_j^{(1)}=0} a_j p_j. \\
S_2 &= 1 - \sum_{j: i_j^{(1)}=0} a_j p_j - a_{k_1} p_{k_1} - a_{k_2} p_{k_2}. \\
S_3 &= - \sum_{\substack{j: i_j^{(1)}=1 \\ j \notin \{k_1, k_2\}}} (1 - a_j) p_j. \\
S_4 &= -\widehat{\mathbf{P}^*}(\mathbf{i}^{(1)} - \mathbf{s}_{k_1}, \{\mathbf{i}^{(1)} - \mathbf{s}_{k_1} - \mathbf{s}_{k_2}\}^{\uparrow}) - \widehat{\mathbf{P}^*}(\mathbf{i}^{(1)} - \mathbf{s}_{k_2}, \{\mathbf{i}^{(1)} - \mathbf{s}_{k_1} - \mathbf{s}_{k_2}\}^{\uparrow}) \\
&= -(1 - \sum_{j: i_j^{(1)}=0} a_j p_j - a_{k_2} p_{k_2}) - (1 - \sum_{j: i_j^{(1)}=0} a_j p_j - a_{k_1} p_{k_1}). \\
S_5 &= \sum_{\substack{j: i_j^{(1)}=1 \\ j \notin \{k_1, \ldots, k_t\}}} (1 - a_j) p_j.
\end{aligned}
$$

Summing up, $S_1 + S_2 + S_3 + S_4 + S_5 = 0$, what is also the case for $t > 2$ (the proof, although longer, is quite similar, we skip the details). This means that for $t \geq 2$

$$
\mathbf{P}(\mathbf{i}^{(1)}, \mathbf{i}^{(1)} - \mathbf{s}_{k_1} - \ldots - \mathbf{s}_{k_t}) = 0.
$$

- **Case 4.** *Increasing one, decreasing another coordinate:* $\mathbf{i}^{(2)} = \mathbf{i}^{(1)} + \mathbf{s}_{k_1} - \mathbf{s}_{k_2}$. We have shown that increasing/decreasing $t \geq 2$ coordinate has probability 0, thus it suffices to consider only changing two coordinates (one increasing, the other decreasing). Then the summands $\sum_{\mathbf{i} \preceq \mathbf{i}^{(1)}} \widehat{\mathbf{P}^*}(\mathbf{i}, \{\mathbf{i}^{(1)} + \mathbf{s}_{k_1} - \mathbf{s}_{k_2}\}^{\uparrow})(-1)^{|\mathbf{i}^{(1)} - \mathbf{i}|}$ are nonzero only for $\mathbf{i} = \mathbf{i}^{(1)}$ or $\mathbf{i} = \mathbf{i}^{(1)} - \mathbf{s}_{k_2}$, we have

$$
\begin{aligned}
\widehat{\mathbf{P}^*}(\mathbf{i}^{(1)}, \{\mathbf{i}^{(1)} + \mathbf{s}_{k_1} - \mathbf{s}_{k_2}\}^{\uparrow})(-1)^{|\mathbf{i}^{(1)} - \mathbf{i}^{(1)}|} &= f(\mathbf{i}^{(1)}, k_1) p_{k_1} = (1 - a_{k_1}) p_{k_1}, \\
\widehat{\mathbf{P}^*}(\mathbf{i}^{(1)} - \mathbf{s}_{k_2}, \{\mathbf{i}^{(1)} + \mathbf{s}_{k_1} - \mathbf{s}_{k_2}\}^{\uparrow})(-1)^{|\mathbf{i}^{(1)} - \mathbf{i}^{(1)} - 1|} &= -f(\mathbf{i}^{(1)} - \mathbf{s}_{k_2}, k_1) p_{k_1} = -(1 - a_{k_1}) p_{k_1},
\end{aligned}
$$

  thus $\mathbf{P}(\mathbf{i}^{(1)}, \mathbf{i}^{(1)} + \mathbf{s}_{k_1} - \mathbf{s}_{k_2}) = 0$.

- **Case 5.** *Staying at the same state:* $\mathbf{i}^{(2)} = \mathbf{i}^{(1)}$. Then we have

$$
\begin{aligned}
\mathbf{P}(\mathbf{i}^{(1)}, \mathbf{i}^{(1)}) &= \sum_{\mathbf{i} \preceq \mathbf{i}^{(1)}} \widehat{\mathbf{P}^*}(\mathbf{i}, \{\mathbf{i}^{(1)}\}^{\uparrow})(-1)^{|\mathbf{i}^{(1)} - \mathbf{i}|} \\
&= \widehat{\mathbf{P}^*}(\mathbf{i}^{(1)}, \{\mathbf{i}^{(1)}\}^{\uparrow}) - \sum_{j: \mathbf{i}_j^{(1)}=1} \widehat{\mathbf{P}^*}(\mathbf{i}^{(1)} - \mathbf{s}_j, \{\mathbf{i}^{(1)}\}^{\uparrow}).
\end{aligned}
$$

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 jth coordinate of $\mathbf{i}^{(1)} - \mathbf{s}_j$ to one:

$$\begin{aligned}\mathbf{P}(\mathbf{i}^{(1)}, \mathbf{i}^{(1)}) &= 1 - \sum_{j:i_j^{(1)}=0} p_j(1 - f(\mathbf{i}^{(1)}, j)) - \sum_{j:i_j^{(1)}=1} f(\mathbf{i}^{(1)} - \mathbf{s}_j, j)p_j \\ &= 1 - \sum_{j:i_j^{(1)}=0} a_j p_j - \sum_{j:i_j^{(1)}=1} (1 - a_j)p_j.\end{aligned}$$

Finally, we obtain matrix $\mathbf{P}$ given in (4.5). □

Remark 5.1. Showing that $\mathbf{P}(\mathbf{i}^{(1)}, \mathbf{i}^{(1)} + \mathbf{s}_{k_1} - \mathbf{s}_{k_2}) = 0$ relied heavily on the fact that for the stationary distribution given in (4.6), we had $f(\mathbf{i}, j) = 1 - a_j$ and it did not depend on $\mathbf{i}$. That is why the terms $f(\mathbf{i}^{(1)}, k_1)p_{k_1}$ and $f(\mathbf{i}^{(1)}, k_1)p_{k_1}$ cancelled out. Similarly, it is the reason why decreasing $t \geq 2$ coordinates has probability 0. For other, not product-form stationary distributions, such transitions are possible.

5.2. Proof of Theorem 4.16

Let $\mathbf{X}^*$ be an absorbing chain on $\mathbb{E} = \{1, \ldots, M\}, M \geq 2$ with the 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 (4.7). We will show that $\mathbf{P}$ is its sharp antidual chain.
We consider the total ordering $\preceq := \leq$. Then the link given in (3.1) reads

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

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

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

Let us calculate

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

Calculating transitions of the antidual chain:

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

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 $s=2$.

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

– $k=M$. We have

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

Thus,

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

– $1<k<M$. 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,

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

Consider three sub-cases:

- $s = k+1$. 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

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

- $s < k$. Then we have

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

For $k\in\{1,M\}$ we obviously have $\sum_{s=1}^M \mathbf{P}(k,s) = 1$. For $1<k<M$ we have

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

$$= \frac{H(k-1)}{\pi(k)}\left[p_{k-1}\left(1-\frac{H(k-1)}{H(k)}\right)-p_k\left(1-\frac{H(k)}{H(k+1)}\right)\right]$$

$$+1+p_k\left[\frac{H(k)}{H(k+1)}\frac{\pi(k+1)}{\pi(k)}-\frac{H(k+1)-H(k)}{H(k+1)}\right]-p_{k-1}\frac{H(k-1)}{H(k)}$$

$$= 1+p_k\left[\frac{H(k)}{H(k+1)}\frac{\pi(k+1)}{\pi(k)}-\frac{\pi(k+1)}{H(k+1)}-\frac{H(k-1)}{\pi(k)}\frac{\pi(k+1)}{H(k+1)}\right]$$

$$+p_{k-1}\left[\frac{H(k-1)}{\pi(k)}\left(1-\frac{H(k-1)}{H(k)}\right)-\frac{H(k-1)}{H(k)}\right]$$

$$= 1+p_k\frac{\pi(k+1)}{H(k+1)}\left[\frac{H(k)}{\pi(k)}-1-\frac{H(k-1)}{\pi(k)}\right]+p_{k-1}\left[\frac{H(k-1)}{\pi(k)}-\frac{H(k-1)}{H(k)}\left(\frac{H(k-1)}{\pi(k)}+1\right)\right]$$

$$= 1+p_k\frac{\pi(k+1)}{H(k+1)}\left[\frac{H(k)-\pi(k)-H(k-1)}{\pi(k)}\right]+p_{k-1}\left[\frac{H(k-1)}{\pi(k)}-\frac{H(k-1)}{H(k)}\frac{H(k)}{\pi(k)}\right]=1.$$

Thus (*cf.* (4.1)) we considered all the cases. The only thing left to calculate is the initial distribution of the antidual chain. Using relation (2.1) 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 $\{1-p_1,\dots,1-p_{M-1},1\}$ are its eigenvalues. Because of the relation (2.1) these are also the eigenvalues of $\mathbf{P}$.

Acknowledgements. The author thanks anonymous reviewers for thorough reviews and appreciates the comments and suggestions, which contributed to improving the quality of the publication.

References

[1] D. Aldous and P. Diaconis, Shuffling cards and stopping times. *Am. Math. Mon.* **93** (1986) 333–348.
[2] D. Aldous and P. Diaconis, Strong uniform times and finite random walks. *Adv. Appl. Math.* **97** (1987) 69–97.
[3] R. Basu, J. Hermon and Y. Peres, Characterization of cutoff for reversible Markov chains. *Ann. Probab.* **45** (2017) 1448–1487.
[4] G.-Y. Chen and L. Saloff-Coste, The cutoff phenomenon for ergodic Markov processes. *Electron. J. Probab.* **13** (2008) 26–78.
[5] G.-Y. Chen and L. Saloff-Coste, Computing cutoff times of birth and death chains. *Electron. J. Probab.* **20** (2015) 1–47.
[6] M.C.H. Choi and P. Patie, A sufficient condition for continuous-time finite skip-free Markov chains to have real eigenvalues, in Mathematical and Computational Approaches in Advancing Modern Science and Engineering, edited by J. Bélair, I. Frigaard, H. Kunze, R. Makarov, R. Melnik and R. Spiteri. Springer, Switzerland (2016) 529–536.
[7] S.B. Connor, Separation and coupling cutoffs for tuples of independent Markov processes. *Lat. Am. J. Probab. Math. Stat.* **7** (2010) 65–77.
[8] P. Diaconis and J.A. Fill, Strong stationary times via a new form of duality. *Ann. Probab.* **18** (1990) 1483–1522.
[9] P. Diaconis and L. Miclo, On times to quasi-stationarity for birth and death processes. *J. Theor. Probab.* **22** (2009) 558–586.
[10] P. Diaconis and L. Saloff-Coste, What do we know about the Metropolis algorithm? *J. Comput. Syst. Sci.* **57** (1998) 20–36.
[11] P. Diaconis and L. Saloff-Coste, Separation cut-offs for birth and death chains. *Ann. Appl. Probab.* **16** (2006) 2098–2122.
[12] P. Diaconis and M. Shahshahani, Generating a random permutation with random transpositions. *Z. Wahrscheinlichkeitstheor. verw. Geb.* **57** (1981) 159–179.
[13] J. Ding, E. Lubetzky and Y. Peres, Total variation cutoff in birth-and-death chains. *Probab. Theory Relat. Fields* **146** (2010) 61–85.

[14] A.V. Doumas and V.G. Papanicolaou, The coupon collector's problem revisited: generalizing the double dixie cup problem of Newman and Shepp. *ESAIM: PS* **20** (2016) 367–399.
[15] P.L. Erdłs and A. Rényi, On a classical problem of probability theory. *Publ. Math. Inst. Hung. Acad. Sci. Ser. A* **6** (1961) 215–220.
[16] W. Feller, An Introduction to Probability Theory and Its Applications, 2nd edn., Vol. 2. John Wiley & Sons, NJ (1971).
[17] J.A. Fill, An exact formula for the move-to-front rule for self-organizing lists. *J. Theor. Probab.* **9** (1996) 113–160.
[18] J.A. Fill, On hitting times and fastest strong stationary times for skip-free and more general chains. *J. Theor. Probab.* **22** (2009) 587–600.
[19] J.A. Fill, The passage time distribution for a birth-and-death chain: strong stationary duality gives a first stochastic proof. *J. Theor. Probab.* **22** (2009) 543–557.
[20] J.A. Fill and V. Lyzinski, Strong stationary duality for diffusion processes. *J. Theor. Probab.* **29** (2016) 1298–1338.
[21] J. Hermon, H. Lacoin and Y. Peres, Total variation and separation cutoffs are not equivalent and neither one implies the other. *Electron. J. Probab.* **21** (2016) 1–36.
[22] L. Holst, Extreme value distributions for random coupon collector and birthday problems. *Extremes* **4** (2001) 129–145.
[23] S. Karlin and J. McGregor, Coincidence properties of birth and death processes. *Pac. J. Math.* **9** (1959) 1109–1140.
[24] J. Keilson, Log-concavity and log-convexity in passage time densities of diffusion and birth-death processes. *J. Appl. Probab.* **8** (1971) 391–398.
[25] H. Lacoin, The cutoff profile for the simple-exclusion process on the circle. *Ann. Probab.* **44** (2016) 3399–3430.
[26] D. Levin, Y. Peres and E. Wilmer, Markov Chains and Mixing Times, 2nd edn. American Mathematical Society, Rhode Island (2017).
[27] P. Lorek, Generalized Gambler's ruin problem: explicit formulas via Siegmund duality. *Methodol. Comput. Appl. Probab.* **19** (2017) 603–613.
[28] P. Lorek, Siegmund duality for Markov chains on partially ordered state spaces. *Probab. Eng. Inf. Sci.* **32** (2018) 495–521.
[29] P. Lorek and P. Markowski, Monotonicity requirements for efficient exact sampling with Markov chains. *Markov Process. Relat. Fields* **23** (2017) 485–514.
[30] P. Lorek and R. Szekli, Strong stationary duality for Möbius monotone Markov chains. *Queueing Syst.* **71** (2012) 79–95.
[31] E. Lubetzky and A. Sly, Cutoff for the Ising model on the lattice. *Invent. Math.* **191** (2013) 719–755.
[32] Y.-H. Mao, C. Zhang and Y.-h. Zhang, Separation cutoff for upward skip-free chains. *J. Appl. Probab.* **1** (2016) 299–306.
[33] L. Miclo, On absorption times and Dirichlet eigenvalues. *ESAIM: PS* **14** (2010) 117–150.
[34] P. Neal, The generalised coupon collector problem. *J. Appl. Probab.* **45** (2008) 621–629.
[35] D.J. Newman, The double dixie cup problem. *Am. Math. Mon.* **67** (1960) 58–61.
[36] I. Pak and V.H. Vu, On mixing of certain random walks, cutoff phenomenon and sharp threshold of random matroid processes. *Discrete Appl. Math.* **110** (2001) 251–272.
[37] G.-C. Rota, On the foundations of combinatorial theory I. Theory of Möbius functions. *Probab. Theory and Relat. Fields* **368** (1964) 340–368.
[38] D. Siegmund, The equivalence of absorbing and reflecting barrier problems for stochastically monotone Markov processes. *Ann. Probab.* **4** (1976) 914–924.