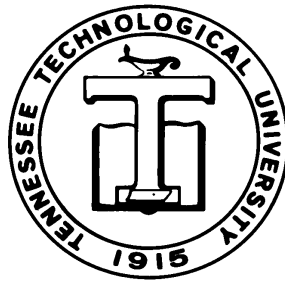

DEPARTMENT OF MATHEMATICS
TECHNICAL REPORT

SET-VALUED DUALITY FOR
FINITE-STATE MARKOV CHAINS

DR. JAMES ALLEN FILL
AND
DR. MOTOYA MACHIDA

OCTOBER 2003

No. 2003-5



TENNESSEE TECHNOLOGICAL UNIVERSITY
Cookeville, TN 38505

SET-VALUED DUALITY FOR FINITE-STATE MARKOV CHAINS

JAMES ALLEN FILL AND MOTOYA MACHIDA

Johns Hopkins University and Tennessee Technological University

Abstract

We discuss set-valued dual chains in a finite state space setting. Several different notions of duality for Markov chains are introduced, and their properties and constructive techniques are investigated. Our goal is to understand perfect rejection sampling algorithms proposed by Fill and successively by Fill, Machida, Murdoch, and Rosenthal in the light of duality, which will culminate in demonstrating their intimate connection with strong stationary duality and presenting a new interpretation in terms of set-valued dual chains.

1 Introduction

1.1 Duality in Markov chains

Duality between two homogeneous Markov processes relates the process of interest with its “dual” one, and has been used in the work of Karlin and McGregor [11], Siegmund [15], Liggett [13], Harris [8], and many others. Among several different formulations which have been proposed, Liggett’s notion of duality has naturally extended others (see, e.g., Cox and Rösler [2] for examples of such duality). Here we will present Liggett’s duality in discrete state space setting (Section 2.3 of [13] for general Markov processes). We denote by $\mathbf{P} = [P(x, y)]$ a Markov transition matrix, and write \mathbf{P}^T for the transpose $[P^T(x, y)] := [P(y, x)]$, and \mathbf{P}^n for the n -step transition matrix $[P^n(x, y)]$ (i.e., the n -th power of the matrix \mathbf{P}).

Definition 1.1. Let \mathbf{P} and $\check{\mathbf{P}}$ be Markov transition matrices on the respective discrete (countably infinite or finite) state spaces \mathcal{S} and $\check{\mathcal{S}}$, and let $\mathbf{\Gamma} = [\Gamma(x^*, x)]$ be a nonnegative bounded matrix on $\mathcal{S} \times \check{\mathcal{S}}$. Then $\check{\mathbf{P}}$ is said to be a *Liggett $\mathbf{\Gamma}$ -dual* of \mathbf{P} if the *duality relation* $\mathbf{\Gamma} \mathbf{P}^T = \check{\mathbf{P}} \mathbf{\Gamma}$ holds, that is, if

$$(1.1) \quad \sum_{z \in \mathcal{S}} \Gamma(x^*, z) P(x, z) = \sum_{z^* \in \check{\mathcal{S}}} \check{P}(x^*, z^*) \Gamma(z^*, x)$$

holds for every $(x^*, x) \in \check{\mathcal{S}} \times \mathcal{S}$.

AMS 2000 subject classifications. Primary 60J10; secondary 65C40, 60G17

Key words and phrases. Duality, Siegmund dual, strong stationary dual, set-valued Markov chain, coupling, perfect sampling.

In Definition 1.1 we call $\check{\mathcal{S}}$ a *dual state space*, and Γ a *duality function*. The duality relation (1.1) clearly implies that $\Gamma(\mathbf{P}^n)^T = \Gamma(\mathbf{P}^T)^n = \check{\mathbf{P}}^n \Gamma$ for every integer $n \geq 1$. Suppose that \mathcal{S} is a finite state space, and that $\check{\mathcal{S}}$ is a subcollection of subsets of \mathcal{S} . Then a Liggett Γ -dual with specific duality function

$$(1.2) \quad \Gamma(x^*, x) := \begin{cases} 1 & \text{if } x \in x^*; \\ 0 & \text{otherwise,} \end{cases}$$

is called a *set-valued dual*.

We denote by $\mathbf{X} = (X_n)_{n=0,1,\dots}$ a Markov chain on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and write $P_x(\cdot)$ for the conditional probability $\mathbb{P}(\cdot | X_0 = x)$ with the underlying Markov chain \mathbf{X} . From now on a Markov chain \mathbf{X} and its Markov transition matrix \mathbf{P} are used interchangeably, and simply called a chain. When $\check{\mathbf{X}}$ is a set-valued dual for \mathbf{X} , the duality relation (1.1) is equivalently characterized by

$$(1.3) \quad P_x(X_n \in x^*) = P_{x^*}(x \in \check{X}_n), \quad (x^*, x) \in \check{\mathcal{S}} \times \mathcal{S}.$$

The dual state space $\check{\mathcal{S}} = \{\mathcal{S}, \emptyset\}$ has a trivial set-valued dual with both \mathcal{S} and \emptyset being necessarily absorbing states. In what follows it is assumed that a dual state space $\check{\mathcal{S}}$ is strictly larger than the trivial $\{\mathcal{S}, \emptyset\}$. In Section 2 we begin with our investigation on set-valued duality for an arbitrary subcollection $\check{\mathcal{S}}$ of subsets.

Another notion of duality was introduced by Diaconis and Fill [4] in the course of their investigation on strong stationary duality. Let \mathcal{S} and \mathcal{S}^* be discrete state spaces. Here we assume that $(\mathbf{X}^*, \mathbf{X}) = (X_n^*, X_n)_{n=0,1,\dots}$ is a bivariate Markov chain, and that the marginal processes $\mathbf{X}^* = (X_n^*)_{n=0,1,\dots}$ and $\mathbf{X} = (X_n)_{n=0,1,\dots}$ take values on the respective state spaces \mathcal{S}^* and \mathcal{S} , where \mathcal{S}^* is considered as a dual state space. A matrix $\Lambda = [\Lambda(x^*, x)]$ is called a *link* from \mathcal{S}^* to \mathcal{S} if it is a probability transition matrix from \mathcal{S}^* to \mathcal{S} , that is, if it is nonnegative and satisfies $\sum_{z \in \mathcal{S}} \Lambda(x^*, z) = 1$ for every $x^* \in \mathcal{S}^*$.

Definition 1.2. Let Λ be a link from \mathcal{S}^* to \mathcal{S} . Then \mathbf{X}^* is said to be a Λ -*linked dual* of \mathbf{X} , if (a) \mathbf{X}^* and \mathbf{X} are Markovian, (b) X_n^* and X_n are conditionally independent given (X_0, \dots, X_n) for every $n \geq 0$, and (c) it satisfies

$$(1.4) \quad \mathbb{P}(X_n \in \cdot | X_0^* = x_0^*, \dots, X_n^* = x_n^*) = \Lambda(x_n^*, \cdot)$$

for every $n \geq 0$ and for every possible value (x_0^*, \dots, x_n^*) of (X_0^*, \dots, X_n^*) .

In the rest of Section 1 we assume that \mathbf{X} is an ergodic chain with stationary distribution π . Then a notion of strong stationary duality can be introduced in terms of Λ -linked duality, though it is not necessarily the one defined by Diaconis and Fill [4] [cf. Definition 2.1 and Theorem 2.4(b) of [4]].

Definition 1.3. A Λ -linked dual \mathbf{X}^* is called a *strong stationary dual* for \mathbf{X} if (a) the chain \mathbf{X}^* is absorbed to a single absorbing state, denoted by ∞ , and (b) the link Λ satisfies $\Lambda(\infty, \cdot) = \pi(\cdot)$.

In Section 3 we discuss connections between Λ -linked and Liggett duality. In Section 4 we lay out a foundation of set-valued strong stationary duality.

1.2 Perfect rejection sampling algorithm

Since Propp and Wilson [14] first introduced the coupling from the past (CFTP) algorithm, a possibility of perfect sampling has become an important consideration in Markov chain Monte Carlo (MCMC) practice (see Wilson [18] for extensive bibliography on the subject). Fill [5] and successively Fill, Machida, Murdoch, and Rosenthal [7] proposed a different perfect sampling algorithm based on rejection sampling method. In this subsection we describe their perfect rejection sampling algorithm in relation with set-valued duality.

Assuming that \mathbf{P} is ergodic with stationary distribution π , we can introduce the *time-reversed* Markov transition matrix $\tilde{\mathbf{P}}$ by $\tilde{P}(x, y) := \pi(y)P(y, x)/\pi(x)$. Let $\xi := (\xi(x) : x \in \mathcal{S})$ be a collection of \mathcal{S} -valued random variables $\xi(x)$ satisfying $\tilde{P}(x, y) = \mathbb{P}(\xi(x) = y)$. Then we call ξ collectively a *transition rule* for \mathbf{P} . By viewing ξ as a random map from \mathcal{S} to itself, we can define the image $\xi(x^*) := \{\xi(z) : z \in x^*\}$ and the inverse image $\xi^{-1}(x^*) := \{z \in \mathcal{S} : \xi(z) \in x^*\}$ for any subset x^* of \mathcal{S} . In Algorithm 1.4 we denote by $\mathbf{X}_n = (X_0, \dots, X_n)$ a Markov chain up to time n .

Algorithm 1.4. If the algorithm stops in Step 3, then we accept the value $X_n = x_n$ in Step 1 as an observation from π ; otherwise, reject the value and restart the routine.

Step 1. Sample a path $\mathbf{X}_n = (x_0, \dots, x_n)$ from the chain \mathbf{P} with initial state $X_0 = x_0$.

Step 2. Impute $\hat{\xi}_i$ from the conditional distribution $\mathcal{L}(\xi \mid \xi(x_i) = x_{i-1})$ independently for $i = 1, \dots, n$, and generate a set-valued path \mathbf{X}_n^* with initial state $X_0 = \{x_0\}$ and recursively by $X_i^* = \hat{\xi}_i^{-1}(X_{i-1}^*)$ for $i = 1, \dots, n$.

Step 3. If $X_n^* = \mathcal{S}$, then stop.

The use of a transition rule ξ for $\tilde{\mathbf{P}}$ is fit into the viewpoint of set-valued duality. When the backward path (X_n, \dots, X_0) is viewed as if it were sampled from the time-reversed $\tilde{\mathbf{P}}$, the path \mathbf{X}_n^* coincides with a set-valued dual for $\tilde{\mathbf{P}}$. In Section 2.1 we present such a construction of dual chain via transition rule ξ . In Section 4.2 we confirm this observation, and conclude that \mathbf{X}_n^* is a set-valued strong stationary dual for \mathbf{X}_n . Here the role of ∞ in Definition 1.3 is played by \mathcal{S} , which immediately indicates that $\mathbb{P}(X_n \in \cdot \mid X_n^* = \mathcal{S}) = \pi(\cdot)$, as desired.

Algorithm 1.5. It executes Algorithm 1.4 with Steps 2' and 3' in the place of Steps 2 and 3.

Step 2'. Impute $\hat{\xi}_i$ from the conditional distribution $\mathcal{L}(\xi \mid \xi(x_i) = x_{i-1})$ independently for $i = 1, \dots, n$, and generate a set-valued path \mathbf{Y}_n with initial state $Y_0 = \mathcal{S}$ and recursively by $Y_i = \hat{\xi}_{n-i+1}^{-1}(Y_{i-1})$ for $i = 1, \dots, n$.

Step 3'. If \mathbf{Y}_n coalesces, that is, if $Y_n = \{x_0\}$, then stop.

In comparing Algorithms 1.4 and 1.5, it is not difficult to see that $\{X_n^* = \mathcal{S}\}$ and $\{Y_n = \{x_0\}\}$ are the same event on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$, having the same effect in their rejection sampling scheme. Notice that Algorithm 1.5 has some advantage by forming a computationally less demanding rejection mechanism in Steps 2'–3'. In fact, Algorithm 1.4 was devised to demonstrate the correctness of Algorithm 1.5 via strong stationary duality (cf. Section 9.1 in [5]).

1.3 Monotone cases

Now we suppose further that \mathcal{S} is equipped with a partial ordering \leq , and that it has a minimum element $\hat{0}$ and a maximum element $\hat{1}$ such that $\hat{0} \leq x \leq \hat{1}$ for all $x \in \mathcal{S}$. Then a transition rule ξ for $\tilde{\mathbf{P}}$ is said to be *monotone* if $\xi(x) \leq \xi(y)$ almost surely whenever $x \leq y$. Using the imputed monotone transition rule ξ_i 's in Algorithm 1.5, we can construct a sample path $\tilde{\mathbf{Y}}_n$ recursively by $\tilde{Y}_i = \hat{\xi}_{n-i+1}(\tilde{Y}_{i-1})$ with initial state $\tilde{Y}_0 = \hat{1}$ so that the state \tilde{Y}_i becomes the least upper bound of Y_i for $i = 1, \dots, n$. By setting $x_0 = \hat{0}$ in Step 1 of Algorithms 1.4 and 1.5, $Y_n = \{\hat{0}\}$ is observed exactly when $\tilde{Y}_n = \hat{0}$ occurs. Thus, the procedure of detecting $\tilde{Y}_n = \hat{0}$ can simplify Steps 2–3 of Algorithm 1.5.

Let $k(\cdot | x', x, y)$ be a probability mass function on \mathcal{S} defined for each $x', x, y \in \mathcal{S}$ such that $x' \geq x$. Then we call k collectively an upward kernel for $\tilde{\mathbf{P}}$ if it satisfies for $x' \geq x$,

$$(1.5) \quad \sum_z k(y' | x', x, z) \tilde{P}(x, z) = \tilde{P}(x', y') \quad \text{for } x', y' \in \mathcal{S},$$

and

$$(1.6) \quad k(y' | x', x, y) = 0 \quad \text{if } y' \not\geq y.$$

Given a monotone transition rule ξ for $\tilde{\mathbf{P}}$, we can devise an upward kernel

$$(1.7) \quad k(\cdot | x', x, y) = \mathbb{P}(\xi(x') \in \cdot | \xi(x) = y),$$

and implement the alternative procedure for Algorithm 1.5.

Algorithm 1.6. The following two-step algorithm produces a value $X_n = x_n$ in Step 1, and return it as an observation from π if $\tilde{Y}_n = \hat{0}$ is attained in Step 2; otherwise, it restart the routine from Step 1.

Step 1. Sample a path $\mathbf{X}_n = (x_0, \dots, x_n)$ from the chain \mathbf{P} with initial state $x_0 = \hat{0}$.

Step 2. Generate a path $\tilde{\mathbf{Y}}_n = (y_n, \dots, y_0)$ by recursively sampling $\tilde{Y}_i = y_{n-i}$ from the probability distribution $k(\cdot | y_{n-i+1}, x_{n-i+1}, x_{n-i})$ for $i = 1, \dots, n$ with initial state $y_n = \hat{1}$.

The existence of upward kernel for $\tilde{\mathbf{P}}$ was studied by Kamae, Krengel and O'Brien [10], and is equivalently characterized by the stochastic monotonicity of $\tilde{\mathbf{P}}$ without assuming a monotone transition rule for $\tilde{\mathbf{P}}$. In Section 5.1 we introduce a set-valued dual for stochastically monotone chain with particular choice of dual state space. In Section 5.2 we discuss the validity of Algorithm 1.6 via strong stationary duality when $\tilde{\mathbf{P}}$ is stochastically monotone.

2 Set-valued duality

2.1 Constructive methods for dual chains

Let \mathbf{P} be a Markov transition matrix on \mathcal{S} , and let $\tilde{\mathcal{S}}$ be a subcollection of subsets of \mathcal{S} . Suppose that there is a transition rule ξ for \mathbf{P} , and that $\xi^{-1}(x^*)$ almost surely

belongs to $\check{\mathcal{S}}$ for every $x^* \in \check{\mathcal{S}}$. Then we obtain a set-valued dual $\check{\mathbf{P}}$ of \mathbf{P} [with the duality function (1.2)] by

$$(2.1) \quad \check{P}(x^*, y^*) := \mathbb{P}(\xi^{-1}(x^*) = y^*), \quad x^*, y^* \in \check{\mathcal{S}}.$$

It clearly follows that for every $(x^*, x) \in \check{\mathcal{S}} \times \mathcal{S}$,

$$\sum_{z^* \in \check{\mathcal{S}}} \check{P}(x^*, z^*) \Gamma(z^*, x) = \mathbb{P}(x \in \xi^{-1}(x^*)) = \mathbb{P}(\xi(x) \in x^*) = \sum_{z \in \mathcal{S}} \Gamma(x^*, z) P(x, z).$$

A transition rule ξ can be constructed with independent random variables $\xi(x)$ each distributed as $\mathbf{P}(x, \cdot)$, and (2.1) can always form a set-valued dual for \mathbf{P} if $\check{\mathcal{S}}$ is the collection $\check{\mathcal{S}}$ of all subsets of \mathcal{S} .

The next method for a set-valued dual has been known as a “greedy construction” by Diaconis and Fill (cf. Section 3 of [4]). For a fixed $x^* \in \check{\mathcal{S}}$, by setting

$$(2.2) \quad P(x, x^*) := \sum_{z \in \mathcal{S}} \Gamma(x^*, z) P(x, z)$$

as a function of x , the duality relation $P(x, x^*) = \sum_{z^* \in \check{\mathcal{S}}} \check{P}(x^*, z^*) \Gamma(z^*, x)$ can be viewed as a mixture of the indicator function $\Gamma(z^*, \cdot)$'s having the coefficient $\check{P}(x^*, z^*)$'s. Recalling that \mathcal{S} is finite, we can form a finite and strictly decreasing sequence $z_1^* \supset \cdots \supset z_k^*$ of the inverse images

$$(2.3) \quad z_i^* = \{x \in \mathcal{S} : P(x, x^*) > r\}, \quad r \in [a_{i-1}, a_i) \quad \text{for } i = 1, \dots, k,$$

accompanying with strictly increasing values $a_0 = 0 < a_1 < \cdots < a_k = 1$, in which we seek the subset z_i^* 's as greedily as possible. Provided that $z_i^* \in \check{\mathcal{S}}$ for every $i = 1, \dots, k$, we can identify the positive coefficient $\check{P}(x^*, z_i^*) = (a_i - a_{i-1})$, and therefore determine the set-valued dual

$$(2.4) \quad \check{P}(x^*, y^*) := \begin{cases} a_i - a_{i-1} & \text{if } y^* = z_i^* \text{ for some } i = 1, \dots, k; \\ 0 & \text{otherwise.} \end{cases}$$

Then the following corollary is a straightforward consequence of the greedy construction.

Corollary 2.1. *Let U be a uniform random variable on $[0, 1)$. Then*

$$(2.5) \quad \check{P}(x^*, y^*) = \mathbb{P}(y^* = \{z \in \mathcal{S} : P(z, x^*) > U\}).$$

is a set-valued dual of \mathbf{P} .

Proof. Comparing (2.3) and (2.5), it is clear that (2.4) and (2.5) are the same. \square

Example 2.2. Let \mathbf{P} be a Markov transition matrix on $\mathcal{S} := \{1, \dots, n\}$ with $P(i, j) = 1/n$ for $i, j \in \mathcal{S}$, and let $\check{\mathcal{S}}$ be the collection of all subsets in \mathcal{S} . And let ξ be a uniform random variable on the set of permutations over \mathcal{S} ; thus, ξ can be viewed as a transition rule for \mathbf{P} . By (2.1) we obtain a set-valued dual

$$(2.6) \quad \check{P}(x^*, y^*) = \begin{cases} \binom{n}{|x^*|}^{-1} & \text{if } |y^*| = |x^*|; \\ 0 & \text{otherwise,} \end{cases}$$

where $|x^*|$ denotes the size of the subset x^* . In applying the method of greedy construction, we have $P_x(X_1 \in x^*) \equiv |x^*|/n$ for each $x^* \in \check{\mathcal{S}}$, and therefore, obtain $z_1^* = \mathcal{S}$ and $z_2^* = \emptyset$ with $a_0 = 0$, $a_1 = |x^*|/n$, and $a_2 = 1$ in (2.3). By (2.4) we get

$$(2.7) \quad \check{P}(x^*, y^*) = \begin{cases} 1 - |x^*|/n & \text{if } y^* = \emptyset; \\ |x^*|/n & \text{if } y^* = \mathcal{S}; \\ 0 & \text{otherwise.} \end{cases}$$

Hence, different methods may yield different set-valued duals. \square

2.2 Some characteristics of constructive methods

Suppose that a set-valued dual $\check{\mathbf{P}}$ is obtained by (2.1) with transition rule ξ for \mathbf{P} . Then we can generate a sequence of independent transition rules ξ_1, \dots, ξ_n identically distributed as ξ , and form a sample path $\check{\mathbf{X}}_n$ of the dual chain up to n steps by $\check{X}_i = \xi_i^{-1}(\check{X}_{i-1})$ for $i = 1, \dots, n$ with initial state $\check{X}_0 = x_0^*$. By using the same sequence of transition rules, we can simultaneously formulate a set-valued path \mathbf{Y}_n recursively by $Y_i = \xi_{n-i+1}(Y_{i-1})$ for $i = 1, \dots, n$ with initial state $Y_0 = \mathcal{S}$. Then we can observe that $\{\check{X}_n = \mathcal{S}\}$ and $\{x_0^* \supset Y_n\}$ are the same event, and that $\{\check{X}_n = \emptyset\}$ and $\{x_0^* \cap Y_n = \emptyset\}$ are the same event. We will apply these observations in Section 4.1 to determine whether the chain $\check{\mathbf{X}}_n$ reaches either one of the absorbing states \emptyset and \mathcal{S} . We say that the transition rule ξ is *coalescing* if the path \mathbf{Y}_n coalesces to a singleton with some integer n . The size $|Y_i|$ of Y_i is non-increasing but not necessarily decreasing. For instance, the transition rule ξ in Example 2.2 is not coalescing since $|Y_i| = n$ for all i .

For the method of greedy construction we will find the following lemma useful in Section 4.1.

Lemma 2.3. *Let \mathbf{X} be a chain on \mathcal{S} . Suppose that a set-valued dual for \mathbf{X} is obtained by (2.3)–(2.4). Then the dual chain $\check{\mathbf{X}}$ with initial state $\check{X}_0 = x_0^*$ can reach both*

$$(2.8) \quad y_n^* := \{y \in \mathcal{S} : P_y(X_n \in x_0^*) > 0\} \text{ and } z_n^* := \{z \in \mathcal{S} : P_z(X_n \in x_0^*) = 1\}$$

in n steps.

Proof. We will show by induction that

$$(2.9) \quad \check{P}^n(x_0^*, y_n^*) > 0 \text{ and } \check{P}^n(x_0^*, z_n^*) > 0.$$

If $n = 0$, $y_0^* = z_0^* = x_0^*$ in (2.8), and therefore, (2.9) is obvious. Now we assume (2.8) and (2.9) with $(n - 1)$ substituting for n by induction hypothesis. By combining y_{n-1}^* and z_{n-1}^* in (2.8) [with $(n - 1)$ substituting for n] with the expression $P_y(X_n \in x_0^*) = \sum_{y' \in \mathcal{S}} P(y, y')P_{y'}(X_{n-1} \in x_0^*)$, we can observe that y_n^* and z_n^* can be equivalently expressed as

$$y_n^* = \{y \in \mathcal{S} : P(y, y_{n-1}^*) > 0\} \text{ and } z_n^* = \{z \in \mathcal{S} : P(z, z_{n-1}^*) = 1\},$$

respectively. Therefore, we obtain $\check{P}(y_{n-1}^*, y_n^*) > 0$ and $\check{P}(z_{n-1}^*, z_n^*) > 0$ by the greedy construction. Together with the induction hypothesis we can find

$$\begin{aligned}\check{P}^n(x_0^*, y_n^*) &\geq \check{P}^{n-1}(x_0^*, y_{n-1}^*) \check{P}(y_{n-1}^*, y_n^*) > 0 \quad \text{and} \\ \check{P}^n(x_0^*, z_n^*) &\geq \check{P}^{n-1}(x_0^*, z_{n-1}^*) \check{P}(z_{n-1}^*, z_n^*) > 0,\end{aligned}$$

as desired. \square

3 Intertwining and Liggett duality

3.1 Intertwining duality

In this subsection we assume the setting of Definition 1.2, and denote by \mathbf{Q} the Markov transition matrix for the bivariate Markov chain $(\mathbf{X}^*, \mathbf{X})$. Then we can recapture Theorem 2.17 and Remark 2.23 of [4].

Proposition 3.1. *Let Λ be a link from \mathcal{S}^* to \mathcal{S} , and let \mathbf{P}^* and \mathbf{P} be Markov transition matrices on the respective state spaces \mathcal{S}^* and \mathcal{S} . Then \mathbf{X}^* is a Λ -linked dual of \mathbf{X} if and only if (a) the initial pair (X_0^*, X_0) satisfies*

$$(3.1) \quad \mathbb{P}(X_0 = x \mid X_0^* = x^*) = \Lambda(x^*, x),$$

and (b) the bivariate Markov chain $(\mathbf{X}^*, \mathbf{X})$ satisfies

$$(3.2) \quad \sum_{z^*} Q((x^*, x), (z^*, y)) = P(x, y) \quad \text{for every possible value } x^*;$$

$$(3.3) \quad \sum_z \Lambda(x^*, z) Q((x^*, z), (y^*, y)) = P^*(x^*, y^*) \Lambda(y^*, y).$$

Proof. If \mathbf{X}^* is a Λ -linked dual for \mathbf{X} , then (3.1) obviously holds, and (3.2) and (3.3) are obtained via $\mathbb{P}(X_1 = y \mid X_0^* = x^*, X_0 = x)$ and $\mathbb{P}(X_1^* = y^*, X_1 = y \mid X_0^* = x^*)$, respectively. Conversely, suppose that (3.1)–(3.3) hold. By (3.2), we obtain

$$(3.4) \quad \begin{aligned}\mathbb{P}(X_{n+1} = x_{n+1}, \dots, X_{n+k} = x_{n+k} \mid X_n^* = x_n^*; X_0 = x_0, \dots, X_n = x_n) \\ = P(x_n, x_{n+1}) \cdots P(x_{n+k-1}, x_{n+k}),\end{aligned}$$

indicating the Markovian property of \mathbf{X} and Definition 1.2(b). By (3.1) and (3.3), we obtain

$$(3.5) \quad \begin{aligned}\mathbb{P}(X_1^* = x_1^*, \dots, X_n^* = x_n^*, X_n = x_n \mid X_0^* = x_0^*) \\ = P^*(x_0^*, x_1^*) \cdots P^*(x_{n-1}^*, x_n^*) \Lambda(x_n^*, x_n),\end{aligned}$$

indicating the Markovian property of \mathbf{X}^* and Definition 1.2(c). \square

In the sense of Proposition 3.1, we call \mathbf{Q} Λ -linked if it satisfies (3.2)–(3.3). By combining (3.4) and (3.5) with (1.4), we can simplify the calculation of $\mathbb{P}(X_1 = y \mid X_0^* = x^*)$

in the following way:

$$(3.6) \quad \begin{aligned} \sum_x \Lambda(x^*, x) P(x, y) &= \sum_x \mathbb{P}(X_1 = y, X_0 = x \mid X_0^* = x^*) \\ &= \sum_{y^*} \mathbb{P}(X_1 = y, X_1^* = y^* \mid X_0^* = x^*) = \sum_{y^*} P^*(x^*, y^*) \Lambda(y^*, y). \end{aligned}$$

From the viewpoint of (3.6) we can formulate another notion of duality.

Definition 3.2. Let $\mathbf{\Lambda}$ be a link from \mathcal{S}^* to \mathcal{S} . Then \mathbf{P}^* is said to be an *intertwining $\mathbf{\Lambda}$ -dual* of \mathbf{P} if the *intertwining relation* $\mathbf{\Lambda} \mathbf{P} = \mathbf{P}^* \mathbf{\Lambda}$ holds.

Clearly a $\mathbf{\Lambda}$ -linked \mathbf{Q} implies the existence of an intertwining $\mathbf{\Lambda}$ -dual. Conversely, suppose that \mathbf{P}^* is an intertwining $\mathbf{\Lambda}$ -dual of \mathbf{P} . Then we can obtain the $\mathbf{\Lambda}$ -linked \mathbf{Q} by

$$Q((x^*, x), (y^*, y)) := \begin{cases} \frac{P(x, y) P^*(x^*, y^*) \Lambda(y^*, y)}{\Delta(x^*, y)} & \text{if } \Delta(x^*, y) > 0; \\ 0 & \text{otherwise,} \end{cases}$$

where

$$\Delta(x^*, y) := \sum_{z \in \mathcal{S}} \Lambda(x^*, z) P(z, y) = \sum_{z^* \in \mathcal{S}^*} P^*(x^*, z^*) \Lambda(z^*, y).$$

And the state space of \mathbf{Q} is given by $\mathcal{S}_Q := \{(x^*, x) \in \mathcal{S}^* \times \mathcal{S} : \Delta(x^*, x) > 0\}$.

3.2 A duality theorem

An intriguing aspect of Liggett duality is its role in reversing entrance and exit laws, which has been demonstrated for continuous time Markov processes by Cox and Rösler [2]. In Lemma 3.3 we describe how Liggett duality converts invariant measures to harmonic functions in the discrete time setting. As a convention we write a measure ν in a form of row vector, and a function h in a form of column vector. Then we call ν *invariant* for \mathbf{P} if $\nu = \nu \mathbf{P}$, and h *harmonic* for $\check{\mathbf{P}}$ if h is nonnegative and $h = \check{\mathbf{P}} h$. The following lemma is due to Vervaat [17].

Lemma 3.3. *Suppose that $\check{\mathbf{P}}$ is a Liggett $\mathbf{\Gamma}$ -dual of \mathbf{P} , and that ν is an invariant measure for \mathbf{P} . Then $h := \mathbf{\Gamma} \nu^T$ is harmonic for $\check{\mathbf{P}}$.*

Proof. The function h is clearly nonnegative. By Definition 1.1 we obtain $\check{\mathbf{P}} h = \check{\mathbf{P}} \mathbf{\Gamma} \nu^T = \mathbf{\Gamma} \mathbf{P}^T \nu^T = \mathbf{\Gamma} (\nu \mathbf{P})^T = \mathbf{\Gamma} \nu^T = h$. \square

Given an invariant measure ν for \mathbf{P} , we introduce a new Markov transition matrix $\check{P}_\nu(x, y) := \nu(y) P(y, x) / \nu(x)$ on the appropriately reduced state space $\{x \in \mathcal{S} : \nu(x) > 0\}$, and call it the *time-reversed* of \mathbf{P} . Correspondingly, given a harmonic function h for $\check{\mathbf{P}}$, we introduce another Markov transition matrix $\check{P}_h(x^*, y^*) := \check{P}(x^*, y^*) h(y^*) / h(x^*)$ on the state space $\{x^* \in \check{\mathcal{S}} : h(x^*) > 0\}$, and call it the *Doob h transform* of $\check{\mathbf{P}}$. Then the notion of Liggett duality can be related to that of intertwining duality by means of these Markov transition matrices.

Theorem 3.4. (Cf. Theorem 5.5 of [4]) *Let ν be a strictly positive invariant measure for \mathbf{P} , and let Γ be a duality function from $\check{\mathcal{S}}$ to \mathcal{S} . Then if $\check{\check{\mathbf{P}}}_\nu$ is a Liggett Γ -dual for the time-reversed $\check{\check{\mathbf{P}}}_\nu$, then*

- (a) $h := \Gamma \nu^T$ is harmonic for $\check{\check{\mathbf{P}}}_\nu$;
- (b) $\Lambda(x^*, x) := \Gamma(x^*, x) \nu(x) / h(x^*)$ is a link from $\mathcal{S}^* := \{x^* \in \check{\mathcal{S}} : h(x^*) > 0\}$ to \mathcal{S} ;
- (c) $(\check{\check{\mathbf{P}}}_\nu)_h$ is an intertwining Λ -dual for \mathbf{P} .

Proof. Since ν is strictly positive, we can observe that ν is also an invariant measure for $\check{\check{\mathbf{P}}}_\nu$. By Lemma 3.3 the function h is harmonic for $\check{\check{\mathbf{P}}}_\nu$. Furthermore, the construction of h indicates that $\sum_{z \in \mathcal{S}} \Lambda(x^*, z) = 1$, and therefore, that Λ is a link.

In verifying (c), we will assume $h(x^*) > 0$ for all x^* for the sake of simplicity; otherwise, one rest its attention in the following calculations to $\{x^* \in \check{\mathcal{S}} : h(x^*) > 0\}$. We denote by \mathbf{D}_ν the diagonal matrix with diagonal entry $D_\nu(x, x) := \nu(x)$, and by \mathbf{D}_h the diagonal matrix with $D_h(x^*, x^*) := h(x^*)$. Then we can write $(\check{\check{\mathbf{P}}}_\nu)_h = \mathbf{D}_h^{-1} \check{\check{\mathbf{P}}}_\nu \mathbf{D}_h$ and $\Lambda = \mathbf{D}_h^{-1} \Gamma \mathbf{D}_\nu$, and obtain

$$\Lambda \mathbf{P} = \mathbf{D}_h^{-1} \Gamma \mathbf{D}_\nu \mathbf{P} = \mathbf{D}_h^{-1} \Gamma (\check{\check{\mathbf{P}}}_\nu)^T \mathbf{D}_\nu = \mathbf{D}_h^{-1} \check{\check{\mathbf{P}}}_\nu \Gamma \mathbf{D}_\nu = (\check{\check{\mathbf{P}}}_\nu)_h \Lambda,$$

as desired. \square

4 Set-valued strong stationary duality

4.1 Set-valued duals for ergodic Markov chains

We now assume that a Markov chain \mathbf{P} is ergodic (i.e., irreducible and aperiodic in a finite state space setting). Then we have a unique stationary distribution (i.e., invariant and strictly positive probability measure) for \mathbf{P} , denoted by π . We call a Markov chain *absorbing* if all of its recurrent states are absorbing. In the setting of finite state space, a Markov chain is absorbing if every non-absorbing state is transient, and reaches one of the absorbing states. However, we note that this notion of absorbing chain may not be the same as others introduced in the literature (see e.g., [9]).

It is natural to relate an ergodic chain \mathbf{X} to an absorbing set-valued dual $\check{\check{\mathbf{X}}}$. Since $P_{\mathcal{S}}(x \in \check{\check{X}}_n) = 1$ and $P_{\emptyset}(x \in \check{\check{X}}_n) = 0$ for every $x \in \mathcal{S}$ by (1.3), both \mathcal{S} and \emptyset are absorbing in $\check{\check{\mathbf{X}}}$. Let $x^* \in \check{\mathcal{S}}$ be neither \emptyset nor \mathcal{S} . By the ergodicity, X_n converges to π in distribution as $n \rightarrow \infty$, and there is some integer n large enough to satisfy

$$(4.1) \quad 0 < P_x(X_n \in x^*) = P_{x^*}(x \in \check{\check{X}}_n) < 1 \quad \text{for all } x \in \mathcal{S}.$$

Thus, x^* is not absorbing. In summary, $\check{\check{\mathbf{X}}}$ has the exactly two absorbing states \mathcal{S} and \emptyset .

Proposition 4.1. *Let \mathbf{P} be ergodic, and let ξ be a transition rule for \mathbf{P} . Then the set-valued dual (2.1) is absorbing if and only if ξ is coalescing.*

Proof. Recall the sample path \mathbf{Y}_n in Section 2.2. We can observe that $P_{x_0^*}(\check{X}_n = \mathcal{S}) = \mathbb{P}(x_0^* \supset Y_n)$, and that $P_{x_0^*}(\check{X}_n = \emptyset) = \mathbb{P}(x_0^* \cap Y_n = \emptyset)$. Let x^* be a non-absorbing state, and let $x, x' \in \mathcal{S}$ be such that $x \in x^*$ and $x' \notin x^*$. If ξ is coalescing, by the ergodicity we have $P_{x^*}(\check{X}_n = \mathcal{S}) = \mathbb{P}(x^* \supset Y_n) \geq \mathbb{P}(Y_n = \{x\}) > 0$, and $P_{x^*}(\check{X}_n = \emptyset) = \mathbb{P}(x^* \cap Y_n = \emptyset) \geq \mathbb{P}(Y_n = \{x'\}) > 0$ for some integer n . Thus, $\check{\mathbf{X}}$ is absorbing. Conversely if $\check{\mathbf{X}}$ is absorbing, then by choosing $x_0^* = \{x\}$ we have $\mathbb{P}(Y_n = \{x\}) = P_{\{x\}}(\check{X}_n = \mathcal{S}) > 0$ for some n , which implies that ξ is coalescing. \square

Suppose that $\check{\mathbf{P}}$ is greedily constructed. If x^* is a non-absorbing state, then by Lemma 2.3 we have $y_n^* = \mathcal{S}$ and $z_n^* = \emptyset$ in (2.8) for some n satisfying (4.1), and therefore, x^* is transient. Thus, we have proved the following corollary.

Corollary 4.2. *Let $\check{\mathbf{P}}$ be a set-valued dual of \mathbf{P} obtained via (2.3)–(2.4). Then if \mathbf{P} is ergodic, $\check{\mathbf{P}}$ is absorbing.*

In Example 2.2 every non-absorbing state in the set-valued dual (2.7) moves to either \emptyset and \mathcal{S} in one step. But the set-valued dual (2.6) is not absorbing since every non-absorbing state x^* of size $|x^*| = k$ stays in the closed class of the subsets of the same size k .

We can find a natural interpretation for Lemma 3.3 when \mathbf{X} is ergodic and $\check{\mathbf{X}}$ is an absorbing set-valued dual for \mathbf{X} . Let T be the time to absorption in the state \mathcal{S} for $\check{\mathbf{X}}$. Then let

$$(4.2) \quad H(x^*) := P_{x^*}(T < \infty)$$

be the probability that the chain $\check{\mathbf{X}}$ starting from $\check{X}_0 = x^*$ is ever absorbed into \mathcal{S} . By using the first step analysis, we have $H(x^*) = \sum_{\xi^* \in \check{\mathcal{S}}} \check{P}(x^*, \xi^*) H(\xi^*)$; thus, H is harmonic for $\check{\mathbf{P}}$. Since the chain $\check{\mathbf{X}}$ will be eventually absorbed into either the state \emptyset or \mathcal{S} , we can observe that for any $x \in \mathcal{S}$,

$$(4.3) \quad H(x^*) = \lim_{n \rightarrow \infty} P_{x^*}(x \in \check{X}_n) = \lim_{n \rightarrow \infty} P_x(X_n \in x^*) = \sum_z \Gamma(x^*, z) \pi(z).$$

So that H is the harmonic function obtained by Lemma 3.3.

4.2 Sample path construction

A ergodic chain \mathbf{P} has a necessarily unique invariant measure π , and $\check{\mathbf{P}}_\pi$ in Theorem 3.4 is simply the time-reversed $\check{\mathbf{P}}$ in Section 1.2. In the setting of set-valued duality Theorem 3.4 can be tailored in a form of the following corollary.

Corollary 4.3. *Suppose that $\check{\check{\mathbf{P}}}$ is a set-valued dual on $\check{\mathcal{S}}$ for the time-reversed $\check{\mathbf{P}}$. Then (a) (4.3) is harmonic for $\check{\check{\mathbf{P}}}$, (b)*

$$(4.4) \quad \Lambda(x^*, x) := \frac{\pi(x)}{H(x^*)} \Gamma(x^*, x)$$

is a link from $\mathcal{S}^* := \check{\mathcal{S}} \setminus \{\emptyset\}$ to \mathcal{S} , and (c)

$$(4.5) \quad P^*(x^*, y^*) := \frac{H(y^*)}{H(x^*)} \check{P}(x^*, y^*), \quad x^*, y^* \in \mathcal{S}^*,$$

is an intertwining $\mathbf{\Lambda}$ -dual for \mathbf{P} . Moreover, if $\check{\mathbf{P}}$ is absorbing, so is \mathbf{P}^* with absorbing state \mathcal{S} .

Proof. In addition to Theorem 3.4, we need to verify the assertion on absorbing. The state \mathcal{S} is clearly an absorbing state in \mathbf{P}^* . From (4.5) we obtain the n -step transition

$$(4.6) \quad P^{*n}(x^*, y^*) = \frac{H(y^*)}{H(x^*)} \check{P}^n(x^*, y^*).$$

Thus, the chain \mathbf{P}^* reaches the absorbing state \mathcal{S} starting from a non-absorbing state x^* if $\check{\mathbf{P}}$ does so, which verifies the claim. \square

Let $q(\cdot | x^*, x, y)$ be a probability mass function on $\check{\mathcal{S}}$ for each $x^* \in \check{\mathcal{S}}$ and each $(x, y) \in \mathcal{S} \times \mathcal{S}$. Then we call q collectively a *dual generating kernel* for $\check{\mathbf{P}}$ if it satisfies

$$(4.7) \quad q(y^* | x^*, x, y) = 0 \quad \text{if } x \notin y^* \text{ and } y \in x^*, \text{ or if } x \in y^* \text{ and } y \notin x^*,$$

and defines a Markov transition probability

$$(4.8) \quad \check{P}(x^*, y^*) := \sum_z q(y^* | x^*, x, z) \check{P}(x, z)$$

without depending on the choice of $x \in \mathcal{S}$. Together it implies that

$$\sum_{z^*} \check{P}(x^*, z^*) \Gamma(z^*, x) = \sum_z \check{P}(x, z) \sum_{z^*} q(y^* | x^*, x, z) \Gamma(z^*, x) = \sum_z \check{P}(x, z) \Gamma(x^*, z),$$

and therefore, that $\check{\mathbf{P}}$ is a set-valued dual for $\check{\mathbf{P}}$. If we have a transition rule ξ for $\check{\mathbf{P}}$ such that $\xi^{-1}(x^*) \in \check{\mathcal{S}}$ for every $x^* \in \check{\mathcal{S}}$, then we can construct a dual generating kernel by

$$(4.9) \quad q(y^* | x^*, x, y) = \mathbb{P}(y^* = \xi^{-1}(x^*) | \xi(x) = y),$$

which clearly satisfies (4.7)–(4.8) with $\check{P} = \mathbb{P}(y^* = \xi^{-1}(x^*))$.

As discussed in Section 3.1, the existence of an intertwining $\mathbf{\Lambda}$ -dual implies that of a $\mathbf{\Lambda}$ -linked dual. Moreover, a dual generating kernel arises naturally for a set-valued $\mathbf{\Lambda}$ -linked dual.

Proposition 4.4. *Let $q(\cdot | x^*, x, y)$ be a dual generating kernel satisfying (4.7)–(4.8), and let $\mathcal{S}_Q := \{(x^*, x) \in \check{\mathcal{S}} \times \mathcal{S} : x \in x^*\}$ be a state space. Then*

$$(4.10) \quad Q((x^*, x), (y^*, y)) := q(y^* | x^*, y, x) P(x, y), \quad (x^*, x), (y^*, y) \in \mathcal{S}_Q,$$

is $\mathbf{\Lambda}$ -linked with link (4.4).

Proof. Since (4.8) is a set-valued dual of $\tilde{\mathbf{P}}$, we can construct the intertwining $\mathbf{\Lambda}$ -dual (4.5) via Corollary 4.3. By Proposition 3.1, it suffices to show that (3.2)–(3.3) hold. By (4.7), for every $(x^*, x) \in \mathcal{S}_Q$ we have $\sum_{z^*} q(z^* | x^*, x, y) = 1$, and therefore, $\sum_{z^*} Q((x^*, x), (z^*, y)) = P(x, y)$, where the sum is over $z^* \in \check{\mathcal{S}}$ such that $y \in z^*$. By (4.8)–(4.7) we obtain

$$\begin{aligned} \sum_z \Lambda(x^*, z) Q((x^*, z), (y^*, y)) &= \frac{\pi(y)}{H(x^*)} \sum_z q(y^* | x^*, y, z) \tilde{P}(y, z) \\ &= \frac{\pi(y)}{H(x^*)} \Gamma(y^*, y) \check{P}(x^*, y^*) = P^*(x^*, y^*) \Lambda(y^*, y), \end{aligned}$$

which completes the proof. \square

By Propositions 3.1 and 4.4, we can sample the path $(\mathbf{X}_n^*, \mathbf{X}_n)$ from the $\mathbf{\Lambda}$ -linked bivariate chain (4.10). Let $\rho_0(x^*, x) = \pi_0(x^*) \Lambda(x^*, x)$ be an initial distribution with probability mass function $\pi_0(x^*)$ on $\check{\mathcal{S}}$.

Algorithm 4.5. Sample (x_0^*, x_0) from the initial distribution ρ_0 , and the following two-step procedure generates a sample path $(\mathbf{X}_n^*, \mathbf{X}_n)$ up to time n .

Step 1. Sample a path $\mathbf{X}_n = (x_0, \dots, x_n)$ from the chain \mathbf{P} starting from $X_0 = x_0$.

Step 2. Generate a set-valued path $\mathbf{X}_n^* = (x_0^*, \dots, x_n^*)$ by recursively sampling $X_i^* = x_i^*$ from the probability $q(\cdot | x_{i-1}^*, x_i, x_{i-1})$ for $i = 1, \dots, n$.

When Algorithm 4.5 starts from a fixed state $(X_0^*, X_0) = (\{x_0\}, x_0)$ and uses the dual generating kernel (4.9), it coincides exactly with Algorithm 1.4. Moreover, if the transition rule ξ in (4.9) is coalescing, by Proposition 4.1 and Corollary 4.3 the chain \mathbf{P}^* is absorbed to the state \mathcal{S} . By Proposition 4.4 \mathbf{X}_n^* is a $\mathbf{\Lambda}$ -linked dual of \mathbf{X}_n up to time n with $\Lambda(\mathcal{S}, \cdot) = \pi(\cdot)$. Hence, by Definition 1.3, we have verified that \mathbf{X}_n^* is a sample path of a strong stationary dual for \mathbf{X}_n .

5 Stochastically monotone case

5.1 Siegmund duality

The idea of duality can be traced back to Lévy [12] who observed the duality relation between a random walk \mathbf{X} on $[0, \infty)$ with absorption at 0 and another $\check{\mathbf{X}}$ with reflection at 0. Here we simplify it to that of random walks on the state space $\mathcal{S} = \check{\mathcal{S}} := \{0, 1, 2, \dots\}$. Siegmund [15] introduced the duality relation

$$(5.1) \quad P_x(X_n \leq y) = P_y(x \leq \check{X}_n), \quad x, y \in \mathcal{S},$$

which is a Liggett duality with duality function $\Gamma(x, y) = 1$ if $x \leq y$; otherwise, $\Gamma(x, y) = 0$. Siegmund [15] showed that \mathbf{X} has a dual chain $\check{\mathbf{X}}$ satisfying (5.1) if and only if (a) \mathbf{X} is stochastically monotone, and (b) the state 0 is absorbing for \mathbf{X} .

Now suppose that \mathcal{S} is partially ordered with partial ordering \leq . Then a real-valued function f on \mathcal{S} is said to be *decreasing* if $f(x) \geq f(y)$ whenever $x \leq y$. Similarly a subset x^* of \mathcal{S} is said to be *decreasing* if $z \in x^*$ whenever $z \leq x$ for some $x \in x^*$, and is often

called a “down-set” or an “order ideal” in the literature (see e.g., Stanley [16]). The empty set \emptyset and \mathcal{S} itself are trivially decreasing. A decreasing function f can be equivalently characterized by decreasing subset $\{z \in \mathcal{S} : f(z) > r\}$ with every real value r . A Markov chain \mathbf{P} on \mathcal{S} is said to be *stochastically monotone* if the function $P(x, x^*)$ in (2.2) is decreasing in x for every decreasing subset x^* of \mathcal{S} . The concept of stochastic monotonicity was introduced by Daley [3] in the linearly ordered setting, and extended by Kamae, Krengel and O’Brien [10] for partially ordered spaces.

Proposition 5.1. *Let \mathcal{S} be a partially ordered finite state space with minimum element $\hat{0}$, and let $\check{\mathcal{S}}$ be the collection of all the nonempty decreasing subsets of \mathcal{S} . Then \mathbf{P} has a set-valued dual $\check{\mathbf{P}}$ if and only if (a) \mathbf{P} is stochastically monotone and (b) the state $\hat{0}$ is absorbing in \mathbf{P} .*

Proof. Suppose that \mathbf{P} satisfies (a)–(b). Then we can claim that a set-valued dual $\check{\mathbf{P}}$ can be greedily constructed via (2.3)–(2.4). Let $x^* \in \check{\mathcal{S}}$ be fixed. The subset z_i^* ’s in (2.3) are all decreasing by (a). Observing that $\hat{0} \in x^*$, and that $P(\hat{0}, x^*) = 1$ by (b), we can verify that z_i^* ’s belongs to $\check{\mathcal{S}}$. Thus, (2.4) becomes a set-valued dual.

Conversely, suppose that \mathbf{X} has a set-valued dual $\check{\mathbf{X}}$. (a) Let $x^* \in \check{\mathcal{S}}$, and let $x \leq y$. By observing that $\{z^* \in \check{\mathcal{S}} : x \in z^*\} \supseteq \{z^* \in \check{\mathcal{S}} : y \in z^*\}$, we obtain $P_x(X_1 \in x^*) = P_{x^*}(x \in \check{X}_1) \geq P_{x^*}(y \in \check{X}_1) = P_y(X_1 \in x^*)$, which implies the stochastic monotonicity of \mathbf{X} . (b) Since $\{\hat{0}\} \in \check{\mathcal{S}}$, we have $P_{\hat{0}}(X_1 = \hat{0}) = P_{\hat{0}}(X_1 \in \{\hat{0}\}) = P_{\{\hat{0}\}}(\hat{0} \in \check{X}_1) = 1$. This completes the proof. \square

By adding the empty set \emptyset to the dual state space $\check{\mathcal{S}}$, we obtain the following corollary to Proposition 5.1.

Corollary 5.2. *Let \mathcal{S} be a partially ordered state space, and let $\check{\mathcal{S}}$ be the collection of all decreasing subsets of \mathcal{S} . Then \mathbf{X} has a set-valued dual $\check{\mathbf{X}}$ on the dual state space $\check{\mathcal{S}}$ if and only if \mathbf{X} is stochastically monotone.*

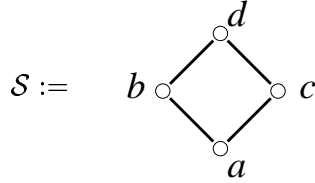
Proof. The assumption that $\emptyset \in \check{\mathcal{S}}$ eliminates the necessity of the absorbing state $\hat{0}$. Otherwise, the proof becomes essentially that of Proposition 5.1. \square

We call $\check{\mathbf{X}}$ in Proposition 5.1 or in Corollary 5.2 a *set-valued Siegmund dual*. If $\mathcal{S} := \{0, 1, \dots, d\}$ is a space with its usual linear ordering, then $\check{\mathcal{S}}$ becomes the collection of subsets of the form $\langle x \rangle := \{z \in \mathcal{S} : z \leq x\}$. Provided that (a)–(b) in Proposition 5.1 hold, a set-valued Siegmund dual is uniquely determined by $\check{P}(\langle x \rangle, \langle y \rangle) = P_y(X_1 \leq x) - P_{y+1}(X_1 \leq x)$, where we set $P_{d+1}(X_1 \leq x) \equiv 0$ for convenience. A sample path construction in this setting has been also discussed by Clifford and Sudbury [1], though they used much more elaborate construction for absorbing and reflecting continuous time Markov processes. A question naturally arises on whether we can adapt the dual state space $\check{\mathcal{S}} = \{\langle x \rangle : x \in \mathcal{S}\}$ even if \mathcal{S} is partially ordered. As remarked in page 87 of Liggett [13], such an extension of Siegmund dual requires “a much more special property” than stochastic monotonicity.

We call \mathbf{P} *realizably monotone* if there exists a monotone transition rule ξ for \mathbf{P} . Fill and Machida [6] showed that the notion of realizable monotonicity is strictly stronger than that of stochastic monotonicity. Provided a monotone transition rule ξ , $\xi^{-1}(x^*)$

is decreasing almost surely for every decreasing subset x^* , and therefore, a set-valued Siegmund dual in Corollary 5.2 can be obtained via (2.1).

Example 5.3. Let



be a partially ordered state space in a Hasse diagram (the standard graphical representation of partial ordering) where $a < b, c$ and $b, c < d$, and let

$$\check{\mathcal{S}} := \{\emptyset, \{a\}, \{a, b\}, \{a, c\}, \{a, b, c\}, \mathcal{S}\}$$

be the collection of all the decreasing subsets. Consider the Markov transition matrix

$$\begin{array}{c} a \\ b \\ c \\ d \end{array} \begin{bmatrix} a & b & c & d \\ 1/2 & 1/4 & 1/4 & 0 \\ 1/2 & 0 & 0 & 1/2 \\ 0 & 1/2 & 1/2 & 0 \\ 0 & 1/4 & 1/4 & 1/2 \end{bmatrix} .$$

Then it is stochastically monotone. By applying (2.3)–(2.4), we obtain a set-valued Siegmund dual

$$\check{\mathbf{P}}_1 := \begin{array}{c} \emptyset \\ \{a\} \\ \{a, b\} \\ \{a, c\} \\ \{a, b, c\} \\ \mathcal{S} \end{array} \begin{bmatrix} \emptyset & \{a\} & \{a, b\} & \{a, c\} & \{a, b, c\} & \mathcal{S} \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 & 0 \\ 1/4 & 1/4 & 0 & 0 & 1/4 & 1/4 \\ 1/4 & 1/4 & 0 & 0 & 1/4 & 1/4 \\ 0 & 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} .$$

Observe that \mathbf{P} is realizably monotone, and that it has the monotone transition rule ξ :

$$\begin{aligned} \mathbb{P}(\xi(a) = b, \xi(b) = d, \xi(c) = b, \xi(d) = d) &= 1/4; \\ \mathbb{P}(\xi(a) = c, \xi(b) = d, \xi(c) = c, \xi(d) = d) &= 1/4; \\ \mathbb{P}(\xi(a) = a, \xi(b) = a, \xi(c) = b, \xi(d) = b) &= 1/4; \\ \mathbb{P}(\xi(a) = a, \xi(b) = a, \xi(c) = c, \xi(d) = c) &= 1/4. \end{aligned}$$

By applying (2.1), we have another set-valued Siegmund dual

$$\check{\mathbf{P}}_2 := \begin{array}{c} \emptyset \\ \{a\} \\ \{a, b\} \\ \{a, c\} \\ \{a, b, c\} \\ \mathcal{S} \end{array} \begin{bmatrix} \emptyset & \{a\} & \{a, b\} & \{a, c\} & \{a, b, c\} & \mathcal{S} \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 & 0 \\ 1/4 & 0 & 1/4 & 1/4 & 0 & 1/4 \\ 1/4 & 0 & 1/4 & 1/4 & 0 & 1/4 \\ 0 & 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} .$$

Thus, a set-valued Siegmund dual is not unique. \square

5.2 Sample path construction

Suppose that the time-reversed $\tilde{\mathbf{P}}$ is stochastically monotone, and that \mathcal{S} has the minimum $\hat{0}$ and the maximum $\hat{1}$ as in the setting of Section 1.3. Let $\check{\mathcal{S}}$ be the collection of all decreasing subsets of \mathcal{S} , and let U be a uniform random variable on $[0, 1)$. Combining Corollaries 2.1 and 5.2 together, we can find that

$$(5.2) \quad \check{\tilde{P}}(x^*, y^*) = \mathbb{P}\left(y^* = \{z \in \mathcal{S} : \tilde{P}(z, x^*) > U\}\right), \quad x^*, y^* \in \check{\mathcal{S}},$$

is a set-valued Siegmund dual for $\tilde{\mathbf{P}}$. Furthermore, we can define a probability mass function $q(y^* | x^*, x, y)$ on $\check{\mathcal{S}}$ for $x^* \in \check{\mathcal{S}}$ and for $x, y \in \mathcal{S}$ by

$$(5.3) \quad \begin{cases} \mathbb{P}\left(y^* = \{z \in \mathcal{S} : \tilde{P}(z, x^*) > U\} \mid 0 \leq U < \tilde{P}(x, x^*)\right) & \text{if } x \in y^* \text{ and } y \in x^*; \\ \mathbb{P}\left(y^* = \{z \in \mathcal{S} : \tilde{P}(z, x^*) > U\} \mid \tilde{P}(x, x^*) \leq U < 1\right) & \text{if } x \notin y^* \text{ and } y \notin x^*; \\ 0 & \text{otherwise.} \end{cases}$$

Then (5.3) becomes a dual generating kernel satisfying (4.7)–(4.8), and generates the dual (5.2) via (4.8). Given the dual generating kernel (5.3), we can run Algorithm 4.5 with the fixed initial state $(x_0^*, x_0) = (\{\hat{0}\}, \hat{0})$.

Algorithm 5.4. Execute Step 1 of Algorithm 1.6 to generate a path $\mathbf{X}_n = (x_0, \dots, x_n)$. Starting from $X_0^* = \{\hat{0}\}$, generate a set-valued path $\mathbf{X}_n^* = (x_0^*, \dots, x_n^*)$ recursively by

$$(5.4) \quad X_i^* = \{z \in \mathcal{S} : \tilde{P}(z, x_{i-1}^*) > \hat{U}_i\}, \quad i = 1, \dots, n,$$

where \hat{U}_i is uniformly distributed on the interval $[0, \tilde{P}(x_i, x_{i-1}^*)]$.

Algorithm 5.4 is an implementation of Algorithm 4.5, and $\check{\tilde{\mathbf{P}}}$ is absorbing by Corollary 4.2. Thus, as discussed in Section 4.2, we can find that the path \mathbf{X}_n^* is a set-valued strong stationary dual for \mathbf{X}_n . Hence, in exactly the same manner as in Algorithm 1.4, we can generate $X_n = x_n$ as a sample from π if we detect $X_n^* = \mathcal{S}$ in Algorithm 5.4. Moreover, it is possible to further modify Algorithm 5.4 so that we can exploit Step 2 of Algorithm 1.6.

Algorithm 5.5. Execute Algorithm 1.6 to generate paths $\mathbf{X}_n = (x_0, \dots, x_n)$ and $\check{\mathbf{Y}}_n = (y_n, \dots, y_0)$. Starting from $X_0^* = \{\hat{0}\}$, generate a set-valued path $\mathbf{X}_n^* = (x_0^*, \dots, x_n^*)$ recursively by (5.4) with random variable \hat{U}_i uniformly distributed on the interval

$$\begin{cases} [0, \tilde{P}(y_i, x_{i-1}^*)] & \text{if } y_{i-1} \in x_{i-1}^*; \\ [\tilde{P}(y_i, x_{i-1}^*), \tilde{P}(x_i, x_{i-1}^*)] & \text{if } y_{i-1} \notin x_{i-1}^*. \end{cases}$$

In Proposition 5.6 we claim that Algorithms 5.4 and 5.5 generate the same sample path.

Proposition 5.6. *In Algorithm 5.5, (a) \hat{U}_i is uniformly distributed on the interval $[0, \tilde{P}(x_i, x_{i-1}^*)]$ given $X_{i-1}^* = x_{i-1}^*$ and $X_i = x_i$ for $i = 1, \dots, n$, and (b) \mathbf{X}_n^* is a strong stationary dual for \mathbf{X}_n .*

Proof. We prove it by induction. Assume (a) for $i = 1, \dots, k - 1$ so that the sample path $\mathbf{X}_{k-1}^* = (x_0^*, \dots, x_{k-1}^*)$ can be alternatively generated by Algorithm 5.4. Then \mathbf{X}_{k-1}^* is a strong stationary dual for \mathbf{X}_{k-1} . In particular we obtain

$$\mathbb{P}(X_{k-1} \in \cdot \mid \mathbf{X}_{k-1}^* = x_{k-1}^*) = \Lambda(x_{k-1}^*, \cdot).$$

Given $X_{k-1}^* = x_{k-1}^*$, $X_k = x_k$ and $\tilde{Y}_{n-k} = y_k$, we can compute

$$\begin{aligned} & \mathbb{P}\left(\tilde{Y}_{n-k+1} \in x_{k-1}^* \mid X_{k-1}^* = x_{k-1}^*, X_k = x_k, \tilde{Y}_{n-k} = y_k\right) \\ &= \frac{\sum_w \Gamma(x_{k-1}^*, w) \sum_z k(w \mid y_k, x_k, z) \Lambda(x_{k-1}^*, z) P(z, x_k)}{\sum_z \Lambda(x_{k-1}^*, z) P(z, x_k)} = \frac{\tilde{P}(y_k, x_{k-1}^*)}{\tilde{P}(x_k, x_{k-1}^*)}. \end{aligned}$$

This immediately implies that \hat{U}_k is uniform on $[0, \tilde{P}(x_k, x_{k-1}^*)]$, and therefore, that \mathbf{X}_k^* is a strong stationary dual for \mathbf{X}_k . \square

When \mathbf{X}_n^* is generated by Algorithm 5.5, it has to be coupled with $\tilde{\mathbf{Y}}_n$ in such a way that $\tilde{Y}_{n-i} \in X_i^*$ holds simultaneously for $i = 0, \dots, n$. In particular, $\{\tilde{Y}_n = \hat{0}\}$ and $\{X_n^* = \mathcal{S}\}$ becomes the same event. This verifies the correctness of Algorithm 1.6.

References

- [1] Clifford, P. and Sudbury, A. (1985). A sample path proof of the duality for stochastically monotone Markov processes. *Ann. Probab.* **13** 558–565.
- [2] Cox, J. T. and Rösler, U. (1983). A duality relation for entrance and exit laws for Markov processes. *Stochastic Process. Appl.* **16** 141–156.
- [3] Daley, D. J. (1968). Stochastically monotone Markov chains. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **10** 305–317.
- [4] Diaconis, P. and Fill, J. A. (1990). Strong stationary times via a new form of duality. *Ann. Probab.* **18** 1483–1522.
- [5] Fill, J. A. (1998). An interruptible algorithm for perfect sampling via Markov chains. *Ann. Appl. Probab.* **8** 131–162.
- [6] Fill, J. A. and Machida, M. (2001). Stochastic monotonicity and realizable monotonicity. *Ann. Probab.*, to appear.
- [7] Fill, J. A., Machida, M., Murdoch, D. J. and Rosenthal, J. S. (2000). Extension of Fill’s perfect rejection sampling algorithm to general chains. *Random Structures and Algorithms* **17** 219–316.
- [8] Harris, T. E. (1976). On a class of set-valued Markov processes. *Ann. Probab.* **4** 175–194.
- [9] Iosifescu, M. (1980). *Finite Markov Processes and their Applications*. John Wiley & Sons, New York.

- [10] Kamae, T., Krengel, U., and O'Brien, G. L. (1977). Stochastic inequalities on partially ordered state spaces. *Ann. Probab.* **5** 899–912.
- [11] Karlin, S. and McGregor, J. (1957). The classification of birth and death processes. *Trans. Amer. Math. Soc.* **86** 366–400.
- [12] Lévy, P. (1948). *Processus Stochastiques et Mouvement Brownien*. Gauthier–Villars, Paris.
- [13] Liggett, T. (1985). *Interacting Particle Systems*. Springer–Verlag, New York.
- [14] Propp, J. G. and Wilson, D. B. (1996). Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Structures and Algorithms* **9** 223–252.
- [15] Siegmund, D. (1976). The equivalence of absorbing and reflecting barrier problems for stochastically monotone Markov processes. *Ann. Probab.* **4** 914–924.
- [16] Stanley, R. P. (1986). *Enumerative Combinatorics*, volume 1. Wadsworth & Brooks/Cole, Monterey, California.
- [17] Vervaat, W. (1988). Algebraic duality of Markov processes. *Mark Kac Seminar on Probability and Physics*, Syllabus 1985–1987. 61–69. CWI Syllabi, Amsterdam.
- [18] Wilson, D. B. (1998). Annotated bibliography of perfectly random sampling with Markov chains. In Aldous, D. and Propp, J., editors, *Microsurveys in Discrete Probability*, volume 41 of *DIMACS Series in Discrete Mathematics and Theoretical Computer Science* 209–220. American Mathematical Society. Latest updated version is posted at <http://dimacs.rutgers.edu/~dbwilson/exact/>.

DEPARTMENT OF MATHEMATICAL SCIENCES
 JOHNS HOPKINS UNIVERSITY
 BALTIMORE, MARYLAND 21218
 E-MAIL: jimfill@jhu.edu

DEPARTMENT OF MATHEMATICS
 TENNESSEE TECHNOLOGICAL UNIVERSITY
 COOKEVILLE, TENNESSEE 38505
 E-MAIL: mmachida@tntech.edu