

DUALITY AND INTERTWINING FOR DISCRETE MARKOV KERNELS: RELATIONS AND EXAMPLES

THIERRY HUILLET,* *Université de Cergy-Pontoise*

SERVET MARTINEZ,** *Universidad de Chile*

Abstract

We supply some relations that establish intertwining from duality and give a probabilistic interpretation. This is carried out in the context of discrete Markov chains, fixing up the background of previous relations established for monotone chains and their Siegmund duals. We revisit the duality for birth-and-death chains and the nonneutral Moran model, and we also explore the duality relations in an ultrametric-type dual that extends the Siegmund kernel. Finally, we discuss the sharp dual, following closely the Diaconis–Fill study.

Keywords: Duality; intertwining; Siegmund dual; generalized ultrametric matrix; birth-and-death chain; Moran model; sharp strong stationary time; sharp dual

2010 Mathematics Subject Classification: Primary 60J10; 60J70; 60J80

1. Introduction

Our work is devoted to the study of duality and intertwining relations between Markov chain kernels. These concepts can be established only as relations between matrices, as we introduce them in the next section, but their analysis has probabilistic consequences that are part of this study.

This study is devoted to the problems of forming a duality relationship between substochastic kernels. Indeed, once the duality is established, then several properties can be deduced when the starting chain is irreducible and positive recurrent: one of these properties is an intertwining relation. We make these properties explicit in Theorems 1 and 2. In Theorem 1 results are stated for countable matrices and in Theorem 2 the same results and objects are detailed for finite state spaces. Both theorems can be viewed as a generalization of Theorem 5.5 of [4], where some relations involving the Siegmund dual and the Doob transform with respect to the cumulative function of the stationary measure of a monotone stochastic kernel are stated. This generalization is possible because there is a pivotal object, the function φ , defined in Theorem 1, resulting from the application of the duality function to the stationary measure. In Theorem 2(i) and (iii), φ is described in probabilistic terms for the finite case. As we note in Corollary 1(ii), the cumulative function in [4] plays the role of φ for the Siegmund dual of a monotone process. We also prove in Theorem 1 an intertwining relation derived from the stochasticity of the link matrix Λ defined in Definition 2. This is a simple fact, but its importance is based on the rich

Received 6 July 2009; revision received 26 January 2011.

* Postal address: Laboratoire de Physique Théorique et Modélisation, CNRS UMR 8089, Université de Cergy-Pontoise, 2 Avenue Adolphe Chauvin, 95302 Cergy-Pontoise, France. Email address: thierry.huillet@u-cergy.fr

** Postal address: Departamento de Ingeniería Matemática and Centro Modelamiento Matemático (CNRS UMI 2807), Universidad de Chile, Casilla 170/3, Correo 3, Santiago, Chile. Email address: smartine@dim.uchile.cl

probabilistic constructions that have been made around intertwining, which can be found in [2], [4], [5], and [6].

The well-known case of the Siegmund kernel for monotone chains [16] and the duality results for this case are summarized in Corollary 1, and, for monotone birth-and-death chains, they are given in Corollary 2. In Proposition 5 we show that the nonneutral Moran model is monotone when its bias mechanism is nondecreasing.

We find additional examples where duality relations between substochastic kernels can be established. We introduce the generalized ultrametric kernel as a case containing the Siegmund kernel. In this general case, the conditions for the existence of a dual kernel are given in Proposition 6. This result summarizes our exploration of potential matrices as duality functions.

Section 5 follows closely the ideas on sharp stationary times and duals developed in [1], [4], and [5]. In Proposition 8 we give a proof of a sharpness result alluded to in Remark 2.39 of [4] and in Theorem 2.1 of [5]. There, we interpret and summarize the hypotheses stated in Remark 2.39 in the unique condition (37). In Proposition 9 we state the condition for sharpness in terms of the duality function. This applies to the intertwining of a monotone chain under the Siegmund dual; in this case both chains can start from the state 0.

In some examples it appears that the duality function has a constant column, for instance, the Siegmund and the Vandermonde kernels. Also, this condition appears in relation with sharp duals; see Sections 5 and 6. This is why we devote Proposition 2 to this case. In Proposition 3, we give a partial answer why some duality functions, namely in the Siegmund case, have constant rows.

Even if duality and intertwining can be set for Markov chains acting on general state spaces and/or with continuous time, we restrict ourselves to the discrete time and space in order to be able to present quickly our main results and avoid introducing additional, overburdened notation.

2. Duality and intertwining

2.1. Notation

Let I be a countable set. We denote by $\mathbf{1}$ the unit function defined on I . A nonnegative matrix $P = (P(x, y) : x, y \in I)$ is called a kernel on I . Sometimes we will emphasize the nonnegativity by saying a nonnegative kernel. A kernel is substochastic if $P\mathbf{1} \leq \mathbf{1}$, stochastic if the equality $P\mathbf{1} = \mathbf{1}$ holds, and strictly substochastic if it is substochastic and there exists some $x \in I$ such that $P\mathbf{1}(x) < 1$.

The kernel P is irreducible when, for any pair $x, y \in I$, there exists $n > 0$ such that $P^n(x, y) > 0$. A point $x_0 \in I$ is an absorbing point of the kernel P when $P(x_0, y) = \delta_{y, x_0}$ for all $y \in I$.

When P is a substochastic kernel, there exists a uniquely defined (in distribution) Markov chain $X = (X_n : n < \mathcal{T}^X)$ taking values in the countable set I , with lifetime \mathcal{T}^X and with transition kernel P . Let ∂ be a point not belonging to I ; we define $X_n = \partial$ for all $n \geq \mathcal{T}^X$.

The substochastic kernel P acts on the set of bounded real functions and in the set of nonnegative functions, that is, for all $x \in I$,

$$Pf(x) = \mathbb{E}(f(X_1) \mathbf{1}(\mathcal{T}^X > 1)),$$

and it generates the semigroup $(P^n : n \geq 1)$ that satisfies, for all $n \geq 1$ and $x \in I$,

$$P^n f(x) = \mathbb{E}(f(X_n) \mathbf{1}(\mathcal{T}^X > n)).$$

We will denote by $T_J^X = \inf\{n \geq 0: X_n \in J\}$ the hitting time of a subset $J \subseteq I$ by the chain X , where, as usual, we put $+\infty = \inf \emptyset$. When $J = \{a\}$ is a singleton, we write T_a^X instead of $T_{\{a\}}^X$.

In the sequel, we will introduce three kernels, denoted by P, \hat{P} , and \tilde{P} ; they will be defined on the countable sets I, \hat{I} , and \tilde{I} , respectively. When these kernels are substochastic, the associated Markov chains will be respectively denoted by X, \hat{X} , and \tilde{X} , and the lifetimes of these chains will be $\mathcal{T}, \hat{\mathcal{T}}$, and $\tilde{\mathcal{T}}$, respectively. To simplify the notation, for the Markov chains X, \hat{X} , and \tilde{X} , the hitting times are respectively denoted by $T_J = T_J^X, \hat{T}_J = T_J^{\hat{X}}$, and $\tilde{T}_J = T_J^{\tilde{X}}$ (when J is a subset of I, \hat{I} , and \tilde{I} , respectively).

2.2. Definitions

We recall the duality and the intertwining relations. As is usual, we denote the transpose of the matrix M by M^\top , that is, $M^\top(x, y) = M(y, x)$ for all $x, y \in I$.

Definition 1. Let P and \hat{P} be two kernels defined on the countable sets I and \hat{I} , and let $H = (H(x, y): x \in I, y \in \hat{I})$ be a matrix. Then \hat{P} is said to be an H -dual of P if it satisfies

$$H \hat{P}^\top = P H. \tag{1}$$

We call H a duality function between (P, \hat{P}) .

In this work we will consider only nonnegative duality functions H , so we assume that $H(x, y) \geq 0$ for all $x \in I$ and $y \in \hat{I}$. We note that some studies consider duality kernels taking negative values; see, for instance, [17]. However, we will restrict ourselves to nonnegative duality kernels mainly because the conditions for the existence of dual kernels are easier to work with and also for probabilistic interpretations.

Note that, for a kernel P , the H -dual \hat{P} exists when (1) holds and $\hat{P} \geq 0$. When $|I| = |\hat{I}|$ is finite and H is nonsingular, we obtain

$$\hat{P}^\top = H^{-1} P H,$$

and so \hat{P}^\top and P are similar matrices; thus, they share the same spectrum.

Duality is a symmetric notion between kernels, because if \hat{P} is an H -dual of P then P is a H^\top -dual of \hat{P} .

We will assume that no row and no column of H vanishes completely. On the other hand, we note that if H is a duality function between (P, \hat{P}) then, for all $c > 0$, cH is also a duality function between these matrices. Then, when it is necessary, we can always multiply all the coefficients of H by a strictly positive constant.

This notion of duality (1) coincides with that between Markov processes that can be found in [4], [11], and [14], among others. Indeed, let P and \hat{P} be substochastic, and let X and \hat{X} be Markov chains with kernels P and \hat{P} , respectively. Then, if \hat{P} is an H -dual of P , \hat{X} is an H -dual of X , which means that, for all $x \in I, y \in \hat{I}$, and $n \geq 0$,

$$E_x(H(X_n, y)) = E_y(H(x, \hat{X}_n)), \tag{2}$$

where we have extended H to $(I \cup \{\partial\}) \times (\hat{I} \cup \{\partial\})$ by setting $H(x, \partial) = H(\partial, y) = H(\partial, \partial) = 0$ for all $x \in I$ and $y \in \hat{I}$. Let us now introduce intertwining.

Definition 2. Let P and \tilde{P} be two kernels defined on the countable sets I and \tilde{I} , and let $\Lambda = (\Lambda(y, x): y \in \tilde{I}, x \in I)$ be a stochastic matrix. We say that \tilde{P} is a Λ -intertwining of P if it satisfies

$$\tilde{P} \Lambda = \Lambda P. \tag{3}$$

We call Λ a link between (P, \tilde{P}) .

When $|I| = |\tilde{I}|$ is finite and Λ is nonsingular, we obtain

$$\tilde{P} = \Lambda P \Lambda^{-1},$$

and so P and \tilde{P} are similar matrices; thus, they share the same spectrum.

Let P and \tilde{P} be substochastic, and denote by X and \tilde{X} the associated Markov chains. If \tilde{P} is a Λ -intertwining of P , we say that \tilde{X} is a Λ -intertwining of X . Obviously, the intertwining is not a symmetric relation because Λ^\top is not necessarily stochastic. But, when Λ is doubly stochastic, \tilde{P} is a Λ -intertwining of P implies that P is a Λ^\top -intertwining of \tilde{P} .

3. Relations between duality and intertwining

Let us first introduce some additional notation. By e_a we denote a column vector with a 1 in the a th entry and 0s elsewhere. When P is an irreducible positive recurrent stochastic kernel, we denote by $\pi = (\pi(x) : x \in I)$ its stationary distribution and we write it as a column vector. So $\pi^\top P = \pi^\top$, where π^\top is the transposed row vector of π .

Now we give a result on intertwining that we will often use.

Proposition 1. *Let P be an irreducible positive recurrent stochastic kernel, and let π be its stationary distribution. Assume that \tilde{P} is a kernel that is Λ -intertwined with P , i.e. $\tilde{P}\Lambda = \Lambda P$. If \tilde{a} is an absorbing state for \tilde{P} then*

$$\pi^\top = e_{\tilde{a}}^\top \Lambda. \tag{4}$$

Proof. Since the chain P is positive recurrent with stationary distribution π and Λ is stochastic, we obtain $\lim_{k \rightarrow \infty} (1/k) \sum_{n=0}^{k-1} (\Lambda P^n)(x, y) = \pi(y)$; in particular,

$$\lim_{k \rightarrow \infty} \frac{1}{k} \sum_{n=0}^{k-1} (\Lambda P^n)(\tilde{a}, y) = \pi(y). \tag{5}$$

On the other hand, from our assumption, we obtain $\tilde{P}^n(\tilde{a}, y) = \delta_{y, \tilde{a}}$ and then

$$(\tilde{P}^n \Lambda)(\tilde{a}, y) = \Lambda(\tilde{a}, y) \quad \text{for all } n \geq 0, y \in I. \tag{6}$$

From (3) we have $\tilde{P}^n \Lambda = \Lambda P^n$ for all $n \geq 1$, and so from (5) and (6) we deduce that $\Lambda(\tilde{a}, y) = \pi(y)$, which is equivalent to $(e_{\tilde{a}}^\top \Lambda)(y) = \pi(y)$. Then (4) is shown.

For a vector $\rho \in \mathbb{R}^I$, we will denote by D_ρ the diagonal matrix with diagonal entries $(D_\rho)(x, x) = \rho(x)$, $x \in I$.

Let P be an irreducible positive recurrent stochastic kernel with stationary distribution π . By irreducibility we have $\pi > 0$. Denote by \overleftarrow{P} the transition kernel of the reversed chain of X , so $\overleftarrow{P}(x, y) = \pi(x)^{-1} P(y, x) \pi(y)$ or, equivalently,

$$\overleftarrow{P}^\top = D_\pi P D_\pi^{-1}. \tag{7}$$

Then, \overleftarrow{P} is in duality with P via $H = D_\pi^{-1}$. Note that \overleftarrow{P} is also irreducible and positive recurrent with stationary distribution π and that $P^\top = D_\pi \overleftarrow{P} D_\pi^{-1}$, so we can exchange the roles of P and \overleftarrow{P} . In the reversible case, $\overleftarrow{P} = P$, relation (7) expresses a self-duality property.

Below we state two of our main results, Theorems 1 and 2.

Theorem 1. *Let P be an irreducible positive recurrent stochastic kernel, and let π be its stationary distribution. Assume that \hat{P} is a (nonnegative) kernel and that it is an H -dual of P , i.e. $H\hat{P}^\top = PH$, where H is such that none of its rows or columns vanishes completely. Then the following assertions hold.*

(i) $\hat{P}H^\top D_\pi = H^\top D_\pi \overleftarrow{P}$.

(ii) *The vector $\varphi := H^\top \pi$ is strictly positive and satisfies*

$$\hat{P}\varphi = \varphi.$$

(iii) $\tilde{P} = D_\varphi^{-1} \hat{P} D_\varphi$ is a stochastic kernel and the matrix $\Lambda := D_\varphi^{-1} H^\top D_\pi$ is stochastic. Moreover, \tilde{P} is a Λ -intertwining of \overleftarrow{P} . Hence, it holds that

$$\tilde{P}\Lambda = \Lambda \overleftarrow{P} \quad \text{satisfying} \quad \tilde{P}\mathbf{1} = \mathbf{1} = \Lambda\mathbf{1}.$$

Moreover, we have the duality relation

$$K\tilde{P}^\top = PK \quad \text{with} \quad K := HD_\varphi^{-1}.$$

Proof. From $H\hat{P}^\top = PH$ we find that

$$\hat{P}H^\top = H^\top D_\pi \overleftarrow{P} D_\pi^{-1}.$$

By multiplying the right-hand side by D_π we obtain (i). Part (iii) follows directly in the finite nonsingular case.

Since $D_\pi \mathbf{1} = \pi$, we obtain $\hat{P}H^\top \pi = H^\top D_\pi \overleftarrow{P} \mathbf{1}$. Since \overleftarrow{P} is stochastic, we obtain $\hat{P}H^\top \pi = H^\top D_\pi \mathbf{1} = H^\top \pi$. Let $\varphi = H^\top \pi$. Since $\pi > 0$ and at each row of H there exists a strictly positive element, then $\varphi > 0$. Thus, (ii) holds. Now define

$$\tilde{P} = D_\varphi^{-1} \hat{P} D_\varphi.$$

Using (i), we obtain

$$\tilde{P}D_\varphi^{-1}H^\top D_\pi = D_\varphi^{-1}\hat{P}H^\top D_\pi = D_\varphi^{-1}H^\top D_\pi \overleftarrow{P}.$$

Then the relation $\tilde{P}\Lambda = \Lambda \overleftarrow{P}$ holds in (iii); moreover,

$$\tilde{P}\mathbf{1} = D_\varphi^{-1}\hat{P}D_\varphi\mathbf{1} = D_\varphi^{-1}\hat{P}\varphi = D_\varphi^{-1}\varphi = \mathbf{1},$$

$$\Lambda\mathbf{1} = D_\varphi^{-1}H^\top D_\pi\mathbf{1} = D_\varphi^{-1}H^\top\pi = D_\varphi^{-1}\varphi = \mathbf{1}.$$

Then \tilde{P} and Λ are Markov kernels. Finally, from the equality

$$HD_\varphi^{-1}\tilde{P}^\top D_\varphi = H\hat{P}^\top = PH,$$

the relation $K\tilde{P}^\top = PK$ is straightforward. Hence, (iii) is satisfied.

In the next result, we examine the finite case under the same hypotheses as Theorem 1. For this purpose, it is useful to recall the structure of a nonirreducible substochastic finite kernel P . The set I can be partitioned into classes as

$$I = \bigcup_{l=1}^{\ell} I_l \quad \text{with } \ell > 1,$$

where, for all $l = 1, \dots, \ell$, the submatrix $P_{I_l \times I_l} := (P(x, y : x, y \in I_l))$ is irreducible, and the

order of the indices of the classes (I_1, \dots, I_ℓ) is chosen such that (see [7, Section 8.3]), for all $x \in I_l$ and $y \in I_{l^\top}$,

$$P(x, y) > 0 \implies l \leq l^\top.$$

This will be the standard order of the indices of the irreducible classes that we will use in this section—mainly in the statements contained in Theorem 2.

If P is stochastic then the last of these submatrices $P_{I_\ell \times I_\ell}$ is stochastic, that is, $P_{I_\ell \times I_\ell} \mathbf{1}_{I_\ell} = \mathbf{1}_{I_\ell}$, and there could be other stochastic submatrices. If P is strictly substochastic then none or some of these submatrices $P_{I_l \times I_l}$, $l = 1, \dots, \ell$, could be stochastic. We set

$$\text{St}(P) = \{I_l : P_{I_l \times I_l} \text{ is stochastic, } l \in \{1, \dots, \ell\}\}.$$

Then, when P is stochastic, $\text{St}(P) \neq \emptyset$, and if P is strictly substochastic then $\text{St}(P)$ could be empty or not. When $\text{St}(P) \neq \emptyset$, it could contain a unique class or not, and by a simple reordering of the indices of the classes we can always assume that it contains I_ℓ (this reordering is not needed when P is stochastic).

In the next result, we investigate some additional properties of duality, when, besides the hypotheses of Theorem 1, we assume that the spaces are finite. We recall that in the finite case the positive recurrence property of a stochastic kernel P follows from its irreducibility.

Theorem 2. *Let I and \hat{I} be finite sets. Let P be a finite irreducible stochastic kernel, let π be its stationary distribution, and assume that \hat{P} is a substochastic kernel which is an H -dual of P , i.e. $H\hat{P}^\top = PH$, where H is such that none of its rows or columns vanishes completely. Let $\varphi = H^\top \pi$. Then the following assertions hold.*

- (i) *If \hat{P} is substochastic and irreducible, then it is stochastic.*
- (ii) *When \hat{P} is stochastic and irreducible, $\varphi = c \mathbf{1}$ for some $c > 0$, and $\tilde{P} = \hat{P}$.*
- (iii) *If \hat{P} is nonirreducible then $\text{St}(\hat{P}) \neq \emptyset$ and there exist some constants $c_l > 0$ for $\hat{I}_l \in \text{St}(\hat{P})$ such that*

$$\begin{aligned} \varphi(x) &= \sum_{\hat{I}_l \in \text{St}(\hat{P})} c_l P_x(\text{there exists an } N \text{ such that, for all } n \geq N, \hat{X}_n \in \hat{I}_l) \\ &= \sum_{\hat{I}_l \in \text{St}(\hat{P})} c_l P_x(\hat{T}_{\hat{I}_l} < \hat{T}). \end{aligned} \tag{8}$$

- (iv) *If \hat{P} has a unique stochastic class \hat{I}_ℓ then*

$$\frac{\varphi(x)}{\varphi(y)} = P_x(\hat{T}_{\hat{I}_\ell} < \hat{T}) \text{ for any } y \in \hat{I}_\ell, \tag{9}$$

and the intertwined Markov chain \tilde{X} is given by the Doob transform

$$P_x(\tilde{X}_1 = y_1, \dots, \tilde{X}_k = y_k) = P_x(\hat{X}_1 = y_1, \dots, \hat{X}_k = y_k \mid \hat{T}_{\hat{I}_\ell} < \hat{T}). \tag{10}$$

- (v) *If \hat{a} is an absorbing state for \hat{P} then \hat{a} is an absorbing state for \tilde{P} and relation (4), i.e. $\pi^\top = e_{\hat{a}}^\top \Lambda$, is satisfied. Moreover, the sets of absorbing points in \hat{P} and \tilde{P} coincide.*
- (vi) *If $|I| = |\hat{I}|$ is finite and H is nonsingular, then $\hat{P} = H^\top D_\pi \overleftarrow{P} D_\pi^{-1} H^{-1^\top}$ and $\tilde{P} = \Lambda \overleftarrow{P} \Lambda^{-1}$. Hence, \hat{P} , \overleftarrow{P} , \tilde{P} , and P are similar matrices.*

Proof. Since I is finite, if \hat{P} is an irreducible strictly substochastic kernel then its spectral radius is necessarily strictly smaller than 1, which contradicts the equality $\hat{P}\varphi = \varphi$, because $\varphi > 0$. In the case when \hat{P} is stochastic and irreducible, the equation $\hat{P}\varphi = \varphi$, $\varphi > 0$, implies that $\varphi = c\mathbf{1}$ for some constant $c > 0$. So (i) and (ii) follow.

Now assume that the matrix \hat{P} is substochastic and nonirreducible. Let $\hat{I} = \bigcup_{l=1}^{\ell} \hat{I}_l$ be the partition in irreducible components $\hat{P}_{\hat{I}_l \times \hat{I}_l}$ such that $x \in \hat{I}_l$, $y \in \hat{I}_\tau$, and $\hat{P}(x, y) > 0$ implies that $l^\top \geq \tau$. The last submatrix $\hat{P}_{\hat{I}_\ell \times \hat{I}_\ell}$ satisfies

$$\hat{P}_{\hat{I}_\ell \times \hat{I}_\ell} \varphi|_{\hat{I}_\ell} = \varphi|_{\hat{I}_\ell},$$

where $\varphi|_{\hat{I}_\ell}$ denotes the restriction of φ to the set \hat{I}_ℓ . Then $\hat{P}_{\hat{I}_\ell \times \hat{I}_\ell}$ is an irreducible substochastic matrix whose Perron–Frobenius eigenvalue is 1, so we deduce that $\hat{P}_{\hat{I}_\ell \times \hat{I}_\ell}$ is stochastic and $\varphi|_{\hat{I}_\ell} = c_\ell \mathbf{1}_{\hat{I}_\ell}$ for some constant $c_\ell > 0$; thus, $\text{St}(\hat{P}) \neq \emptyset$. The same argument applies to all the irreducible stochastic classes ($\hat{I}_l \in \text{St}(\hat{P})$) and we obtain $\varphi|_{\hat{I}_l} = c_l \mathbf{1}_{\hat{I}_l}$ for some quantity $c_l > 0$.

Let $\hat{X} = (\hat{X}_t : t < \hat{\mathcal{T}})$ be the Markov chain with kernel \hat{P} . It is known that all the trajectories that are not killed are attracted by $\bigcup_{\hat{I}_l \in \text{St}(\hat{P})} \hat{I}_l$, that is,

$$P_x \left(\text{there exists an } N \text{ such that, for all } n \geq N, \hat{X}_n \in \bigcup_{\hat{I}_l \in \text{St}(\hat{P})} \hat{I}_l \mid \hat{\mathcal{T}} = \infty \right) = 1.$$

On the other hand the equality $\hat{P}\varphi = \varphi$ expresses that φ is a harmonic function for the chain \hat{X} . Hence, for all $n \geq 0$, it is satisfied that

$$\begin{aligned} \varphi(x) &= E_x(\varphi(\hat{X}_n), \hat{\mathcal{T}} > n) \\ &= \sum_{\hat{I}_l \in \text{St}(\hat{P})} E_x(\varphi(\hat{X}_n), \hat{\mathcal{T}} > n, \hat{T}_{\hat{I}_l} < \hat{\mathcal{T}}) \\ &\quad + E_x(\varphi(\hat{X}_n), \hat{\mathcal{T}} > n, \hat{\mathcal{T}} < \min\{\hat{T}_{\hat{I}_l} : \hat{I}_l \in \text{St}(\hat{P})\}). \end{aligned}$$

Then, by taking $n \rightarrow \infty$ in the above expression and since $\lim_{n \rightarrow \infty} P_x(\min\{\hat{T}_{\hat{I}_l} : \hat{I}_l \in \text{St}(\hat{P})\} > \hat{\mathcal{T}} > n) = 0$, we obtain relation (8), i.e.

$$\varphi(x) = \sum_{\hat{I}_l \in \text{St}(\hat{P})} c_l P_x(\hat{T}_{\hat{I}_l} < \hat{\mathcal{T}}).$$

Let us prove part (iv). Since there is a unique stochastic class, equality (9) follows straightforwardly. Then the transition probabilities of \tilde{P} are given by the Doob h -transform

$$\begin{aligned} \tilde{P}(x, y) &= P_x(\hat{T}_{\hat{I}_\ell} < \hat{\mathcal{T}})^{-1} \hat{P}(x, y) P_y(\hat{T}_{\hat{I}_\ell} < \hat{\mathcal{T}}) \\ &= P_x(\hat{X}_1 = y \mid \hat{T}_{\hat{I}_\ell} < \hat{\mathcal{T}}) \quad \text{for all } x, y \in \hat{I}. \end{aligned}$$

The Markov property gives the formula for every cylinder.

Finally, let us show part (v). Since the chain \tilde{P} is positive recurrent with stationary distribution π , it suffices to show that \hat{a} is an absorbing state for \tilde{P} . This follows straightforwardly from the equality $\tilde{P} = D_\varphi^{-1} \hat{P} D_\varphi$. Indeed, it implies that $\tilde{P}(\hat{a}, y) = \hat{P}(\hat{a}, y)\varphi(y)/\varphi(\hat{a}) = \delta_{y, \hat{a}}$. Also, this proves the equality of the set of absorbing points for both kernels \hat{P} and \tilde{P} .

Remark 1. We can exchange the roles of P and \overleftarrow{P} in the irreducible and positive recurrent case. Thus, in the hypotheses of Theorems 1 and 2 we can take \overleftarrow{P} instead of P , so \hat{P} is an H -dual of \overleftarrow{P} , i.e. $H\hat{P}^\top = \overleftarrow{P}H$, and in all the statements of both theorems we must change P to \overleftarrow{P} .

Remark 2. A probabilistic explanation of the result $\varphi := H^\top \pi > 0$ can be given when \hat{P} is substochastic and H is bounded. In this case the duality relation $H\hat{P}^\top = PH$ is equivalent to (2), i.e. for all $x \in I, y \in \hat{I}$, and $n \geq 0$,

$$E_x(H(X_n, y)) = E_y(H(x, \hat{X}_n)).$$

Since, by hypothesis, X is an irreducible and positive recurrent Markov chain, then φ appears as the following limit on the left-hand side:

$$\lim_{k \rightarrow \infty} \frac{1}{k} \sum_{n \leq k} E_x(H(X_n, y)) = \sum_{u \in I} \pi(u)H(u, y) = \varphi(y).$$

Remark 3. We have

$$\Lambda(x, y) = \frac{1}{\varphi(x)} H(y, x)\pi(y),$$

in particular, $\Lambda(x, y) = 0$ if and only if $H(y, x) = 0$.

Remark 4. The formulae in Theorem 1 state that Λ, \hat{P} , and \tilde{P} are invariant when H is multiplied by a strictly positive constant. Then, we can fit $c > 0$ and take cH in order to have $\varphi(x) = 1$ for all $x \in \hat{I}$ in Theorem 2. This is equivalent to setting $c_l = 1$ for some fixed stochastic class $\hat{I}_l \in \text{St}(\hat{P})$.

Remark 5. When the starting equality between stochastic kernels is the intertwining relation $\tilde{P}\Lambda = \Lambda\tilde{P}$, then we have the duality relation $H\tilde{P}^\top = PH$ with $H = D_\pi^{-1}\Lambda^\top$ and $\hat{P} = \tilde{P}$. In this case $\varphi = \mathbf{1}$.

We note the equality $\hat{I} = \tilde{I}$ of the sets, where the kernels \hat{P} and \tilde{P} are defined in Theorem 1.

In the next result we study what happens when the duality function H has a constant column, a case which appears as a condition in the study of sharp duals; see Proposition 9 of Section 5.

Proposition 2. Assume that H is nonsingular and has a constant column which is strictly positive, that is, there exists an $\hat{a} \in \hat{I}$ such that $He_{\hat{a}} = c\mathbf{1}$ for some $c > 0$. Then

- (i) \hat{a} is an absorbing state for \hat{P} (so $\{\hat{a}\}$ is a stochastic class);
- (ii) under the hypotheses of Theorem 2, $\pi^\top = e_{\hat{a}}^\top \Lambda$ holds and if \hat{P} is strictly substochastic and $\{\hat{a}\}$ is the unique stochastic class, then $\mathbb{P}_y(\hat{T}_{\hat{a}} < \hat{T}) = \varphi(y)/\varphi(\hat{a})$ and relation (10) is satisfied.

Proof. (i) From $He_{\hat{a}} = c\mathbf{1}$ we obtain

$$\begin{aligned} e_{\hat{a}}^\top \hat{P} &= e_{\hat{a}}^\top H^\top D_\pi \overleftarrow{P} D_\pi^{-1} H^{-1\top} \\ &= (He_{\hat{a}})^\top D_\pi \overleftarrow{P} D_\pi^{-1} H^{-1\top} \\ &= c\pi^\top \overleftarrow{P} D_\pi^{-1} H^{-1\top} \\ &= (H^{-1}c\mathbf{1})^\top \\ &= e_{\hat{a}}^\top. \end{aligned}$$

Then

$$\hat{P}(\hat{a}, y) = (e_{\hat{a}}^\top \hat{P})(y) = e_{\hat{a}}^\top(y) = \delta_{\hat{a},y}.$$

So \hat{a} is an absorbing state for \hat{P} .

(ii) From Theorem 2(v), \hat{a} is an absorbing state of \tilde{P} and $\pi^\top = e_{\hat{a}}^\top \Lambda$. The proof of the rest of this part follows straightforwardly.

When P does not satisfy positive recurrence, let us consider only the following special case where we explain why in some duality functions there appear constant rows (as is the case of the Siegmund kernel).

Proposition 3. *Let $x_0 \in I$ be an absorbing point of the kernel P , and let \hat{P} be a substochastic kernel that is an H -dual of P , i.e. $H\hat{P}^\top = PH$. Then $h(y) := H(x_0, y)$, $y \in \hat{I}$, is a nonnegative \hat{P} -harmonic function. When H is bounded and \hat{P} is a stochastic recurrent kernel, the x_0 row, $H(x_0, \cdot)$, is constant.*

Proof. It suffices to show that the function h is \hat{P} -harmonic. Since $P(x_0, z) = \delta_{z,x_0}$ for all $z \in I$, we obtain $(PH)(x_0, y) = H(x_0, y)$. Therefore, if \hat{P} satisfies the duality equality (1), we obtain $(H\hat{P}^\top)(x_0, y) = H(x_0, y) = h(y)$. Then

$$(\hat{P}h)(y) = \sum_{z \in \hat{I}} \hat{P}(y, z)H(x_0, z) = \sum_{z \in \hat{I}} H(x_0, z)\hat{P}^\top(z, y) = (H\hat{P}^\top)(x_0, y) = h(y),$$

and the result is shown.

4. Classes of dual matrices

We consider the finite set case. We assume that $I = \hat{I} = \tilde{I} = \{0, \dots, N\}$, so the kernels are nonnegative $I \times I$ matrices and, when they are substochastic, the associated Markov chains take values in I .

We will study some classes of nonnegative matrices H for which there exist substochastic kernels P and \hat{P} in duality relation (1). So, in these cases we would be able to apply the results established in Theorems 1 and 2, and in Proposition 2.

The Siegmund kernel is an example of a dual function that is a potential matrix. We will explore below other examples of potential matrices as dual functions, such as the generalized ultrametric matrices (see Subsection 4.5). Let us state an elementary weak condition on potential matrices H ensuring the existence of substochastic kernels that are H -duals. Let R be a strictly substochastic kernel with no stochastic classes. Then it has a well-defined finite potential

$$H = (\text{Id} - R)^{-1} = \sum_{n \geq 0} R^n \geq 0.$$

So $H^{-1} = \text{Id} - R$. (In particular, no column or row of H vanishes.) Let P be a substochastic kernel. Define

$$\hat{P}^\top = H^{-1}PH = (\text{Id} - R)P(\text{Id} - R)^{-1}.$$

Assume that the transpose matrix R^\top is substochastic. Then, $\hat{P}\mathbf{1} \geq 0$ and there exists a stochastic kernel P for which $\hat{P} \geq 0$ holds. Indeed, the constant stochastic kernel $P = (1/(N + 1))\mathbf{1}\mathbf{1}^\top$ fulfills the property.

4.1. The Siegmund kernel

A well-known case of a kernel H arising as a potential of a strict substochastic kernel R as above is the Siegmund kernel. Let $R(x, y) = \mathbf{1}(x + 1 = y)$, so it is strictly substochastic (because the N th row vanishes) and has no stochastic classes. Its transposed matrix $R^\top(x, y) = \mathbf{1}(x = y + 1)$ is also substochastic.

By direct computation we find that $H_S = (\text{Id} - R)^{-1}$ satisfies

$$H_S(x, y) = \mathbf{1}(x \leq y),$$

so it is the Siegmund kernel [16]. We have $H_S^{-1} = \text{Id} - R$. Then $H_S^{-1}(x, y) = \mathbf{1}(x = y) - \mathbf{1}(x + 1 = y)$.

The Siegmund kernel has been widely studied, a relevant reference is [16], where the equivalence between absorbing and reflecting barrier problems for stochastically monotone chains was proved. The main tools are equalities (18) and (19), below, which will be exploited in Subsection 4.3.

The Siegmund kernel was studied in detail in [4, Section 5]. Let us summarize some well-known relations. We have $(H_S \hat{P}^\top)(x, y) = \sum_{z \geq x} \hat{P}(y, z)$ and $(PH_S)(x, y) = \sum_{z \leq y} P(x, z)$. Then, the equation $H_S \hat{P}^\top = PH_S$ gives

$$\hat{P}(y, x) = \sum_{z \geq x} \hat{P}(y, z) - \sum_{z > x} \hat{P}(y, z) = \sum_{z \leq y} (P(x, z) - P(x + 1, z)). \tag{11}$$

In particular, $\hat{P} \geq 0$ requires the monotonicity of P : for all $y \in I$,

$$\sum_{z \leq y} P(x, z) \text{ decreases in } x \in I. \tag{12}$$

From $\hat{P}(N, x) = \sum_{z \leq N} (P(x, z) - P(x + 1, z))$, it follows that if P is stochastic then $\hat{P}(N, x) = \delta_{x,N}$, that is, N is an absorbing state for \hat{P} . Also, from (11) we obtain

$$\hat{P}(N - 1, N) = \sum_{z \leq N-1} P(N, z) = 1 - P(N, N). \tag{13}$$

From

$$\sum_{x \leq N} \hat{P}(y, x) = \sum_{z \leq y} P(0, z),$$

it follows that \hat{P} is substochastic and also $\sum_{x \leq N} \hat{P}(0, x) = P(0, 0)$. Then, when $P(0, 0) = 1$, the matrix \hat{P} is stochastic, and if $P(0, 0) < 1$ then $\sum_{x \leq N} \hat{P}(0, x) < 1$. This last case occurs for any irreducible stochastic kernel P with $N \geq 1$. Also, we find that $P(0, 0) + P(0, 1) = 1$ implies that \hat{P} does not lose mass through $\{1, \dots, N\}$. When the finite matrix P is irreducible, we can apply Theorem 2, and in this case

$$\varphi(x) = (H_S^\top \pi)(x) = \sum_{y \in I} \mathbf{1}(y \leq x) \pi(y) = \sum_{y \leq x} \pi(y) =: \pi^c(x),$$

where π^c is the cumulative distribution of π . The cumulative distribution is not constant because $\pi > 0$.

From (11), we deduce that the irreducibility of P implies that N is the unique absorbing state of \hat{P} . We obtain the following result, within which we assume that $N \geq 1$ in order to avoid the trivial case when $N = 0$ and $P(0, 0) = 1$.

Corollary 1. *Let H_S be the Siegmund kernel, and let P be a monotone finite irreducible stochastic kernel with stationary distribution π . Let $H_S \hat{P}^\top = PH_S$ with $\hat{P} \geq 0$. Then the following assertions hold.*

- (i) \hat{P} is a strictly substochastic kernel that loses mass through 0, and Theorem 2(ii) and (iii) are satisfied.
- (ii) $\varphi = \pi^c$ and the stochastic intertwining kernel Λ satisfies

$$\Lambda(x, y) = \mathbf{1}(x \geq y) \frac{\pi(y)}{\pi^c(x)}. \tag{14}$$

The intertwining matrix \tilde{P} of \overleftarrow{P} is given by

$$\tilde{P}(x, y) = \hat{P}(x, y) \frac{\pi^c(y)}{\pi^c(x)}, \quad x, y \in I.$$

- (iii) N is the unique absorbing state for \hat{P} , and Theorem 2(iv) and (v) are satisfied with $\hat{I}_\ell = \{N\}$ and $\hat{a} = N$. In particular, $\pi^\top = e_N^\top \Lambda$ holds.
- (iv) It holds that

$$\Lambda e_N = \pi(N) e_N. \tag{15}$$

Proof. The first three parts are direct consequences of Theorems 1 and 2, and the above relations. Finally, (15) is a direct computation from (14) and $\varphi(N) = 1$.

4.2. Duality for finite state space birth-and-death chains

Recall that $I = \hat{I} = \tilde{I} = \{0, \dots, N\}$. Let $X = (X_n : n \geq 0)$ be a discrete birth-and-death (BD) chain with transition Markov kernel $P = (P(x, y) : x, y \in I)$. Then $P(x, y) = 0$ if $|x - y| > 1$ and

$$P(x, x + 1) = p_x, \quad P(x, x - 1) = q_x, \quad P(x, x) = r_x, \quad x \in I,$$

with $q_x + r_x + p_x = 1$ for all $x \in I$ and boundary conditions $q_0 = p_N = 0$.

We always take

$$q_x, p_x > 0, \quad x \in \{1, \dots, N - 1\}.$$

We will consider the irreducible case, which in this case is equivalent to the conditions that

$$p_0 > 0, \quad q_N > 0. \tag{16}$$

(A unique exception will be given in Subsection 4.3, where we will explicitly assume that (16) is not satisfied.) The stationary distribution $\pi = (\pi(x) : x \in I)$ satisfies $\pi(y) = \pi(0) \prod_{z < y} p_z / (q_{z+1}) > 0$, $y \in \{1, \dots, N\}$, where $\pi(0)$ is such that $\sum_{y \in I} \pi(y) = 1$.

The matrix P satisfies $\pi(x)P(x, y) = \pi(y)P(y, x)$ for all x, y . So, $P = \overleftarrow{P}$, where $\overleftarrow{P} = D_\pi^{-1} P^\top D_\pi$ is the transition matrix of the time-reversed process and P has real eigenvalues.

The unique constraint in (11) to obtain $\hat{P} \geq 0$ is the condition $\hat{P}(x, x) \geq 0$. It reads

$$p_x + q_{x+1} \leq 1 \quad \text{for all } x \in \{0, \dots, N - 1\}. \tag{17}$$

This is the monotonicity condition (12) for BD chains. So, P is monotone when (17) is satisfied. In this case, the Siegmund dual \hat{P} exists and it is a BD kernel with

$$\hat{P}(x, x - 1) = p_x, \quad \hat{P}(x, x) = 1 - (p_x + q_{x+1}), \quad \hat{P}(x, x + 1) = q_{x+1}.$$

Note that $\hat{P}(0, 0) = 1 - (p_0 + q_1)$ and $\hat{P}(0, 1) = q_1$. Then the Markov chain \hat{X} loses mass through the state 0 if and only if $p_0 = 0$. On the other hand, $\hat{P}(N, N) = 1 - (p_N + q_{N+1}) = 1$ (because $p_N = q_{N+1} = 0$). From this analysis, Corollary 1, and the reversibility relation $P = \overleftarrow{P}$, we obtain (see also Subsection 4.3 of [4]) the following result.

Corollary 2. *Let H_S be the Siegmund kernel, and let P be a finite irreducible stochastic BD chain with monotone kernel P with transition parameters p_x and q_x . Let π be the stationary distribution of P . Then, the dual matrix \hat{P} defined by $\hat{P}^\top = H_S^{-1} P H_S$ is a strictly substochastic kernel that loses mass through the state 0. Moreover, the following assertions hold.*

- (i) $\phi = \pi^c$, and Theorem 2(ii) and (iii) are satisfied.
- (ii) N is an absorbing state of \hat{P} and $\{N\}$ is the unique stochastic class of \hat{P} , all the other states in I are transient, and Theorem 2(iv) is satisfied with $\hat{I}_\ell = \{N\}$.
- (iii) Let Λ be the stochastic kernel given by (14). Then, the Λ -intertwining matrix \tilde{P} of P is a BD stochastic matrix with entries given by

$$\begin{aligned} \tilde{P}(x, x - 1) &= p_x \frac{\pi^c(x - 1)}{\pi^c(x)}, & \tilde{P}(x, x) &= 1 - (p_x + q_{x+1}), \\ \tilde{P}(x, x + 1) &= q_{x+1} \frac{\pi^c(x + 1)}{\pi^c(x)}. \end{aligned}$$

4.3. Absorbing points for the BD kernels

Let us modify the BD kernel P by taking 0 as an absorbing state. That is, instead of the irreducibility conditions (16) we take $p_0 = 0$ and no restriction on q_N , it could be 0 or greater than 0. Assume that P is monotone, so (17) holds. Then the BD kernel \hat{P} is stochastic and N is the unique absorbing state for \hat{P} .

Let us describe what happens by exploiting the special form of the Siegmund dual. By evaluating (2) at $y = 0$ we obtain

$$P_x(X_n \leq 0) = P_0(x \leq \hat{X}_n), \tag{18}$$

and by evaluating (2) at $x = N$ we obtain

$$P_N(X_n > y) = P_y(\hat{X}_n < N). \tag{19}$$

Now there are two cases to consider.

Case (i). If $q_N > 0$ then 0 is the unique absorbing state for P . By (13) we obtain $\hat{P}(N - 1, N) = q_N > 0$, so N is an absorbing state that attracts all the trajectories of the chain, $P_x(\lim_{n \rightarrow \infty} X_n = N) = 1$ for $x \in I$.

Case (ii). If $q_N = 0$ then 0 and N are absorbing states for P . From (13) we obtain $\hat{P}(N - 1, N) = q_N = 0$, so N , besides being an absorbing state for \hat{P} , is an isolated state for \hat{P} (that is, $\hat{P}(y, N) = 0$ for all $y < N$). Therefore, it does not attract any of the trajectories starting from a state different from N . Hence, (19) is simply the equality 1=1 when $y < N$.

Let us summarize the picture for case (ii): P has 0 and N as absorbing states that attract all the trajectories of its associated Markov chain X , \hat{P} is stochastic, N is a \hat{P} -absorbing isolated state, and $\hat{P}|_{I \setminus \{N\} \times I \setminus \{N\}}$ is stochastic and irreducible. Let $\hat{\pi}_* = (\hat{\pi}_*(z) : z \in I \setminus \{N\})$ be the stationary distribution of the submatrix $\hat{P}|_{I \setminus \{N\} \times I \setminus \{N\}}$.

Let $\phi(x) = P_x(\lim_{n \rightarrow \infty} X_n = 0)$ be the absorption probability at 0 of the chain X starting from x . We have the following result.

Proposition 4. *If $p_0 = 0$ and $q_N = 0$, then 0 and N are absorbing states for P , and $\hat{P}|_{I \setminus \{N\} \times I \setminus \{N\}}$ is stochastic and has 0 as an absorbing point. Let $\phi(x) = \mathbb{P}_x(\lim_{n \rightarrow \infty} X_n = 0)$, and let $\hat{\pi}_* = (\hat{\pi}_*(z) : z \in I \setminus \{N\})$ be the stationary distribution of $\hat{P}|_{I \setminus \{N\} \times I \setminus \{N\}}$. Then*

$$\phi(x) = 1 - \frac{\eta(x)}{\eta(N)} = 1 - \hat{\pi}_*^c(x + 1),$$

where $\hat{\pi}_*^c$ is the cumulative distribution of $\hat{\pi}_*$ and $\eta(x) := \sum_{y=0}^{x-1} \prod_{z=1}^y q_z/p_z$ is the scale function of P .

Proof. The first equality follows from the fact that η is a martingale and $\eta(0) = 0$. For the second relation, we take $x < N$ and, letting $n \rightarrow \infty$ in (18) gives

$$\phi(x) = \sum_{z \geq x} \hat{\pi}_*(z) = 1 - \hat{\pi}_*^c(x + 1).$$

4.4. Conditions for monotonicity in the nonneutral Moran model

Let us introduce the 2-allele Moran model with bias mechanism p . Let $p : [0, 1] \rightarrow [0, 1]$ be continuous with $0 \leq p(0)$ and $p(1) \leq 1$. Define $q(u) := 1 - p(u)$. The Moran model is a BD Markov chain X characterized by the quadratic transition probabilities p_x, r_x , and $q_x, x \in I = \{0, \dots, N\}$, where

$$q_x = \frac{x}{N}q\left(\frac{x}{N}\right), \quad r_x = \frac{x}{N}p\left(\frac{x}{N}\right) + \left(1 - \frac{x}{N}\right)q\left(\frac{x}{N}\right), \quad p_x = \left(1 - \frac{x}{N}\right)p\left(\frac{x}{N}\right).$$

We assume that $p_0 = p(0) > 0, q_N = 1 - p(1) > 0$, and $p_y > 0, q_y > 0$ for all $y \in \{1, \dots, N\}$, so the BD chain is irreducible.

Proposition 5. *Assume that in the Moran model the bias p is nondecreasing. Then the BD chain is monotone, that is, condition (17), $p_x + q_{x+1} \leq 1$, is fulfilled (and so the Siegmund dual exists).*

Proof. First, since $p_N = q_{N+1} = 0$, we have nothing to check for $x = N$. Let us see what happens with $x = 0$. We need to guarantee that $1 - p_0 - q_1 \geq 0$, but this is true because $p(1/N) \geq p(0) \geq Np(0) - (N - 1)$. Let us consider the case $x \in \{1, \dots, N - 1\}$. We have the following relations, where in the first inequality we use the fact that p is nondecreasing:

$$\begin{aligned} p_x + q_{x+1} &= p\left(\frac{x}{N}\right) - \frac{x}{N}p\left(\frac{x}{N}\right) + \left(\frac{x+1}{N}\right) - \left(\frac{x+1}{N}\right)p\left(\frac{x+1}{N}\right) \\ &\leq p\left(\frac{x}{N}\right)\left(1 - \frac{x}{N} - \left(\frac{x+1}{N}\right)\right) + \left(\frac{x+1}{N}\right) \\ &= \frac{1}{N}\left(p\left(\frac{x}{N}\right)((N - 1 - 2x) + (x + 1))\right) \\ &\leq 1. \end{aligned} \tag{20}$$

Now, the last inequality in (20) is fulfilled because if $x = (N - 1)/2$, it reduces to $(x + 1)/N \leq 1$; if $x < (N - 1)/2$, it reduces to $p(x/N) \leq (N - x - 1)/(N - 1 - 2x)$, and this is satisfied because the right-hand side of this expression is greater than 1; if $(N - 1)/2 < x \leq N - 1$, it is satisfied because $N - 1 - x \geq 0$ and $N - 1 - 2x < 0$.

4.5. Generalized ultrametric case

We will examine another triangular matrix H which is a potential matrix, and which generalizes the Siegmund kernel. It belongs to the class of generalized ultrametric matrices (see [13] and [15]), a class that contains the ultrametric matrices introduced in [12]. We will search conditions on a kernel P in order that it has an H -dual.

Let C be a nonempty set strictly contained in $I = \{0, \dots, N\}$. Define $C' = I \setminus C$. We set $C(x) = C$ when $x \in C$ and $C(x) = C'$ otherwise. Take $\alpha, \beta \geq 0$, and set $\gamma(x) = \alpha$ if $x \in C$ and $\gamma(x) = \beta$ otherwise. Now, define the matrix $H_{\alpha,\beta}$ by

$$H_{\alpha,\beta}(x, y) = \mathbf{1}(x \leq y) + \gamma(x) \mathbf{1}(x \leq y) \mathbf{1}(C(x) = C(y)),$$

which is a clear generalization of the Siegmund dual because $H_{0,0} = H_S$. It is straightforward to check that $H_{\alpha,\beta}$ belongs to the class of potential matrices; indeed, $H_{\alpha,\beta} = (\text{Id} - R)^{-1}$ with

$$R(x, y) = \mathbf{1}(x = y) - \frac{1}{1 + \gamma(x)} \mathbf{1}(x = y) + \frac{1}{1 + \gamma(x)} \mathbf{1}(x + 1 = y).$$

As is easily checked, R is an irreducible strictly substochastic matrix that loses mass through the state N . Then $H_{\alpha,\beta}^{-1} = \text{Id} - R$, so

$$H_{\alpha,\beta}^{-1}(x, y) = \frac{1}{1 + \gamma(x)} \mathbf{1}(x = y) - \frac{1}{1 + \gamma(x)} \mathbf{1}(x + 1 = y).$$

In this case we are able to compute the inverse matrix $H_{\alpha,\beta}^{-1}$; the description of the inverse of any generalized ultrametric matrices can be found in [3]. We point out that R^\top is substochastic only when $\alpha \geq \beta$ and in this case it is an irreducible strictly substochastic matrix that loses mass through the state 0. In the rest of this subsection, we will set $H = H_{\alpha,\beta}$ to avoid overburdened notation. We have

$$\begin{aligned} (H\hat{P}^\top)(x, y) &= \sum_{z \geq x} H(x, z) \hat{P}(y, z) = \sum_{z \geq x} \hat{P}(y, z) + \gamma(x) \sum_{z \geq x, z \in C(x)} \hat{P}(y, z), \\ (PH)(x, y) &= \sum_{z \leq y} P(x, z) H(z, y) = \sum_{z \leq y} P(x, z) + \gamma(y) \sum_{z \leq y, z \in C(y)} P(x, z). \end{aligned}$$

By permuting I we can always assume that C is an interval, that is, $C = \{1, \dots, k\}$ for some $0 \leq k < N$, and so $C' = \{k + 1, \dots, N\}$. With this choice, each $x \notin \{k, N\}$ satisfies $C(x) = C(x + 1)$ and so $\gamma(x) = \gamma(x + 1)$.

Case (i): let $x \neq k$. From the above equalities we find that

$$(H\hat{P}^\top)(x, y) - (H\hat{P}^\top)(x + 1, y) = (1 + \gamma(x))\hat{P}(y, x).$$

(the case $x = N$ follows from $(H\hat{P}^\top)(N + 1, y) = 0$). Then the equality $H\hat{P}^\top = PH$ implies that

$$(1 + \gamma(x))\hat{P}(y, x) = \sum_{z \leq y} (P(x, z) - P(x + 1, z)) + \gamma(y) \sum_{z \leq y, z \in C(y)} (P(x, z) - P(x + 1, z)).$$

Subcase (i.1): let $x \neq k$ and $y \leq k$. In this case we have $\gamma(y) = \alpha$, and $z \leq y$ implies that $z \in C(y)$. So, we find that

$$(1 + \gamma(x))\hat{P}(y, x) = (1 + \alpha) \sum_{z \leq y} (P(x, z) - P(x + 1, z)).$$

Then, a necessary and sufficient condition for $\hat{P}(y, x) \geq 0$ is that

$$\sum_{z \leq y} P(x + 1, z) \leq \sum_{z \leq y} P(x, z), \tag{21}$$

and we obtain

$$\hat{P}(y, x) = \frac{1 + \alpha}{1 + \gamma(x)} \sum_{z \leq y} (P(x, z) - P(x + 1, z)). \tag{22}$$

Subcase (i.2): let $x \neq k$ and $y > k$. In this case we have $\gamma(y) = \beta$, and $C(z) = C(y)$ if and only if $z > k$. Then,

$$(1 + \gamma(x))\hat{P}(y, x) = \sum_{z \leq y} (P(x, z) - P(x + 1, z)) + \beta \sum_{k < z \leq y} (P(x, z) - P(x + 1, z)),$$

and so

$$\hat{P}(y, x) = \frac{1}{1 + \gamma(x)} \sum_{z \leq k} (P(x, z) - P(x + 1, z)) + \frac{1 + \beta}{1 + \gamma(x)} \sum_{k < z \leq y} (P(x, z) - P(x + 1, z)). \tag{23}$$

Then, a necessary and sufficient condition to ensure that $\hat{P}(y, x) \geq 0$ for $x \neq k$ is

$$\sum_{z \leq k} P(x + 1, z) + (1 + \beta) \sum_{k < z \leq y} P(x + 1, z) \leq \sum_{z \leq k} P(x, z) + (1 + \beta) \sum_{k < z \leq y} P(x, z). \tag{24}$$

We can summarize subcases (i.1) and (i.2) as, for all $x \neq k$,

$$\begin{aligned} \hat{P}(y, x) &= \frac{1 + \gamma(y)}{1 + \gamma(x)} \sum_{z \leq y} (P(x, z) - P(x + 1, z)) \\ &\quad - \mathbf{1}(y > k) \frac{\beta}{1 + \gamma(x)} \sum_{z \leq k} (P(x, z) - P(x + 1, z)). \end{aligned}$$

The necessary and sufficient condition to ensure that $\hat{P}(y, x) \geq 0$ for $x \neq k$ is constituted by (21) and (24).

Case (ii): assume that $x = k$. Recall that $\gamma(k) = \alpha$ and $\gamma(k + 1) = \beta$, so

$$\begin{aligned} (H\hat{P}^\top)(k, y) &= \sum_{z \geq k} \hat{P}(y, z) + \alpha \sum_{z \geq k, z \in C(k)} \hat{P}(y, z) = \sum_{z > k} \hat{P}(y, z) + (1 + \alpha)\hat{P}(y, k), \\ (H\hat{P}^\top)(k + 1, y) &= \sum_{z > k} \hat{P}(y, z) + \beta \sum_{z > k, z \in C(x)} \hat{P}(y, z) = (1 + \beta) \sum_{z > k} \hat{P}(y, z). \end{aligned}$$

Then, using $H\hat{P}^\top = PH$,

$$\begin{aligned} (1 + \alpha)\hat{P}(y, k) &= (H\hat{P}^\top)(k, y) - (H\hat{P}^\top)(k + 1, y) + \beta \sum_{z > k} \hat{P}(y, z) \\ &= (PH)(k, y) - \frac{1}{1 + \beta} (PH)(k + 1, y), \end{aligned}$$

and so

$$(1 + \alpha)\hat{P}(y, k) = \sum_{z \leq y} \left(P(k, z) - \frac{1}{1 + \beta} P(k + 1, z) \right) + \gamma(y) \sum_{z \leq y, z \in C(y)} \left(P(k, z) - \frac{1}{1 + \beta} P(k + 1, z) \right). \tag{25}$$

From (25) we deduce that

$$\hat{P}(y, k) = \sum_{z \leq y} \left(P(k, z) - \frac{1}{1 + \beta} P(k + 1, z) \right), \quad y \leq k, \tag{26}$$

and

$$\hat{P}(y, k) = \frac{1}{1 + \alpha} \sum_{z \leq k} \left(P(k, z) - \frac{1}{1 + \beta} P(k + 1, z) \right) + \frac{1 + \beta}{1 + \alpha} \sum_{k < z \leq y} \left(P(k, z) - \frac{1}{1 + \beta} P(k + 1, z) \right), \quad y > k. \tag{27}$$

Hence, (21) and (24) imply that $\hat{P}(y, k) \geq 0$ for all y , and so they are necessary and sufficient for $\hat{P} \geq 0$.

From (22) we find that, for all $y \leq k$,

$$\sum_{x < k} \hat{P}(y, x) = \sum_{z \leq y} (P(0, z) - P(k, z)), \quad \sum_{x > k} \hat{P}(y, x) = \frac{1 + \alpha}{1 + \beta} \sum_{z \leq y} P(k + 1, z).$$

So, using (26), we obtain, for all $y \leq k$,

$$\sum_{x \leq N} \hat{P}(y, x) = \sum_{z \leq y} P(0, z) + \frac{\alpha}{1 + \beta} \sum_{z \leq y} P(k + 1, z). \tag{28}$$

On the other hand, from (23) we obtain, for all $y > k$,

$$\begin{aligned} \sum_{x < k} \hat{P}(y, x) &= \frac{1}{1 + \alpha} \sum_{z \leq k} (P(0, z) - P(k, z)) + \frac{1 + \beta}{1 + \alpha} \sum_{k < z \leq y} (P(0, z) - P(k, z)), \\ \sum_{x > k} \hat{P}(y, x) &= \frac{1}{1 + \beta} \sum_{z \leq k} P(k + 1, z) + \sum_{k < z \leq y} P(k + 1, z). \end{aligned}$$

Using (27), we obtain, for all $y > k$,

$$\begin{aligned} \sum_{x \leq N} \hat{P}(y, x) &= \frac{1}{1 + \alpha} \sum_{z \leq k} P(0, z) + \frac{1 + \beta}{1 + \alpha} \sum_{k < z \leq y} P(0, z) + \frac{\alpha}{(1 + \alpha)(1 + \beta)} \sum_{z \leq k} P(k + 1, z) \\ &\quad + \frac{\alpha}{1 + \alpha} \sum_{k < z \leq y} P(k + 1, z). \end{aligned} \tag{29}$$

Proposition 6. *Let P be a stochastic kernel, and let \hat{P} be an $H_{\alpha,\beta}$ -dual of P . Then, a sufficient condition to ensure that $\hat{P} \geq 0$ is the following. There exists a $\delta \in (0, 1)$ such that, for all $x \in \{0, \dots, N\}$,*

$$\sum_{z \leq k} P(x, z) = \delta, \tag{30}$$

where, for all $y \leq k$,

$$\sum_{z \leq y} P(x, z) \text{ decreases in } x \in \{1, \dots, k\} \tag{31}$$

and, for all $y > k$,

$$\sum_{k < z \leq y} P(x, z) \text{ decreases in } x \in \{k + 1, \dots, N\}. \tag{32}$$

Moreover, under conditions (30), (31), and (32), \hat{P} is substochastic if and only if $\delta = (1 + \beta)/(1 + \alpha + \beta)$. In this case \hat{P} is conservative at sites k and N .

Proof. Relations (30), (31), and (32), are sufficient for $\hat{P} \geq 0$ because they imply conditions (21) and (24). Now set

$$L(y) = \sum_{x \leq N} \hat{P}(y, x).$$

From (28) we find that $\{L(y) : y \leq k\}$ attains its maximum at $y = k$ and, by using (30), this maximum becomes $L(k) = \delta + \alpha\delta/(1 + \beta)$. So, this last quantity must be at most 1 in order that \hat{P} is substochastic. On the other hand, from (29), it follows that $\{L(y) : y > k\}$ attains its maximum at $y = N$ and that this maximum is

$$L(N) = \frac{\delta}{1 + \alpha} + \frac{(1 + \beta)(1 - \delta)}{1 + \alpha} + \frac{\alpha\delta}{(1 + \alpha)(1 + \beta)} + \frac{\alpha(1 - \delta)}{1 + \alpha}.$$

By straightforward computations, it follows that

$$L(N) = \frac{1}{1 + \alpha} (1 + \alpha + \beta(1 - L(k))).$$

Then, by using $L(k) \leq 1$ we deduce that $L(N) \leq 1$ if and only if $L(k) = 1$, in which case $L(N) = 1$. This completes the proof.

If the ultrametric dual is seen as a perturbation of the Siegmund dual then there is a rigidity result for the BD chains.

Proposition 7. *Let P be the stochastic kernel of an irreducible BD chain on $I = \{0, \dots, N\}$. Assume that there exists a substochastic kernel \hat{P} that is an $H_{\alpha,\beta}$ -dual of P , i.e. $H_{\alpha,\beta} \hat{P}^\top = P H_{\alpha,\beta}$.*

Then we necessarily have $\beta = 0$ and the monotone property (17) holds. Moreover, if $k \geq 1$ then $\alpha = \beta = 0$ and $H_{\alpha,\beta} = H_{0,0} = H_S$ is the Siegmund dual.

If $k = 0$ then $\alpha \leq p_0/q_1$. If $\alpha = p_0/q_1$, the kernel \hat{P} is stochastic, and when $\alpha < p_0/q_1$, the kernel \hat{P} is substochastic and it only loses mass through $\{0\}$.

Proof. From (23) we have

$$\begin{aligned} \hat{P}(k + 2, k - 1) &= \frac{1}{1 + \alpha} \sum_{z \leq k} (P(k - 1, z) - P(k, z)) \\ &\quad + \frac{1 + \beta}{1 + \alpha} \sum_{k < z \leq k+2} (P(k - 1, z) - P(k, z)) \\ &= \frac{1}{1 + \alpha} (1 - (1 - P(k, k + 1))) - \frac{1 + \beta}{1 + \alpha} P(k, k + 1). \end{aligned}$$

So $\hat{P}(k + 2, k - 1) = -P(k, k + 1)(\beta/(1 + \alpha))$, and we must necessarily have $\beta = 0$. Since $\beta = 0$, it follows from relations (21) and (24) that the condition to ensure that $\hat{P} \geq 0$ is (17), that is, $p_x + q_{x+1} \leq 1$ for all $x \in \{0, \dots, N - 1\}$. On the other hand, if $k \geq 1$, from (28) we obtain, for $y = k$,

$$\sum_{x \leq N} \hat{P}(k, x) = \sum_{z \leq k} P(0, z) + \frac{\alpha}{1 + \beta} \sum_{z \leq k} P(k + 1, z) = 1 + \frac{\alpha}{1 + \beta} P(k + 1, k).$$

So, we must necessarily have $\alpha = 0$.

In the $k = 0$ case, it follows from relation (29) that $\sum_{y \in I} \hat{P}(x, y) = 1$ for all $y > 0$. The only case we must examine is (28) for $k = 0$, where the condition $\sum_{y \in I} \hat{P}(0, y) = (1 - p_0) + \alpha q_1 \leq 1$ implies that $\alpha \leq p_0/q_1$.

5. Strong stationary times

Let P be an irreducible positive recurrent stochastic kernel on the countable set I , and let $X = (X_n : n \geq 0)$ be a Markov chain with kernel P . Let π be the stationary probability measure of X . We denote by π_0 the initial distribution of X and by π_n the distribution of X_n , $\pi_n(\cdot) = P_{\pi_0}(X_n = \cdot)$. It satisfies $\pi_n^\top = \pi_0^\top P^n$.

A random time T is called a strong stationary time for X , if X_T has distribution π and it is independent of T ; see [1]. The separation discrepancy is defined by

$$\text{sep}(\pi_n, \pi) := \sup_{y \in I} \left[1 - \frac{\pi_n(y)}{\pi(y)} \right].$$

The separation discrepancy satisfies $\text{sep}(\pi_n, \pi) \geq \|\pi_n - \pi\|_{\text{TV}}$, where

$$\|\pi_n - \pi\|_{\text{TV}} = \frac{1}{2} \sum_{y \in I} |\pi_n(y) - \pi(y)|$$

is the total variation distance between π_n and π ; see [1] and [4]. In Proposition 2.10 of [1] it was proved that every stationary time T satisfies

$$\text{sep}(\pi_n, \pi) \leq P_{\pi_0}(T > n) \quad \text{for all } n \geq 0. \tag{33}$$

Based on this result, the strong stationary time T is called sharp when there is equality in (33), that is, when

$$\text{sep}(\pi_n, \pi) = P_{\pi_0}(T > n) \quad \text{for all } n \geq 0.$$

In Proposition 3.2 of [1] it was shown that a sharp strong stationary time always exists.

Let \tilde{P} be a stochastic kernel on the countable set \tilde{I} such that \tilde{P} is a Λ -intertwining of P , where Λ is a nonsingular stochastic kernel, so $\tilde{P}\Lambda = \Lambda P$. Let $\tilde{X} = (\tilde{X}_n : n \geq 0)$ be the Markov chain with kernel \tilde{P} .

Recall that when we are in the framework of Theorem 1, we have $\tilde{P}\Lambda = \Lambda \overleftarrow{P}$, so \tilde{P} is a Λ -intertwining of the reversal kernel \overleftarrow{P} . Hence, when the intertwining is constructed from a duality relation, \overleftarrow{P} and the reversed chain \overleftarrow{X} will play the role of P and X in the intertwining relation. In the reversible case $\overleftarrow{P} = P$ and the notation coincides, that is, $\overleftarrow{P} = P$ and we can take $\overleftarrow{X} = X$; this occurs, for instance, when P is the kernel of an irreducible BD chain.

The initial probability distributions of the chains X and \tilde{X} will be respectively π_0 and $\tilde{\pi}_0$, that is, $X_0 \stackrel{D}{\sim} \pi_0$ and $\tilde{X}_0 \stackrel{D}{\sim} \tilde{\pi}_0$. We assume that the initial distributions are linked, that is,

$$\pi_0^\top = \tilde{\pi}_0^\top \Lambda. \tag{34}$$

When this relation holds, we say that π_0^\top and π_0 form an admissible condition. Let π_n and $\tilde{\pi}_n$ be the distributions of X_n and \tilde{X}_n . By the intertwining relation $\tilde{P}^n \Lambda = \Lambda P^n$ for all $n \geq 1$, and the initial condition (34), we obtain

$$\pi_n^\top = \tilde{\pi}_n^\top \Lambda \quad \text{for all } n \geq 0.$$

5.1. The coupling

Consider the following stochastic kernel \bar{P} defined on $I \times \tilde{I}$, which was introduced in [4]:

$$\bar{P}((x, \tilde{x}), (y, \tilde{y})) = \frac{P(x, y)\tilde{P}(\tilde{x}, \tilde{y})\Lambda(\tilde{y}, y)}{(\Lambda P)(\tilde{x}, y)} \mathbf{1}((\Lambda P)(\tilde{x}, y) > 0).$$

Let $\bar{X} = (\bar{X}_n : n \geq 0)$ be the chain taking values in $I \times \tilde{I}$, evolving with the kernel \bar{P} and having as initial distribution the vector $(\pi_0, \tilde{\pi}_0)$, where $\pi_0^\top = \tilde{\pi}_0^\top \Lambda$. In [4] it was proved that \bar{X} is a coupling of the chains X and \tilde{X} . Then, in the sequel we will write X and \tilde{X} for the components of \bar{X} , so $\bar{X}_n = (X_n, \tilde{X}_n)$ for all $n \geq 0$. In the above construction it can be also checked that

$$\Lambda(\tilde{x}, x) = P(X_n = x \mid \tilde{X}_n = \tilde{x}) \quad \text{for all } n \geq 0. \tag{35}$$

(For this equality, see also [2].) In [4] this coupling was characterized as the unique coupling that satisfies (35) and three other properties on conditional independence. These properties imply that the coupling also satisfies

$$\Lambda(\tilde{x}_n, x_n) = P(X_n = x_n \mid \tilde{X}_0 = \tilde{x}_0, \dots, \tilde{X}_n = \tilde{x}_n) \quad \text{for all } n \geq 0.$$

The following concept was introduced in [4].

Definition 3. The Markov chain \tilde{X} will be called a strong stationary dual of the Markov chain X if \tilde{X} has an absorbing state $\tilde{\delta}$ that satisfies

$$\pi(x) = P(X_n = x \mid \tilde{X}_0 = \tilde{x}_0, \dots, \tilde{X}_{n-1} = \tilde{x}_{n-1}, \tilde{X}_n = \tilde{\delta}) \quad \text{for all } x \in I, n \geq 0, \tag{36}$$

and where $\tilde{x}_0, \dots, \tilde{x}_{n-1} \in \tilde{I}$ satisfy $P(\tilde{X}_0 = \tilde{x}_0, \dots, \tilde{X}_{n-1} = \tilde{x}_{n-1}, \tilde{X}_n = \tilde{\delta}) > 0$.

In Theorem 2.4 of [1] it was shown that when condition (36) holds, then the absorption time $\tilde{T}_{\tilde{\delta}}$ at $\{\tilde{\delta}\}$ is a strong stationary time for X . Moreover, Diaconis and Fill [4, Remark 2.8] give a specific dual process \tilde{X} having an absorbing state $\tilde{\delta}$ and whose absorption time $\tilde{T}_{\tilde{\delta}}$ is sharp.

Assume that $\tilde{\delta}$ is an absorbing state for \tilde{X} . From (4) we obtain $\pi^\top = e_{\tilde{\delta}}^\top \Lambda$. When the initial conditions are linked by relation (34), it follows that $\tilde{T}_{\tilde{\delta}}$ is a strong stationary time for X . Indeed, from $\Lambda(\tilde{\delta}, x) = P(X_n = x \mid \tilde{X}_0 = \tilde{x}_0, \dots, \tilde{X}_n = \tilde{\delta})$ and $\pi^\top = e_{\tilde{\delta}}^\top \Lambda$, it follows that $\pi(x) = P(X_n = x \mid \tilde{X}_0 = \tilde{x}_0, \dots, \tilde{X}_n = \tilde{\delta})$. Observe that, for the Siegmund dual and monotone kernels, the absorbing state is $\tilde{\delta} = N$.

5.2. Choice of the initial conditions

Let $\tilde{\delta}$ be an absorbing state of \tilde{X} . Assume (34) holds. Then we are able to perform the duality construction and find that the absorption time $\tilde{T}_{\tilde{\delta}}$ is a strong stationary time for X .

Assume that $I = \tilde{I}$. Since Λ is a stochastic matrix, it has a left probability eigenvector π_Λ^\top satisfying $\pi_\Lambda^\top = \pi_\Lambda^\top \Lambda$. So, we can choose $\tilde{X}_0 \stackrel{D}{\sim} X_0 \stackrel{D}{\sim} \pi_\Lambda$ because (34) holds. Then, when \tilde{X} is initially distributed as π_Λ , $\tilde{T}_{\tilde{\delta}}$ is a strong stationary time for the chain X starting from π_Λ . If Λ is nonirreducible then π_Λ could fail to be strictly positive. This is the case for the Siegmund kernel because, from (14), it can be checked that e_0 is the unique left eigenvector satisfying $e_0^\top = e_0^\top \Lambda$ and so $\pi_\Lambda = e_0$. Then, for the Λ -intertwining given by (14), the initial conditions $\tilde{X}_0 \stackrel{D}{\sim} \delta_0$ and $X_0 \stackrel{D}{\sim} \delta_0$ are admissible.

When $\tilde{b} \in \tilde{I}, b \in I, \tilde{b} \neq \tilde{\delta}$ satisfy $e_{\tilde{b}}^\top \Lambda = e_b^\top$, then $\tilde{X}_0 \stackrel{D}{\sim} \delta_{\tilde{b}}$ and $X_0 \stackrel{D}{\sim} \delta_b$ form an admissible initial condition (for which (34) holds). Then $\tilde{T}_{\tilde{\delta}}$ starting from \tilde{b} is a strong stationary time for X starting from b , and it is strictly positive. In this case, both \tilde{X}_0 and X_0 start at a single point. We observe that the condition $e_{\tilde{b}}^\top \Lambda = e_b^\top$ is equivalent to the following condition on the duality function: $He_{\tilde{b}} = ce_b$ for some $c > 0$. Indeed, if H satisfies this condition and since $\Lambda = D_\varphi H^\top D_\pi$ (see Theorem 1(iii)), we obtain $e_{\tilde{b}}^\top \Lambda = c'e_b^\top$ with $c' = c\pi(b)/\varphi(\tilde{b})$. Since Λ is stochastic, we obtain $c' = 1$, and so $c = \varphi(\tilde{b})/\pi(b)$. This gives $He_{\tilde{b}} = \varphi(\tilde{b})e_b/\pi(b)$, which is exactly $e_{\tilde{b}}^\top \Lambda = e_b^\top$.

For the Siegmund kernel and monotone P , Λ is given by (14). Condition (34) requires that $\pi_0(x)/\pi(x)$ decreases with $x \in I$ and in this case $\tilde{\pi}_0(x) = \pi^c(x)(\pi_0(x)/\pi(x) - \pi_0(x + 1)/\pi(x + 1))$. These statements are respectively condition (4.7) and Formula (4.10) of [4].

We recall that every monotone kernel P satisfies the condition that $\pi^\top = e_N^\top \Lambda$ (see (4)). The Λ -intertwining \tilde{P} is that of \tilde{P} , and in this case \tilde{X} and \tilde{X} denote the Markov chains associated to \tilde{P} and \tilde{P} , respectively.

5.3. Conditions for sharpness

We now give a proof of the sharpness result alluded to in Remark 2.39 of [4] and in Theorem 2.1 of [5]. The hypotheses stated in Remark 2.39 are understood as condition (37) below and we give a complete proof. Owing to the deep consequences of the result that \tilde{X} is a sharp dual, in Proposition 9 we also write it in terms of conditions on the duality function H .

Proposition 8. *Let X be an irreducible positive recurrent Markov chain, and let \tilde{X} be a Λ -intertwining of X having $\tilde{\delta}$ as an absorbing state. Assume that there exists $d \in I$ such that*

$$\Lambda e_d = \pi(d)e_{\tilde{\delta}}. \tag{37}$$

Then \tilde{X} is a sharp dual to X , that is, for $\tilde{X}_0 \stackrel{D}{\sim} \tilde{\pi}_0$ and $X_0 \stackrel{D}{\sim} \pi_0$ with $\pi_0^\top = \tilde{\pi}_0^\top \Lambda$, we have

$$\text{sep}(\pi_n, \pi) = P_{\tilde{\pi}_0}(\tilde{T}_{\tilde{\delta}} > n) \text{ for all } n \geq 0. \tag{38}$$

Proof. From the condition that $\Lambda e_d = \pi(d)e_{\tilde{\delta}}$ we obtain

$$\pi_n(d) = \pi_n^\top e_d = \tilde{\pi}_n^\top \Lambda e_d = \pi(d)\tilde{\pi}_n(\tilde{\delta}). \tag{39}$$

Since $\pi > 0$, the above equalities imply that

$$\pi_n(d) > 0 \iff \tilde{\pi}_n(\tilde{\partial}) > 0. \tag{40}$$

On the other hand, the condition that $\pi^\top = e_{\tilde{\partial}}^\top \Lambda$ means that the $\tilde{\partial}$ -row of Λ satisfies $\Lambda(\tilde{\partial}, \cdot) = \pi^\top(\cdot) > 0$. Then, if, for some n , we have $\tilde{\pi}_n(\tilde{\partial}) > 0$, from $\pi_n^\top = \tilde{\pi}_n^\top \Lambda$ we deduce that $\pi_n > 0$. Moreover,

$$\pi_n(x) = \sum_{\tilde{x} \in \tilde{I}} \tilde{\pi}(\tilde{x}) \Lambda(\tilde{x}, x) \geq \tilde{\pi}(\tilde{\partial}) \Lambda(\tilde{\partial}, x) = \tilde{\pi}(\tilde{\partial}) \pi(x).$$

Therefore, from (39) we obtain

$$\min_{x \in I} \frac{\pi_n(x)}{\pi(x)} = \tilde{\pi}(\tilde{\partial}) = \frac{\pi_n(d)}{\pi(d)}.$$

Then, $\text{sep}(\pi_n, \pi) = 1 - \tilde{\pi}(\tilde{\partial})$. Since $\tilde{\partial}$ is an absorption state implies that $\tilde{\pi}_n(\tilde{\partial}) = P_{\tilde{\pi}_0}(\tilde{T}_{\tilde{\partial}} \leq n)$, we obtain the desired relation:

$$\text{sep}(\pi_n, \pi) = P_{\tilde{\pi}_0}(\tilde{T}_{\tilde{\partial}} > n) \quad \text{for all } n \geq n_+, \text{ with } n_+ = \inf\{n \geq 0: \tilde{\pi}_n(\tilde{\partial}) > 0\}.$$

Let us show that relation (38) holds for $n < n_+$. First, we note that in this case $\tilde{\pi}_n(\tilde{\partial}) = 0$, which, by (40), implies that $\pi_n(d) = 0$. Then $\text{sep}(\pi_n, \pi) = 1$ and so the equality $\text{sep}(\pi_n, \pi) = P_{\tilde{\pi}_0}(\tilde{T}_{\tilde{\partial}} > n) = 1$ holds. We have proved that \tilde{X} is a sharp dual to X .

Proposition 9. (i) Assume that the hypotheses of Theorem 2 are satisfied and that \hat{P} is a substochastic kernel having \hat{a} as an absorbing state in \hat{P} . Then, if there exists some $d \in I$ such that

$$e_d^\top H = ce_{\hat{a}}^\top \quad \text{for some } c > 0 \tag{41}$$

then \hat{a} is an absorbing state for \tilde{X} and \tilde{X} is a sharp dual to \overleftarrow{X} . That is, when $\pi_0^\top = \tilde{\pi}_0^\top \Lambda$, relation (38) holds.

(ii) Assume that the hypotheses of Theorem 2 are satisfied and that \hat{P} is a substochastic kernel for which there exist $\hat{a} \in \hat{I}$ and $d \in I$ such that, for some constants $c' > 0$ and $c > 0$, we have

$$He_{\hat{a}} = c' \mathbf{1} \quad \text{and} \quad e_d^\top H = ce_{\hat{a}}^\top. \tag{42}$$

Then part (i) holds, and \tilde{X} is a sharp dual to \overleftarrow{X} .

Proof. (i) From Theorem 2(v), it follows that \hat{a} is an absorbing state for \tilde{P} . From Proposition 8, it suffices to show that d satisfies (37): $\Lambda e_d = \pi(d)e_{\hat{a}}$. Since the hypothesis is $H(d, y) = c\delta_{y, \hat{a}}$ for some $c > 0$ and all $y \in \tilde{I}$, Remark 3 implies that $\Lambda(x, d) = c''\delta_{x, \hat{a}}$ for some $c'' > 0$. Now, from Theorem 2(v) and (4), we have $\pi(d) = \Lambda(\hat{a}, d)$, and we deduce that $c'' = \pi(d)$. Therefore, $\Lambda(x, d) = \pi(d)\delta_{x, \hat{a}}$, which is equivalent to (37).

(ii) From Proposition 2, the first relation in (42) guarantees that \hat{a} is an absorbing state for \hat{P} . So, we are under the hypotheses of part (i) and the result follows.

Corollary 3. (i) For a monotone irreducible stochastic kernel P , the Λ -intertwining Markov chain \tilde{X} has N as an absorbing state and it is a sharp dual of \overleftarrow{X} . Moreover, both chains \overleftarrow{X} and \tilde{X} can start at the state 0.

(ii) For a monotone irreducible stochastic BD kernel P , the BD chain \tilde{X} is a sharp dual to X .

Proof. For part (i), the properties required for the sharpness for the Siegmund intertwining of BD chains follow straightforwardly because the N th row of H_S satisfies (41) with $d = N$. Also, relation (15) in Corollary 1 is exactly (37). The fact that the state 0 is admissible for both X and \tilde{X} is a consequence of $e_0^\top = e_0^\top \Lambda$. In part (ii) the only novelty is that, for BD chains, $P = \tilde{P}$.

We note that, by definition, for an absorbing point \hat{a} , there is a unique state d satisfying (41), as it occurs for the Siegmund kernel.

When d satisfies property (40), we say that d is a witness state in X that \tilde{X} hits $\tilde{\delta}$. It reflects the following more general situation. Assume that Λ fulfills the equivalence $\Lambda(x, y) > 0 \Leftrightarrow x \geq y$. Then, from $\pi^\top = \tilde{\pi}^\top \Lambda$ we obtain

$$\tilde{\pi}_0(x) > 0 \implies \pi_0(y) > 0 \text{ for all } y \leq x.$$

Then if N is an absorbing state of \tilde{P} , and $P(y, y + 1) > 0$ and $\tilde{P}(y, y + 1) > 0$ for all $y \in \{0, \dots, N - 1\}$, the equivalence $\pi_n(N) > 0 \Leftrightarrow \tilde{\pi}_n(N) > 0$ is satisfied, and so N will be a witness state in X that \tilde{X} hits the state N .

From Proposition 9, for the BD chains, the random time \tilde{T}_N starting from the state 0 gives information on the speed of convergence to its invariant measure of the original BD chain X . Let $\tilde{T}_{N;0}$ be a random variable distributed as the hitting time \tilde{T}_N when starting from 0. We denote its variance by $\text{var}(\tilde{T}_{N;0})$.

For BD chains absorbed at N , the probability generating function of $\tilde{T}_{N;0}$ can be written in terms of the eigenvalues of both \tilde{P} and P ; see [5] and [10]. Then, when these eigenvalues are known explicitly, it is possible to compute $E(\tilde{T}_{N;0})$ and $\text{var}(\tilde{T}_{N;0})$. But, in general, the whole set of eigenvalues are not known. Then, below we give an elementary computation of the mean and the variance of the absorption time $\tilde{T}_{N;0}$ for BD chains, in terms of the coefficients of the transition matrix \tilde{P} . Let us consider the sequence independent random variables:

$$(S_y : y = 0, \dots, N - 1) \text{ with law } P(S_y = n) = P_y(\tilde{T}_{y+1} = n) \text{ for all } n \geq 0.$$

Assume that the sequence $(S_y : y = 0, \dots, N - 1)$ is independent of the Markov chain \tilde{X} . We have $\sum_{y=0}^{N-1} S_y \stackrel{D}{=} \tilde{T}_{N;0}$. Using this classical observation, we easily obtain the mean absorption time as

$$E(\tilde{T}_{N;0}) = \sum_{y=0}^{N-1} E(S_y) = \sum_{y=0}^{N-1} \left(\sum_{z=0}^y \frac{1}{\tilde{p}_y} \prod_{r=z+1}^y \frac{\tilde{q}_r}{\tilde{p}_r} \right).$$

We also obtain the variance of the hitting time of N starting from 0 in the form

$$\text{var}(\tilde{T}_{N;0}) = \sum_{y=0}^{N-1} \text{var}(S_y) = \sum_{y=0}^{N-1} \left(\sum_{z=0}^y A_z \prod_{x=z+1}^y \frac{\tilde{q}_x}{\tilde{p}_x} \right),$$

where the coefficients A_z are given by

$$A_z = \frac{\tilde{p}_z - 1}{\tilde{p}_z^2} + \frac{2(1 - \tilde{p}_z)}{\tilde{p}_z} E(S_z) + \frac{2\tilde{q}_z(\tilde{p}_z - 1)}{\tilde{p}_z^2} E(S_{z-1}) + \frac{2\tilde{q}_z}{\tilde{p}_z} E(S_{z-1}) E(S_z) - \frac{\tilde{q}_z(\tilde{q}_z - \tilde{p}_z)}{\tilde{p}_z^2} E(S_{z-1})^2.$$

Remark 6. The above expressions of the mean and the variance are difficult to handle in terms of the transition probabilities of \tilde{P} , and often they cannot be used to describe the behavior of the mean and the variance when N grows large.

6. The hypergeometric dual

For $I = \{0, \dots, N\}$, let us suggest other potentially interesting examples of nonsingular duality functions H for which there exists a column of H which is constant so that Proposition 2 can be applied. For these examples, H^{-1} is known explicitly, which turns out to be useful to decide whether for a given irreducible stochastic kernel the H -dual defines a substochastic matrix. If this occurs, the problem of interpreting the intertwining chain given by Theorem 1 remains a challenging problem for each specific case.

The Vandermonde dual and the hypergeometric duals were first introduced in [14] in the context of neutral population genetics. In this context and also in nonneutral situations, the hypergeometric kernel plays a central role.

Vandermonde dual: $H(x, y) = (x/N)^y$. In this case the 0th column is constant.

Hypergeometric dual: $H(x, y) = \binom{N-x}{y} / \binom{N}{y}$. In this case $H = H^\top$, H is upper-left triangular, and (42) is fulfilled with $\hat{a} = 0$ and $d = N$, that is,

$$He_0 = \mathbf{1} \quad \text{and} \quad e_N^\top H = e_0^\top. \tag{43}$$

Let us comment on the hypergeometric dual H .

Assume that P is given by the reversible Moran model with nonneutrality bias mechanism p . Assume that $q = 1 - p$ is completely monotone, that is,

$$(-1)^k q^{(k)}(u) \geq 0 \quad \text{for all } k \geq 0 \text{ and } u \in (0, 1).$$

In particular, p is nondecreasing. Then the H -dual kernel \hat{P} exists and can be interpreted in terms of a multisex backward process. As shown in [9], for the Moran model with bias p satisfying $p(0) \in (0, 1)$, we have $\hat{P}\mathbf{1}(0) = 1$ and $0 < \hat{P}\mathbf{1}(x) = 1 - (x/N)p(0) < 1$ for all $x \neq 0$. From $p(0) \neq 0$, all the states but $a = 0$ of \hat{P} are mass-defective. The intertwining matrix \tilde{P} is the transition kernel of a skip-free-to-the-left BD chain that can be easily obtained from [9], and 0 is the unique absorbing state for \tilde{X} . Relation (43) fulfills the hypotheses of Proposition 9 with $d = N$; then in the above Moran model the sharpness property is satisfied.

On the other hand, the link matrix Λ is upper-left triangular, stochastic, and irreducible. Then there exists a probability vector π_Λ that satisfies $\pi_\Lambda^\top \Lambda = \pi_\Lambda^\top$, so $\pi_0 = \tilde{\pi}_0 = \pi_\Lambda$ is an admissible initial condition for X and \tilde{X} . Also, from $e_N^\top \Lambda = e_0^\top$ we obtain another admissible initial condition, $\pi_0 = \delta_0$ and $\tilde{\pi}_0 = \delta_N$. We can summarize this discussion in the following result.

Corollary 4. *Let X be the nonneutral Moran chain with transition matrix P fulfilling the above monotonicity conditions on $q = 1 - p$. Then, the construction of the intertwining kernel \tilde{P} in Theorem 1 starting from the hypergeometric dual H can be achieved and the Markov chain \tilde{X} is well defined. The absorbing state of \tilde{X} is 0, the process \tilde{X} is a sharp dual of X , and \tilde{X} can be started at N while X starts at 0.*

Hence, the time $\tilde{T}_{0;N}$ that \tilde{X} reaches 0 when it starts from N is the stochastically smallest time at which $X_{\tilde{T}_{0;N}} \stackrel{D}{\sim} \pi$ given $X_0 = 0$ and $\tilde{X}_0 = N$. We point out that the latter time $\tilde{T}_{0;N}$ that \tilde{X} needs to reach 0 when it starts from N is distributed like the time $\tilde{T}_{N;0}$ to reach N starting from 0 for the Siegmund intertwined BD chain to the same Moran model. This is in accordance with Theorem 1.2 of [6], which states that, for a skip-free-to-the-right Markov chain absorbed at N , the law of the time it takes to hit N starting from 0 is equal to the law

of $\tilde{T}_{N;0}$ for the BD chain. This result can be transferred to our skip-free-to-the-left BD chain case, while exchanging the boundaries $\{0, N\}$.

Let us finally consider the Wright–Fisher transition matrix P given by

$$P(x, y) = \binom{N}{y} p\left(\frac{x}{N}\right)^y \left(1 - p\left(\frac{x}{N}\right)\right)^{N-y},$$

whose bias $p(u)$ is such that $q = 1 - p$ is again a completely monotone function, satisfying $p(0) > 0$. This process is not reversible, nor is it in the BD class. However, using the hypergeometric duality function it was shown in [8] that the H -dual \hat{P} to P in (1) defines a substochastic matrix. From Theorem 1 we conclude that the corresponding \tilde{P} is Λ -linked to \tilde{P} .

Acknowledgements

The authors acknowledge partial support given by the CONICYT BASAL-CMM project. S. Martínez thanks the Guggenheim Fellowship and the hospitality of Laboratoire de Physique Théorique et Modélisation at the Université de Cergy-Pontoise. T. Huillet acknowledges support from the ANR Modélisation Aléatoire en Ecologie, Génétique et Evolution (ANR-Manège-09-BLAN-0215). We are indebted to an anonymous referee because his/her comments allowed us to correct several mistakes appearing in the first version, to focus our results, to push further some of our discussions, and to improve the whole presentation of this work.

References

- [1] ALDOUS, D. AND DIACONIS, P. (1987). Strong uniform times and finite random walks. *Adv. Appl. Math.* **8**, 69–97.
- [2] CARMONA, P., PETTIT, F. AND YOR, M. (1998). Beta-gamma random variables and intertwining relations between certain Markov processes. *Rev. Mat. Iberoamericana* **14**, 311–367.
- [3] DELLACHERIE, C., MARTÍNEZ, S. AND SAN MARTÍN, J. (2000). Description of the sub-Markov kernel associated to generalized ultrametric matrices. An algorithmic approach. *Linear Algebra Appl.* **318**, 1–21.
- [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. (2009). The passage time distribution for a birth-and-death chain: strong stationary duality gives a first stochastic proof. *J. Theoret. Probab.* **22**, 543–557.
- [6] FILL, J. A. (2009). On hitting times and fastest strong stationary times for skip-free and more general chains. *J. Theoret. Probab.* **22**, 587–600.
- [7] HORN, R. A. AND JOHNSON, C. R. (1985). *Matrix Analysis*. Cambridge University Press.
- [8] HUILLET, T. E. (2009). A duality approach to the genealogies of discrete non-neutral Wright–Fisher models. *J. Prob. Statist.* **2009**, 22pp.
- [9] HUILLET, T. E. AND MÖHLE, M. (2009). Duality and asymptotics for a class of nonneutral discrete Moran models. *J. Appl. Probab.* **46**, 866–893.
- [10] KEILSON, J. (1979). *Markov Chain Models—Rarity and Exponentiality* (Appl. Math. Sci. **28**). Springer, New York.
- [11] LIGGETT, T. M. (1985). *Interacting Particle Systems* (Fund. Principles Math. Sci. **276**). Springer, New York.
- [12] MARTÍNEZ, S., MICHON, G. AND SAN MARTÍN, J. (1994). Inverse of ultrametric matrices are of Stieltjes type. *SIAM J. Matrix Anal. Appl.* **15**, 98–106.
- [13] McDONALD, J. J., NEUMANN, M., SCHNEIDER, H. AND TSATSOMEROS, M. J. (1995). Inverse M -matrix inequalities and generalized ultrametric matrices. *Linear Algebra Appl.* **220**, 321–341.
- [14] MÖHLE, M. (1999). The concept of duality and applications to Markov processes arising in neutral population genetics models. *Bernoulli* **5**, 761–777.
- [15] NABBEN, R. AND VARGA, R. S. (1995). Generalized ultrametric matrices—a class of inverse M -matrices. *Linear Algebra Appl.* **220**, 365–390.
- [16] SIEGMUND, D. (1976). The equivalence of absorbing and reflecting barrier problems for stochastically monotone Markov processes. *Ann. Probab.* **4**, 914–924.
- [17] SUDBURY, A. AND LLOYD, P. (1995). Quantum operators in classical probability theory. II. The concept of duality in interacting particle systems. *Ann. Probab.* **23**, 1816–1830.