

Spectral methods for bivariate Markov processes with diffusion and discrete components and a variant of the Wright–Fisher model[☆]

Manuel D. de la Iglesia

Departamento de Análisis Matemático, Universidad de Sevilla, Apdo (P.O. Box) 1160, 41080 Sevilla, Spain

ARTICLE INFO

Article history:

Received 22 August 2011

Available online 17 April 2012

Submitted by Magda Peligrad

Keywords:

Bivariate Markov processes

Switching diffusions

Matrix-valued orthogonal functions

Wright–Fisher models

ABSTRACT

The aim of this paper is to study differential and spectral properties of the infinitesimal operator of two dimensional Markov processes with diffusion and discrete components. The infinitesimal operator is now a second-order differential operator with matrix-valued coefficients, from which we can derive backward and forward equations, a spectral representation of the probability density, study recurrence of the process and the corresponding invariant distribution. All these results are applied to an example coming from group representation theory which can be viewed as a variant of the Wright–Fisher model involving only mutation effects.

© 2012 Elsevier Inc. All rights reserved.

1. Introduction

The connection between stochastic processes and orthogonal polynomials is very well known. For a very good account, see [1]. Among them we can find diffusion processes where the state space is an interval of the real line, i.e. $\mathcal{S} \subset \mathbb{R}$, and the parameter set is $\mathcal{T} = [0, \infty)$, i.e. continuous time. In this case, the corresponding infinitesimal operator \mathcal{A} comes in terms of a second-order differential operator. It is possible to calculate an expression of the transition probability density by *spectral methods*, i.e. in terms of the eigenfunctions of the infinitesimal operator \mathcal{A} (see Section 8, Chapter V of [2] or Section 13, Chapter 15 of [3]), among other properties. In some cases, the invariant distribution can also be given in terms of the associated orthogonality measure of the eigenfunctions. Prominent examples connected with spectral methods are the Orstein–Uhlenbeck process, population growth models or Wright–Fisher models.

Continuous-time bivariate Markov processes with diffusion and discrete components are a natural extension, where now the state space is two dimensional, i.e. $\mathcal{S} = S \times \{1, 2, \dots, N\}$, where $S \subset \mathbb{R}$ and N is a positive integer. The first component is continuous, which we will call the *diffusion* component, while the second one is discrete and which we will call the *phase*. We will denote these bivariate Markov processes by (X_t, Y_t) , $t \geq 0$. At first it would seem that this hybrid process is just blending a regular scalar diffusion process and a continuous time Markov process, but there are more subtleties, especially in the case when both components are dependent on each other.

The ideas behind these coupled stochastic processes may be traced back to what is known in the literature as *random evolutions*. These were introduced by Griego and Hersh in 1969 [4], although the seminal work was given by Kac several years before (see [5] for an historical account). The main idea of a random evolution is a model for a dynamical system whose equation of state is subject to random variation. Other authors like Papanicolaou, Pinsky, Kurtz or Varadhan, to mention a few, were involved in this field during the 1960s and the 1970s. For a very good account see the Hersh survey [6] and the books [7,8]. In the 1980s, the Kiev school led by Korolyuk and Swishchuk put these processes in the framework of semi-Markov processes (see [9] for a recent book on this subject).

[☆] The work of the author is partially supported by MTM2009-12740-C03-02 (Ministerio de Economía y Competitividad), by FQM-262, P09-FQM-4643, FQM-7276 (Junta de Andalucía), by Feder Funds (European Union) and by Subprograma de estancias de movilidad posdoctoral en el extranjero, MICINN, ref. -2008-0207.

E-mail address: mdi29@us.es.

When the equation of state is governed by diffusion processes, we are in the case of bivariate Markov processes, the objects we will be interested in in this paper. Now the infinitesimal operator \mathcal{A} comes in terms of a second-order differential operator with matrix-valued coefficients. The basic definitions and properties of these bivariate Markov processes were introduced in [10] (see also Section 2 below). Here the author focused on the calculation of the first-passage time distribution out of the phases using the resolvent of the corresponding infinitesimal operator. This model was shown to be applicable to describing the progression of HIV in an infected individual. These processes have received other names in the literature, like *diffusions in random environments* or *switching diffusion models* (see for instance [11]).

It is not difficult to foresee the applications that these bivariate Markov processes may have, as in random evolutions or switching diffusion models, in fields like population dynamics, insurance or financial mathematics.

In this paper we will focus on differential and spectral properties of the infinitesimal operator generated by a bivariate Markov process with diffusion and discrete components. Surprisingly enough, this has not been explored in depth so far. The aim of this paper is to fill this gap using spectral analysis of differential operators with matrix-valued coefficients.

In the first part, after some preliminary considerations in Section 2 (already introduced in [10]), we will derive, in Section 3, the corresponding backward and forward equations, a representation of the (matrix-valued) probability density in terms of the eigenfunctions of the infinitesimal operator \mathcal{A} and some differential equations associated with some functionals, including the invariant distribution of these processes.

In the second part, in Section 4, we will study an example (coming from group representation theory [12,13]) of these bivariate Markov processes which can be viewed as a variant of the Wright–Fisher model involving only mutation effects. We will use the results in Sections 2 and 3 to study the probabilistic aspects of this example, especially the spectral representation of the probability density and an explicit expression of the invariant distribution.

2. Preliminaries

Let (X_t, Y_t) , $t \geq 0$ be a time-homogeneous Markov process assuming values in $S \times \{1, 2, \dots, N\}$, where $S \subset \mathbb{R}$ and N is a positive integer. We use preferably boldface capital letters to denote matrices, boldface lowercase letters to denote vectors and standard font for scalars. We define the transition probability density as a matrix-valued function $\mathbf{P}(t; x, A) = (P_{ij}(t; x, A))$, defined for every $t \geq 0$, $x \in S$, and real Borel set A , as a matrix whose entry (i, j) is given by

$$P_{ij}(t; x, A) = \Pr\{X_t \in A, Y_t = j | X_0 = x, Y_0 = i\}. \quad (2.1)$$

For simplicity, we will assume that every entry $P_{ij}(t; x, dy)$ has a density, also called $P_{ij}(t; x, y)$, in such a way that we can write

$$P_{ij}(t; x, A) = \int_A P_{ij}(t; x, y) dy.$$

Observe that $\mathbf{P}(t; x, S)$ is not a probability density in the common sense, but we have that all entries are positive and for every real Borel set $A \subset S$

$$\left(\int_A \mathbf{P}(t; x, y) dy \right) \mathbf{e}_N \leq \mathbf{e}_N,$$

where \mathbf{e}_N is N -dimensional column vector with all entries equal to 1. The equality holds when $A = S$.

It is assumed that $\mathbf{P}(t; x, y)$ satisfies the Chapman–Kolmogorov equation

$$\mathbf{P}(s + t; x, y) = \int_S \mathbf{P}(s; x, z) \mathbf{P}(t; z, y) dz. \quad (2.2)$$

In a similar way as the classical theory of diffusion processes we have to make the following assumptions:

1. For every $\epsilon > 0$

$$\lim_{h \rightarrow 0} \frac{1}{h} \mathbf{P}(h; x, \{y : |y - x| > \epsilon\}) = \mathbf{0}. \quad (2.3)$$

2. There exist $N \times N$ matrix-valued functions $\mathbf{A}(x)$ (with positive entries) and $\mathbf{B}(x)$, $x \in S$, such that for any $\epsilon > 0$

$$\lim_{h \rightarrow 0} \frac{1}{h} \left[\int_{\{y : |y - x| \leq \epsilon\}} (y - x) \mathbf{P}(h; x, y) dy \right] = \mathbf{B}(x) \quad (2.4)$$

and

$$\lim_{h \rightarrow 0} \frac{1}{h} \left[\int_{\{y : |y - x| \leq \epsilon\}} (y - x)^2 \mathbf{P}(h; x, y) dy \right] = \mathbf{A}(x). \quad (2.5)$$

3. There exists an $N \times N$ matrix-valued function $\mathbf{Q}(x) = (Q_{ij}(x))$ with Q_{ij} real-valued such that

$$\lim_{h \rightarrow 0} \frac{1}{h} [\mathbf{P}(h; x, S) - \mathbf{I}] = \mathbf{Q}(x). \quad (2.6)$$

Observe that (2.3) gives continuous sample paths for the process X_t while (2.4) and (2.5) are matrix-valued versions of the corresponding hypothesis employed in diffusion processes (see, for example, [2,3]). Note that (2.6) is a more general form of a corresponding hypothesis employed in continuous time Markov chains. The formulation here is different because \mathbf{Q} depends on x . In particular we have that, for fixed x ,

$$\begin{aligned} Q_{ij}(x) &\leq 0 \quad \text{for } i = j, \\ Q_{ij}(x) &\geq 0 \quad \text{for } i \neq j; \end{aligned}$$

and for each i

$$-Q_{ii}(x) = \sum_{i \neq j} Q_{ij}(x),$$

i.e., every off diagonal entry of $\mathbf{Q}(x)$ is a positive function and the sum of every row of $\mathbf{Q}(x)$ is 0, for every x , i.e., $\mathbf{Q}(x)\mathbf{e}_N = \mathbf{0}$.

We will denote by $\mathfrak{B}(S^N)$ the set of all column vector-valued functions where all components are real-valued, bounded and Borel measurable functions on S .

For any $\mathbf{f}(x) \in \mathfrak{B}(S^N)$ we define the transition operator as

$$\mathbf{T}_t \mathbf{f}(x) = \mathbf{E}[\mathbf{f}(X_t) | X_0 = x] \doteq \int_S \mathbf{P}(t; x, y) \mathbf{f}(y) dy. \quad (2.7)$$

The transition operator (2.7) can be applied to matrix-valued functions $\mathbf{F}(x) \in \mathfrak{B}(S^{N \times N})$, with similar properties as $\mathfrak{B}(S^N)$. We just have to perform the operator column by column. For simplicity we will use vector-valued analysis, but sometimes it will be more convenient to use a matrix-valued approach. Both are vector spaces with a norm over \mathbb{R} .

The family of transition operators $\{\mathbf{T}_t : t > 0\}$ has the semigroup property

$$\mathbf{T}_{s+t} = \mathbf{T}_s \mathbf{T}_t,$$

as a consequence of the Chapman–Kolmogorov equation (2.2). In particular, this implies that the transition operators commute. Therefore, the behavior of \mathbf{T}_t near $t = 0$ completely determines the semigroup, as in the scalar situation.

The infinitesimal generator \mathcal{A} of the process is defined for each smooth vector-valued function $\mathbf{f}(x)$ as

$$(\mathcal{A}\mathbf{f})(x) = \lim_{h \rightarrow 0} \frac{1}{h} [\mathbf{T}_h \mathbf{f}(x) - \mathbf{f}(x)] = \lim_{h \rightarrow 0} \frac{1}{h} \left[\int_S \mathbf{P}(h; x, y) \mathbf{f}(y) dy - \mathbf{f}(x) \right]. \quad (2.8)$$

The following two important results were proved in [10]:

Lemma 2.1 (Lemma 2.1 of [10]). $\mathbf{A}(x)$ and $\mathbf{B}(x)$, $x \in S$, are diagonal matrices.

Theorem 2.1 (Theorem 2.1 of [10]). For any vector-valued function $\mathbf{f}(x)$, $x \in S$, whose entries are real-valued, bounded and have continuous second derivatives, the infinitesimal operator \mathcal{A} defined in (2.8) is of the form

$$(\mathcal{A}\mathbf{f})(x) = \frac{1}{2} \mathbf{A}(x) \mathbf{f}''(x) + \mathbf{B}(x) \mathbf{f}'(x) + \mathbf{Q}(x) \mathbf{f}(x). \quad (2.9)$$

For simplicity, since the coefficients $\mathbf{A}(x)$ and $\mathbf{B}(x)$ are diagonal matrices, we will denote the corresponding diagonal entries by $\sigma_i^2(x)$ and $\tau_i(x)$, $i = 1, \dots, N$, respectively.

Now we will give a stochastic representation of these processes, already given in Section 3 of [10]. We refer to Fig. 1 for a better comprehension of how these processes work.

Suppose the process (X_t, Y_t) starts at state $X_0 = x$ and phase $Y_0 = i$. Let X_t^i be the diffusion process starting at the point x and evolving according to the infinitesimal generator

$$\frac{1}{2} \sigma_i^2(x) \frac{d^2}{dx^2} + \tau_i(x) \frac{d}{dx} + Q_{ii}(x) \frac{d^0}{dx^0}. \quad (2.10)$$

X_t^i is a diffusion process with diffusion and drift coefficients $\sigma_i^2(x)$ and $\tau_i(x)$, respectively, and a killing coefficient $-Q_{ii}(x) \geq 0$. Since the diffusion X_t^i has a lifetime t_i (chosen according to exponential holding time of $-Q_{ii}(X_t)$, see p. 314 of [3]), the process is defined on $[0, t_i]$ as $X_t = X_t^i$, $Y_t = i$ (if $t_i = \infty$, then $X_t = X_t^i$ and $Y_t = i$ for all $t > 0$). At time $t = t_i$, Y_t changes from phase i to another phase $j \neq i$. If $-Q_{ij}(X_{t_i}) > 0$, there is a probability distribution over the phases $j \neq i$, defined as $-Q_{ij}(X_{t_i})/Q_{ii}(X_{t_i})$, $j \neq i$, and Y_t is selected in accordance with this distribution.

At the moment t_i , the diffusion component X_t is changed from X_t^i to X_t^j , starting at $X_{t_i}^i$ and evolving according to the generator (2.10) with index j in the place of i . The process is killed at the time $t_i + t_j$, where t_j is the lifetime of X_t^j and the process is defined on $[t_i, t_i + t_j]$ as $X_t = X_t^j$, $Y_t = j$. If $-Q_{jk}(X_{t_i+t_j}) > 0$, there is a probability distribution over the phases $k \neq j$, defined as $-Q_{jk}(X_{t_i+t_j})/Q_{jj}(X_{t_i+t_j})$, $k \neq j$, and Y_t is selected in accordance with this distribution. Then the diffusion changes from X_t^j to X_t^k at time $t_i + t_j$, where X_t^k starts at $X_{t_i+t_j}^j$ and has generator (2.10) with k in the place of i , and so on.

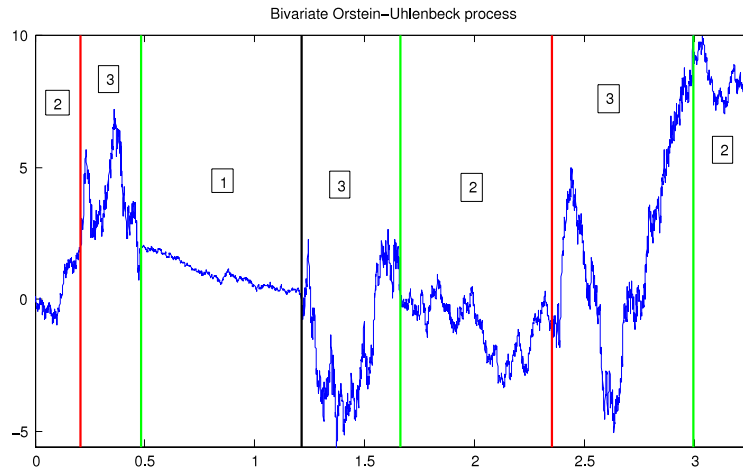


Fig. 1. Sample path of a bivariate Markov process with 3 phases.

Observe that the sample path of the process is defined as a piecewise function and is always *continuous* for $t > 0$. The only difference with respect to a regular diffusion process is that this process can move through N different phases with different diffusion and drift coefficients in general. The transitions through all the phases and the lifetime spent at each phase, which depend on the position X_t , are derived in terms of the coefficient $\mathbf{Q}(x)$.

As an illustration, in Fig. 1 we display a sample path of one example of this kind of processes. The example is conveniently chosen so that it points out the effect of the diffusion coefficients.

We take $N = 3$ phases and $S = \mathbb{R}$. The simulation of these processes are made according to the stochastic differential equation

$$dX_t = \tau_{Y_t}(X_t)dt + \sigma_{Y_t}(X_t)dB_t, \quad Y_t = 1, 2, 3,$$

where

$$\sigma_i^2(x) = i^2, \quad \tau_i(x) = -ix, \quad i = 1, 2, 3.$$

Here B_t denotes the standard Brownian motion.

The three phases evolve as Orstein-Uhlenbeck processes with different parameters. The diffusion coefficients grow with the number of phase Y_t and the drift coefficients are more intense with the number of phase Y_t . The transition of phases goes according to the matrix

$$\mathbf{Q}(x) = \begin{pmatrix} -\frac{2+x^2}{1+x^2} & \frac{5+3x^2}{6(1+x^2)} & \frac{7+3x^2}{6(1+x^2)} \\ 1+\frac{x^2}{400} & -2-\frac{x^2}{100} & 1+\frac{3x^2}{400} \\ \frac{1}{2}+\frac{1}{2}e^{-x^2} & \frac{1}{2}+\frac{1}{2}e^{-x^2} & -1-e^{-x^2} \end{pmatrix}.$$

In Fig. 1 we clearly see the evolution of the particle at each phase, starting at $x = 0$ and phase 2. We observe that the intensity of the process is higher with the number of phase Y_t .

3. Differential and spectral properties of \mathcal{A}

In this section we will show the main theoretical results of this paper. We will start deriving the corresponding matrix-valued backward and forward equations associated with these processes. Then we will give an explicit formula of the matrix-valued transition density $\mathbf{P}(t; x, y)$ defined in (2.1) in terms of the matrix-valued eigenfunctions of the infinitesimal generator \mathcal{A} defined in (2.9). Finally, we will study some differential equations associated with some functionals related to the recurrence and the invariant distribution of these processes.

3.1. The matrix-valued backward and forward equations

For all $\mathbf{f}(x) \in \mathcal{B}(S^N)$ that are twice continuous differentiable we have that the matrix-valued expectation defined in (2.7) satisfies the backward differential equation

$$\frac{\partial}{\partial t} \mathbf{T}_t \mathbf{f}(x) = \mathcal{A} \mathbf{T}_t \mathbf{f}(x), \quad (3.1)$$

with initial condition $T_0 \mathbf{f}(x) = \mathbf{f}(x)$, where \mathcal{A} is defined in (2.9). This is easy to see by definition of \mathcal{A} in (2.8) applied to $\mathbf{f} = T_t \mathbf{f}$, using the Chapman–Kolmogorov equation (2.2). The initial condition is a consequence of the assumption (2.6). We also have that T_t and \mathcal{A} commute on all functions in the domain of \mathcal{A} .

As a consequence, the transition density $\mathbf{P}(t; x, y)$ satisfies the *Kolmogorov backward equation*

$$\frac{\partial \mathbf{P}(t; x, y)}{\partial t} = \frac{1}{2} \mathbf{A}(x) \frac{\partial^2 \mathbf{P}(t; x, y)}{\partial x^2} + \mathbf{B}(x) \frac{\partial \mathbf{P}(t; x, y)}{\partial x} + \mathbf{Q}(x) \mathbf{P}(t; x, y), \quad (3.2)$$

applicable for $t > 0$ and $x, y \in S$.

The *matrix-valued forward, evolution or Fokker–Planck equation* for $\mathbf{P}(t; x, y)$ can be regarded as dual to (3.2) and the pertinent variables are t and y (the state variable at time t rather than the initial value x).

For any integrable vector-valued function \mathbf{g} , define the operator

$$T_t^* \mathbf{g}(y) = \int_S \mathbf{P}^*(t; x, y) \mathbf{g}(x) dx.$$

Observe that the integration of this operator is with respect to x , rather than y , as in (2.7). This operator is *adjoint* to T_t in the sense that

$$\begin{aligned} \langle T_t \mathbf{f}, \mathbf{g} \rangle &= \int_S \mathbf{g}^*(x) \left(\int_S \mathbf{P}(t; x, y) \mathbf{f}(y) dy \right) dx = \int_S \int_S \mathbf{g}^*(x) \mathbf{P}(t; x, y) \mathbf{f}(y) dx dy \\ &= \int_S \left(\int_S \mathbf{P}^*(t; x, y) \mathbf{g}(x) dx \right)^* \mathbf{f}(y) dy = \langle \mathbf{f}, T_t^* \mathbf{g} \rangle. \end{aligned} \quad (3.3)$$

Here $\langle \mathbf{f}, \mathbf{g} \rangle = \int_S \mathbf{g}^*(x) \mathbf{f}(x) dx$ is a real-valued inner product for any two integrable vector-valued functions. Observe that in case we are using matrix-valued functions (as we will do in Section 3.2) we have to use the same inner product $\langle \mathbf{F}, \mathbf{G} \rangle = \int_S \mathbf{G}^*(x) \mathbf{F}(x) dx$, but now this inner product is matrix-valued.

If \mathbf{f} is twice continuously differentiable and vanishes outside a compact subset of S , then one may differentiate with respect to t in (3.3) and interchange the orders of integration and differentiation to get, using (3.1),

$$\left\langle \mathbf{f}, \frac{\partial}{\partial t} T_t^* \mathbf{g} \right\rangle = \left\langle \frac{\partial}{\partial t} T_t \mathbf{f}, \mathbf{g} \right\rangle = \langle T_t \mathcal{A} \mathbf{f}, \mathbf{g} \rangle = \langle \mathcal{A} \mathbf{f}, T_t^* \mathbf{g} \rangle = \langle \mathbf{f}, \mathcal{A}^* T_t^* \mathbf{g} \rangle. \quad (3.4)$$

Here the formal adjoint \mathcal{A}^* of \mathcal{A} with respect to the inner product $\langle \cdot, \cdot \rangle$ is given by

$$(\mathcal{A}^* \mathbf{g})(y) = \frac{1}{2} \frac{\partial^2}{\partial y^2} (\mathbf{A}^*(y) \mathbf{g}(y)) - \frac{\partial}{\partial y} (\mathbf{B}^*(y) \mathbf{g}(y)) + \mathbf{Q}^*(y) \mathbf{g}(y). \quad (3.5)$$

From (3.4) we get the forward differential equation

$$\frac{\partial}{\partial t} T_t^* \mathbf{g}(y) = \mathcal{A}^* T_t^* \mathbf{g}(y),$$

or, equivalently, for the transition density $\mathbf{P}(t; x, y)$

$$\frac{\partial \mathbf{P}^*(t; x, y)}{\partial t} = \frac{1}{2} \frac{\partial^2}{\partial y^2} (\mathbf{A}^*(y) \mathbf{P}^*(t; x, y)) - \frac{\partial}{\partial y} (\mathbf{B}^*(y) \mathbf{P}^*(t; x, y)) + \mathbf{Q}^*(y) \mathbf{P}^*(t; x, y).$$

In other words, taking adjoint

$$\frac{\partial \mathbf{P}(t; x, y)}{\partial t} = \frac{1}{2} \frac{\partial^2}{\partial y^2} (\mathbf{P}(t; x, y) \mathbf{A}(y)) - \frac{\partial}{\partial y} (\mathbf{P}(t; x, y) \mathbf{B}(y)) + \mathbf{P}(t; x, y) \mathbf{Q}(y). \quad (3.6)$$

Observe that the forward differential equation (3.6) has the coefficients of the infinitesimal operator multiplied *on the right*, while the backward differential equation (3.2) *on the left*.

In principle we can try to solve explicitly the backward (or forward) differential equation with specific boundary values to obtain a general expression of the transition probability density $\mathbf{P}(t; x, y)$. Nevertheless, this task is not straightforward at all.

There is a very simple case, when $\mathbf{A}(x) = \sigma^2(x) \mathbf{I}$, $\mathbf{B}(x) = \tau(x) \mathbf{I}$ and $\mathbf{Q}(x) = \mathbf{Q}$, i.e., a constant matrix. In this case it is easy to see that

$$\mathbf{P}(t; x, y) = p(t; x, y) e^{t\mathbf{Q}},$$

where $p(t; x, y)$ is the transition probability density of the one dimensional diffusion process with drift $\tau(x)$ and diffusion coefficient $\sigma^2(x)$. In this case both the continuous and discrete components are *independent* and the transition of phases and the lifetime spent on them is independent of the position of X_t .

We can try to generalize this separation of variables in the same case as before, but with $\mathbf{Q}(x)$ depending on x . In this case substituting $\mathbf{P}(t; x, y) = p(t; x, y) e^{t\mathbf{Q}(x)}$ in (3.2) leads to a nonlinear second-order differential equation satisfied by $\mathbf{Q}(x)$, which is not straightforward to solve. Things get more complicated assuming $\mathbf{A}(x)$ and $\mathbf{B}(x)$ diagonal with different entries. Even when $\mathbf{Q}(x)$ is constant it is not easy to solve (3.2).

Therefore, we need different approaches of obtaining or approximating the transition probability density $\mathbf{P}(t; x, y)$. One of them, as we will see in the next section, is via the eigenfunctions of the infinitesimal operator \mathcal{A} .

3.2. The spectral representation of the matrix-valued transition density

As we have seen in the last section, the explicit computation of $\mathbf{P}(t; x, y)$ can be difficult to handle. In this section we show one way of obtaining a formula for $\mathbf{P}(t; x, y)$ in terms of the matrix-valued eigenfunctions of the infinitesimal operator \mathcal{A} .

Consider now the space $L^2_{\mathbf{W}}(S, \mathbb{C}^{N \times N})$ of all matrix-valued functions $\mathbf{F}(x)$ such that

$$\langle \mathbf{F}, \mathbf{F} \rangle_{\mathbf{W}} = \int_S \mathbf{F}^*(x) \mathbf{W}(x) \mathbf{F}(x) dx < \infty,$$

where $d\mathbf{W}$ is a *weight matrix* with a smooth density \mathbf{W} with respect to the Lebesgue measure (see [14] for a more concise definition of a weight matrix). In the above definition we mean that the integral is finite entry by entry. This induces a matrix-valued inner product for any two matrix-valued functions $\mathbf{F}, \mathbf{G} \in L^2_{\mathbf{W}}(S, \mathbb{C}^{N \times N})$, denoted by

$$\langle \mathbf{F}, \mathbf{G} \rangle_{\mathbf{W}} = \int_S \mathbf{G}^*(x) \mathbf{W}(x) \mathbf{F}(x) dx. \quad (3.7)$$

This is not an inner product in the common sense, but it has properties similar to the usual scalar inner products. It is also possible to define a scalar product between two matrix-valued functions (see [15]), given by

$$(\mathbf{F}, \mathbf{G}) = \text{Tr}(\langle \mathbf{F}, \mathbf{G} \rangle_{\mathbf{W}}).$$

Therefore, $L^2_{\mathbf{W}}(S, \mathbb{C}^{N \times N})$ with the norm $\|\mathbf{F}\|_{\mathbf{W}} = \text{Tr}(\langle \mathbf{F}, \mathbf{F} \rangle_{\mathbf{W}})^{1/2}$ is a Hilbert space and (3.7) is the inner product (in fact, the set of equivalence classes $\mathbf{F} \sim \mathbf{G}$ if $\|\mathbf{F} - \mathbf{G}\|_{\mathbf{W}} = 0$).

The idea behind this method is to find a representation of the transition operator $\mathbf{T}_t \mathbf{F}(x)$ (see (2.7)) using the method of separation of variables as a superposition (linear combination) involving eigenfunctions of the infinitesimal operator \mathcal{A} .

Suppose that we know a set of countable normalized matrix-valued eigenfunctions $(\Phi_n(x))_n$ of \mathcal{A} with corresponding eigenvalues Γ_n (Hermitian), i.e. $\mathcal{A} \Phi_n(x) = \Phi_n(x) \Gamma_n$. Normalization means $\langle \Phi_n, \Phi_m \rangle_{\mathbf{W}} = \delta_{nm} \mathbf{I}$ for certain weight matrix \mathbf{W} .

If the set of finite linear combinations of eigenfunctions is complete in $L^2_{\mathbf{W}}(S, \mathbb{C}^{N \times N})$, then each $\mathbf{F} \in L^2_{\mathbf{W}}(S, \mathbb{C}^{N \times N})$ has a Fourier expansion of the form

$$\mathbf{F}(x) = \sum_{n=0}^{\infty} \Phi_n(x) \langle \mathbf{F}, \Phi_n \rangle_{\mathbf{W}}.$$

Consider the superposition defined by

$$\mathbf{U}_{\mathbf{F}}(t, x) = \sum_{n=0}^{\infty} \Phi_n(x) e^{t\Gamma_n} \langle \mathbf{F}, \Phi_n \rangle_{\mathbf{W}}.$$

Here the exponential for matrix-valued argument is defined using the usual Taylor expansion. It is easy to see that $\mathbf{U}_{\mathbf{F}}(t, x)$ satisfies the backward differential equation

$$\frac{\partial}{\partial t} \mathbf{U}_{\mathbf{F}}(t, x) = \mathcal{A} \mathbf{U}_{\mathbf{F}}(t, x),$$

with initial condition $\mathbf{U}_{\mathbf{F}}(0, x) = \mathbf{F}(x)$. So if there is uniqueness for a sufficiently large class of initial functions \mathbf{F} then we get

$$\mathbf{T}_t \mathbf{F}(x) = \mathbf{U}_{\mathbf{F}}(t, x),$$

which means

$$\int_S \mathbf{P}(t; x, y) \mathbf{F}(y) dy = \int_S \left(\sum_{n=0}^{\infty} \Phi_n(x) e^{t\Gamma_n} \Phi_n^*(y) \mathbf{W}(y) \right) \mathbf{F}(y) dy.$$

Therefore, $\mathbf{P}(t; x, y)$ is given by

$$\mathbf{P}(t; x, y) = \sum_{n=0}^{\infty} \Phi_n(x) e^{t\Gamma_n} \Phi_n^*(y) \mathbf{W}(y). \quad (3.8)$$

Remark 3.1. It is well known that $\Psi_n(x) = \Phi_n(x) \mathbf{U}_n$ with $\mathbf{U}_n^* \mathbf{U}_n = \mathbf{I}$ is always another family of normalized eigenfunctions with eigenvalues $\mathbf{U}_n^* \Gamma_n \mathbf{U}_n$. However, it is straightforward to see that the spectral representation (3.8) of $\mathbf{P}(t; x, y)$ is invariant under this change.

Observe that the condition of having eigenfunctions with Hermitian eigenvalues forces the infinitesimal operator \mathcal{A} to be *self-adjoint* with respect to the inner product (3.7), i.e. $\langle \mathcal{A} \mathbf{F}, \mathbf{G} \rangle_{\mathbf{W}} = \langle \mathbf{F}, \mathcal{A} \mathbf{G} \rangle_{\mathbf{W}}$ for all $\mathbf{F}, \mathbf{G} \in L^2_{\mathbf{W}}(S, \mathbb{C}^{N \times N})$. This means, using the formal definition of the adjoint of \mathcal{A} given in (3.5), that the coefficients of the infinitesimal operator \mathcal{A} are subject

to certain symmetry equations with respect to the weight matrix \mathbf{W} (with the corresponding boundary conditions)

$$\begin{aligned}\mathbf{A}^*(x)\mathbf{W}(x) &= \mathbf{W}(x)\mathbf{A}(x), \\ \mathbf{B}^*(x)\mathbf{W}(x) &= (\mathbf{W}(x)\mathbf{A}(x))' - \mathbf{W}(x)\mathbf{B}(x), \\ \mathbf{Q}^*(x)\mathbf{W}(x) &= \frac{1}{2}(\mathbf{W}(x)\mathbf{A}(x))'' - (\mathbf{W}(x)\mathbf{B}(x))' + \mathbf{W}(x)\mathbf{Q}(x).\end{aligned}\tag{3.9}$$

These symmetry equations were already derived in [14,16]. The goal in those papers was to find matrix-valued orthogonal polynomials satisfying second-order differential equations of Sturm–Liouville type (i.e. the differential coefficients are matrix polynomials of degree less than or equal to the order of differentiation). Observe that our infinitesimal operator \mathcal{A} does not have to be in principle of Sturm–Liouville type. In fact, we are more interested in examples with $\mathbf{Q}(x)$ depending on x , along with diagonal coefficients $\mathbf{A}(x)$ and $\mathbf{B}(x)$. In this line, we are closer to differential operators as the ones introduced in [12]. Later it was shown (see [17]) that the operators introduced in [12] are intimately related to second-order differential equations of Sturm–Liouville type.

3.3. Differential equations associated with some functionals

In this section we will study some differential equations associated with some functionals related to the recurrence and the invariant distribution of these processes.

The *hitting time* of a one dimensional diffusion process $\{X_t, 0 \leq t < \zeta\}$ to the state z is defined by $T_z = \inf\{t \geq 0 : X_t = z\}$ or $T_z = \infty$ if $X_t \neq z$ for $0 \leq t < \zeta$. We use the notation

$$T^* = \min\{T_c, T_d\},$$

for the hitting time to c or d , the first time $X_t = c$ or $X_t = d$, where $c, d \in S = (a, b)$. For processes starting at $X_0 = x \in (c, d)$, this is the same as the *exit time* of the interval (c, d) . We have to accommodate this definition with our bivariate Markov processes. The solution will be introducing matrix-valued analysis.

We will assume that our process is *regular* in the interior of S , i.e.

$$\Pr\{T_y < \infty, Y_{T_y} = j | X_0 = x, Y_0 = i\} > 0,$$

for any x, y in the interior of S and any phases i, j .

Consider the matrix-valued function $\mathbf{U}(x) = (U_{ij}(x))$, where every entry (i, j) is given by

$$U_{ij}(x) = \Pr\{T_d < T_c, Y_{T_d} = j | X_0 = x, Y_0 = i\}, \quad c < x < d,\tag{3.10}$$

i.e., starting at $x \in (c, d)$ and phase i , $U_{ij}(x)$ is the probability that the process reaches d before c and at that time the process is in phase j .

Theorem 3.1. *The matrix-valued function $\mathbf{U}(x)$ defined in (3.10) possesses two bounded derivatives for $c < x < d$ and satisfies the following differential equation*

$$\frac{1}{2}\mathbf{A}(x)\mathbf{U}''(x) + \mathbf{B}(x)\mathbf{U}'(x) + \mathbf{Q}(x)\mathbf{U}(x) = \mathbf{0}, \quad c < x < d, \quad \mathbf{U}(c) = \mathbf{0}, \quad \mathbf{U}(d) = \mathbf{I}.\tag{3.11}$$

Proof. The proof is essentially the same as in the scalar case. For a heuristic justification, see p. 193 of [3]. For more technical details the reader should consult Proposition 9.1 of Chapter V in [2]. \square

Remark 3.2. In [2,3] we can find an explicit solution of this problem in the scalar situation (where $\mathbf{Q}(x) = \mathbf{0}$) in terms of what they call the *scale function* and *speed density* of the process. The presence of $\mathbf{Q}(x)$, not zero in general, leads to a more complicate system of second-order differential equations with initial conditions

$$\frac{1}{2}\sigma_i^2(x)U_{ij}''(x) + \tau_i(x)U_{ij}'(x) + Q_{ii}(x)U_{ij}(x) + \sum_{k \neq i}^N Q_{ik}(x)U_{kj}(x) = 0, \quad i, j = 1, \dots, N.\tag{3.12}$$

With the above considerations we can define *recurrence* and *transience* of bivariate Markov processes (in the same line as Definition 9.1 of Chapter V in [2]). Write the matrix-valued function $\mathbf{R}_{x,y} = (R_{x,y})_{i,j}$ where every entry (i, j) is given by

$$(R_{x,y})_{i,j} = \Pr\{X_t = y, Y_t = j | X_0 = x, Y_0 = i\}, \quad x, y \in S,$$

i.e., the probability that starting at x and phase i , the process ever reaches y and phase j .

For $y > x$, $\mathbf{R}_{x,y}$ will be the solution of the differential equation (3.11) letting $c \rightarrow a$ and $d \rightarrow y$ and, for $y < x$, $\mathbf{R}_{x,y}$ will be the solution of the differential equation (3.11) letting $c \rightarrow b$ and $d \rightarrow y$.

Definition 3.1. A state y is called recurrent if $\mathbf{R}_{x,y}\mathbf{e}_N = \mathbf{e}_N$ for all $x \in S$, where \mathbf{e}_N is an N -dimensional column vector with all entries equal to 1, and transient otherwise. If all states in S are recurrent, then the bivariate Markov process is said to be recurrent.

Consider now, for a continuous matrix-valued function $\mathbf{G}(x)$, finding $\mathbf{V}(x)$ given by

$$\mathbf{V}(x) = \mathbf{E} \left[\int_0^{T^*} \mathbf{G}(X_s) ds | X_0 = x \right], \quad c < x < d, \quad (3.13)$$

where \mathbf{E} is defined in (2.7).

Theorem 3.2. The matrix-valued function $\mathbf{V}(x)$ defined in (3.13) possesses two bounded derivatives for $c < x < d$ and satisfies the following differential equation

$$\frac{1}{2}\mathbf{A}(x)\mathbf{V}''(x) + \mathbf{B}(x)\mathbf{V}'(x) + \mathbf{Q}(x)\mathbf{V}(x) + \mathbf{G}(x) = \mathbf{0}, \quad c < x < d, \quad \mathbf{V}(c) = \mathbf{V}(d) = \mathbf{0}. \quad (3.14)$$

Proof. As before, the proof is essentially the same as in the scalar case. For a heuristic justification see p. 194 of [3]. For more technical details the reader should consult Proposition 10.1 of Chapter V in [2]. \square

In particular, in the case where

$$\mathbf{G}(x) = \mathbf{e}_N \mathbf{e}_N^* = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{pmatrix}, \quad (3.15)$$

we have that the matrix-valued function $\mathbf{V}(x) = (V_{ij}(x))$ gives

$$V_{ij}(x) = E \left[T^*, Y_{T^*} = j | X_0 = x, Y_0 = i \right], \quad (3.16)$$

i.e., starting at $X_0 = x$ and phase i , $V_{ij}(x)$ is the mean time to reach either c or d and at that time the process is at phase j .

Remark 3.3. Again, in [2,3] we can find an explicit solution of this problem in the scalar situation in terms of the Green function of the process, but the presence of $\mathbf{Q}(x)$ will lead to a system of differential equations similar to (3.12), which is more difficult to solve.

Definition 3.2. Assume our process is recurrent. By letting $c \rightarrow a$ and $d \rightarrow b$ in (3.14), we say that the bivariate Markov process is positive recurrent if all entries of $\mathbf{V}(x)$ in (3.16), with $\mathbf{G}(x)$ given by (3.15), are finite. Otherwise it is null recurrent.

Finally, let us discuss the *invariant distribution* of a bivariate Markov process, which is the behavior of $\mathbf{P}(t; x, y)$ as $t \rightarrow \infty$. This distribution should be independent of the initial state $X_0 = x$ and the initial phase $Y_0 = i$. Therefore we should expect a row vector-valued invariant distribution

$$\boldsymbol{\psi}(y) = (\psi_1(y), \psi_2(y), \dots, \psi_N(y)),$$

satisfying

$$\boldsymbol{\psi}(y) = \int_S \boldsymbol{\psi}(x) \mathbf{P}(t; x, y) dx, \quad t > 0.$$

It is also expected, as in the scalar situation (see Theorem 12.2 of Chapter V in [2] or p. 241 of [3]), that the existence and unicity of such an invariant distribution is restricted to the case when the process is positive recurrent.

It is easy to see, mimicking the derivation of the forward differential equation (3.6), that $\boldsymbol{\psi}(y)$ satisfies

$$\frac{1}{2}(\boldsymbol{\psi}(y)\mathbf{A}(y))'' - (\boldsymbol{\psi}(y)\mathbf{B}(y))' + \boldsymbol{\psi}(y)\mathbf{Q}(y) = \mathbf{0}. \quad (3.17)$$

In order to have a vector-valued probability distribution we must impose that $0 \leq \psi_j(y) \leq 1$ and

$$\left(\int_S \boldsymbol{\psi}(y) dy \right) \mathbf{e}_N = \mathbf{1}. \quad (3.18)$$

If the infinitesimal operator \mathcal{A} is self-adjoint with respect to certain weight matrix \mathbf{W} then we can solve directly (3.17). This equation resembles the third symmetry equation in (3.9). Substituting

$$\boldsymbol{\psi}(y) = c \mathbf{e}_N^* \mathbf{W}(y), \quad c > 0,$$

in (3.17) and using the third equation in (3.9) give

$$c \mathbf{e}_N^* \left[\frac{1}{2} (\mathbf{W}(y) \mathbf{A}(y))'' - (\mathbf{W}(y) \mathbf{B}(y))' + \mathbf{W}(y) \mathbf{Q}(y) \right] = c \mathbf{e}_N^* \mathbf{Q}^*(y) \mathbf{W}(y) = \mathbf{0},$$

as a consequence of $\mathbf{Q}(y) \mathbf{e}_N = \mathbf{0}$, for all y . The normalization constant c is given by (3.18):

$$c = \left(\int_S \mathbf{e}_N^* \mathbf{W}(y) \mathbf{e}_N dy \right)^{-1}.$$

Therefore an explicit formula for the invariant distribution in this case is given by

$$\psi(y) = \left(\int_S \mathbf{e}_N^* \mathbf{W}(y) \mathbf{e}_N dy \right)^{-1} \mathbf{e}_N^* \mathbf{W}(y). \quad (3.19)$$

4. A variant of the Wright–Fisher diffusion model involving only mutation effects

The Wright–Fisher diffusion model involving only mutation effects considers a big population of constant size composed of two types A and B. The intensities of mutation from A to B and from B to A are given by two positive constants, $\frac{1+\beta}{2}$ and $\frac{1+\alpha}{2}$, respectively ($\alpha, \beta > -1$). As the size of the population tends to infinity, it is very well known (see p. 177 of [3]) that this model can be described by a diffusion process whose state space is the unit interval $S = [0, 1]$ with drift and diffusion coefficient given by

$$\tau(x) = \alpha + 1 - x(\alpha + \beta + 2), \quad \sigma^2(x) = 2x(1 - x), \quad \alpha, \beta > -1. \quad (4.1)$$

The variable x is the rate of the number of A-types with respect to the total population. These coefficients correspond with the Jacobi diffusion model, and the probability density can be described in terms of the Jacobi polynomials on $[0, 1]$.

In this section we will study a variation of the Wright–Fisher model with only mutation effects given by a bivariate Markov process whose state space is $[0, 1] \times \{1, 2, \dots, N\}$, so that we can apply all the analysis introduced in the previous sections. In our case, there will be N possible phases with different drift coefficients (the diffusion coefficient will always be the same). Also, there will be an extra positive parameter k , depending on β , which measures how the process moves through all the phases. All these considerations will be studied deeply in Sections 4.2 and 4.3.

The example we study comes from group representation theory, and it is related to matrix-valued spherical functions associated with the complex projective space. It was first introduced in [12] for a special choice of the parameters (see also [17]). Later in [13] it was related to matrix-valued orthogonal polynomials. In Section 4.1 we will give the coefficients relevant in this example to treat it as a bivariate Markov process. The recurrence coefficients of this example have also been related to certain quasi-birth-and-death processes in [18]. Recently, new applications of these processes have been related to urn or Young diagram models [19]. In both cases the differential operator played no role. In this paper we will give a meaning of the differential operator in terms of bivariate Markov processes.

4.1. A family of examples arising from the complex projective space

In what follows we will use \mathbf{E}_{ij} to denote the matrix with 1 at entry (i, j) and 0 otherwise. Let $N \in \{1, 2, \dots\}$, $\alpha, \beta > -1$ and $0 < k < \beta + 1$. \mathbf{M} will denote the nilpotent upper triangular matrix

$$\mathbf{M} = \sum_{i=1}^{N-1} \mathbf{E}_{i, i+1},$$

while \mathbf{J} and $\check{\mathbf{J}}$ the following diagonal matrices

$$\mathbf{J} = \sum_{i=1}^N (\mathbf{N} - \mathbf{i}) \mathbf{E}_{ii}, \quad \check{\mathbf{J}} = \sum_{i=1}^N (\mathbf{i} - \mathbf{1}) \mathbf{E}_{ii} = (\mathbf{N} - \mathbf{1}) \mathbf{I} - \mathbf{J}.$$

Finally, the diagonal matrix \mathbf{H} will denote

$$\mathbf{H} = \sum_{i=1}^N \omega_i \mathbf{E}_{ii},$$

where

$$\omega_i = \binom{\beta - k + i - 1}{i - 1} \binom{N + k - i - 1}{N - i}.$$

Observe that the weights ω_i correspond to the weights of the *Hahn polynomials* supported on $\{1, 2, \dots, N\}$ with parameters $\beta - k > -1$ and $k - 1 > -1$ (see for instance p. 345 of [20]).

The pair $\{\mathbf{W}, \mathcal{A}\}$, where

$$\mathbf{W}(x) = x^\alpha (1-x)^\beta \mathbf{H} x^{\mathbf{J}}, \quad x \in (0, 1), \quad (4.2)$$

and

$$\mathcal{A} = \frac{1}{2} \mathbf{A}(x) \frac{d^2}{dx^2} + \mathbf{B}(x) \frac{d}{dx} + \mathbf{Q}(x) \frac{d^0}{dx^0}, \quad (4.3)$$

where

$$\begin{aligned} \mathbf{A}(x) &= 2x(1-x)\mathbf{I}, & \mathbf{B}(x) &= (\alpha+1)\mathbf{I} + \mathbf{J} - x((\alpha+\beta+2)\mathbf{I} + \mathbf{J}), \\ \mathbf{Q}(x) &= \frac{1}{1-x} \left(\mathbf{J} \left((\beta-k+1)\mathbf{I} + \check{\mathbf{J}} \right) (\mathbf{M} - \mathbf{I}) + x \left(\check{\mathbf{J}}(k\mathbf{I} + \mathbf{J})(\mathbf{M}^* - \mathbf{I}) \right) \right), \end{aligned} \quad (4.4)$$

is a *symmetric pair*, meaning that \mathcal{A} is self-adjoint with respect to \mathbf{W} (with respect to the inner product (3.7)). See [16,13] for definitions.

The *tridiagonal* matrix $\mathbf{Q}(x)$ can be written as

$$\mathbf{Q}(x) = \sum_{i=2}^N \mu_i(x) \mathbf{E}_{i,i-1} - \sum_{i=1}^N (\lambda_i(x) + \mu_i(x)) \mathbf{E}_{ii} + \sum_{i=1}^{N-1} \lambda_i(x) \mathbf{E}_{i,i+1}, \quad (4.5)$$

where

$$\begin{aligned} \lambda_i(x) &= \frac{1}{1-x} (N-i)(i+\beta-k), \\ \mu_i(x) &= \frac{x}{1-x} (i-1)(N-i+k). \end{aligned} \quad (4.6)$$

We remark that $\lambda_i(x)$ and $\mu_i(x)$ are x -dependent variations of the coefficients of the second-order difference equation satisfied by the Hahn polynomials (see for instance p. 346 of [20]).

Observe that $\mathbf{A}(x)$ and $\mathbf{B}(x)$ are diagonal matrices and the diagonal entries of $\mathbf{A}(x)$ are positive for all $x \in (0, 1)$. Also $\lambda_i(x)$ and $\mu_i(x)$ are always positive and $\mathbf{Q}(x)\mathbf{e}_N = \mathbf{0}$ for every $x \in (0, 1)$. Therefore $\mathbf{Q}(x)$ is the intensity matrix of a classical continuous time Markov chain with finite state space. All these requirements fit perfectly with the definition of bivariate Markov processes introduced in Section 2.

We have many formulas and properties associated with this example at our disposal in the literature. In particular we can find a set of normalized eigenfunctions $(\Phi_n(x))_n$ of the differential operator \mathcal{A} (4.3) with the corresponding eigenvalues $(\Gamma_n)_n$. This will allow us to get a formula for the matrix-valued transition density $\mathbf{P}(t; x, y)$ of this process using formula (3.8).

The eigenfunctions $(\Phi_n(x))_n$ of \mathcal{A} (also known as the matrix-valued spherical functions associated with the complex projective space, see [12]) are very close related to matrix-valued orthogonal polynomials and they have been studied in several papers. We use the family of matrix-valued orthogonal polynomials given in [18], since it contains most of the structural formulas we need for this paper. First we describe the relation between the symmetric pair $\{\mathbf{W}, \mathcal{A}\}$ given in (4.2) and (4.3) and the symmetric pair $\{\mathbf{W}, \mathcal{A}\}$ given in [18]. This transformation was introduced for the first time in [17] (see also [13]). Denote $\Psi(x)$ the following matrix

$$\Psi(x) = e^{\check{\mathbf{J}}\mathbf{M}^*} (1-x)^{\check{\mathbf{J}}} \mathbf{T},$$

where \mathbf{T} is the constant upper triangular matrix given at the beginning of Section 7 in [18].

The relation between the pair $\{\mathbf{W}, \mathcal{A}\}$ and $\{\mathbf{W}, \mathcal{A}\}$ is the following

$$\tilde{\mathbf{W}}(x) = \Psi^*(x) \mathbf{W}(x) \Psi(x), \quad \tilde{\mathcal{A}}\mathbf{F}(x) = \Psi^{-1}(x) \mathcal{A}(\Psi(x)\mathbf{F}(x)). \quad (4.7)$$

Take now the family of matrix-valued orthogonal polynomials introduced in [18], which we will denote by $(\mathbf{R}_n(x))_n$ (in [18] they are denoted by $(Q_n^*(x))_n$). Then the family $(\mathbf{R}_n(x))_n$ is an eigenfunction of the differential operator \mathcal{A} , i.e. $\mathcal{A}\mathbf{R}_n(x) = \mathbf{R}_n(x)\Gamma_n$, where Γ_n is the *diagonal* matrix

$$\Gamma_n = -n^2\mathbf{I} - n((\alpha+\beta+N)\mathbf{I} + \check{\mathbf{J}}) - \check{\mathbf{J}}((\alpha+\beta-k+1)\mathbf{I} + \check{\mathbf{J}}). \quad (4.8)$$

This family has very special properties. In particular it is very easy to normalize since the norms $\|\mathbf{R}_n\|_{\tilde{\mathbf{W}}}^2$, given in formula (7.3) of [18], are *diagonal* matrices.

Therefore, define the family

$$\Phi_n(x) = \Psi(x) \mathbf{R}_n(x) \|\mathbf{R}_n\|_{\tilde{\mathbf{W}}}^{-1}.$$

We have that this family is eigenfunction of the infinitesimal generator \mathcal{A} , i.e.

$$\mathcal{A}\Phi_n(x) = \Phi_n(x)\Gamma_n,$$

with Γ_n defined in (4.8), and orthonormal with respect to $\mathbf{W}(x)$ (4.2). Observe that $(\Phi_n(x))_n$ are also matrix-valued polynomials, but not of degree exactly n (more precisely of degree $n + N - 1$).

Therefore, according to formula (3.8), we have

$$\mathbf{P}(t; x, y) = \sum_{n=0}^{\infty} \Phi_n(x) e^{t\Gamma_n} \Phi_n^*(y) \mathbf{W}(y), \quad (4.9)$$

or, in terms of the pair $\{\tilde{\mathbf{W}}, \tilde{\mathbf{A}}\}$,

$$\mathbf{P}(t; x, y) = \Psi(x) \left(\sum_{n=0}^{\infty} \mathbf{R}_n(x) e^{t\Gamma_n} \|\mathbf{R}_n\|_{\tilde{\mathbf{W}}}^{-2} \mathbf{R}_n^*(y) \tilde{\mathbf{W}}(y) \right) \Psi^{-1}(y).$$

4.2. Probabilistic interpretation

Consider the Wright–Fisher diffusion model involving only mutation effects for a big population of constant size composed of two types A and B. The intensity of mutation from A to B and from B to A are described at the beginning of Section 4 and are given in terms of the parameters $\alpha, \beta > -1$ of the example introduced in Section 4.1. Now there will be an indicator denoted by the extra parameter k , depending on the intensity β , which helps the population of A to survive against the population of B. We will give more details in Section 4.3.

First we remark the importance of having a representation of the matrix-valued probability density $\mathbf{P}(t; x, y)$ given by (4.9), which can be very well approximated by the first eigenfunctions $\Phi_n(x)$ and the corresponding eigenvalues Γ_n . We can calculate very accurately what the probability is of reaching any Borel set $B \subset (0, 1)$ and certain phase $Y_t = j$ in time t given that initially we start from state $X_0 = x$ and phase $Y_0 = i$. As an illustration, for $N = 4$ phases, starting at $x = 1/2$, at time $t = 1$ we have

$$\Pr \left\{ X_1 \in \left(\frac{3}{4}, 1 \right) \middle| X_0 = \frac{1}{2} \right\} = \begin{pmatrix} 0.12410905 & 0.08138740 & 0.08920446 & 0.1633878 \\ 0.11006872 & 0.07334748 & 0.08181737 & 0.1527569 \\ 0.09668381 & 0.06555764 & 0.07453306 & 0.1420720 \\ 0.08394494 & 0.05801744 & 0.06735379 & 0.1313385 \end{pmatrix},$$

where the entries (i, j) represent the initial phase $Y_0 = i$ and final phase $Y_1 = j$. Here we have taken the values of the parameters $\alpha = \beta = 0$ and $k = \frac{1}{2}$. In this case only 3 eigenfunctions are needed in order to get a very accurate approximation of the probabilities.

Our family of bivariate Markov processes (X_t, Y_t) with state space $[0, 1] \times \{1, 2, \dots, N\}$ can be described in terms of the stochastic differential equation

$$dX_t = \tau_{Y_t}(X_t)dt + \sigma(X_t)dB_t,$$

where (see (4.4))

$$\tau_i(x) = \alpha + 1 + N - i - x(\alpha + \beta + 2 + N - i), \quad \sigma^2(x) = 2x(1 - x), \quad Y_t = i = 1, \dots, N. \quad (4.10)$$

The description of how the process moves through the different phases is given by $\mathbf{Q}(x)$ (4.5). Observe that $\mathbf{Q}(x)$ is tridiagonal, meaning that transitions of phases are only allowed between adjacent phases (like a birth-and-death process).

The behavior of X_t depends on the phase Y_t and the transition of phases and the waiting times at each phase Y_t depend also on the position of X_t . Concerning the continuous part X_t , we are interested in studying the behavior at the boundaries' values, in this case 0 and 1. On the contrary, from the discrete part Y_t , we will be interested in analyzing how the process moves through the different phases, including the waiting times at each phase. The behavior of the bivariate Markov processes (X_t, Y_t) will be a combination of both aspects.

The diffusion coefficient $\sigma^2(x)$ is always the same, but the drift coefficients $\tau_i(x)$ depend on the phase i . For a given phase i , moving to the next phase $i + 1$ means decreasing one unity of the intensity of mutation $B \rightarrow A$ (in terms of the parameter α), and moving to the previous phase $i - 1$ means increasing one unity of the same intensity of mutation $B \rightarrow A$. The limiting situation is at phase N , where we recover the regular Jacobi diffusion (4.1). The intensity of mutation $A \rightarrow B$ (in terms of the parameter β) is not affected in these coefficients.

In p. 239 of [3] one can find a classification of the boundary states 0 and 1 of the Wright–Fisher model involving only mutation pressures for Jacobi diffusions in terms of the parameters. Both states 0 and 1 are *regular* or *absorbing* boundaries when $-1 < \alpha, \beta < 0$, respectively. While both states 0 and 1 are *entrance* or *reflecting* boundaries when $\alpha, \beta \geq 0$.¹

For our bivariate Markov process observe that the boundary state 1 preserves this behavior for any phase, but the boundary state 0 only preserves this behavior at phase N . For the rest of the phases $i \in \{1, \dots, N - 1\}$ the boundary

¹ A reflecting boundary cannot be reached from the interior of the state space, but it is possible to consider processes that begin there. An absorbing boundary can both be entered and left.

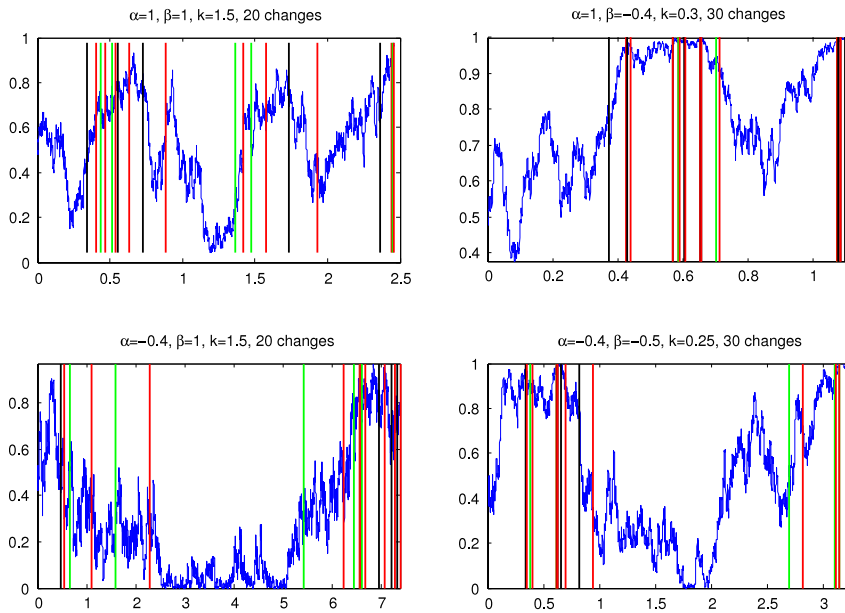


Fig. 2. Different behaviors of the boundaries 0 and 1 for different values of α, β and k . In the first plot 0 and 1 are reflecting. In the second plot, 1 is absorbing and 0 reflecting. In the third plot, 1 is reflecting and 0 is absorbing (only at phase 3). Finally in the fourth plot, 0 (only at phase 3) and 1 are absorbing.

state 0 is always reflecting since $\alpha + N - i > 0$ (see (4.10)). Therefore

$$\begin{aligned} 1 & \begin{cases} \text{is an absorbing boundary} & \text{for } -1 < \beta < 0, \\ \text{is a reflecting boundary} & \text{for } \beta \geq 0. \end{cases} \\ 0 & \begin{cases} \text{is an absorbing boundary} & \text{for } -1 < \alpha < 0 \text{ AND phase } N, \\ \text{is a reflecting boundary} & \text{for } \alpha \geq 0. \end{cases} \end{aligned}$$

In Fig. 2 we can check this behavior for four different situations. In all of them we have taken $N = 3$ phases. In this figure we have not marked with numbers the transitions of phases (in vertical lines) since we are only interested in the behavior at the boundaries 0 and 1. In the first plot the parameters are taken so that the boundaries 0 and 1 are both reflecting, meaning that the sample path is never going to reach 0 or 1. In the second plot, 1 is absorbing, while 0 is reflecting. We observe that 1 can be reached in principle at any phase. In the third plot, on the contrary, 0 is absorbing while 1 is reflecting. In this case the only phase where the sample path reaches 0 is *only on the last third phase*. In the rest of the phases, 0 becomes reflecting. Finally, in the fourth plot, the parameters are taken so that 0 and 1 are both absorbing. The process can reach boundary 1 at any phase, but, again, boundary 0 only when the process is at the third phase.

According to Definition 3.1 our process is always *recurrent* for $\alpha, \beta \geq 0$, while *transient* otherwise. This fact has been checked by looking extensively at the numerical solutions of $\mathbf{R}_{x,y}$ via the second-order differential equation (3.11). In a similar way, we can state, using Definition 3.2, that our process is always *positive recurrent* for $\alpha, \beta \geq 0$ by looking the numerical solutions of $\mathbf{V}(x)$ via the second-order differential equation (3.14) with $\mathbf{G}(x)$ given by (3.15).

Let us now study a couple of aspects from the discrete component Y_t . First we will study the *waiting times* at each phase depending on the parameters and secondly we will study the *tendency* of moving forward or backward in phases. All these facts depend on the position of X_t .

For the waiting times we have to look at the diagonal entries of $\mathbf{Q}(x)$ given by (see (4.5) and (4.6))

$$\begin{cases} Q_{11}(x) = -\frac{1}{1-x}(N-1)(\beta-k+1) & \text{for } i = 1, \\ Q_{ii}(x) = -\frac{1}{1-x}[(N-i)(i+\beta-k)+x(i-1)(N-i+k)] & \text{for } i = 2, \dots, N-1, \\ Q_{NN}(x) = -\frac{x}{1-x}(N-1)k & \text{for } i = N. \end{cases}$$

We have that if x is close to 1, the diagonal coefficients are very large, meaning that all phases are *instantaneous*, i.e., the waiting times at each phase are very short until x is far from 1. This can be checked in the graphs of Fig. 2. If x is close to 0, we have that $Q_{NN}(x)$ is very small, meaning that phase N becomes *absorbing*, i.e., if the process enters this phase and the position of X_t is close to 0, then it tends to spend long periods of time there (see the last two plots of Fig. 2). Also phase N becomes absorbing (no matter of the position of X_t) if the parameter k is close to 0 (see first plot in Fig. 3). The middle phases (from 2 to $N-1$) are never absorbing, since $Q_{ii}(x)$, $i = 2, \dots, N-1$, are never close to 0. Finally, if k is close to $\beta+1$, we

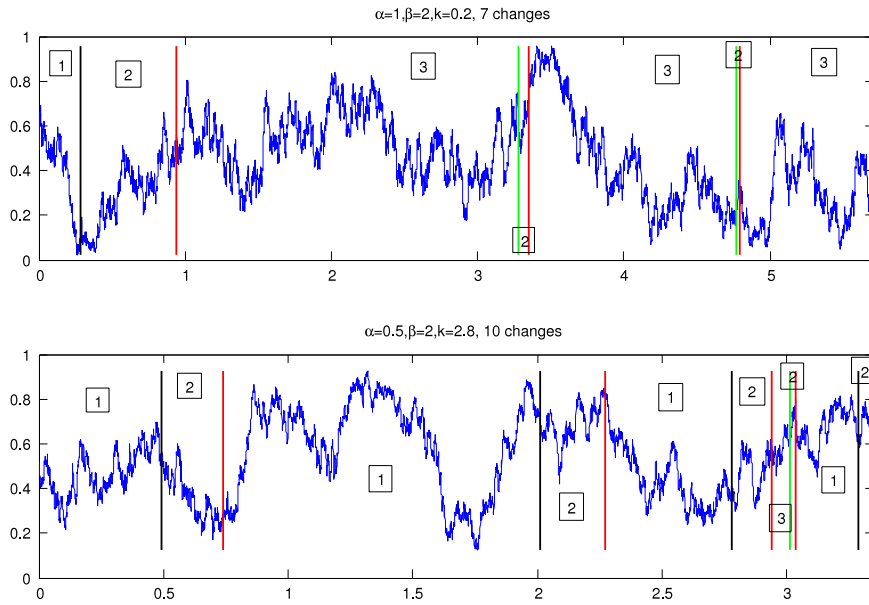


Fig. 3. Different behaviors of the phases 1 and 3 for different values of k ($N = 3$). In the first plot k is close to 0, so the process tends to spend more time in the last absorbing phase 3. In the second plot, k is close to $\beta + 1$, so phase 1 is absorbing.

have that phase 1 is absorbing (see the second plot in Fig. 3). For the rest of the cases, the value of these diagonal entries and the position of X_t determine the waiting time in that phase.

For the tendency we have to study, given that the process is at phase $Y_t = i$, the probability of moving to the next or to the previous phase. Obviously, when the process is at one of the boundary phases 1 or N it can only go forward or backward, respectively. This probability is given by

$$\Pr\{Y_t = i \rightarrow Y_t = i + 1\} = \frac{\lambda_i(x)}{\lambda_i(x) + \mu_i(x)},$$

$$\Pr\{Y_t = i \rightarrow Y_t = i - 1\} = \frac{\mu_i(x)}{\lambda_i(x) + \mu_i(x)},$$

where the coefficients $\lambda_i(x)$ and $\mu_i(x)$ are given by (4.6). Here x is the state X_t of the process at the time t of transition of phases. The forward or backward tendency depends on if $\mu_i(x) < \lambda_i(x)$ or $\mu_i(x) > \lambda_i(x)$, respectively. Define the threshold values

$$x_0(i) = \frac{(N-i)(i+\beta-k)}{(i-1)(N-i+k)}, \quad i = 2, \dots, N-1. \quad (4.11)$$

These values are the x -solutions of the algebraic equation $\lambda_i(x) = \mu_i(x)$. Therefore, at the moment of changing from phase $Y_t = i$ to the next or previous phase, we will have a *forward tendency* when $x < x_0(i)$, and a *backward tendency* when $x > x_0(i)$. In order to be relevant, this threshold value $x_0(i)$ must be less than 1. Otherwise, there is always a forward tendency.

Let us study the cases for which the threshold values have a meaning. If $x_0(i) > 1$ for all $i = 1, \dots, N-1$, then we will always have a forward tendency. This happens when

$$k < \frac{(\beta+1)}{N-1}.$$

Nevertheless, if

$$k > \frac{(\beta+1)(N-i)}{N-1}, \quad (4.12)$$

for some i there can be forward and backward tendency at the same run, depending on the position of X_t at the moment of transition. We have then a maximum backward tendency for $i = 2$ in (4.12). The range of the values of k depending on the phase $Y_t = i$ are described in Fig. 4. Therefore we have

$$\begin{cases} \text{Maximum forward tendency} & \text{for } k < \frac{(\beta+1)}{N-1}, \\ \text{Maximum backward tendency} & \text{for } k > \frac{(\beta+1)(N-2)}{N-1}, \\ \text{Backward/forward tendency} & \text{otherwise.} \end{cases}$$

This also explains why phases 1 or N are absorbing in terms of the values of k .

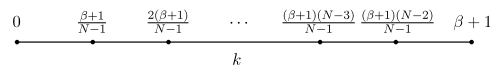


Fig. 4. Range of k for the study or the tendency of the process.

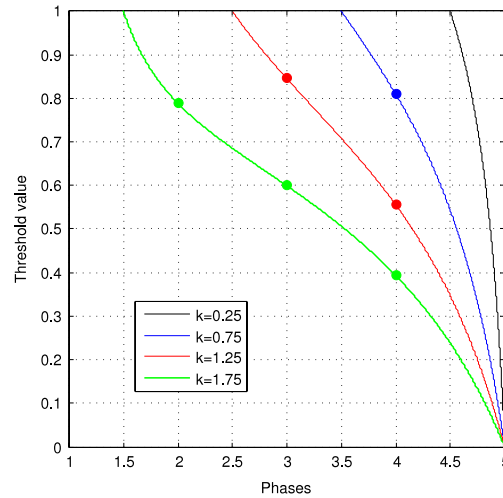


Fig. 5. Threshold values (4.11) (in big dots) for different values of k for $N = 5$ phases, $\alpha = \frac{1}{2}$, $\beta = 1$.

Let us study closer this tendency behavior and the threshold values (4.11) for an specific case. We fix $N = 5$ phases and the parameters $\alpha = \frac{1}{2}$, $\beta = 1$. The interval $(0, 2)$ in Fig. 4 is split in four subintervals $(0, 1/2] \cup (1/2, 1] \cup (1, 3/2] \cup (3/2, 1)$. We study four values of k taken in the middle of each subinterval. In Fig. 5 we have plotted the threshold function (4.11) (depending on the phase i) for these four values of k . We observe that the threshold value at boundary phase 5 is 0, meaning that the next phase of the process is always 4, and the threshold value at boundary phase 1 is ∞ , meaning that the next phase of the process is always 2.

We observe, for $k \in (0, 1/2)$ in Fig. 5, that there is no threshold between 0 and 1, i.e. always a forward tendency. In the following table we can confirm this behavior for a sample path of the process. Here in the first row we have written the sequence of transitions of phases and in the second row the position of X_t at the end of that phase.

Phases	1	2	3	2	3	4	5	4	5	4	5	4	5	4	5
Position	.68	.79	.82	.80	.69	.71	.76	.55	.86	.86	.85	.71	.59	.76	.69

For $k \in (1/2, 1)$ in Fig. 5, only phase 4 has a threshold value (around 0.8). That means that for phases 2 and 3 there is always a forward tendency. At the moment of transition at the end of phase 4, if the position of X_t is greater that the threshold value 0.8, then there is a backward tendency. Otherwise, there is a forward tendency. In the following table, as before, we have run a sample path for this situation:

Phases	1	2	3	2	3	4	3	4	5	4	5	4	5	4	5
Position	.49	.53	.56	.51	.46	.87	.52	.45	.12	.13	.20	.04	.13	.12	.37

For $k \in (1, 3/2)$ in Fig. 5, phases 3 and 4 have a threshold value (around 0.85 and 0.55 respectively). Again, phase 2 has a forward tendency, while phases 3 and 4 depend on the position of X_t as before. In the following table we have run a sample path for this situation:

Phases	1	2	3	4	5	4	3	2	3	2	3	4	5	4	3
Position	.71	.83	.78	.43	.83	.81	.88	.78	.64	.86	.69	.72	.42	.58	.31

Finally, for $k \in (3/2, 1)$ in Fig. 5, all middle phases have a threshold value (around 0.8, 0.6 and 0.4 respectively). This is the situation with the maximum backward tendency, but it depends strongly on the position of X_t . Again, in the following table we have run a sample path for this situation:

Phases	1	2	1	2	3	2	3	4	3	2	1	2	3	4	3
Position	.56	.72	.43	.47	.75	.66	.64	.87	.55	.62	.65	.17	.57	.62	.43

Finally, we will give the explicit expression of the vector-valued invariant distribution $\psi(y)$ (see the end of Section 3 for definitions). We can get an explicit formula in this case since \mathcal{A} , given by (4.3), is self-adjoint and we have an explicit

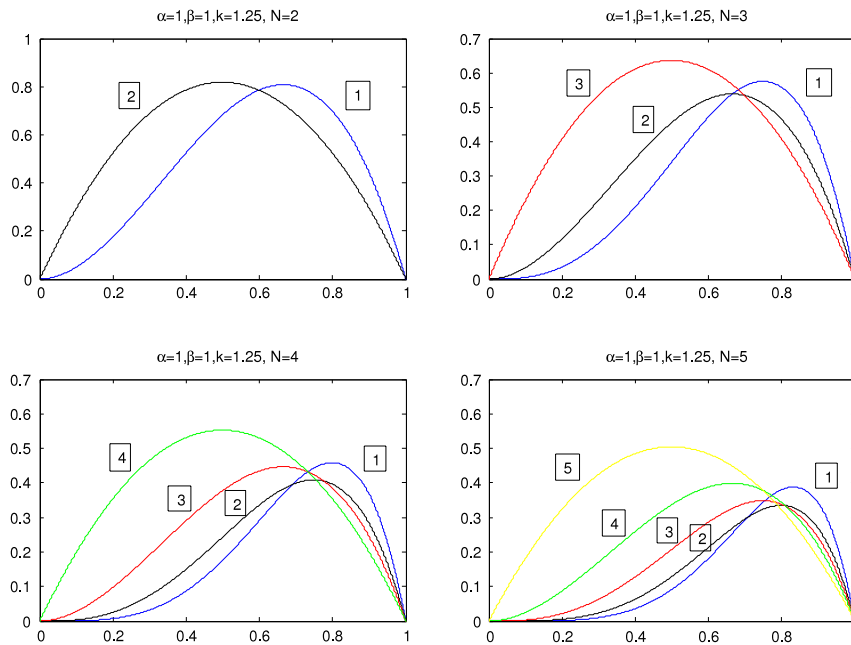


Fig. 6. The components of the vector-valued invariant distribution $\psi_j(y)$, $j = 1, \dots, N$ (see (4.14)), for $N = 2, 3, 4, 5$, and $\alpha, \beta = 1, k = 5/4$.

expression of the orthogonality measure \mathbf{W} . Formula (3.19) gives

$$\psi(y) = \left(\int_0^1 \mathbf{e}_N^* \mathbf{W}(y) \mathbf{e}_N dy \right)^{-1} \mathbf{e}_N^* \mathbf{W}(y) \quad (4.13)$$

where \mathbf{W} is given by (4.2). A closer look to the 0-norm and \mathbf{H} shows, for $j = 1, \dots, N$,

$$\psi_j(y) = y^{\alpha+N-j} (1-y)^\beta \binom{N-1}{j-1} \binom{\alpha+\beta+N}{\alpha} \frac{(\beta+N)(k)_{N-j}(\beta-k+1)_{j-1}}{(\alpha+\beta-k+2)_{N-1}}, \quad (4.14)$$

where $(a)_n$ denotes the Pochhammer symbol defined by $(a)_n = a(a+1) \cdots (a+n-1)$ for $n > 0$, $(a)_0 = 1$. Another possibility of obtaining (4.13) is by looking the spectral representation of $\mathbf{P}(t; x, y)$ (4.9). From this representation we observe that the limit as $t \rightarrow \infty$ only depends on the eigenvalues Γ_n , given in (4.8). All diagonal entries of Γ_n are strictly negative for the range of all parameters n, α, β, k and N involved, except for $n = 0$ and entry $(1, 1)$ of Γ_0 , which is 0. Therefore taking limits in (4.9) when $t \rightarrow \infty$, we have that the only relevant element is the first term of the sum, given by

$$\lim_{t \rightarrow \infty} \mathbf{P}(t; x, y) = \Phi_0(x) \mathbf{E}_{11} \Phi_0^*(y) \mathbf{W}(y).$$

A straightforward computation of the expression above, using the notation given in Section 4.1, leads to a matrix where all rows are equal to (4.13).

This vector-valued invariant distribution (4.13) is valid only when the process is positive recurrent, i.e. $\alpha, \beta \geq 0$. For $-1 < \alpha, \beta < 0$, (4.13) is also meaningful, but the boundary points of the process are now absorbing, meaning that the correct vector-valued invariant distribution of such cases involves mass jumps at the boundaries 0 and 1 plus a density portion of the form (4.13).

As an illustration, in Fig. 6, we have plotted $\psi_j(y)$, $j = 1, \dots, N$, in (4.14) for $N = 2, 3, 4, 5$. We have fixed the parameters to $\alpha, \beta = 1, k = 5/4$. We remark that $\int_0^1 \psi_j(y) dy < 1$ for each component, but $\sum_{j=1}^N \int_0^1 \psi_j(y) dy = 1$.

From these plots we immediately see that, for a large time, if the process is at one state in $(0, 1/2)$, then there is a very high probability that the process is located at one of the last phases. Nevertheless if the process is at one state in $(1/2, 1)$, there is more or less the same probability that the process is located at any phase (higher probability for the first phases).

This interpretation may change depending on the value of k . If k is close to $\beta + 1$ then we have the maximum backward tendency and the process is more likely to be at one of the initial phases. Nevertheless, if k is close to 0, then we have maximum forward tendency, and the process is more likely to be located at one of the final phases.

4.3. Final remarks

From the considerations in Section 4.2 we are ready to give an interpretation of this bivariate Markov process in relation to the Wright–Fisher diffusion model involving only mutation effects. Consider a big population of constant size composed

of two types A and B, where the intensity of mutation from A to B is given in terms of the parameter β and the intensity of mutation from B to A is given in terms of the parameter α . The value that the process X_t takes in the interval $(0, 1)$ is the rate $\#A/(\#A + \#B)$, i.e. if $X_t < 1/2$ then $\#B > \#A$ and if $X_t > 1/2$ then $\#A > \#B$. We assume that the intensity of mutations α and β , are more or less the same. If $\alpha > \beta$ then the population of A's tends to survive against the population of B's (the opposite for $\alpha < \beta$).

Now the process can move through N different phases. At phase N the process behaves as a regular Wright–Fisher model (in terms of the Jacobi diffusion). At phase $N - 1$ the intensity $B \rightarrow A$ increases in one unity, at phase $N - 2$ the intensity $B \rightarrow A$ increases in two unities, and so on, until the first phase, where the intensity $B \rightarrow A$ increases in $N - 1$ unities. Therefore moving through the first phases means that the intensity of mutation $B \rightarrow A$ is much higher if we compare with the intensity of mutation $A \rightarrow B$, i.e. it helps to increase the population of A's against B's.

The meaning of the indicator parameter k (depending on β) is the following. If k is close to 0, then we have a forward tendency, i.e. the process tends to spend more time at the last phases, i.e. the populations of A's and B's “fight” more or less in the same conditions (remember that we are assuming that α is more or less equal to β). On the contrary, if this indicator k is close to $\beta + 1$, then we have a backward tendency, i.e. the process tends to spend more time at the initial phases, i.e. it helps the population of A's to survive against the population of B's. For the middle values of k there can be forward/backward tendency, and it depends on the number of A's and B's, i.e. the position of X_t .

5. Conclusions

In this paper we have studied bivariate Markov processes with diffusion and discrete components from its matrix-valued infinitesimal operator properties. This infinitesimal operator is a second-order differential operator with matrix-valued coefficients with certain properties. We have given some theoretical results related to these processes, like the backward/forward equations, the spectral representation in terms of the eigenfunctions of the infinitesimal operator, recurrence considerations and the invariant distribution. All these results have been applied to an example coming from group representation theory and related to matrix-valued spherical functions, which can be viewed as a variant of the Wright–Fisher diffusion model involving only mutation effects. Since we have an explicit set of eigenfunctions with the corresponding orthogonality measure, we can approximate very accurately the matrix-valued probability density $\mathbf{P}(t; x, y)$ and get an explicit expression of the vector-valued invariant distribution, among other properties.

Certainly, new examples of bivariate Markov processes may be possible, especially when the diffusion state space is other than a bounded interval. In [12] one can find other types of examples. The example studied in this paper, also known as the *one-step example*, is just one class of a very large class of matrix-valued spherical functions associated with the complex projective space. The second-order differential operators associated with these matrix-valued spherical functions seem to fit very well in the theory of bivariate Markov processes. For instance, in [17], one can find a case of the *two-steps example* (4×4). The second-order differential operator is also the infinitesimal operator of a bivariate Markov process. In this case, apart from the parameters $\alpha, \beta > -1$, there are two new extra parameters, k_1 and k_2 with the condition that $0 < k_1 < k_2 < \beta + 1$. Therefore this example can be viewed as another variant of the Wright–Fisher model involving only mutation effects, but with two parameters to study, which decide how the process moves through all the phases.

Eigenfunctions of second-order differential operators with matrix-valued coefficients have been studied in the context of matrix-valued orthogonal polynomials, see [14,16,17]. These examples are potential candidates of bivariate Markov processes. The difference is that these differential operators do not exactly fit into the framework of bivariate Markov processes. An inverse transformation of the type (4.7) is going to be needed in order to get an infinitesimal operator of a bivariate Markov process, which in principle is not an easy task. This consideration, however, goes beyond the scope of this paper, and will be pursued elsewhere.

Acknowledgments

The author would like to thank Professor Simeon M. Berman for introducing reference [10] as well as Professor F. Alberto Grünbaum for pointing out the references about random evolutions.

References

- [1] W. Schoutens, Stochastic Processes and Orthogonal Polynomials, in: Lecture Notes in Statistics, vol. 146, Springer-Verlag, New York, 2000.
- [2] R.N. Bhattacharya, E.C. Waymire, Stochastic processes with applications, in: Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics, John Wiley & Sons, Inc., New York, 1990.
- [3] S. Karlin, H.M. Taylor, A Second Course in Stochastic Processes, Academic Press, Inc. Harcourt Brace Jovanovich, Publishers, New York–London, 1981.
- [4] R.J. Griego, R. Hersh, Random evolutions, Markov chains, and systems of partial differential equations, Proc. Nat. Acad. Sci. USA 62 (1969) 305–308.
- [5] R. Hersh, The birth of random evolutions, Math. Intell. 25 (2003) 53–60.
- [6] R. Hersh, Random evolutions: a survey of results and problems, Rocky Mt. J. Math. 4 (1974) 443–477.
- [7] G.C. Papanicolaou, Random Media, Springer-Verlag, Berlin, 1987.
- [8] M.A. Pinsky, Lectures on Random Evolutions, World Scientific, Singapore, 1991.
- [9] A. Swishchuk, Random Evolutions and their Applications, New Trends, Kluwer AP, Dordrecht, 2000.
- [10] S.M. Berman, A bivariate Markov process with diffusion and discrete components, Comm. Statist. Stochastic Models 10 (2) (1994) 271–308.
- [11] M.K. Ghosh, A. Arapostathis, S.I. Marcus, Ergodic control of switching diffusions, SIAM J. Control Optim. 35 (6) (1997) 1952–1988.

- [12] F.A. Grünbaum, I. Pacharoni, J.A. Tirao, Matrix valued spherical functions associated to the complex projective plane, *J. Funct. Anal.* 188 (2002) 350–441.
- [13] I. Pacharoni, J.A. Tirao, Matrix valued orthogonal polynomials arising from the complex projective space, *Constr. Approx.* 25 (2007) 177–192.
- [14] A.J. Durán, F.A. Grünbaum, Orthogonal matrix polynomials satisfying second order differential equations, *Internat. Math. Res. Not.*, 2004 10 (2004) 461–484.
- [15] D. Damanik, A. Pushnitski, B. Simon, The analytic theory of matrix orthogonal polynomials, *Surv. Approx. Theory* 4 (2008) 1–85.
- [16] F.A. Grünbaum, I. Pacharoni, J.A. Tirao, Matrix valued orthogonal polynomials of the Jacobi type, *Indag. Mathem.* 14 (3,4) (2003) 353–366.
- [17] F.A. Grünbaum, I. Pacharoni, J.A. Tirao, Matrix valued orthogonal polynomials of the Jacobi type: the role of group representation theory, *Ann. Inst. Fourier* 55 (5) (2005) 1–18.
- [18] F.A. Grünbaum, M.D. de la Iglesia, Matrix valued orthogonal polynomials arising from group representation theory and a family of quasi-birth-and-death processes, *SIAM J. Matrix Anal. Appl.* 30 (2) (2008) 741–761.
- [19] F.A. Grünbaum, I. Pacharoni, J.A. Tirao, Two stochastic models of a random walk in the $U(n)$ -spherical duals of $U(n+1)$, *Ann. Mat. Pura Appl.* (2012) <http://dx.doi.org/10.1007/s10231-011-0232-z>.
- [20] G.E. Andrews, R. Askey, R. Roy, *Special functions*, *Encyclopedia Math. Appl.* (71) (1999) Cambridge.