An explicit transition density expansion for a multi-allelic Wright-Fisher diffusion with general diploid selection

Matthias Steinrücken[a], **Y. X. Rachel Wang**[a], and **Yun S. Song**[a,b]

Matthias Steinrücken: steinrue@stat.berkeley.edu; Y. X. Rachel Wang: rachelwang@stat.berkeley.edu; Yun S. Song: yss@cs.berkeley.edu

[a]Department of Statistics, University of California, Berkeley, CA 94720, USA

[b]Computer Science Division, University of California, Berkeley, CA 94720, USA

Abstract

Characterizing time-evolution of allele frequencies in a population is a fundamental problem in population genetics. In the Wright-Fisher diffusion, such dynamics is captured by the transition density function, which satisfies well-known partial differential equations. For a multi-allelic model with general diploid selection, various theoretical results exist on representations of the transition density, but finding an explicit formula has remained a difficult problem. In this paper, a technique recently developed for a diallelic model is extended to find an explicit transition density for an arbitrary number of alleles, under a general diploid selection model with recurrent parent-independent mutation. Specifically, the method finds the eigenvalues and eigenfunctions of the generator associated with the multi-allelic diffusion, thus yielding an accurate spectral representation of the transition density. Furthermore, this approach allows for efficient, accurate computation of various other quantities of interest, including the normalizing constant of the stationary distribution and the rate of convergence to this distribution.

1. Introduction

Diffusion processes can be used to describe the evolution of population-wide allele frequencies in large populations, and they have been successfully applied in various population genetic analyses in the past. Karlin and Taylor (1981), Ewens (2004), and Durrett (2008) provide excellent introduction to the subject. The diffusion approximation captures the key features of the underlying evolutionary model and provides a concise framework for describing the dynamics of allele frequencies, even in complex evolutionary scenarios. However, finding explicit expressions for the transition density function (TDF) is a challenging problem for most models of interest. Although a partial differential equation (PDE) satisfied by the TDF can be readily obtained from the standard diffusion theory, few models admit analytic solutions.

Since closed-form transition density functions are unknown for general diffusions, approaches such as finite difference methods (Bollback et al., 2008; Gutenkunst et al., 2009) and series expansions (Lukić et al., 2011) have been adopted recently to obtain approximate

Correspondence to: Yun S. Song, yss@cs.berkeley.edu.

solutions. In a finite difference scheme, one needs to discretize the state space, but since the TDF depends on the parameters of the model (e.g. the selection coefficients), the suitability of a given discretization might depend strongly on the parameter values, and whether a particular discretization would produce accurate solutions is difficult to predict *a priori*. Series expansions allow one to circumvent the problem of choosing an appropriate discretization for the state space. However, if the chosen basis functions in the representation are not the eigenfunctions of the diffusion generator, as in Lukić et al. (2011), then one has to solve a system of coupled ordinary differential equations (ODE) to obtain the transition density. Lukić et al. (2011) solve this system of ODEs numerically, which may introduce potential errors of numerical approximations.

If the eigenvalues and eigenfunctions of the diffusion generator can be found, the spectral representation (which in some sense provides the optimal series expansion) of the TDF can be obtained. For the one-locus Wright-Fisher diffusion with an arbitrary number K of alleles evolving under a *neutral* parent-independent mutation (PIM) model, Shimakura (1977) and Griffiths (1979) derived an explicit spectral representation of the TDF using orthogonal polynomials. More recently, Baxter et al. (2007) derived the same solution by diagonalizing the associated PDE using a suitable coordinate transformation, followed by solving for each dimension independently. In a related line of research, Griffiths and Li (1983) and Tavaré (1984) expressed the time-evolution of the allele frequencies in terms of a stochastic process dual to the diffusion, and showed that the resulting expression is closely related to the spectral representation.

The duality approach was later extended by Barbour et al. (2000) to incorporate a general selection model. Although theoretically very interesting, this approach does not readily lead to efficient computation of the TDF because of the following reason: Computation under the dual process requires evaluating the moments of the stationary distribution. Although the functional form of this distribution is known (Ethier and Kurtz, 1994; Barbour et al., 2000), the normalization constant and moments can only be computed analytically in special cases (Genz and Joyce, 2003), and numerical computation under a general model of diploid selection is difficult. Incidentally, this issue arises in various applications (e.g., Buzbas et al. 2009, 2011), and it has therefore received significant attention in the past; see, for example, Donnelly et al. (2001) and Buzbas and Joyce (2009).

Many decades ago, Kimura (1955, 1957) addressed the problem of finding an *explicit* spectral representation of the TDF for models with selection. Specifically, in the case of a *diallelic* model with special selection schemes, he employed a perturbation method to find the required eigenvalues and eigenfunctions of the diffusion generator. Being perturbative in the selection coefficient, this approach is accurate only for small selection parameters. Recently, Song and Steinrücken (2012) revisited this problem and developed an alternative method of deriving an explicit spectral representation of the TDF for the diallelic Wright-Fisher diffusion under a *general* diploid selection model with recurrent mutation. In contrast to Kimura's approach, this new approach is non-perturbative and is applicable to a broad range of parameter values. The goal of the present paper is to extend the work of Song and Steinrücken (2012) to an arbitrary number K of alleles, assuming a PIM model with general diploid selection.

The rest of this paper is organized as follows. In Section 2, we lay out the necessary mathematical background and review the work of Song and Steinrücken (2012) in the case of a diallelic ($K = 2$) model with general diploid selection. In Section 3, we describe the spectral representation for the *neutral* PIM model with an arbitrary number K of alleles. Then, in Section 4, we generalize the method of Song and Steinrücken (2012) to an arbitrary K-allelic PIM model with general diploid selection. We demonstrate in Section 5 that the

quantities involved in the spectral representation converge rapidly. Further, we discuss the computation of the normalization constant of the stationary distribution under mutation-selection balance and the rate of convergence to this distribution. We conclude in Section 6 with potential applications and extensions of our work.

2. Background

2.1. The Wright-Fisher diffusion

In this paper, we consider a single locus with K distinct possible alleles. The dynamics of the allele frequencies in a large population is commonly approximated by the Wright-Fisher diffusion on the $(K-1)$-simplex

$$\Delta_{K-1} := \{\mathbf{x} \in \mathbb{R}_{\geq 0}^{K-1} : 1 - |\mathbf{x}| \geq 0\},$$

where $|\mathbf{x}| = \sum_{i=1}^{K-1} x_i$. For a given $\mathbf{x} = (x_1, \ldots, x_{K-1}) \in \Delta_{K-1}$, the component x_i denotes the population frequency of allele $i \in \{1, \ldots, K-1\}$. The frequency of allele K is given by $x_K = 1 - |\mathbf{x}|$.

The associated diffusion generator $\mathscr{L}$ is a second order differential operator of form

$$\mathscr{L}f(\mathbf{x}) = \frac{1}{2}\sum_{i,j=1}^{K-1} b_{i,j}(\mathbf{x})\frac{\partial^2}{\partial x_i \partial x_j}f(\mathbf{x}) + \sum_{i=1}^{K-1} a_i(\mathbf{x})\frac{\partial}{\partial x_i}f(\mathbf{x}), \quad (1)$$

which acts on twice continuously differentiable functions $f: \Delta_{K-1} \rightarrow \mathbb{R}$. The diffusion coefficient $b_{i,j}(\mathbf{x})$ is given by

$$b_{i,j}(\mathbf{x}) = x_i(\delta_{i,j} - x_j),$$

where the Kronecker delta $\delta_{i,j}$ is equal to 1 if $i = j$ and 0 otherwise. For a neutral PIM model, we use $\theta_i = 4Nu_i$ to denote the population-scaled mutation rate associated with allele i, where u_i is the probability of mutation producing allele i per individual per generation and N is the effective population size. Under this model, the drift coefficient $a_i(\mathbf{x})$ is given by

$$a_i(\mathbf{x}) = \frac{1}{2}(\theta_i - |\theta|x_i),$$

where $\theta = (\theta_1, \ldots, \theta_K) \in \mathbb{R}_{>0}^{K}$ and $|\theta| = \sum_{i=1}^{K} \theta_i$.

Consider a general diploid selection model in which the relative fitness of a diploid individual with one copy of allele i and one copy of allele j is given by $1 + 2s_{i,j}$. We measure fitness relative to that of an individual with two copies of allele K, thus $s_{K,K} = 0$. The diffusion generator in this case is given by $\mathscr{L} = \mathscr{L}_0 + \mathscr{L}_\sigma$, where $\mathscr{L}_0$ denotes the diffusion generator under neutrality and the additional term $\mathscr{L}_\sigma$, which captures the contribution from selection to the drift coefficient $a_i(\mathbf{x})$, is given by

$$\mathscr{L}_\sigma = \sum_{i=1}^{K-1} x_i \left[\sigma_i(\mathbf{x}) - \bar{\sigma}(\mathbf{x})\right]\frac{\partial}{\partial x_i},$$

where $\sigma_i(\mathbf{x})$ denotes the marginal fitness of type i and $\bar{\sigma}(\mathbf{x})$ denotes the mean fitness of the population with allele frequencies $\mathbf{x}$. More precisely,

$$\sigma_i(\mathbf{x}) = \sum_{j=1}^{K} \sigma_{i,j} x_j \quad (2)$$

and

$$\bar{\sigma}(\mathbf{x}) := \sum_{i,j=1}^{K} \sigma_{i,j} x_i x_j, \quad (3)$$

where $\sigma_{i,j} = 2N s_{i,j}$. Intuitively, the population frequency of a given allele tends to increase if its marginal fitness is higher than the mean fitness of the population. The selection scheme is specified by a symmetric matrix $\boldsymbol{\sigma} = (\sigma_{i,j})_{1 \le i,j \le K} \in \mathbb{R}^{K \times K}$ of population-scaled selection coefficients, with $\sigma_{K,K} = 0$.

The operators $\mathcal{L}_0$ and $\mathcal{L}$ are elliptic inside the simplex Δ_{K-1}, but not on the boundaries. Thus, the precise domain of the generator is not straightforward to describe, but Epstein and Mazzeo (2011) give a suitable characterization.

2.2. Spectral representation of the transition density function

For $t \ge 0$, the time evolution of a diffusion $\mathbf{X}_t$ on the simplex Δ_{K-1} is described by the transition density function $p(t; \mathbf{x}, \mathbf{y}) d\mathbf{y} = \mathbb{P}[\mathbf{X}_t \in d\mathbf{y} \mid \mathbf{X}_0 = \mathbf{x}]$, where $\mathbf{x}, \mathbf{y} \in \Delta_{K-1}$. The transition density function satisfies the Kolmogorov backward equation

$$\frac{\partial}{\partial t} p(t; \mathbf{x}, \mathbf{y}) = \mathcal{L} p(t; \mathbf{x}, \mathbf{y}), \quad (4)$$

where $\mathcal{L}$, the generator associated with the diffusion, is a differential operator in $\mathbf{x}$.

We briefly review the framework underlying the spectral representation of the transition density function. The operator $\mathcal{L}$ is said to be symmetric with respect to a density $\pi : \Delta_{K-1} \to \mathbb{R}_{\ge 0}$ if, for all twice continuously differentiable functions $f : \Delta_{K-1} \to \mathbb{R}$ and $g : \Delta_{K-1} \to \mathbb{R}$ that belong to the domain of the operator, the following equality holds:

$$\int_{\Delta_{K-1}} [\mathcal{L} f(\mathbf{x})] g(\mathbf{x}) \pi(\mathbf{x}) d\mathbf{x} = \int_{\Delta_{K-1}} f(\mathbf{x}) [\mathcal{L} g(\mathbf{x})] \pi(\mathbf{x}) d\mathbf{x}.$$

A straightforward calculation using integration by parts yields that the diffusion generators described in Section 2.1 are symmetric with respect to their associated stationary densities.

Theorem 1.4.4 of Epstein and Mazzeo (2011) guarantees that an unbounded symmetric operator $\mathcal{L}$ of the kind defined in Section 2.1 of this paper has countably many eigenvalues $\{-\Lambda_0, -\Lambda_1, -\Lambda_2, \ldots\}$, which are real and non-positive, satisfying

$$0 \le \Lambda_0 \le \Lambda_1 \le \Lambda_2 \le \cdots,$$

with $\Lambda_n \to \infty$ as $n \to \infty$. An eigenfunction $B_n : \Lambda_{K-1} \to \mathbb{R}$ with eigenvalue $-\Lambda_n$ satisfies

$$\mathcal{L} B_n(\mathbf{x}) = -\Lambda_n B_n(\mathbf{x}), \quad (5)$$

and, furthermore, $B_n(\mathbf{x})$ is an element of the Hilbert space $L^2(\Delta_{K-1}, \pi(\mathbf{x}))$ of functions square integrable with respect to the density $\pi(\mathbf{x})$, equipped with the canonical inner product $\langle \cdot, \cdot \rangle_\pi$. If $\mathcal{L}$ is symmetric with respect to $\pi(\mathbf{x})$, then its eigenfunctions are orthogonal with respect to $\pi(\mathbf{x})$:

$$\langle B_n, B_m \rangle_\pi := \int_{\Delta_{K-1}} B_n(\mathbf{x}) B_m(\mathbf{x}) \pi(\mathbf{x}) d\mathbf{x} = \delta_{n,m} d_n,$$

where $\delta_{n,m}$ is the Kronecker delta and d_n are some constants. In the cases considered in this paper, the eigenfunctions form a basis of the Hilbert space $L^2(\Delta_{K-1}, \pi(\mathbf{x}))$.

It follows from equation (5) that $\exp(-\Lambda_n t) B_n(\mathbf{x})$ is a solution to the Kolmogorov backward equation (4). By linearity of (4), the sum of two solutions is again a solution. Combining the initial condition $p(0; \mathbf{x}, \mathbf{y}) = \delta(\mathbf{x} - \mathbf{y})$ with the fact that $\{B_n(\mathbf{x})\}$ form a basis yields the following spectral representation of the transition density:

$$p(t; \mathbf{x}, \mathbf{y}) = \sum_{n=0}^{\infty} \frac{1}{d_n} e^{-\Lambda_n t} B_n(\mathbf{x}) B_n(\mathbf{y}) \pi(\mathbf{y}). \quad (6)$$

The initial condition being the Dirac delta $\delta(\mathbf{x} - \mathbf{y})$ corresponds to the frequency at time zero being $\mathbf{x}$.

2.3. Univariate Jacobi polynomials

The univariate Jacobi polynomials play an important role throughout this paper. Here we review some key facts about this particular type of classical orthogonal polynomials. An excellent treatise on univariate orthogonal polynomials can be found in Szegö (1939) and a comprehensive collection of useful formulas can be found in Abramowitz and Stegun (1965, Chapter 22).

The Jacobi polynomials $p_n^{(a,b)}(z)$, for $z \in [-1, 1]$, satisfy the differential equation

$$(1 - z^2) \frac{d^2 f(z)}{dz^2} + [b - a - (a+b+2)z] \frac{df(z)}{dz} + n(n+a+b+1) f(z) = 0. \quad (7)$$

For given $a, b > -1$, the set $\left\{ p_n^{(a,b)}(z) \right\}_{n=0}^{\infty}$ forms an orthogonal system on the interval $[-1, 1]$ with respect to the weight function $(1 - z)^a (1 + z)^b$. For a more convenient correspondence with the diffusion parameters, we define the following modified Jacobi polynomials, for $x \in [0, 1]$ and $a, b > 0$:

$$R_n^{(a,b)}(x) = p_n^{(b-1,a-1)}(2x - 1).$$

This definition is slightly different from that adopted by Griffiths and Spanò (2010).

Equation (7) implies that the modified Jacobi polynomials $R_n^{(a,b)}(x)$, for $x \in [0, 1]$, satisfy the differential equation

$$x(1 - x) \frac{d^2 f(x)}{dx^2} + [a - (a+b)x] \frac{df(x)}{dx} + n(n+a+b-1) f(x) = 0. \quad (8)$$

For fixed *a, b* > 0, the set $\left\{R_n^{(a,b)}(x)\right\}_{n=0}^{\infty}$ forms an orthogonal system on [0, 1] with respect to the weight function $x^{a-1}(1-x)^{b-1}$. More precisely,

$$\int_0^1 R_n^{(a,b)}(x)R_m^{(a,b)}(x)x^{a-1}(1-x)^{b-1}dx=\delta_{n,m}c_n^{(a,b)},$$

where $\delta_{n,m}$ denotes the Kronecker delta and

$$c_n^{(a,b)}=\frac{\Gamma(n+a)\Gamma(n+b)}{(2n+a+b-1)\Gamma(n+a+b-1)\Gamma(n+1)}. \quad (9)$$

Note that $\left\{R_n^{(a,b)}(x)\right\}_{n=0}^{\infty}$ form a complete basis of the Hilbert space $L^2([0,1], x^{a-1}(1-x)^{b-1})$.

For *n* ≥ 1, the modified Jacobi polynomial $R_n^{(a,b)}(x)$ satisfies the recurrence relation

$$xR_n^{(a,b)}(x)=\frac{(n+a-1)(n+b-1)}{(2n+a+b-1)(2n+a+b-2)}R_{n-1}^{(a,b)}(x)+\left[\frac{1}{2}-\frac{b^2-a^2-2(b-a)}{2(2n+a+b)(2n+a+b-2)}\right]R_n^{(a,b)}(x)+\frac{(n+1)(n+a+b-1)}{(2n+a+b)(2n+a+b-1)}R_{n+1}^{(a,b)}(x), \quad (10)$$

while, for *n* = 0,

$$xR_0^{(a,b)}(x)=\frac{a}{a+b}R_0^{(a,b)}(x)+\frac{1}{a+b}R_1^{(a,b)}(x). \quad (11)$$

Also, note that $R_0^{(a,b)}(x) \equiv 1$. These recurrence relations play an important role in the work of Song and Steinrücken (2012), and the multivariate analogues, discussed later in Section 3.2, are similarly important for the present work.

The modified Jacobi polynomials satisfy other interesting relations, one of them being the following:

$$R_n^{(a,b)}(x)=\frac{n+a+b-1}{2n+a+b-1}R_n^{(a,b+1)}(x) - \mathbb{1}_{\{n>0\}}\frac{n+a-1}{2n+a+b-1}R_{n-1}^{(a,b+1)}(x). \quad (12)$$

Using this identity, polynomials with parameter *b* can be related to polynomials with parameter *b* + 1. We utilize this relation later.

2.4. A review of the K = 2 case

To motivate the approach to be employed in the general case, we briefly review the work of Song and Steinrücken (2012) for deriving the transition density function in the diallelic (*K* = 2) case. The vector of mutation rates is given by $\theta = (\alpha, \beta)$, while the symmetric matrix describing the general diploid selection scheme can be parametrized as

$$\sigma=\begin{pmatrix} 2\sigma & 2\sigma h \\ 2\sigma h & 0 \end{pmatrix},$$

where σ is the selection strength and *h* the dominance parameter. For *K* = 2, the diffusion is one dimensional and the simplex Δ_1 is equal to the unit interval [0, 1]. With *x* denoting x_1, the generator (1) reduces to

$$\mathscr{L} f(x) = \frac{1}{2} x(1-x) \frac{\partial^2}{\partial x^2} f(x) + \left\{ \frac{1}{2} [\alpha - (\alpha+\beta)x] + 2\sigma x(1-x)[x + h(1-2x)] \right\} \frac{\partial}{\partial x} f(x).$$

In the neutral case (i.e., $\sigma = 0$), the modified Jacobi polynomials $R_n^{(\alpha,\beta)}(x)$ are eigenfunctions of the diffusion generator with eigenvalues $\lambda_n^{(\alpha,\beta)} = \frac{1}{2} n(n-1+\alpha+\beta)$. Hence, a spectral representation of the transition density function can be readily obtained via (6).

In the non-neutral case (i.e., $\sigma \neq 0$), consider the functions $S_n^\theta(x) = e^{-\bar{\sigma}(x)/2} R_n^{(\alpha,\beta)}(x)$, which form an orthogonal basis of the Hilbert space $L^2([0, 1], e^{\bar{\sigma}(x)} x^{\alpha-1}(1-x)^{\beta-1})$, where $e^{\bar{\sigma}(x)} x^{\alpha-1}(1-x)^{\beta-1}$ corresponds to the stationary distribution of the non-neutral diffusion, up to a multiplicative constant. Since the eigenfunctions $B_n(x)$ of the diffusion generator are elements of this Hilbert space, we can pose an expansion $B_n(x) = \sum_{m=0}^\infty w_{n,m} S_m^\theta(x)$ in terms of the basis functions $S_n^\theta(x)$, where $w_{n,m}$ are to be determined. Then, the eigenvalue equation $\mathscr{L} B_n(x) = -\Lambda_n B_n(x)$ implies the algebraic equation

$$\sum_{m=0}^\infty w_{n,m} [\lambda_m^{(\alpha,\beta)} + Q(x;\alpha, \beta, \sigma, h)] R_m^{(\alpha,\beta)}(x) = \Lambda_n \sum_{m=0}^\infty w_{n,m} R_m^{(\alpha,\beta)}(x),$$

where $Q(x; \alpha, \beta, \sigma, h)$ is a polynomial in x of degree four. Utilizing the recurrence relations in (10) and (11), one can then arrive at a linear system $M w_n = \Lambda_n w_n$, where $w_n = (w_{n,0}, w_{n,1}, w_{n,2}, \ldots)$ is an infinite-dimensional vector of variables and M is a sparse infinite-dimensional matrix with entries that depend on the index n and the parameters α, β, σ, h of the model. The infinite linear system $M w_n = \Lambda_n w_n$ is approximated by a finite-dimensional truncated linear system

$$M^{[D]} w_n^{[D]} = \Lambda_n^{[D]} w_n^{[D]},$$

where $w_n^{(D)} = (w_{n,0}^{[D]}, w_{n,1}^{[D]}, \ldots, w_{n,D-1}^{[D]})$ and $M^{[D]}$ is the submatrix of M consisting of its first D rows and D columns. This finite-dimensional linear system can be easily solved using standard linear algebra to obtain the eigenvalues $\Lambda_n^{[D]}$ and the eigenvectors $w_n^{[D]}$ of $M^{[D]}$. Song and Steinrücken observed that $\Lambda_n^{[D]}$ and $w_{n,m}^{[D]}$ converge very rapidly as the truncation level D increases. Finally, the coefficients $w_{n,m}^{[D]}$ can be used to approximate the eigenfunctions $B_n(x)$, and, together with the eigenvalues $\Lambda_n^{[D]}$, an efficient approximation of the transition density function can be obtained via (6).

3. The Neutral Case with an Arbitrary Number of Alleles

In this section, we describe the spectral representation of the transition density of a neutral PIM model with an arbitrary number K of alleles. As in the case of $K = 2$, reviewed in Section 2.4, for an arbitrary K the eigenfunctions in the neutral case can be used to construct the eigenfunctions in the case with selection. The latter case is considered in Section 4.

3.1. Multivariate Jacobi polynomials

In what follows, let $\mathbb{N}_0 = \{0, 1, 2, \ldots\}$ denote the set of non-negative integers. As in Griffiths and Spanò (2011), we define the following system of multivariate orthogonal polynomials in $K - 1$ variables:

Definition 1. *For each vector* $\mathbf{n} = (n_1, \ldots, n_{K-1}) \in \mathbb{N}_0^{K-1}$ *and* $\theta = (\theta_1, \ldots, \theta_K) \in \mathbb{R}_{\geq 0}^K$, *the orthogonal polynomial* $P_{\mathbf{n}}^{\theta}(\mathbf{x})$ *is defined as*

$$P_{\mathbf{n}}^{\theta}(\mathbf{x}) = \prod_{j=1}^{K-1} \left[\left(1 - \frac{x_j}{1 - \sum_{i=1}^{j-1} x_i} \right)^{N_j} R_{n_j}^{(\theta_j, \Theta_j + 2N_j)} \left(\frac{x_j}{1 - \sum_{i=1}^{j-1} x_i} \right) \right],$$

where $N_j = \sum_{i=j+1}^{K-1} n_i$ *and* $\Theta_j = \sum_{i=j+1}^{K} \theta_i$.

For $\mathbf{x} = (x_1, \ldots, x_{K-1}) \in \Delta_{K-1}$, let $\Pi_0(\mathbf{x})$ denote an unnormalized density of the Dirichlet distribution with parameter $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_K)$:

$$\Pi_0(\mathbf{x}) = \prod_{i=1}^{K} x_i^{\theta_i - 1}, \quad (13)$$

where $x_K = 1 - |\mathbf{x}|$. The following lemma, a proof of which is provided in Appendix A, states that the above multivariate Jacobi polynomials $P_{\mathbf{n}}^{\theta}(\mathbf{x})$ are orthogonal with respect to $\Pi_0(\mathbf{x})$:

Lemma 2. *For* $\mathbf{n}, \mathbf{m} \in \mathbb{N}_0^{K-1}$,

$$\int_{\Delta_{K-1}} P_{\mathbf{n}}^{\theta}(\mathbf{x}) \, P_{\mathbf{m}}^{\theta}(\mathbf{x}) \Pi_0(\mathbf{x}) d\mathbf{x} = \delta_{\mathbf{n},\mathbf{m}} C_{\mathbf{n}}^{\theta},$$

where $\delta_{\mathbf{n},\mathbf{m}} = \prod_{i=1}^{K-1} \delta_{n_i, m_i}$ *and*

$$C_{\mathbf{n}}^{\theta} := \prod_{i=1}^{K-1} c_{n_i}^{(\theta_i, \Theta_i + 2N_i)}, \quad (14)$$

with $C_n^{(a,b)}$ *defined in* (9).

Remark: The multivariate Jacobi polynomials form a complete basis of $L^2(\Delta_{K-1}, \Pi_0(\mathbf{x}))$, the Hilbert space of functions on Δ_{K-1} square integrable with respect to the unnormalized Dirichlet density $\Pi_0(\mathbf{x})$.

3.2. Recurrence relation for multivariate Jacobi polynomials

Recall that the univariate Jacobi polynomials satisfy the recurrence relation (10). Theorem 3.2.1 of Dunkl and Xu (2001) guarantees that the multivariate Jacobi polynomials satisfy a similar recurrence relation. More precisely, we have the following lemma, the proof of which is provided in Appendix A:

Lemma 3. *Given* $\mathbf{n}=(n_1,\ldots,n_{K-1}) \in \mathbb{N}_0^{K-1}$ *and* $\mathbf{m}=(m_1,\ldots,m_{K-1}) \in \mathbb{N}_0^{K-1}$, *define*
$N_j=\sum_{i=j+1}^{K-1} n_i$ *and* $\mathbf{M}_j=\sum_{i=j+1}^{K-1} m_i$. *For given* $i \in \{1,\ldots,K-1\}$ *and* $\mathbf{n}$, $P_{\mathbf{n}}^{\theta}(\mathbf{x})$ *satisfies the recurrence relation*

$$x_i P_{\mathbf{n}}^{\theta}(\mathbf{x})= \sum_{\mathbf{m}\in \mathscr{M}_i(\mathbf{n})} r_{\mathbf{n},\mathbf{m}}^{(\theta,i)} P_{\mathbf{m}}^{\theta}(\mathbf{x}), \quad (15)$$

where $r_{\mathbf{n},\mathbf{m}}^{(\theta,i)}$ *are known constants (provided in Appendix A) and*

$$\mathscr{M}_i(\mathbf{n}) := \left\{ \mathbf{m} \in \mathbb{N}_0^{K-1}: M_j=N_j \text{ for all } j>i \text{ and } |M_j - N_j| \leq 1 \text{ for all } j \leq i\right\}. \quad (16)$$

Impose an ordering on the $(K-1)$-dimensional index vectors $n \in \mathbb{N}_0^{K-1}$. Then, the recurrence (15) can be represented as

$$x_i P_{\mathbf{n}}^{\theta}(\mathbf{x})= \sum_{\mathbf{m}\in \mathscr{M}_i(\mathbf{n})} [\mathscr{G}_i^{\theta}]_{\mathbf{n},\mathbf{m}} P_{\mathbf{m}}^{\theta}(\mathbf{x}),$$

where $\mathscr{G}_i^{\theta}$ corresponds to an infinite dimensional matrix in which columns and rows are indexed by the ordered $(K-1)$-tuples, and the $(\mathbf{n},\mathbf{m})$-th entry is defined as

$$[\mathscr{G}_i^{\theta}]_{\mathbf{n},\mathbf{m}}= \begin{cases} r_{\mathbf{n},\mathbf{m}}^{(\theta,i)} & \text{if } \mathbf{m} \in \mathscr{M}_i(\mathbf{n}), \\ 0, & \text{otherwise.} \end{cases} \quad (17)$$

Note that for each given $\mathbf{n}$, the number of non-zero entries in every row of $\mathscr{G}_i^{\theta}$ is finite. One can deduce the following corollary from the new representation:

Corollary 4. *Let* $a=(a_{\mathbf{n}})_{\mathbf{n}\in\mathbb{N}_0^{K-1}}$ *be such that* $\sum_{\mathbf{n}\in\mathbb{N}_0^{K-1}} a_{\mathbf{n}}^2 C_{\mathbf{n}}^{\theta}<\infty$. *Then*

$$x_i \cdot \sum_{\mathbf{n}\in\mathbb{N}_0^{K-1}} a_{\mathbf{n}} P_{\mathbf{n}}^{\theta}(\mathbf{x})= \sum_{\mathbf{n}\in\mathbb{N}_0^{K-1}} b_{\mathbf{n}} P_{\mathbf{n}}^{\theta}(\mathbf{x}),$$

where $(b_{\mathbf{n}})_{\mathbf{n}\in\mathbb{N}_0^{K-1}}=a \cdot \mathscr{G}_i^{\theta}$.

Remark: Since under multiplication x_i commutes with x_j for $1 \quad i, j \quad K-1$, the corresponding matrices $\mathscr{G}_i^{\theta}$ and $\mathscr{G}_j^{\theta}$ also commute.

3.3. Eigenfunctions of the neutral generator $\mathcal{L}_0$

It is well known that the stationary distribution of the Wright-Fisher diffusion under a *neutral* PIM model is the Dirichlet distribution (Wright, 1949). The density of the Dirichlet distribution is a weight function with respect to which the associated diffusion generator $\mathcal{L}_0$ is symmetric. As discussed in Section 3.1, the multivariate Jacobi polynomials $P_{\mathbf{n}}^{\theta}(\mathbf{x})$ are orthogonal with respect to the weight function $\Pi_0(\mathbf{x})$, which is equal to the density of the Dirichlet distribution up to a multiplicative constant. Given the discussion in Section 2.2, one might then suspect that $P_{\mathbf{n}}^{\theta}(\mathbf{x})$ are potential eigenfunctions of $\mathcal{L}_0$. The following lemma establishes that this is indeed the case:

Lemma 5. *For all* $\mathbf{n} \in \mathbb{N}_0^{K-1}$, *the multivariate Jacobi polynomials* $P_{\mathbf{n}}^{\theta}(\mathbf{x})$ *satisfy*

$$\mathscr{L}_0 P_{\mathbf{n}}^{\theta}(\mathbf{x}) = -\lambda_{|\mathbf{n}|}^{\theta} P_{\mathbf{n}}^{\theta}(\mathbf{x}),$$

where

$$\lambda_{|\mathbf{n}|}^{\theta} = \frac{1}{2}|\mathbf{n}|(|\mathbf{n}| - 1 + |\theta|). \quad (18)$$

That is, $P_{\mathbf{n}}^{\theta}(\mathbf{x})$ *are eigenfunctions of* $\mathscr{L}_0$ *with eigenvalues* $-\lambda_{|\mathbf{n}|}^{\theta}$.

A proof of this lemma is deferred to Appendix A. We conclude this section with a few comments.

Remarks:

    i. Substituting the eigenvalues and eigenfunctions into the spectral representation (6), we obtain

$$p(t; \mathbf{x}, \mathbf{y}) = \sum_{\mathbf{n} \in \mathbb{N}_0^{K-1}} \frac{1}{C_{\mathbf{n}}^{\theta}} e^{-\lambda_{|\mathbf{n}|}^{\theta} t} P_{\mathbf{n}}^{\theta}(\mathbf{x}) P_{\mathbf{n}}^{\theta}(\mathbf{y}) \Pi_0(\mathbf{y}). \quad (19)$$

    ii. For every $\mathbf{n} \in \mathbb{N}_0^{K-1}$, note that $-\lambda_{|\mathbf{n}|}^{\theta}$ only depends on the norm $|\mathbf{n}|$, which implies degeneracy in the spectrum of $\mathscr{L}_0$. Griffiths (1979) constructed orthogonal kernel polynomials indexed by $|\mathbf{n}|$, that is the sum over all orthogonal polynomials with index summing to $|\mathbf{n}|$, and obtained the transition density expansion (19).

4. A General Diploid Selection Case with an Arbitrary Number of Alleles

In this section, we derive the spectral representation of the transition density function of the Wright-Fisher diffusion under a *K*-allelic PIM model with general diploid selection. This work extends the work of Song and Steinrücken (2012), the special case of $K = 2$ briefly summarized in Section 2.4, to an arbitrary number *K* of alleles. The recurrence relation presented in Lemma 3 plays a crucial role in the following derivation.

Recall that the backward generator for the full model is $\mathscr{L} = \mathscr{L}_0 + \mathscr{L}_{\sigma}$, where $\mathscr{L}_0$ corresponds to the generator under neutrality and $\mathscr{L}_{\sigma}$ corresponds to the contribution from selection. The diffusion has a unique stationary density [see Ethier and Kurtz (1994) or Barbour et al. (2000)] proportional to

$$\Pi(\mathbf{x}) := e^{\bar{\sigma}(\mathbf{x})} \Pi_0(\mathbf{x}), \quad (20)$$

where $\Pi_0(\mathbf{x})$ is defined in (13) and $\bar{\sigma}(\mathbf{x})$ is the mean fitness defined in (3). As mentioned in Section 2.2, $\mathscr{L}$ is symmetric with respect to $\Pi(\mathbf{x})$. For $n \in \mathbb{N}_0$, we aim to find the eigenvalues $-\Lambda_n$ and the eigenfunctions B_n of $\mathscr{L}$ such that

$$\mathscr{L} B_n(\mathbf{x}) = -\Lambda_n B_n(\mathbf{x}). \quad (21)$$

By convention, we place Λ_n in non-decreasing order. The symmetry of $\mathscr{L}$ implies that $\{B_n(\mathbf{x})\}$ form an orthogonal system with respect to $\Pi(\mathbf{x})$, that is

$$\int_{\Delta_{K-1}} B_n(\mathbf{x})B_m(\mathbf{x})\Pi(\mathbf{x})d\mathbf{x} \propto \delta_{n,m}.$$

Such a system of orthogonal functions, however, is not unique. The orthogonality of $\{P_{\mathbf{n}}^\theta\}$ with respect to Π_0, established in Lemma 2, can be used to show that the functions

$$S_{\mathbf{n}}^\theta(\mathbf{x}) := P_{\mathbf{n}}^\theta(\mathbf{x})e^{-\bar{\sigma}(\mathbf{x})/2} \quad (22)$$

are orthogonal with respect to Π, as are $B_n(\mathbf{x})$. Furthermore, the fact that $\{P_{\mathbf{n}}^\theta(\mathbf{x})\}$ form a complete basis of $L^2(\Delta_{K-1}, \Pi_0(\mathbf{x}))$ means that $\{S_{\mathbf{n}}^\theta(\mathbf{x})\}$ is a complete basis of $L^2(\Delta_{K-1}, \Pi(\mathbf{x}))$. Since $B_n \in L^2(\Delta_{K-1}, \Pi(\mathbf{x}))$, we thus seek to represent $B_n(\mathbf{x})$ as linear combination of the basis $S_{\mathbf{n}}^\theta(\mathbf{x})$:

$$B_n(\mathbf{x}) = \sum_{\mathbf{m}\in\mathbb{N}_0^{K-1}} u_{n,\mathbf{m}} S_{\mathbf{m}}^\theta(\mathbf{x}), \quad (23)$$

where $u_{n,\mathbf{m}}$ are some constants to be determined.

Define an index set $\mathscr{I} = \cup_{L=0}^4 \{1, \ldots, K-1\}^L$ and for $\mathbf{i} = (i_1, \ldots, i_L) \in \mathscr{I}$, define $x_{\mathbf{i}} = x_{i_1} \cdots x_{i_L}$. We have the following theorem for solving the eigensystem associated with the full generator $\mathcal{L}$:

Theorem 6. *For all $n \in \mathbb{N}_0$, the eigenfunction $B_n(\mathbf{x})$ of $\mathcal{L}$ can be represented by* (23). *The corresponding eigenvalues $-\Lambda_n$ and the coefficients $u_{n,\mathbf{m}}$ can be found by solving the infinite-dimensional eigensystem*

$$\mathbf{u}_n\mathbf{M} = \mathbf{u}_n\Lambda_n, \quad (24)$$

where $\mathbf{u}_n = (u_{n,\mathbf{m}})_{\mathbf{m}\in\mathbb{N}_0^{K-1}}$ *and*

$$\mathbf{M} = \mathrm{diag}\left(\{\lambda_{|\mathbf{m}|}^\theta\}_{\mathbf{m}\in\mathbb{N}_0^{K-1}}\right) + \sum_{i\in\mathscr{I}} q(\mathbf{i})\mathscr{G}_{\mathbf{i}}^\theta.$$

Here, $\lambda_{|\mathbf{m}|}^\theta$ is defined as in (18) *and, for* $\mathbf{i} = (i_1, \ldots, i_L) \in \mathscr{I}$, *we define* $\mathscr{G}_{\mathbf{i}}^\theta = \mathscr{G}_{i_1}^\theta \cdots \mathscr{G}_{i_L}^\theta$, *where $\mathscr{G}_i^\theta$ is given in* (17). *When $L = 0$, $\mathscr{G}_{\mathbf{i}}^\theta$ is defined to be the identity matrix. Explicit expressions of the constants $q(\mathbf{i})$ are provided in* Appendix C.

Remarks:

   i. Although $\mathbf{M}$ is infinite dimensional, it is in fact sparse with only finitely many non-zero entries in every row and column.

   ii. Equation (24) implies that Λ_n and $\mathbf{u}_n$ are in fact the left eigenvalues and eigenvectors of $\mathbf{M}$. To solve the eigensystem in practice requires some truncation of the matrix to finite dimensions, as in the $K = 2$ case described in Section 2.4. For a given $n \in \mathbb{N}_0$, we would like both Λ_n and $u_{n,\mathbf{m}}$ to converge as the truncation level increases. In Section 5, we demonstrate that this is indeed the case using empirical examples.

We now provide a proof of the theorem.

Proof of Theorem 6. Substituting (23) into (21) we obtain

$$\sum_{\mathbf{k}\in\mathbb{N}_0^{K-1}} u_{n,\mathbf{k}} \mathscr{L} S_{\mathbf{k}}^{\theta}(\mathbf{x}) = -\sum_{\mathbf{k}\in\mathbb{N}_0^{K-1}} \Lambda_n u_{n,\mathbf{k}} S_{\mathbf{k}}^{\theta}(\mathbf{x}).$$

It is shown in Appendix D that

$$\mathscr{L} S_{\mathbf{k}}^{\theta}(\mathbf{x}) = -e^{-\bar{\sigma}(\mathbf{x})/2}\, [\lambda_{|\mathbf{k}|}^{\theta} P_{\mathbf{k}}^{\theta}(\mathbf{x}) + Q(\mathbf{x};\sigma,\theta) P_{\mathbf{k}}^{\theta}(\mathbf{x})], \quad (25)$$

where

$$Q(\mathbf{x};\sigma,\theta) = \frac{1}{2}\left[\sum_{i=1}^{K} x_i \sigma_i^2(\mathbf{x}) + \sum_{i=1}^{K} \theta_i \sigma_i(\mathbf{x}) + \sum_{i=1}^{K} x_i \sigma_{i,i} - (1+|\theta|)\bar{\sigma}(\mathbf{x}) - \bar{\sigma}(\mathbf{x})^2\right],$$

with $\sigma_i(\mathbf{x})$ and $\bar{\sigma}(\mathbf{x})$ defined as in (2) and (3), respectively. Thus, one arrives at the following equation:

$$\sum_{\mathbf{k}\in\mathbb{N}_0^{K-1}} \Lambda_n u_{n,\mathbf{k}} P_{\mathbf{k}}^{\theta}(\mathbf{x}) = \sum_{\mathbf{k}\in\mathbb{N}_0^{K-1}} u_{n,\mathbf{k}}\, [\lambda_{|\mathbf{k}|}^{\theta} P_{\mathbf{k}}^{\theta}(\mathbf{x}) + Q(\mathbf{x};\sigma,\theta) P_{\mathbf{k}}^{\theta}(\mathbf{x})]. \quad (26)$$

We solve the equation by first representing $Q(\mathbf{x}; \sigma, \theta) P_{\mathbf{k}}^{\theta}(\mathbf{x})$ as a finite linear combination of $\left\{ P_{\mathbf{k}}^{\theta}(\mathbf{x}) \right\}_{\mathbf{k}\in\mathbb{N}_0^{K-1}}$. Observe that Q is in fact a fourth-order polynomial in $\mathbf{x}$. Collecting terms, Q can be written in the form

$$Q(\mathbf{x};\sigma,\theta) = \sum_{\mathbf{i}\in\mathscr{I}} q(\mathbf{i})\mathbf{x_i}, \quad (27)$$

for the constants $q(\mathbf{i})$ given in Appendix C. Applying Corollary 4 recursively, we obtain

$$Q(\mathbf{x};\sigma,\theta)\, P_{\mathbf{k}}^{\theta}(\mathbf{x}) = \sum_{\mathbf{i}\in\mathscr{I}} q(\mathbf{i}) \sum_{\mathbf{l}\in\mathbb{N}_0^{K-1}} [\mathscr{G}_{\mathbf{i}}^{\theta}]_{\mathbf{k},\mathbf{l}} P_{\mathbf{l}}^{\theta}(\mathbf{x}).$$

Finally, substituting this equation into (26), multiplying both sides of (26) by $P_{\mathbf{m}}^{\theta}(\mathbf{x})$, and integrating with respect to $\Pi_0(\mathbf{x})$ over the simplex Δ_{K-1} yields the matrix equation (24).

5. Empirical Results and Applications

In this section, we study the convergence behavior of the eigenvalues and eigenvectors as we approximate the solutions of (24). Further, we show how the spectral representation can be employed to obtain the transient and stationary density explicitly (especially the normalizing constant), and to characterize the convergence rate of the diffusion to stationarity. A Mathematica implementation of the relevant formulas for computing the spectral representation is available from the authors upon request.

5.1. Convergence of the eigenvalues and eigenvectors

In what follows we order the Jacobi polynomials according to the *graded lexicographic ordering* of their corresponding indices. Thus $P^\theta_{\mathbf{n}_1} < P^\theta_{\mathbf{n}_2}$ if

- $|\mathbf{n}_1| < |\mathbf{n}_2|$, or

- $|\mathbf{n}_1| = |\mathbf{n}_2|$ and $\mathbf{n}_1$ is lexicographically smaller than $\mathbf{n}_2$.

Fix K and note that, for a given *truncation level* $D \in \mathbb{N}_0$ and $l \in \mathbb{N}_0$, there are $\binom{l+K-2}{K-2}$ polynomials $P^\theta_{\mathbf{n}}$ with $|\mathbf{n}| = l$, and $\mathcal{U}(D) := \binom{D+K-1}{K-1}$ polynomials with index $|\mathbf{n}| \le D$. For the computations in the rest of this section we chose $K = 3$, unless otherwise stated.

Now, one can obtain a finite-dimensional linear system approximating (24) by truncation, that is, taking only those entries in $\mathbf{M}$ and $\mathbf{u}_n$ whose associated index vectors satisfy $|\mathbf{n}| \le D$. More explicitly, with $\mathbf{M}^{[D]} = ([\mathbf{M}]_{\mathbf{k},\mathbf{l}}) \in \mathbb{R}^{\mathcal{U}(D) \times \mathcal{U}(D)}$ and $\mathbf{u}_n^{[D]} = (u_{n,\mathbf{k}}) \in \mathbb{R}^{\mathcal{U}(D)}$, where $\mathbf{k}, \mathbf{l} \in \mathbb{N}_0^{K-1}$ such that $|\mathbf{k}|, |\mathbf{l}| \le D$, the solutions of

$$\mathbf{u}_n^{[D]} \mathbf{M}^{[D]} = \mathbf{u}_n^{[D]} \Lambda_n^{[D]}$$

should approximate the solutions of the infinite system Λ_n and $\mathbf{u}_n$. The convergence patterns of $\Lambda_n^{[D]}$ and $u_{n,\mathbf{k}}^{[D]}$ as D increases are exemplified in Figure 1 for the parameters

i)
$$K = 3,\ \boldsymbol{\theta} = (0.01, 0.02, 0.03),\quad \sigma = \sigma_1 := \begin{pmatrix} 12 & 14 & 15 \\ 14 & 11 & 13 \\ 15 & 13 & 0 \end{pmatrix}; \text{ and}$$

ii)
$$K = 3,\ \boldsymbol{\theta} = (0.01, 0.02, 0.03),\quad \sigma = \sigma_2 := \begin{pmatrix} 120 & 140 & 150 \\ 140 & 110 & 130 \\ 150 & 130 & 0 \end{pmatrix}.$$

Figure 2 displays the convergence behavior for the parameters

iii) $K = 3,\ \boldsymbol{\theta} = (10, 20, 30),\ \boldsymbol{\sigma} = \boldsymbol{\sigma}_1$; and

iv)
$$K = 4,\ \boldsymbol{\theta} = (0.01, 0.02, 0.03, 0.04),\quad \sigma_3 = \begin{pmatrix} 12 & 14 & 15 & 16 \\ 14 & 11 & 10 & 13 \\ 15 & 10 & 9 & 14 \\ 16 & 13 & 14 & 0 \end{pmatrix}.$$

In all cases, $\Lambda_n^{[D]}$ and $u_{n,\mathbf{k}}^{[D]}$ converge with increasing truncation level to empirical limits. The eigenvalues $\Lambda_n^{[D]}$ decrease towards the empirical limit, whereas the coefficients $u_{n,\mathbf{k}}^{[D]}$ show oscillatory behavior before ultimately stabilizing. The rate of convergence is faster for smaller selection intensity. Varying the mutation parameters does not influence convergence behavior significantly. As expected, $\Lambda_0^{[D]}$ converges rapidly to zero in all cases, consistent with the fact that the diffusion has a stationary distribution. For a fixed n, $\Lambda_n^{[D]}$ and its associated coefficients $u_{n,\mathbf{k}}^{[D]}$ roughly converge at similar truncation levels.

Figure 3 shows $\Lambda_n^{[D]}$ for $D = 24$ and $0 \leq n \leq 35$ under neutrality ($\boldsymbol{\sigma} = \mathbf{0}$) and selection ($\boldsymbol{\sigma} = \boldsymbol{\sigma}_1$ and $\sigma = \frac{1}{4}\sigma_2$). Upon inspection all of the eigenvalues displayed have converged properly. Under neutrality, the eigenvalues are functions of $|\mathbf{n}|$, thus they are degenerate and cluster into groups. In the presence of selection, however, we empirically observe that all of the eigenvalues are distinct. For moderate selection intensity, the group structure is less prominent. In general, increasing the selection parameters evens out the group structure and shifts the entire spectrum upward.

Computing the transition density function for large selection coefficients requires combining terms of substantially different orders of magnitude, because of the exponential weighting factors in the density (20) and in the expansion (22). Therefore, the coefficients $u_{n,\mathbf{k}}^{[D]}$ have to be calculated with high precision to obtain accurate numerical results under strong selection.

5.2. Transient and stationary densities

The approximations to the eigenvalues $\Lambda_n^{[D]}$ and the eigenfunctions B_n (via the eigenvectors $\mathbf{u}_n^{[D]}$ and equation (23)) can be used in the spectral representation (6) to approximate the transition density function at arbitrary times t. Examples with $\boldsymbol{\sigma} = \boldsymbol{\sigma}_1$ for different times are given in Figure 4. At first, the density is concentrated around the initial frequencies $\mathbf{x} = (0.02, 0.02, 0.96)$, but as time increases, the frequencies of the first and second allele increases, since these have a higher relative fitness. Eventually, the transition density converges to the stationary distribution (similar to distribution at $t = 2$), where the bulk of the mass is concentrated at high frequencies for the first and second allele.

The eigenvalues $\Lambda_n^{[D]}$ and coefficients $\mathbf{u}_{n,\mathbf{k}}^{[D]}$ can also be employed to approximate the constant that normalizes the stationary distribution $\Pi(\mathbf{x})$ to a proper probability distribution. Following the same line of argument as Song and Steinrücken (2012), the orthogonal relations enable us to circumvent the difficulty involved in directly evaluating a multivariate integral over the simplex Δ_{K-1}. First, note that since $\mathcal{L}$ maps constant functions to zero, any constant function is an eigenfunction with associated eigenvalue $\Lambda_0 = 0$, thus $B_0(\mathbf{x}) = B_0(\mathbf{y})$ = const. In (6), taking $t \to \infty$, we get

$$\lim_{t \to \infty} p(t; \mathbf{x}, \mathbf{y}) = \Pi(\mathbf{y}) \frac{B_0(\mathbf{x})B_0(\mathbf{y})}{\langle B_0, B_0 \rangle_\Pi} =: \frac{1}{C_\Pi}\Pi(\mathbf{y}).$$

Then for $\mathbf{x} = \mathbf{y} = \mathbf{0}$, by (23) we have

$$C_\Pi = \int_{\Delta_{K-1}} \Pi(\mathbf{z})\, d\mathbf{z} = \frac{\langle B_0, B_0 \rangle_\Pi}{B_0(\mathbf{0})^2} = \frac{\sum_{\mathbf{m} \in \mathbb{N}_0^{K-1}} u_{0,\mathbf{m}}^2 \langle P_\mathbf{m}^\theta, P_\mathbf{m}^\theta \rangle_{\Pi_0}}{e^{-\bar{\sigma}(\mathbf{0})}\left(\sum_{\mathbf{m} \in \mathbb{N}_0^{K-1}} u_{0,\mathbf{m}} P_\mathbf{m}^\theta(\mathbf{0})\right)^2} = \frac{\sum_{\mathbf{m} \in \mathbb{N}_0^{K-1}} u_{0,\mathbf{m}}^2 C_\mathbf{m}^\theta}{\left(\sum_{\mathbf{m} \in \mathbb{N}_0^{K-1}} u_{0,\mathbf{m}} \prod_{j=1}^{K-1} \frac{\Gamma(n_j+\theta_j)}{\Gamma(n_j+1)\Gamma(\theta_j)}\right)^2}, \quad (28)$$

since $\bar{\sigma}(\mathbf{0}) = \sigma_{K,K} = 0$ and $R_{n_j}^{(\theta_j, \Theta_j + 2N_j)}(0) = (-1)^{n_j} \frac{\Gamma(n_j+\theta_j)}{\Gamma(n_j+1)\Gamma(\theta_j)}$. Here $C_\mathbf{m}^\theta$ is the constant defined in (14). The purely algebraic form of the right hand side in equation (28) allows to compute an accurate approximation of the normalizing constant C_Π by replacing the infinite sums by sums over all indices $\mathbf{m}$ such that $|\mathbf{m}|$ is less or equal then a given truncation level. This offers an attractive alternative to other computationally intensive methods (Donnelly et al., 2001; Genz and Joyce, 2003; Buzbas and Joyce, 2009).

Figure 5 shows two examples of stationary distributions for different selection coefficients. In Figure 5(a) the stationary density is concentrated in the interior of the simplex, since all homozygotes are less fit then the heterozygotes. This situation is referred to as heterozygote advantage, resulting in a balancing selection pattern, and the different alleles co-exist at stationarity. In Figure 5(b), allele number 1 is strongly favored by the given selection coefficients, and thus the stationary density is concentrated at high frequencies for this allele.

We can also use (6) to investigate the rate of convergence of the diffusion process to the stationary distribution. Denote the difference between the transition density and the stationary density by

$$d(t;\mathbf{x},\mathbf{y}) := p(t;\mathbf{x},\mathbf{y}) - \frac{1}{C_\Pi}\Pi(\mathbf{y}) = \sum_{n=1}^{\infty} e^{-\Lambda_n t}\Pi(\mathbf{y})\frac{B_n(\mathbf{x})B_n(\mathbf{y})}{\langle B_n, B_n\rangle_\Pi}.$$

We measure the magnitude of $d(t;\mathbf{x},\mathbf{y})$ by the square of its L^2 norm with respect to the weight function $1/\Pi(\mathbf{y})$, that is,

$$\|d(t;\mathbf{x},\cdot)\|_{1/\Pi}^2 := \langle d, d\rangle_{1/\Pi} = \sum_{n=1}^{\infty} e^{-2\Lambda_n t}\frac{B_n(\mathbf{x})^2}{\langle B_n, B_n\rangle_\Pi} = \sum_{n=1}^{\infty} e^{-2\Lambda_n t}\frac{e^{-\tilde{\sigma}(\mathbf{x})}\left(\sum_{\mathbf{k}\in\mathbb{N}_0^{K-1}} u_{n,\mathbf{k}}P_{\mathbf{k}}^\theta(\mathbf{x})\right)^2}{\sum_{\mathbf{m}\in\mathbb{N}_0^{K-1}} u_{n,\mathbf{m}}^2 C_{\mathbf{m}}^\theta}. \quad (29)$$

Again, the sums in this expression can be approximated by truncating at a given level.

Figure 6 shows $\|d(t;\mathbf{x},\cdot)\|_{1/\Pi}^2$ as a function of time t, for $\boldsymbol{\sigma} = \boldsymbol{\sigma}_1$, $\boldsymbol{\sigma} = 0.5\boldsymbol{\sigma}_1$, $\boldsymbol{\sigma} = 0.1\boldsymbol{\sigma}_1$. The initial frequencies were $\mathbf{x} = (0.02, 0.02, 0.96)$. As expected, the distance to the stationary distribution decreases over time. Further, the rate of convergence is faster if the values in $\boldsymbol{\sigma}$ get larger, which was observed by Song and Steinrücken (2012) too. We note that the spectral representation can also be readily employed to study convergence rates measured by other metrics such as the total variation distance or relative entropy.

6. Discussion

In this paper, we have extended the method of Song and Steinrücken (2012) to obtain an explicit spectral representation of the transition density function for the multi-dimensional Wright-Fisher diffusion under a PIM model with general diploid selection and an arbitrary number of alleles. We have demonstrated the fidelity and fast convergence of the approximations. Further, as an example application of our work, we have computed the normalization constant of the stationary distribution and quantified the rate at which the transition density approaches this distribution.

Efficient approximations of the eigensystem and the transition density function lead to a number of important applications. Combining the stationary distribution discussed in Section 5.2 with the recurrence relation shown in Lemma 3, one can calculate algebraically the probability of observing a given genetic configuration of individuals sampled from the stationary distribution of the non-neutral diffusion. This kind of algebraic approach would complement previous works (Evans et al., 2007; Živkovi and Stephan, 2011) on sample allele frequency spectra that involve solving ODEs satisfied by the moments of the diffusion. Further, the algebraic approach is potentially more efficient than computationally expensive Monte Carlo methods (Donnelly et al., 2001) and more generally applicable than methods relying on the selection coefficients being of a certain form (Genz and Joyce,

2003). Note that, by discretizing time and space, our representation of the transition density function can be used for approximate simulation of frequencies from stationarity as well as frequency trajectories, which can in turn be employed in the aforementioned Monte Carlo frameworks.

The sampling probability can be applied, for example, to estimate evolutionary parameters via maximum likelihood or Bayesian inference frameworks. Furthermore, the notion of sampling probability can be combined with the spectral representation of the transition density function in a hidden Markov model framework as in Bollback et al. (2008), to calculate the probability of observing a series of configurations sampled at different times. The method developed in this paper would allow for such an analysis in a model with multiple alleles subject to recurrent parent-independent mutation and general diploid selection.

An important, albeit very challenging, future direction is to extend our current approach to analyze the dynamics of multi-locus diffusions with recombination and selection. Such an extension would allow for the incorporation of additional data at closely linked loci, which has the potential to significantly improve the inference of evolutionary parameters, especially the strength and localization of selection. We have only considered Wright-Fisher diffusions in a single panmictic population of a constant size. As achieved in the alternative approaches of Gutenkunst et al. (2009) and Lukić et al. (2011), mentioned in Introduction, it would be desirable to generalize our approach to incorporate subdivided populations exchanging migrants, with possibly fluctuating population sizes. Another possible extension is to relax the PIM assumption and consider a more general mutation model.

We note that our present technique relies on the diffusion generator being symmetric. This symmetry does not hold in some of the scenarios mentioned above, making a direct application of the ideas developed here difficult. However, we believe that it is worthwhile investigating whether one could apply our approach to devise approximations to the transition density function that are sufficiently accurate for practical applications.

Acknowledgments

Appendix A

Proofs of lemmas

Proof of Lemma 2. For two indices $\mathbf{n}, \mathbf{m} \in \mathbb{N}_0^{K-1}$, consider the integral

$$\int_{\Delta_{K-1}} P_{\mathbf{n}}^{\theta}(\mathbf{x}) P_{\mathbf{m}}^{\theta}(\mathbf{x}) \Pi_0(\mathbf{x}) d\mathbf{x} = \int_{[0,1]^{K-1}} \tilde{P}_{\mathbf{n}}^{\theta}(\boldsymbol{\xi}) \tilde{P}_{\mathbf{m}}^{\theta}(\boldsymbol{\xi}) \Pi_0(\mathbf{x}(\boldsymbol{\xi})) \, |\det(D\mathbf{x})(\boldsymbol{\xi})| \, d\boldsymbol{\xi}, \quad \text{(A.1)}$$

where the right hand side can be obtained by the coordinate transformation introduced in Appendix B and the multivariate integration through substitution rule. Using

$$\Pi_0(\mathbf{x}(\boldsymbol{\xi})) = \prod_{i=1}^{K-1} \xi_i^{\theta_i - 1} (1 - \xi_i)^{\Theta_i - (K-i)},$$

the determinant of the Jacobian

$$|\det(D\mathbf{x})(\boldsymbol{\xi})| = \prod_{i=1}^{K-2} (1 - \xi_i)^{K-(i+1)},$$

see Baxter et al. (2007)[Equation B.1], and the transformed Jacobi polynomials (B.1), it can be shown that

$$\prod_{j=1}^{K-1} \int_0^1 R_{n_j}^{(\theta_j,\Theta_j+2N_j)}(\xi_j) R_{m_j}^{(\theta_j,\Theta_j+2M_j)}(\xi_j) \xi_j^{\theta_j-1} (1-\xi_j)^{\Theta_j+N_j+M_j-1} d\xi_j = C_{\mathbf{n}}^{\theta} \delta_{\mathbf{n},\mathbf{m}}$$

holds, with

$$C_{\mathbf{n}}^{\theta} = \prod_{j=1}^{K-1} c_{n_j}^{(\theta_j,\Theta_j+2N_j)}.$$

In the case $\mathbf{n} = \mathbf{m}$ this can be seen immediately. If $\mathbf{n} \neq \mathbf{m}$ without loss of generality let $1 \leq l \leq K-1$ be the largest l such that $n_l < m_l$ and $n_k = m_k$ for all $k = l+1, \dots, K-1$. Then $N_l = M_l$ (recall $N_{K-1} = M_{K-1} = 0$) and $R_{m_l}^{(\theta_l,\Theta_l+2M_l)}(\xi_l)$ is orthogonal to all polynomials of lesser degree with respect to the weight function $\xi_l^{\theta_l-1}(1-\xi_l)^{\Theta_l+2M_l-1}$, and thus the l-th factor and the whole product is zero.

Proof of Lemma 3. We found it most convenient to derive a recurrence relation for

$$x_i P_{\mathbf{n}}^{\theta}(\mathbf{x}) \quad \text{(A.2)}$$

by projecting expression (A.2) onto the orthogonal basis $\left\{ P_{\mathbf{m}}^{\theta}(\mathbf{x}) \right\}$, and investigate the respective coefficients. First, note that the coordinate transformation introduced in Appendix B yields $x_i = \xi_i \prod_{j<i}(1 - \xi_j)$, so

$$x_i P_{\mathbf{n}}^{\theta}(\mathbf{x}) = \xi_i \prod_{j<i} (1 - \xi_j) \tilde{P}_{\mathbf{n}}^{\theta}(\boldsymbol{\xi}). \quad \text{(A.3)}$$

Further, integrate expression (A.3) against the base function $\tilde{P}_{\mathbf{m}}^{\theta}(\xi)$ times the weight function Π_0 to get the respective coefficient in the basis representation. Using the integration by substitution rule again, as in equation (A.1), this yields

$$\frac{1}{C_{\mathbf{m}}^{\theta}} \int_{[0,1]^{K-1}} \xi_i \prod_{j<i} (1$$

$$- \xi_i) \tilde{P}_{\mathbf{n}}^{\theta}(\xi) \tilde{P}_{\mathbf{m}}^{\theta}(\xi) \prod_{k=1}^{K-1} \xi_k^{\theta_k-1} (1-\xi_k)^{\Theta_k-1} d\xi$$

$$= \prod_{j=i+1}^{K-1} \frac{1}{c_{m_j}^{(\theta_j,\Theta_j+2M_j)}} \int_0^1 R_{n_j}^{(\theta_j,\Theta_j+2N_j)}(\xi_j) R_{m_j}^{(\theta_j,\Theta_j+2M_j)}(\xi_j) \xi_j^{\theta_j-1}(1-\xi_j)^{\Theta_j+N_j+M_j-1} d\xi_j$$

$$\times \frac{1}{c_{m_i}^{(\theta_i,\Theta_i+2M_i)}} \int_0^1 R_{n_i}^{(\theta_i,\Theta_i+2N_i)}(\xi_i) R_{m_i}^{(\theta_i,\Theta_i+2M_i)}(\xi_i) \xi_i^{\theta_i}(1-\xi_j)^{\Theta_i+N_i+M_i-1} d\xi_i$$

$$\times \prod_{j=1}^{i-1} \frac{1}{c_{m_j}^{(\theta_j,\Theta_j+2M_j)}} \int_0^1 R_{n_j}^{(\theta_j,\Theta_j+2N_j)}(\xi_j) R_{m_j}^{(\theta_j,\Theta_j+2M_j)}(\xi_j) \xi_j^{\theta_j-1}(1-\xi_j)^{\Theta_j+N_j+M_j} d\xi_j.$$

The first term on the right hand side yields zero, unless $m_j = n_j$ for all $j > i$, thus $M_i = N_i$. In this case the term is equal to 1. Since $m_j = n_j$ for $j > i$, note that the second term on the right hand side is of the form

$$\frac{1}{c_{m_i}^{(\alpha,\beta)}} \int_0^1 R_{n_i}^{(\alpha,\beta)}(\xi) R_{m_i}^{(\alpha,\beta)}(\xi)\, \xi\, w_{\alpha,\beta}(\xi) d\xi = G_{n_i,m_i}^{(\alpha,\beta)} \delta_{n_i+1,m_i} + G_{n_i,m_i}^{(\alpha,\beta)} \delta_{n_i,m_i} + G_{n_i,m_i}^{(\alpha,\beta)} \delta_{n_i-1,m_i},$$

with $w_{\alpha,\beta}(\xi) = \xi^{\alpha-1}(1-\xi)^{\beta-1}$, $\alpha = \theta_i$, and $\beta = \Theta_i + 2N_i$. Here we applied the recurrence relation (10) to $\xi R_{n_i}^{(\alpha,\beta)}(\xi)$ and used the orthogonality of the Jacobi polynomials. The constants $G_{n,m}^{(a,b)}$ are given by

$$G_{n,m}^{(a,b)} = \begin{cases} \frac{(n+a-1)(n+b-1)}{(2n+a+b-1)(2n+a+b-2)}, & \text{if } n-m=1 \text{ and } n>0, \\ \frac{1}{2} - \frac{b^2-a^2-2(b-a)}{2(2n+a+b)(2n+a+b-2)}, & \text{if } n-m=0 \text{ and } n \geq 0, \\ \frac{(n+1)(n+a+b-1)}{(2n+a+b)(2n+a+b-1)}, & \text{if } n-m=-1 \text{ and } n \geq 0. \end{cases}$$

This expression is non-zero for $-1 \leq n_i - m_i \leq 1$. Furthermore, the form of the integral for $j = i - 1$ depends on this difference, or rather the difference between N_j and M_j. Depending on the difference $N_j - M_j$ we have to consider the integrals

$$-1: \frac{1}{c_{m_j}^{(\alpha,\beta+2)}} \int_0^1 R_{n_j}^{(\alpha,\beta)}(\xi) R_{m_j}^{(\alpha,\beta+2)}(\xi) w_{\alpha,\beta+2}(\xi) d\xi, \quad \text{(A.4)}$$

$$0: \frac{1}{c_{m_j}^{(\alpha,\beta)}} \int_0^1 R_{n_j}^{(\alpha,\beta)}(\xi) R_{m_j}^{(\alpha,\beta)}(\xi) (1-\xi) w_{\alpha,\beta}(\xi) d\xi, \quad \text{(A.5)}$$

$$+1: \frac{1}{c_{m_j}^{(\alpha,\beta+2)}} \int_0^1 R_{n_j}^{(\alpha,\beta)}(\xi) R_{m_j}^{(\alpha,\beta-2)}(\xi) w_{\alpha,\beta}(\xi), d\xi, \quad \text{(A.6)}$$

with $\alpha = \theta_j$ and $\beta = \Theta_j + 2N_j$. In expression (A.6) we have to assume $\beta > 2$, which is equivalent to N_j 1. This holds true, because if $N_j = 0$, this case would not have to be considered.

Applying relation (12) twice to the polynomial $R_{n_j}^{(\alpha,\beta)}(\xi)$ in equation (A.4) and using orthogonality yields

$$H_{n_j,m_j}^{(\alpha,\beta)}\delta_{n_j,m_j}+H_{n_j,m_j}^{(\alpha,\beta)}\delta_{n_j-1,m_j}+H_{n_j,m_j}^{(\alpha,\beta)}\delta_{n_j-2,m_j},$$

for some constants $H_{n,m}^{(\alpha,\beta)}$. Here $H_{0,-1}^{(\alpha,\beta)}=H_{0,-2}^{(\alpha,\beta)}=H_{1,-1}^{(\alpha,\beta)}=0$. Thus, in the case $M_j = N_j + 1$, the expression for j is non-zero for $m_j = n_j, n_j - 1$, and $n_j - 2$. Furthermore, relation (10) can be applied to the term $R_{n_j}^{(\alpha,\beta)}(\xi)(1 - \xi)$, together with orthogonality to get

$$I_{n_j,m_j}^{(\alpha,\beta)}\delta_{n_j+1,m_j}+I_{n_j,m_j}^{(\alpha,\beta)}\delta_{n_j,m_j}+I_{n_j,m_j}^{(\alpha,\beta)}\delta_{n_j-1,m_j},$$

for given constants $I_{n,m}^{(\alpha,\beta)}$, with $I_{-1,0}^{(\alpha,\beta)}=I_{0,-1}^{(\alpha,\beta)}=0$. In the case $M_j = N_j$, the expression is non-zero for $m_j = n_j - 1, n_j$, and $n_j + 1$. Finally, applying relation (12) to the term $R_{m_j}^{(\alpha,\beta-2)}(\xi)$ in expression (A.6) combined with orthogonality yields

$$J_{n_j,m_j}^{(\alpha,\beta)}\delta_{n_j,m_j}+J_{n_j,m_j}^{(\alpha,\beta)}\delta_{n_j+1,m_j}+J_{n_j,m_j}^{(\alpha,\beta)}\delta_{n_j+2,m_j},$$

for given constants $J_{n,m}^{(\alpha,\beta)}$. Again $J_{-1,0}^{(\alpha,\beta)}=J_{-2,0}^{(\alpha,\beta)}=J_{-1,1}^{(\alpha,\beta)}=0$. Thus this expression is non-zero for $m_j = n_j, n_j + 1$, and $n_j + 2$. The constants $H_{n,m}^{(a,b)}, I_{n,m}^{(a,b)}, J_{n,m}^{(a,b)}$ are given by

$$H_{n,m}^{(a,b)}=\begin{cases} \frac{(n+a+b-1)(n+a+b)}{(2n+a+b-1)(2n+a+b)}, & \text{if } m-n=0 \text{ and } n\geq 0, \\ -\frac{2a}{a+b+2}, & \text{if } m-n=-1 \text{ and } n=1 \\ -\frac{2(n+a-1)(n+a+b-1)}{(2n+a+b-2)(2n+a+b)}, & \text{if } m-n=-1 \text{ and } n>1, \\ \frac{(n+a-2)(n+a-1)}{(2n+a+b-2)(2n+a+b-1)}, & \text{if } m-n=-2 \text{ and } n>1, \end{cases}$$

$$I_{n,m}^{(a,b)}=\begin{cases} -\frac{1}{a+b}, & \text{if } m-n=1 \text{ and } n=0, \\ -\frac{(n+1)(n+a+b-1)}{(2n+a+b-1)(2n+a+b)}, & \text{if } m-n=1 \text{ and } n>0, \\ \frac{b}{a+b}, & \text{if } m-n=0 \text{ and } n=0, \\ \frac{b^2+a(b+2)}{(a+b)(a+b+2)}, & \text{if } m-n=0 \text{ and } n=1, \\ \frac{b^2+2n(n+a-1)+b(2n+a-2)}{(2n+a+b-2)(2n+a+b)}, & \text{if } m-n=0 \text{ and } n>1, \\ -\frac{ab}{(a+b)(a+b+1)}, & \text{if } m-n=-1 \text{ and } n=1 \\ -\frac{(n+a-1)(n+b-1)}{(2n+a+b-2)(2n+a+b-1)}, & \text{if } m-n=-1 \text{ and } n>1, \end{cases}$$

$$J_{n,m}^{(a,b)} = \begin{cases} \frac{(b-1)(b-2)}{(a+b-1)(a+b-2)}, & \text{if } m-n=0 \text{ and } n=0, \\ \frac{(n+b-2)(n+b-1)}{(2n+a+b-2)(2n+a+b-1)}, & \text{if } m-n=0 \text{ and } n>0, \\ -\frac{2(n+1)(n+b-1)}{(2n+a+b-2)(2n+a+b)}, & \text{if } m-n=1 \text{ and } n \geq 0, \\ \frac{(n+1)(n+2)}{(2n+a+b-1)(2n+a+b)}, & \text{if } m-n=2 \text{ and } n \geq 0. \end{cases}$$

Now considering all three possible values for $N_j - M_j$, and all possible implications for the difference $n_j - m_j$, it can be shown that $1 \quad N_{j-1} - M_{j-1} \quad 1$ has to hold as well. Using induction shows that $1 \quad N_j - M_j \quad 1$ holds for all $j < i$. Thus for all $j < i$ the same integrals (A.4), (A.5), and (A.6), with adjusted parameters $\alpha = \theta_j$ and $\beta = \Theta_j + 2N_j$, have to be considered.

Combining these results shows that for fixed i and $\mathbf{n}$ the polynomials with a non-zero contribution to the recurrence relation for $x_i P_{\mathbf{n}}^\theta(\mathbf{x})$ are exactly those with indices from the set

$$\mathscr{M}_i(\mathbf{n}) := \left\{ \mathbf{m} \in \mathbb{N}_0^{K-1} \middle| m_j \geq 0 \forall_j, M_j = N_j \forall_{j>i}, |M_j - N_j| \leq 1 \forall_{j \leq i} \right\},$$

defined in (16). Thus,

$$x_i P_{\mathbf{n}}^\theta(\mathbf{x}) = \sum_{\mathbf{m} \in \mathscr{M}_i(\mathbf{n})} r_{\mathbf{n},\mathbf{m}}^{(\theta,i)} P_{\mathbf{m}}^\theta(\mathbf{x}),$$

where the coefficients $r_{\mathbf{n},\mathbf{m}}^{(\theta,i)}$ are given by

$$r_{\mathbf{n},\mathbf{m}}^{(\theta,i)} = G_{n_i,m_i}^{(\theta_i,\Theta_i+2N_i)} \prod_{j<i} \begin{cases} H_{n_j,m_j}^{(\theta_j,\Theta_j+2N_j)}, & \text{if } N_j - M_j = -1, \\ I_{n_j,m_j}^{(\theta_j,\Theta_j+2N_j)}, & \text{if } N_j - M_j = 0, \\ J_{n_j,m_j}^{(\theta_j,\Theta_j+2N_j)}, & \text{if } N_j - M_j = +1. \end{cases}$$

Proof of Lemma 5. Using the coordinate transformation introduced in Appendix B, and applying $\mathscr{L}_0$, given in equation (B.2), to $\tilde{P}_{\mathbf{n}}^\theta(\xi)$ from equation (B.1) yields

$$\begin{aligned} \mathscr{L}_0 \tilde{P}_{\mathbf{n}}^\theta(\xi) \\ = \frac{1}{2} \sum_{i=1}^{K-1} \frac{1}{\Pi_{k<i}(1-\xi_k)} \prod_{j=1, j \neq i}^{K-1} R_{n_j}^{(\theta_j,\Theta_j+2N_j)}(\xi_j)(1-\xi_j)^{N_j} \\ \times (\xi_i(1 \\ -\xi_i)\frac{\partial^2}{\partial\xi_i^2}\{R_{n_i}^{(\theta_i,\Theta_i+2N_i)}(\xi_i)(1-\xi_i)^{N_i}\} + (\theta_i - \Theta_{i-1}\xi_i)\frac{\partial}{\partial\xi_i}\{R_{n_i}^{(\theta_i,\Theta_i+2N_i)}(\xi_i)(1-\xi_i)^{N_i}\}). \end{aligned}$$

(A.7)

Employing equation (8), one can show that the terms in the brackets on the right hand side of equation (A.7) reduce to

$$(1-\xi_i)^{N_i} R_{n_i}^{(\theta_i,\Theta_i+2N_i)}(\xi_i)\left(-N_{i-1}(N_{i-1}-1+\Theta_{i-1}) + \frac{1}{1-\xi_i}N_i(N_i-1+\Theta_i)\right),$$

and substitution yields

$$\mathscr{L}_0 \tilde{P}_{\mathbf{n}}^{\theta}(\xi) = \frac{1}{2} \tilde{P}_{\mathbf{n}}^{\theta}(\xi) \left(-\sum_{i=1}^{K-1} \frac{1}{\Pi_{k<i}(1-\xi_k)} N_{i-1}(N_{i-1}-1+\Theta_{i-1}) + \sum_{i=2}^{K} \frac{1}{\Pi_{k<i}(1-\xi_k)} N_{i-1}(N_{i-1}-1+\Theta_{i-1}) \right) = -\lambda_{|\mathbf{n}|}^{\theta} \tilde{P}_{\mathbf{n}}^{\theta}(\xi)$$

with $\lambda_{|\mathbf{n}|}^{\theta} = \frac{1}{2}|\mathbf{n}|(|\mathbf{n}|-1+|\theta|)$, since $\Theta_0 = |\mathbf{\Theta}|$, $N_0 = |\mathbf{n}|$, and $N_{K-1} = 0$.

Appendix B

Change of coordinates

Working with the multivariate Jacobi polynomials and the neutral diffusion, it is convenient, for some derivations, to transform the equations to a different coordinate system. This transformation maps the simplex Δ_{K-1} to the $K-1$-dimensional unit cube $[0, 1]^{K-1}$. It is implicitly used in Griffiths and Spanò (2011, Section 3), but more explicitly introduced and used as a transformation in Baxter et al. (2007). The vector $\mathbf{\xi}(\mathbf{x}) = (\xi_1(\mathbf{x}), \dots, \xi_{K-1}(\mathbf{x}))$ is obtained from the vector of population frequencies $\mathbf{x}$ via the transformation

$$\xi_i = \frac{x_i}{1 - \sum_{j<i} x_j}$$

for $1 \leq i \leq K-1$. The inverse of this transformation is given by

$$x_i = \xi_i \prod_{j<i}(1-\xi_j)$$

for $1 \leq i \leq K-1$. The inverse relation can be derived by noting that $1 - \sum_{j<i} x_j = \Pi_{j<i}(1-\xi_j)$ holds.

Definition 1 yields immediately that the multivariate Jacobi polynomials $P_{\mathbf{n}}^{\theta}(\mathbf{x})$ take the form

$$\tilde{P}_{\mathbf{n}}^{\theta}(\xi) = P_{\mathbf{n}}^{\theta}(\mathbf{x}(\xi)) = \prod_{j=1}^{K-1} R_{n_j}^{(\theta_j, \Theta_j + 2N_j)}(\xi_j)(1-\xi_j)^{N_j} \quad \text{(B.1)}$$

in the transformed coordinates. The neutral diffusion generator $\mathscr{L}_0$ in the transformed coordinate system is given by the following lemma.

Lemma 7. *Using variables in the new coordinate system, the backward generator of the diffusion under neutrality $\mathscr{L}_0$ can be written as*

$$\mathscr{L}_0 \tilde{f}(\xi) = \frac{1}{2} \sum_{i=1}^{K-1} \tilde{b}_{i,i}(\xi) \frac{\partial^2}{\partial \xi_i^2} \tilde{f}(\xi) + \sum_{i=1}^{K-1} \tilde{a}_i(\xi) \frac{\partial}{\partial \xi_i} \tilde{f}(\xi), \quad \text{(B.2)}$$

with

$$\tilde{b}_{i,j}(\xi) = \delta_{i,j} \left(\frac{\xi_i(1-\xi_i)}{\Pi_{k<i}(1-\xi_k)} \right)$$

and

$$\tilde{a}_i(\boldsymbol{\xi}) = \frac{1}{2} \frac{\theta_i - \Theta_{i-1}\xi_i}{\Pi_{k<i}(1-\xi_k)}.$$

The proof of this lemma is paraphrased in Appendix B of Baxter et al. (2007). The transformation diagonalizes the operator by removing all the mixed second order partial derivatives.

Appendix C

Coefficients of the polynomial Q(x; σ, θ)

$$q = \frac{1}{2}\left(\sum_{j=1}^{K} \theta_j \sigma_{K,j} - |\theta|\sigma_{K,K}\right) \text{ when } \mathbf{i}=\varnothing,$$

$$q(i_1) = \frac{1}{2}\left(\sum_{j=1}^{K} \theta_j (\sigma_{i_1,j} - \sigma_{t,K}) + \sigma_{i_1,K}^2 + \sigma_{K,K}^2 - 2\sigma_{K,K}\sigma_{i_1,K} - 2(1+|\theta|)\sigma_{i_1,K} + (1+2|\theta|)\sigma_{K,K} + \sigma_{i_1,i_1}\right),$$

$$q(i_1,i_2) = \frac{1}{2}\left(2\sigma_{i_1,K}\sigma_{i_1,i_2} - 3\sigma_{i_1,K}\sigma_{i_2,K} + 8\sigma_{i_2,K}\sigma_{K,K} - 2\sigma_{K,K}\sigma_{i_1,i_2} - 2\sigma_{i_1,K}^2 - 3\sigma_{K,K}^2 - (1+|\theta|)(\sigma_{i_1,i_2} + \sigma_{K,K} - 2\sigma_{i_2,K})\right),$$

$$q(i_1,i_2,i_3) = \frac{1}{2}\left((\sigma_{i_1,i_3} - \sigma_{i_1,K})(\sigma_{i_1,i_2} - \sigma_{i_1,K}) - (\sigma_{i_3,K} - \sigma_{K,K})(\sigma_{i_2,K} - \sigma_{K,K}) - 4(\sigma_{i_2,i_3} + \sigma_{K,K} - 2\sigma_{i_3,K})(\sigma_{i_1,K} - \sigma_{K,K})\right),$$

$$q(i_1,i_2,i_3,i_4) = -\frac{1}{2}\left((\sigma_{i_1,i_2} + \sigma_{K,K} - 2\sigma_{i_2,K})(\sigma_{i_3,i_4} + \sigma_{K,K} - 2\sigma_{i_4,K})\right).$$

Appendix D

Derivation of equation (25)

Applying $\mathcal{L}$ to $S_n^{\theta}(\mathbf{x})$,

$$\mathscr{L}\,S_{\mathbf{n}}^{\theta}(\mathbf{x})=(\mathscr{L}_0+\mathscr{L}_\sigma)(P_{\mathbf{n}}^{\theta}(\mathbf{x})e^{-\bar{\sigma}(\mathbf{x})/2})$$

$$=e^{-\frac{\bar{\sigma}(\mathbf{x})}{2}}\mathscr{L}_0 P_{\mathbf{n}}^{\theta}(\mathbf{x})$$

$$+P_{\mathbf{n}}^{\theta}(\mathbf{x})\mathscr{L}_0 e^{-\frac{\bar{\sigma}(\mathbf{x})}{2}}$$

$$+\sum_{i,j=1}^{K-1}x_i(\delta_{i,j}$$

$$-x_j)\frac{\partial}{\partial x_i}\{e^{-\frac{\bar{\sigma}(\mathbf{x})}{2}}\}\frac{\partial}{\partial x_j}\{P_{\mathbf{n}}^{\theta}(\mathbf{x})\}+P_{\mathbf{n}}^{\theta}(\mathbf{x})\mathscr{L}_\sigma e^{-\frac{\bar{\sigma}(\mathbf{x})}{2}}+e^{-\frac{\bar{\sigma}(\mathbf{x})}{2}}\mathscr{L}_\sigma P_{\mathbf{n}}^{\theta}(\mathbf{x})=-\lambda_{\mathbf{n}}^{\theta}e^{-\frac{\bar{\sigma}(x)}{2}}P_{\mathbf{n}}^{\theta}(\mathbf{x})+P_{\mathbf{n}}^{\theta}(\mathbf{x})\mathscr{L}e^{-\frac{\bar{\sigma}(\mathbf{x})}{2}}+\sum_{i,j=1}^{K-1}x_i(\delta_{i,j}-x_j)\frac{\partial}{\partial x_i}\{e^{-\frac{\bar{\sigma}(\mathbf{x})}{2}}$$

$$+e^{-\frac{\bar{\sigma}(\mathbf{x})}{2}}\mathscr{L}_\sigma P_{\mathbf{n}}^{\theta}(\mathbf{x}).$$

It can be shown that the last two terms in the above expression sum up to 0. Note that for 1 i,j $K-1$,

$$\frac{\partial}{\partial x_i}\bar{\sigma}(\mathbf{x})=2\sum_{k=1}^{K}\sigma_{k,i}x_k - 2\sum_{l=1}^{K}\sigma_{l,K}x_l, \quad \text{(D.1)}$$

$$\frac{\partial^2}{\partial x_j\partial x_i}\bar{\sigma}(\mathbf{x})=2(\sigma_{i,j}-\sigma_{j,K}-\sigma_{i,K}+\sigma_{K,K}). \quad \text{(D.2)}$$

It follows that

$$\sum_{i,j=1}^{K-1}x_i(\delta_{i,j}-x_j)\frac{\partial}{\partial x_i}\left\{e^{-\frac{\bar{\sigma}(\mathbf{x})}{2}}\right\}\frac{\partial}{\partial x_j}\left\{P_{\mathbf{n}}^{\theta}(\mathbf{x})\right\}+e^{-\frac{\bar{\sigma}(\mathbf{x})}{2}}\mathscr{L}_\sigma P_{\mathbf{n}}^{\theta}(\mathbf{x})$$

$$=e^{-\frac{\bar{\sigma}(\mathbf{x})}{2}}\left[\sum_{i=1}^{K-1}x_i\frac{\partial}{\partial x_i}\left\{-\frac{\bar{\sigma}(\mathbf{x})}{2}\right\}\frac{\partial}{\partial x_i}\left\{P_{\mathbf{n}}^{\theta}(\mathbf{x})\right\}-\sum_{i,j=1}^{K-1}x_i x_j\frac{\partial}{\partial x_i}\left\{-\frac{\bar{\sigma}(\mathbf{x})}{2}\right\}\frac{\partial}{\partial x_j}\left\{P_{\mathbf{n}}^{\theta}(\mathbf{x})\right\}\right.$$

$$\left.+\sum_{i=1}^{K-1}x_i\frac{\partial}{\partial x_i}\left\{P_{\mathbf{n}}^{\theta}(\mathbf{x})\right\}\sum_{j=1}^{K}\sigma_{i,j}x_j - \bar{\sigma}(\mathbf{x})\sum_{i=1}^{K-1}x_i\frac{\partial}{\partial x_i}\left\{P_{\mathbf{n}}^{\theta}(\mathbf{x})\right\}\right]$$

$$=e^{-\frac{\bar{\sigma}(\mathbf{x})}{2}}\left[-\sum_{i=1}^{K-1}x_i\left(\sum_{k=1}^{K}\sigma_{k,i}x_k - \sum_{l=1}^{K}\sigma_{l,K}x_l\right)\frac{\partial}{\partial x_i}\left\{P_{\mathbf{n}}^{\theta}(\mathbf{x})\right\}\right.$$

$$+\sum_{i,j=1}^{K-1}x_i x_j\left(\sum_{k=1}^{K}\sigma_{k,i}x_k - \sum_{l=1}^{K}\sigma_{l,K}x_l\right)\frac{\partial}{\partial x_j}\left\{P_{\mathbf{n}}^{\theta}(\mathbf{x})\right\}$$

$$\left.+\sum_{i=1}^{K-1}x_i\frac{\partial}{\partial x_i}\left\{P_{\mathbf{n}}^{\theta}(\mathbf{x})\right\}\sum_{j=1}^{K}\sigma_{i,j}x_j - \bar{\sigma}(\mathbf{x})\sum_{i=1}^{K-1}x_i\frac{\partial}{\partial x_i}\left\{P_{\mathbf{n}}^{\theta}(\mathbf{x})\right\}\right]$$

$$=e^{-\frac{\bar{\sigma}(\mathbf{x})}{2}}\left[\sum_{i=1}^{K-1}x_i\frac{\partial}{\partial x_i}\left\{P_{\mathbf{n}}^{\theta}(\mathbf{x})\right\}\sum_{l=1}^{K}\sigma_{l,K}x_l\right.$$

$$\left.-\sum_{j=1}^{K-1}x_j\frac{\partial}{\partial x_j}\left\{P_{\mathbf{n}}^{\theta}(\mathbf{x})\right\}\left(\sum_{l=1}^{K}\sigma_{l,K}x_K x_l+\sum_{i=1}^{K-1}x_i\sum_{l=1}^{K}\sigma_{l,K}x_l\right)\right]=0,$$

where we used equation (D.1) for the second equality and $\sum_{i=1}^{K} x_i = 1$ for the last equality.

Further, using equation (D.1) and (D.2) one can show that

$$\mathscr{L}e^{-\frac{\bar{\sigma}(\mathbf{x})}{2}} = \frac{1}{2}e^{-\frac{\bar{\sigma}(\mathbf{x})}{2}}\left(-\sum_{i=1}^{K}x_i\sigma_i^2(\mathbf{x}) - \sum_{i=1}^{K}x_i\sigma_{ii} + (1+|\theta|)\bar{\sigma}(\mathbf{x}) + \bar{\sigma}(\mathbf{x})^2 - \sum_{i=1}^{K}\theta_i\sigma_i(\mathbf{x})\right) = e^{-\frac{\bar{\sigma}(\mathbf{x})}{2}}Q(\mathbf{x};\sigma,\theta),$$

where Q takes the form (27), that is $Q(\mathbf{x};\sigma,\theta) = {}_{\mathbf{i}\in\mathscr{Q}}\, q(\mathbf{i})\mathbf{x_i}$, with the constants $q(\mathbf{i})$ given in Appendix C.

References

Abramowitz, M.; Stegun, IA. Handbook of Mathematical Functions. Dover Publications; 1965.

Barbour AD, Ethier SN, Griffiths RC. A transition function expansion for a diffusion model with selection. Ann. Appl. Probab. 2000; 10:123–162.

Baxter G, Blythe R, McKane A. Exact solution of the multi-allelic diffusion model. Math. Biosci. 2007; 209:124–170. [PubMed: 17320915]

Bollback JP, York TL, Nielsen R. Estimation of $2N_e s$ from temporal allele frequency data. Genetics. 2008; 179(1):497–502. [PubMed: 18493066]

Buzbas EO, Joyce P. Maximum likelihood estimates under the k-allele model with selection can be numerically unstable. Ann. Appl. Stat. 2009; 3(3):1147–1162.

Buzbas EO, Joyce P, Abdo Z. Estimation of selection intensity under overdominance by Bayesian methods. Stat. Appl. Genet. Mol. Biol. 2009; 8(1) Article 32.

Buzbas EO, Joyce P, Rosenberg NA. Inference on the strength of balancing selection for epistatically interacting loci. Theor. Popul. Biol. 2011; 79(3):102–113. [PubMed: 21277883]

Donnelly P, Nordborg M, Joyce P. Likelihoods and simulation methods for a class of nonneutral population genetics models. Genetics. 2001; 159(2):853–867. [PubMed: 11606558]

Dunkl, C.; Xu, Y. Orthogonal Polynomials of Several Variables. Cambridge University Press; 2001.

Durrett, R. Probability Models for DNA Sequence Evolution. Springer; 2008.

Epstein, CL.; Mazzeo, R. Degenerate diffusion operators arising in population biology. 2011. arXiv preprint: http://arxiv.org/abs/1110.0032

Ethier SN, Kurtz TG. Convergence to Fleming-Viot processes in the weak atomic topology. Stoch. Proc. Appl. 1994; 54(1):1–27.

Evans SN, Shvets Y, Slatkin M. Non-equilibrium theory of the allele frequency spectrum. Theor. Popul. Biol. 2007; 71(1):109–119. [PubMed: 16887160]

Ewens, W. Mathematical Population Genetics. 2nd edition. Vol. volume I. Springer; 2004. Theoretical introduction

Genz A, Joyce P. Computation of the normalizing constant for exponentially weighted dirichlet distribution integrals. Comp. Sci. Stat. 2003; 35:181–212.

Griffiths R. A transition density expansion for a multi-allele diffusion model. Adv. Appl. Prob. 1979; 11:310–325.

Griffiths RC, Li W-H. Simulating allele frequencies in a population and the genetic differentiation of populations under mutation pressure. Theor. Popul. Biol. 1983; 23(1):19–33. [PubMed: 6857548]

Griffiths, RC.; Spanò, D. Probability and Mathematical Genetics, Papers in Honour of Sir John Kingman. LMS Lecture Note Series 378. Vol. chapter 15. Cambridge University Press; 2010. Diffusion processes and coalescent trees; p. 358-375.

Griffiths RC, Spanò D. Multivariate Jacobi and Laguerre polynomials, infinite-dimensional extensions, and their probabilistic connections with multivariate Hahn and Meixner polynomials. Bernoulli. 2011; 17(3):1095–1125.

Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. PLoS Genet. 2009; 5:e1000695. [PubMed: 19851460]

Karlin, S.; Taylor, H. A Second Course in Stochastic Processes. Academic Press; 1981.

Kimura M. Stochastic processes and distribution of gene frequences under natural selection. Cold Spring Harb. Symp. Quant. Biol. 1955; 20:33–53. [PubMed: 13433553]

Kimura M. Some problems of stochastic processes in genetics. Ann. Math. Stat. 1957; 28:882–901.

Luki S, Hey J, Chen K. Non-equilibrium allele frequency spectra via spectral methods. Theor. Popul. Biol. 2011; 79(4):203–219. [PubMed: 21376069]

Shimakura N. Equations différentielles provenant de la génétique des populations. Tohoku Math. J. 1977; 29:287–318.

Song YS, Steinrücken M. A simple method for finding explicit analytic transition densities of diffusion processes with general diploid selection. Genetics. 2012; 190(3):1117–1129. [PubMed: 22209899]

Szegö, G. Orthogonal Polynomials. American Mathematical Society; 1939.

Tavaré S. Line-of-descent and genealogical processes, and their applications in population genetics models. Theor. Popul. Biol. 1984; 26:119–164. [PubMed: 6505980]

Wright, S. Adaptation and selection. In: Jepson, GL.; Mayr, E.; Simpson, GG., editors. Genetics, Paleontology and Evolution. Princeton, New Jersey: Princeton Univ. Press; 1949. p. 365-389.

Živkovi D, Stephan W. Analytical results on the neutral non-equilibrium allele frequency spectrum based on diffusion theory. Theor. Popul. Biol. 2011; 79(4):184–191. [PubMed: 21426909]

(a)

(b)

(c)

(d)
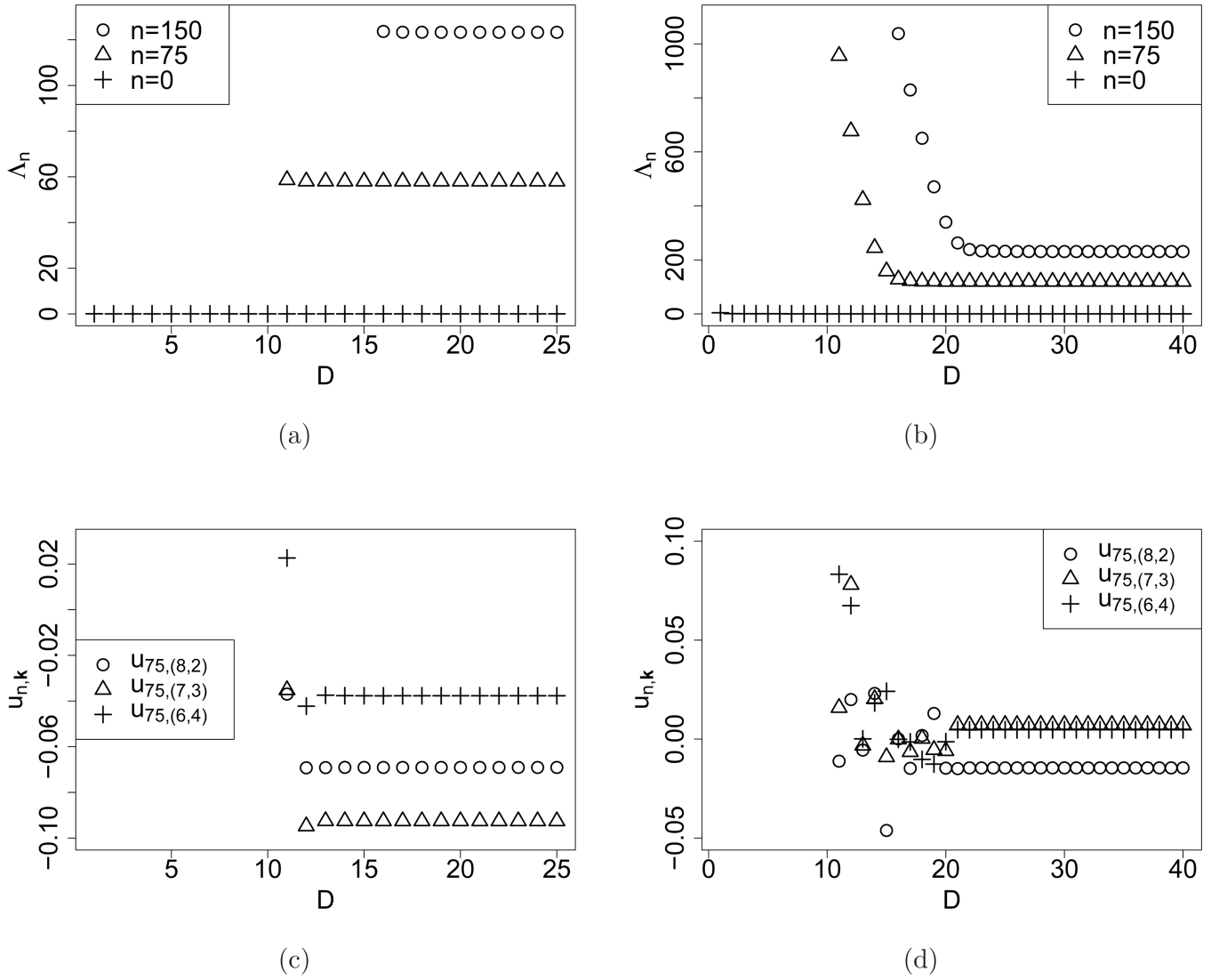
Figure 1.
Convergence of the truncated eigenvalues Λ_n and coefficients of the eigenvectors $\mathbf{u}_n$ as the truncation level D increases, for $K = 3$ with low mutation rates. Subfigures (a) and (b) show $\Lambda_0^{[D]}, \Lambda_{75}^{[D]}$, and $\Lambda_{150}^{[D]}$ for $\boldsymbol{\sigma} = \boldsymbol{\sigma}_1$ and $\boldsymbol{\sigma} = \boldsymbol{\sigma}_2$, respectively. Subfigures (c) and (d) show $u_{75,(8,2)}^{[D]}, u_{75,(7,3)}^{[D]}$, and $u_{75,(6,4)}^{[D]}$ for $\boldsymbol{\sigma} = \boldsymbol{\sigma}_1$ and $\boldsymbol{\sigma} = \boldsymbol{\sigma}_2$, respectively. The mutation rates were set to $\boldsymbol{\theta} = (0.01, 0.02, 0.03)$ for all computations.
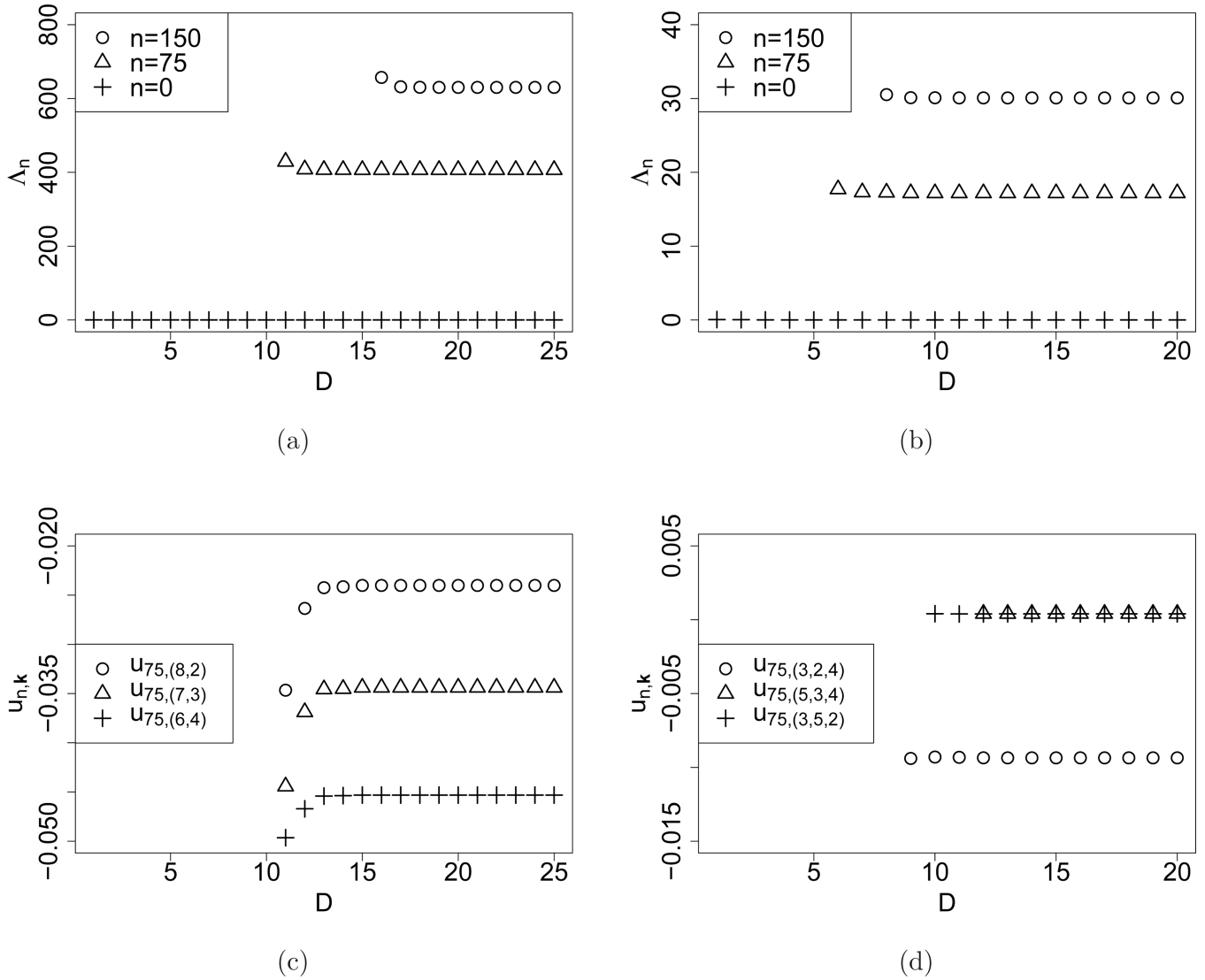
Figure 2.
Convergence of the truncated eigenvalues Λ_n and coefficients of the eigenvectors $\mathbf{u}_n$ as the truncation level D increases, for $K = 3$ with high mutation rates and for $K = 4$ with low mutation rates. Subfigures (a) and (c) show $\Lambda_n^{[D]}$ for $n = 0$, 75, 150, and $u_{75,\mathbf{m}}^{[D]}$ for $\mathbf{m} = (8, 2)$, (7, 3), (6, 4), respectively, for mutation rates $\boldsymbol{\theta} = (10, 20, 30)$ and selection coefficients $\boldsymbol{\sigma} = \boldsymbol{\sigma}_1$. The convergence behavior for $K = 4$ is shown in subfigures (b) and (d) for $\Lambda_n^{[D]}$ with $n = 0$, 75, 150, and $u_{75,\mathbf{m}}^{[D]}$ with $\mathbf{m} = (3, 2, 4)$, (5, 3, 4), (3, 5, 2), respectively; the mutation rates were set to $\boldsymbol{\theta} = (0.01, 0.02, 0.03, 0.04)$ and the selection coefficients to $\boldsymbol{\sigma} = \boldsymbol{\sigma}_3$.
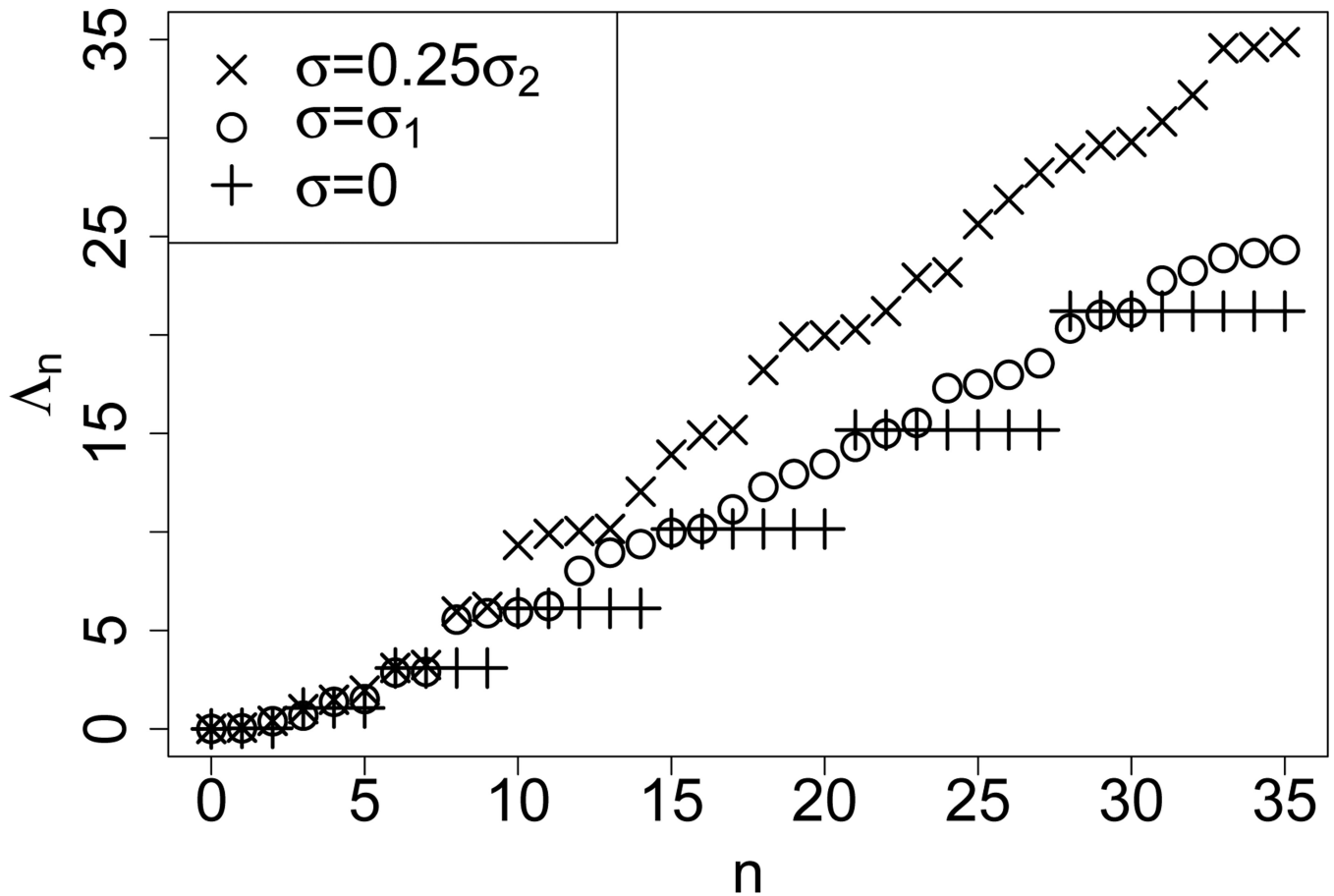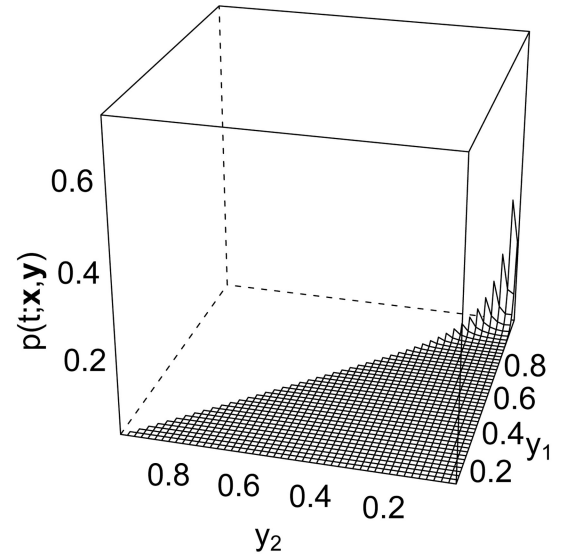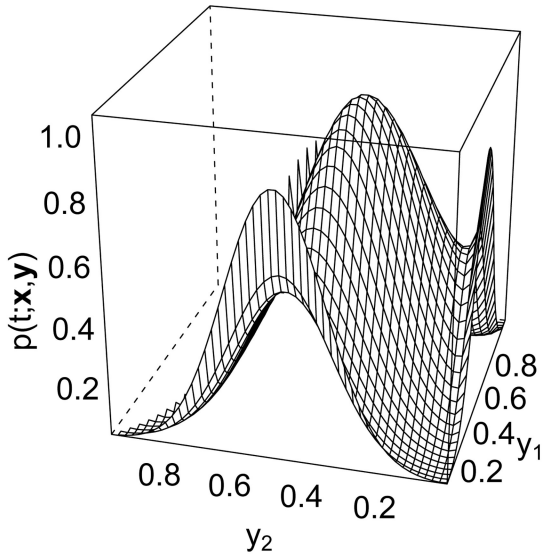
Figure 3.
The first 36 eigenvalues of the different spectra for the selection parameters $\sigma = 0$, σ_1 and $\frac{1}{4}\sigma_2$, respectively. The latter was chosen so that the ranges of the eigenvalues are comparable. The truncation level D was set to 24 and mutation rates $\theta = (0.01, 0.02.0.03)$ were used.

t=0.04

t=0.2

(a)

(b)

t=1.0

t=2.0

(c)

(d)
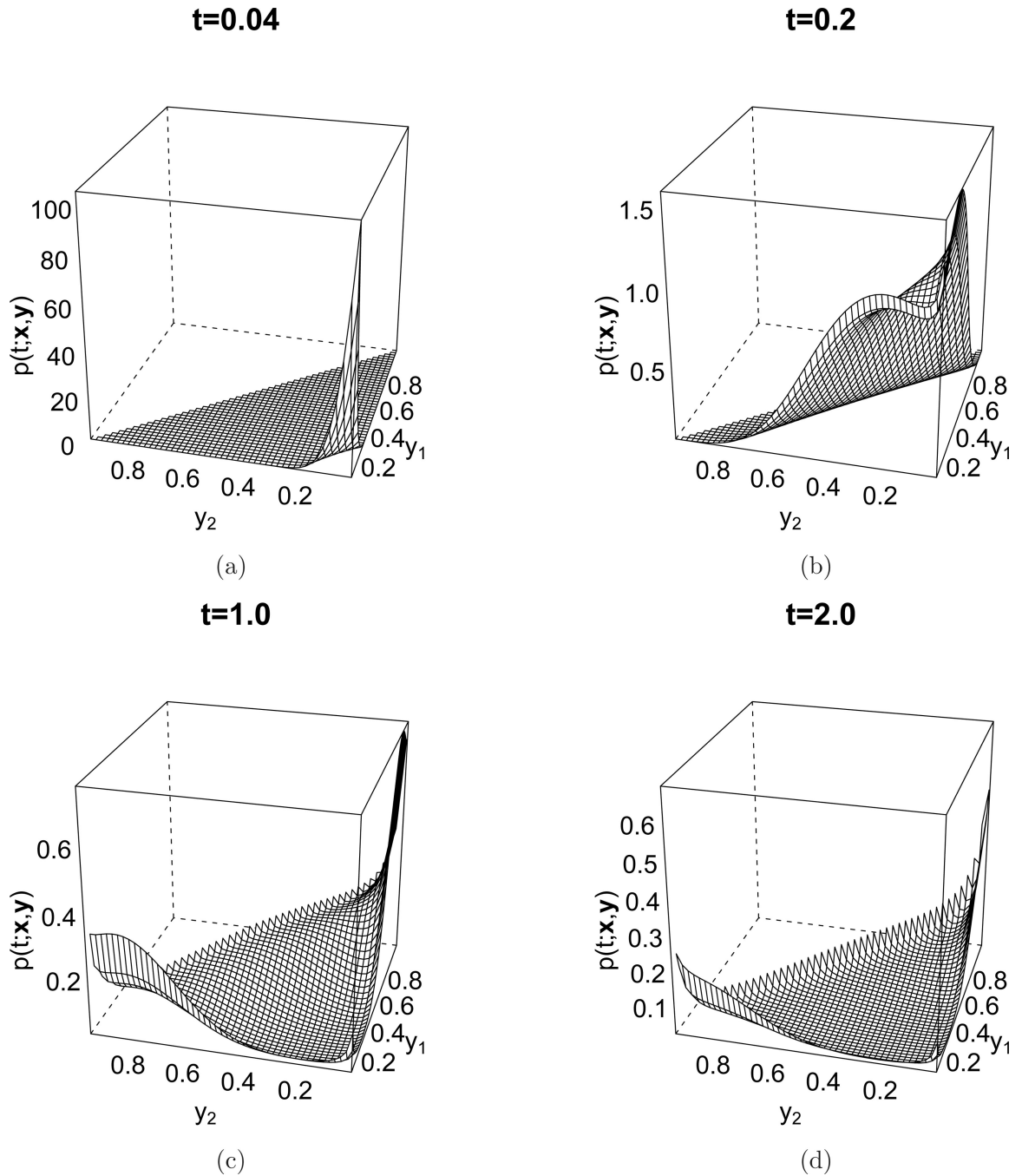
Figure 4.
Approximation of the transition density function (6) for different times $t \in \{0.04, 0.2, 1.0, 2.0\}$. Selection was governed by the matrix of coefficients $\boldsymbol{\sigma} = \boldsymbol{\sigma}_1$ and $\mathbf{x} = (0.02, 0.02, 0.96)$ was used as initial condition. The truncation level was set to $D = 40$, whereas the summation in equation (23) ranged over all $\mathbf{m}$ such that $0 \leq |\mathbf{m}| \leq 36$, and all eigenfunctions and eigenvalues with $0 \leq n \leq 561$ were included in equation (6). The mutation rates were set to $\boldsymbol{\theta} = (0.01, 0.02, 0.03)$. The plots only vary in y_1 and y_2, since $y_3 = 1 - y_1 - y_2$. (a) $t = 0.04$. (b) $t = 0.2$. (c) $t = 1.0$. (d) $t = 2.0$.

(a) Selection coefficients: $\boldsymbol{\sigma} = \begin{pmatrix} 0 & 15 & 15 \\ 15 & 0 & 15 \\ 15 & 15 & 0 \end{pmatrix}$ (b) Selection coefficients: $\boldsymbol{\sigma} = \begin{pmatrix} 10 & 6 & -5 \\ 6 & -6 & 6 \\ -5 & 6 & 0 \end{pmatrix}$

Figure 5.
Two examples of the stationary distribution for different selection parameters. The mutation rates were set to $\boldsymbol{\theta} = (0.01, 0.02.0.03)$ in both cases. Again, a truncation level of $D = 40$ was used, the summation in equation (28) ranged over all $\mathbf{m}$ such that $0 \leq |\mathbf{m}| \leq 36$. The plots only vary in y_1 and y_2, since $y_3 = 1 - y_1 - y_2$.
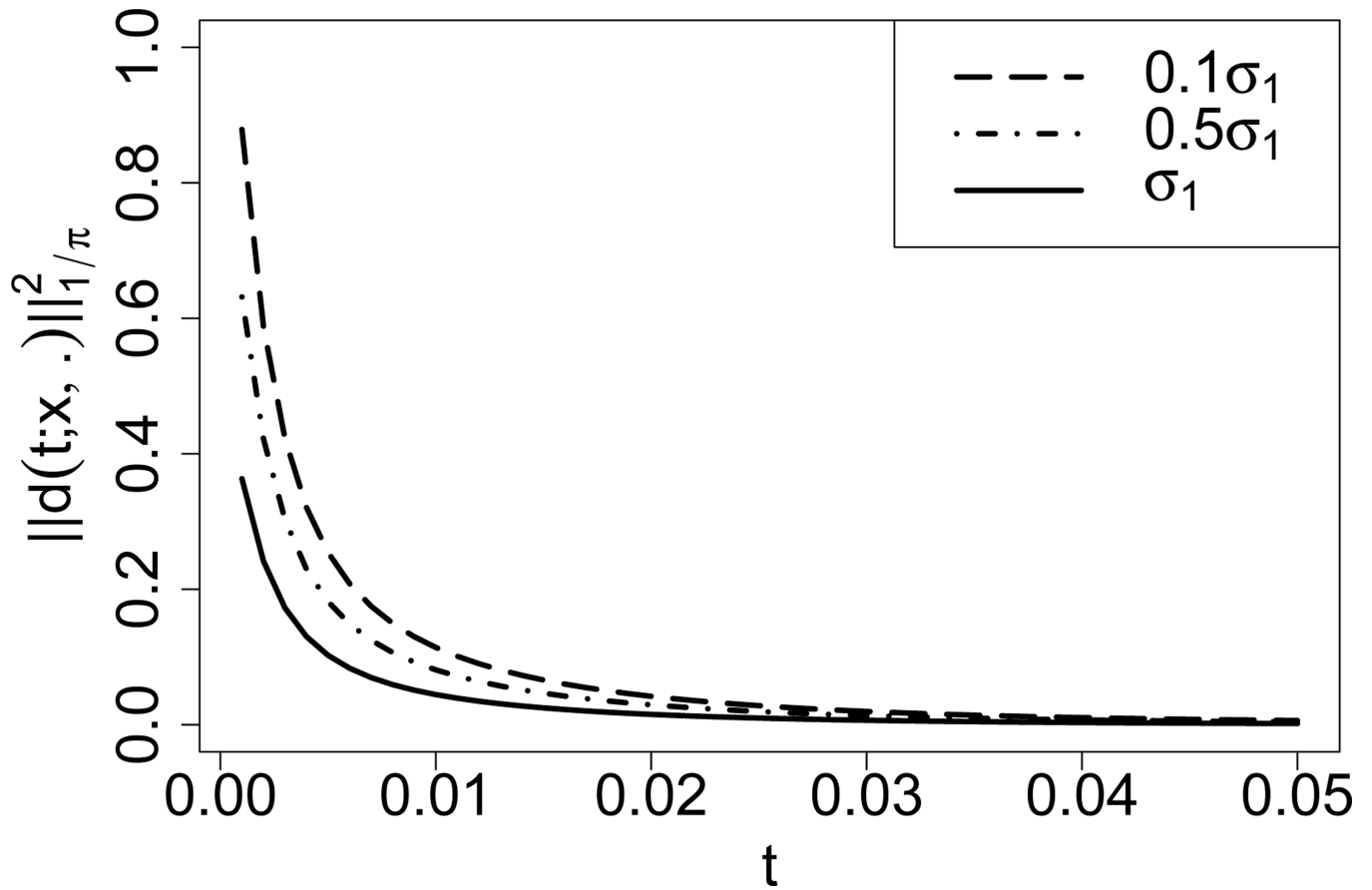
Figure 6.
Convergence of the transition density to stationarity as time evolves, for initial frequencies $\mathbf{x}$ = $(0.02, 0.02, 0.96)^T$. Deviation from the stationary density is measured by $\|d(t;\mathbf{x}, \cdot)\|^2_{1/\Pi}$, defined in (29). The mutation rates were chosen to be $\boldsymbol{\theta}$ = $(0.01, 0.02, 0.03)$ and the selection parameters were $\boldsymbol{\sigma}$ = $0.1\sigma_1$, $\boldsymbol{\sigma}$ = $0.5\sigma_1$ and $\boldsymbol{\sigma}$ = $0.1\sigma_1$, respectively. The truncation level was set to D = 40, and (29) was approximated by summing over 0 ≤ n ≤ 561 and $\mathbf{m}, \mathbf{k}$ such that 0 ≤ |$\mathbf{m}$|, |$\mathbf{k}$| ≤ 36.