The Quasistationary Distributions of Level-Independent Quasi-Birth-and-Death Processes

Nigel BEAN[1], **Phil POLLETT**[2] **and Peter TAYLOR**[1]

1 Department of Applied Mathematics, University of Adelaide, Australia 5005.

2 Department of Mathematics, University of Queensland, Australia 4072.

Abstract For evanescent Markov processes with a single transient communicating class, it is often of interest to examine the stationary probabilities that the process resides in the various transient states, conditional on absorption not having taken place. Such distributions are known as quasistationary distributions. In this paper we consider the determination of a family of quasistationary distributions of a general level-independent quasi-birth-and-death process (QBD). These distributions are shown to have a form analogous to the quasistationary distributions exhibited by birth-and-death processes. We briefly discuss methods for the computation of these quasistationary distributions.

1 INTRODUCTION

Consider a discrete-time Markov chain $(X_n; n \in \mathbb{Z}_+)$ on a countable state space $\mathcal{S} = \{0, 1, \ldots\}$ with transition matrix P. Assume (X_n) has an absorbing state 0 and an irreducible and aperiodic communicating class $\mathcal{C} \equiv \mathcal{S} \setminus \{0\}$. We assume that the expected time to absorption is finite from one (and then all) states $i \in \mathcal{C}$. Let T denote the time until absorption of the process.

For many evanescent Markov chains, T can be very large, and over any reasonable period of time the chain appears to exist in an equilibrium. This leads us to consider quasistationary distributions. A quasistationary distribution, $\boldsymbol{\pi}$, is a stationary distribution of the process conditioned to stay in the transient class; that is, if $P(X_0 = j) = \pi_j$, $j \in \mathcal{C}$, then

$$P(X_n = j | T > n) = \pi_j, \qquad j \in \mathcal{C},$$

for all $n \geq 1$. In other words, conditional on the chain being in $\mathcal{C}$ the state probabilities do not vary with time.

A nontrivial, nonnegative row vector $\boldsymbol{m}(\beta)$ that satisfies

$$\boldsymbol{m}(\beta) = \beta \boldsymbol{m}(\beta) \widehat{P} \tag{1.1}$$

is called a β-invariant measure. Here, and throughout, $\widehat{P}$ denotes the restriction of P to $\mathcal{C}$. It is elementary to show that $\boldsymbol{\pi}$ is a quasistationary distribution if and only if, for some $\beta > 1$, it is a β-invariant measure, in which case

$$\beta^{-1} = 1 - \sum_{i \in \mathcal{C}} \pi_i p_{i0},$$

represents the probability (under the quasistationary distribution) that the process remains within the transient class at the next time step.

For each Markov process there is a maximum value of β for which a quasistationary distribution can exist. This critical parameter is called the convergence radius and is denoted by α. In certain circumstances an α-invariant measure may also have a limiting-conditional interpretation: that is,

$$\lim_{n \to \infty} P(X_n = j | X_0 = i, T > n) = \pi_j, \qquad j \in \mathcal{C},$$

no matter what the initial state i.

The convergence radius can be rigorously characterized as follows. For $z \in \mathbb{R}$, let $N_{ij}(z)$ be defined by

$$N_{ij}(z) = \sum_{n=0}^{\infty} z^n \widehat{P}_{ij}^{(n)} \,, \tag{1.2}$$

where $\widehat{P}_{ij}^{(n)}$ is the (i,j)th entry of $\widehat{P}^{(n)}$. Theorem 6.1 of Seneta [12] states that, for a given value of z, either $N_{ij}(z)$ is finite for all (i,j) or $N_{ij}(z)$ is infinite for all (i,j). The convergence radius associated with $\widehat{P}$ is defined as

$$\alpha = \sup\left\{z : N_{ij}(z) \text{ is finite}\right\}.$$

There are very few substochastic chains for which a full quasistationary analysis is available. Historical exceptions have been finite-state processes, the Galton-Watson branching process, simple birth-and-death chains and the work of Kyprianou [6] on GI/M/1 queues (see also Kyprianou [7] for an analysis under conditions of heavy traffic).

Recently, a notable advance was made by Kijima [3] who gave an algebraic equation for the convergence radius of PH/PH/1 queues (in fact, more generally, for processes of M/G/1 and GI/M/1 type). In the queueing context considered in [3], this equation can be solved by use of the Laplace-Stieltjes transform of the interarrival and service time distributions. Kijima also gave the form of the limiting-conditional distribution for the special cases of the M/PH/1 and PH/M/1 queues. This work was extended by Makimoto [9] who gave an explicit representation of the limiting-conditional distribution for PH/PH/c queues in terms of solutions to a matrix equation. Makimoto did not, however, discuss methods of solution for this equation in the general case. For a nice survey of this area see Kijima and Makimoto [4].

In Bean *et al.* [1] the results of Kijima [3] and Makimoto [9] were extended by examining the limiting-conditional behaviour of general level-independent quasi-birth-and-death processes (QBDs), which includes the PH/PH/c queues as a subclass. An algorithm for the explicit numerical computation of the convergence radius, α, and the limiting-conditional distribution was also presented.

In this paper we extend the results of Bean *et al.* [1] by finding quasistationary distributions of a level-independent QBD for all $\beta \le \alpha$, not just the limiting-conditional distribution. In Bean *et al.* [1] the limiting-conditional distribution is written in such a way that it is not an obvious generalization of the limiting-conditional distribution for a birth-and-death process. In contrast, here we present all the quasistationary distributions as natural extensions of the quasistationary distributions for ordinary birth-and-death processes. We also discuss methods for their computation.

The results of this paper can also be applied to continuous-time QBDs. For details, see Bean *et al.* [1].

2 ABSORBING QBDs AND THE CONVERGENCE RADIUS

In this section we summarise the results of Sections 2, 3 and 4 of Bean *et al.* [1] in order to define the convergence radius α and establish some fundamental concepts. These sections extend the matrix geometric theory of QBDs, as developed by Neuts [10], to absorbing QBDs. Throughout, a matrix is termed finite if all its entries are finite.

Assume that (X_n) is a level-independent quasi-birth-and-death process. This can be regarded as a two-dimensional Markov chain with $\mathcal{C} = \{(k,j) : k \geq 1, 1 \leq j \leq M\}$ and whose transition matrix is of the block-partitioned form

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & \cdots \\ A_2\boldsymbol{e} & A_1 & A_0 & 0 & 0 & \cdots \\ 0 & A_2 & A_1 & A_0 & 0 & \cdots \\ 0 & 0 & A_2 & A_1 & A_0 & \cdots \\ 0 & 0 & 0 & A_2 & A_1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

and so

$$\widehat{P} = \begin{pmatrix} A_1 & A_0 & 0 & 0 & \cdots \\ A_2 & A_1 & A_0 & 0 & \cdots \\ 0 & A_2 & A_1 & A_0 & \cdots \\ 0 & 0 & A_2 & A_1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

Here the partitioning corresponds to distinguishing subsets of states called levels. Level k is defined by $l(k) = \{(k,j) : 1 \leq j \leq M\}$ for $k \geq 1$ and level 0 is the absorbing state 0. Throughout, $\boldsymbol{e}$ denotes a column vector of ones.

The assumption that (X_n) is irreducible on $\mathcal{C}$ implies that the matrix $A = A_0 + A_1 + A_2$ is irreducible and that the matrices A_0 and A_2 are nonzero (the converse is not true).

The equations (1.1) which define the β-invariant measures can now be written

$$\boldsymbol{m}_1(\beta) = \beta\left[\boldsymbol{m}_1(\beta)A_1 + \boldsymbol{m}_2(\beta)A_2\right], \tag{2.1}$$

$$\boldsymbol{m}_k(\beta) = \beta\left[\boldsymbol{m}_{k-1}(\beta)A_0 + \boldsymbol{m}_k(\beta)A_1 + \boldsymbol{m}_{k+1}(\beta)A_2\right], \quad k \geq 2, \tag{2.2}$$

where the M-vector $\boldsymbol{m}_k(\beta)$ is the restriction of $\boldsymbol{m}(\beta)$ to level k.

Let $N_{11}(\beta)$ denote the $M \times M$ matrix whose (i,j)th entry is $N_{(1,i)(1,j)}(\beta)$ as defined in (1.2). Define

$$R(\beta) = \beta A_0 N_{11}(\beta). \tag{2.3}$$

Since the process is level-independent, the transition structure above level 2 is the same as that above level 1 and so, the entry $R_{ij}(\beta)$ can be interpreted as the expected total discounted reward for visits to state $(2, j)$ before returning to level 1, conditional on starting in state $(1, i)$ with a discount factor β. In the rest of this paper, we shall consider only the situation where β is greater than or equal to one. The interpretation given above also holds when levels 1 and 2 are replaced by levels k and $k + 1$, respectively.

The following lemma follows from Theorem 2 of Bean *et al.* [1].

Lemma 1 *The matrix $R(\beta)$ is finite if and only if there exists a nonnegative solution to the matrix-quadratic equation*

$$S = \beta \left[A_0 + SA_1 + S^2 A_2\right], \tag{2.4}$$

in which case $R(\beta)$ is the minimal *nonnegative solution. Here, and throughout, a minimal solution is elementwise minimal.*

A slight modification of the argument in the proof of Lemma 1.3.2 of Neuts [10] shows that the maximal eigenvalue $\eta(\beta)$ of $R(\beta)$ is positive and has geometric multiplicity one. In fact, it is possible to conclude that the maximal eigenvalue $\eta(\beta)$ of $R(\beta)$ also has *algebraic* multiplicity one. This follows from a modified version of the proof of Theorem 2.1 in Latouche and Taylor [8].

For $0 \le z \le 1$, let $\chi(z)$ be the maximal eigenvalue of the irreducible matrix

$$A(z) = A_0 + zA_1 + z^2 A_2 \tag{2.5}$$

and $\boldsymbol{u}(z)$ and $\boldsymbol{v}(z)$ the corresponding left and right eigenvectors normalized so that $\boldsymbol{u}(z)\boldsymbol{e} = 1 = \boldsymbol{u}(z)\boldsymbol{v}(z)$. Theorem 4(ii) of Bean *et al.* [1] shows that if $\beta \le \alpha$ then the vector $\boldsymbol{u}(\eta(\beta))$ is also the left eigenvector of $R(\beta)$ with eigenvalue $\eta(\beta) \in (0, 1]$. Premultiplying equation (2.4) by $\boldsymbol{u}(\eta(\beta))$ we find that

$$\eta(\beta) = \beta\chi(\eta(\beta)). \tag{2.6}$$

The function $\chi(z)$ plays a crucial role in the following analysis. Its fundamental properties are (Neuts [10], Kingman [5]):

1 $\chi(z)$ is analytic on $(0, 1)$, continuous at $z = 1$ and may be defined by continuity at $z = 0$,

2 $\chi(0) \ge 0$, $\chi(1) = 1$ and $\chi(z)$ is strictly increasing, and

3 $\log \chi(e^{-s})$ is convex for $s \ge 0$.

Under the assumptions that the expected time to absorption is finite and that the class C is irreducible and aperiodic, we have

4 $\chi'(1-) > 1$, and

5 there exists $z^* \in (0,1)$ such that

$$\chi'(z)z < \chi(z), \quad \text{for all} \quad z \in (0, z^*) \ . \tag{2.7}$$

It was shown in Bean *et al.* [1] that consequences of these properties are,

6 there exists a minimal solution, z_0, in $(0,1)$ to the equation

$$\chi'(z)z = \chi(z),$$

7 the equation

$$z = \beta\chi(z)$$

has two solutions $z \in (0,1]$ if $1 \leq \beta < \alpha$, one solution if $\beta = \alpha$ and no solutions if $\beta > \alpha$.

In Kijima [3] (see also Bean *et al.* [1]) the following theorem was established.

Theorem 2 *The convergence radius α associated with (X_n) is given by*

$$\alpha = \Big[\boldsymbol{u}(z_0)\,[A_1 + 2z_0A_2]\,\boldsymbol{v}(z_0)\Big]^{-1}, \tag{2.8}$$

where z_0 is the minimal solution to

$$\chi'(z)z = \chi(z) \tag{2.9}$$

in the interval (0,1). Moreover, $z_0 = \eta(\alpha)$.

Now consider the one-dimensional manifold of solutions to

$$z = \beta\chi(z), \tag{2.10}$$

which is equation (2.6) with $z = \eta(\beta)$. From the properties above it can be deduced that this manifold has the general form given in Figure 1.

For $\beta < \alpha$, label the two points of solution as $z_1(\beta)$ and $z_2(\beta)$ with the understanding that $z_1(\beta) < z_2(\beta)$. For $\beta = \alpha$ we have $z_1(\alpha) = z_2(\alpha) = z_0$. For $z \in [z_1(1), 1]$ let $\beta(z)$ be the unique value of β for which equation (2.10) is obeyed.

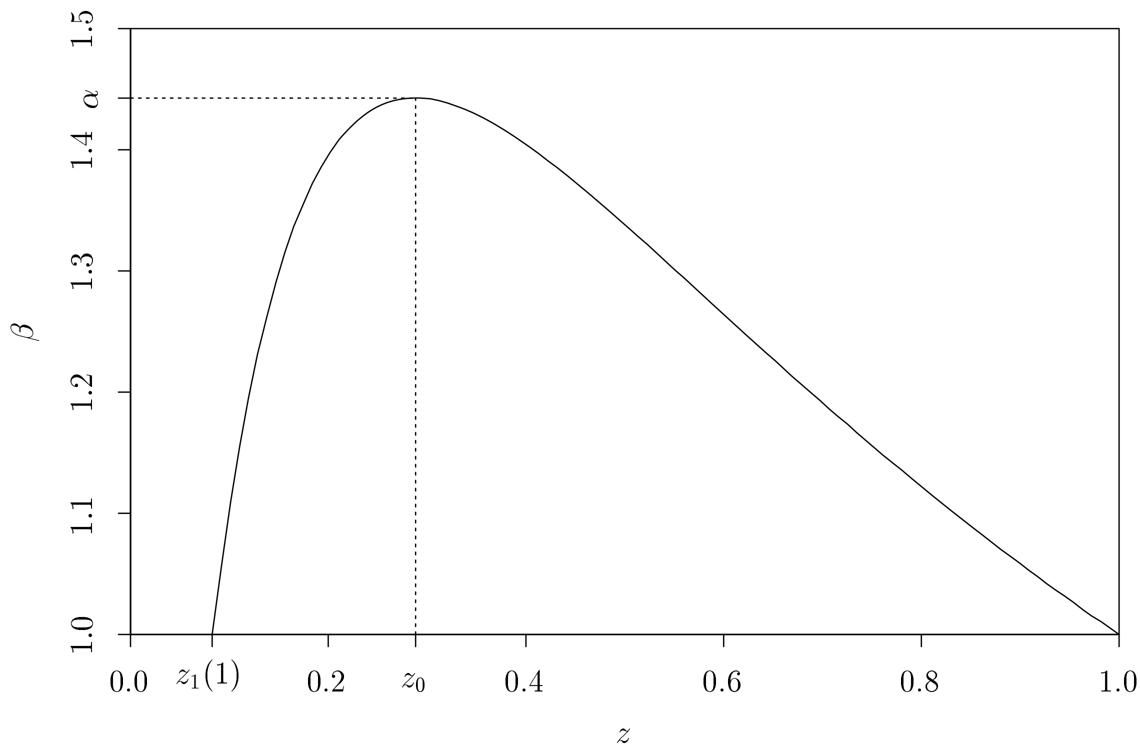

Figure 1: Graph of the one-dimensional manifold of solutions to equation (2.10).

Thus for $z \in [z_1(1), 1]$, $\beta(z)$ is defined by the function

$$\beta(z) = \frac{z}{\chi(z)}. \tag{2.11}$$

For $z \in [z_1(1), z_0]$ write $R_1(z)$ for $R(\beta(z))$. Note that $R_1(z)$ is the minimal nonnegative solution to equation (2.4) for $\beta = \beta(z)$ and has maximal eigenvalue z. We have seen before that the left eigenvector of $R_1(z)$ corresponding to the eigenvalue z is $\boldsymbol{u}(z)$. Denote by $\boldsymbol{w}(z)$ the right eigenvector of $R_1(z)$ corresponding to the eigenvalue z.

In the next section we show that for $y \in (z_0, 1]$ there is a nonnegative solution $R_2(y)$ to equation (2.4) for $\beta = \beta(y)$ with maximal eigenvalue y.

It is very important to note that $R_1(z)$ is the matrix that arises probabilistically as defined in equation (2.3), due to the minimality requirement. In contrast, it seems difficult to find a physical interpretation for $R_2(z)$.

3 THE QUASISTATIONARY DISTRIBUTIONS

In this section we show that the distribution $\boldsymbol{m}(\beta) = (\boldsymbol{m}_1(\beta), \ldots)$ given by

$$\boldsymbol{m}_j(\beta) = c\boldsymbol{x}\left[R_2(z_2(\beta))^j - R_1(z_1(\beta))^j\right] , \tag{3.1}$$

with $\boldsymbol{x}$ a suitable nonnegative vector and c a normalising constant, is a β-invariant measure, and hence quasistationary distribution, for (X_n) when $\beta < \alpha$. We also show that $\boldsymbol{m}(\alpha) = (\boldsymbol{m}_1(\alpha), \ldots)$ given by

$$\boldsymbol{m}_j(\alpha) = c\boldsymbol{x}\frac{d}{dz}\left[R_1(z)^j\right]_{z=z_0}, \tag{3.2}$$

is the α-invariant measure, and hence quasistationary and limiting-conditional distribution, for (X_n).

The major contribution of this paper is that this is the first presentation of a set of β-invariant measures for all $\beta \leq \alpha$. Furthermore, these distributions are the obvious matrix analogues of the well-known scalar results for the ordinary birth-and-death process. This is in contrast to the form of the α-invariant measure presented in Makimoto [9] and Bean *et al.* [1].

3.1 CALCULATION OF THE MATRIX $R_2(y)$

For $x \in [z_1(1), z_0]$ define $y(x) = z_2(\beta(x))$ and for $x \in [z_0, 1]$ define $z(x) = z_1(\beta(x))$. Where there can be no confusion, we usually write y instead of $y(x)$ and z instead of $z(x)$. Thus z and y are $z_1(\beta)$ and $z_2(\beta)$, respectively, where $\beta = \beta(z) = \beta(y)$ and we will freely interchange the notation when convenient.

Throughout this section, we assume that $z \in [z_1(1), z_0)$ and hence $y = y(z) \in (z_0, 1]$.

Lemma 1 *The matrix*

$$S(y) = R_1(z) + \boldsymbol{w}(z)\boldsymbol{u}(y)\left[yI - R_1(z)\right], \tag{3.3}$$

where $\boldsymbol{u}(y)$ is normalised so that

$$\boldsymbol{u}(y)\boldsymbol{w}(z) = 1, \tag{3.4}$$

has the following spectral properties:

1. *The maximal eigenvalue of $S(y)$ is y.*
2. *The left and right eigenvectors of $S(y)$ associated with the eigenvalue y are given by $\boldsymbol{u}(y)$ and $\boldsymbol{w}(z)$, respectively.*
3. *The eigenvalues of $S(y)$ that are not equal to y and the eigenvalues of $R_1(z)$ that are not equal to z are identical. Moreover, the Jordan blocks associated with these eigenvalues are also identical.*
4. *For eigenvalues of $S(y)$ not equal to y, the associated left eigenvectors (and generalized left eigenvectors) of $S(y)$ and $R_1(z)$ are identical. The right eigenvectors (and generalized right eigenvectors) are, in general, different from those of $R_1(z)$.*

Proof: Throughout this proof we shall rely on the definition of $\boldsymbol{u}(y)$, $\boldsymbol{w}(z)$ and $S(y)$ and the normalization condition for $\boldsymbol{u}(y)$ given in equation (3.4). We shall first prove the result in the case where $R_1(z)$ is diagonalisable and then proceed to the case where $R_1(z)$ is not diagonalisable as a little more care is required.

Consider the diagonalisable matrix $R_1(z)$. Let the M linearly independent left and right eigenvectors be denoted by $\boldsymbol{\ell}_i$ and $\boldsymbol{r}_i$ and associated with the eigenvalues $\lambda_i,\ i = 1, 2, \ldots, M$. In particular, order the indices such that $\lambda_M = z$, $\boldsymbol{\ell}_M = \boldsymbol{u}(z)$ and $\boldsymbol{r}_M = \boldsymbol{w}(z)$.

The following equations are basic consequences of the definition of eigenvalues and eigenvectors.

$$\boldsymbol{\ell}_i R_1(z) = \lambda_i \boldsymbol{\ell}_i, \quad i = 1, 2, \ldots, M, \tag{3.5}$$

$$R_1(z)\boldsymbol{w}(z) = z\boldsymbol{w}(z), \tag{3.6}$$

$$\boldsymbol{\ell}_i \boldsymbol{w}(z) = 0, \quad i = 1, 2, \ldots, M-1. \tag{3.7}$$

It is now simple to show that

$$\begin{aligned} \boldsymbol{u}(y)S(y) &= \boldsymbol{u}(y)R_1(z) + \boldsymbol{u}(y)\boldsymbol{w}(z)\boldsymbol{u}(y)\left[yI - R_1(z)\right], \\ &= y\boldsymbol{u}(y), \end{aligned} \tag{3.8}$$

by equation (3.4) and

$$\begin{aligned} S(y)\boldsymbol{w}(z) &= R_1(z)\boldsymbol{w}(z) + \boldsymbol{w}(z)\boldsymbol{u}(y)\left[yI - R_1(z)\right]\boldsymbol{w}(z), \\ &= z\boldsymbol{w}(z) + y\boldsymbol{w}(z) - z\boldsymbol{w}(z), \\ &= y\boldsymbol{w}(z), \end{aligned}$$

by equations (3.4) and (3.6). Hence y is an eigenvalue of S and its associated eigenvectors are $\boldsymbol{u}(y)$ and $\boldsymbol{w}(z)$ and so part 2 is proved.

For each $i = 1, 2, \ldots, M-1$ it is easy to see that

$$\boldsymbol{\ell}_i S(y) = \boldsymbol{\ell}_i R_1(z) + \boldsymbol{\ell}_i \boldsymbol{w}(z)\boldsymbol{u}(y)\left[yI - R_1(z)\right],$$

$$= \boldsymbol{\ell}_i R_1(z), \tag{3.9}$$

$$= \lambda_i \boldsymbol{\ell}_i, \tag{3.10}$$

by equations (3.7) and (3.5), respectively. Hence for each $i = 1, 2, \ldots, M-1$, λ_i is an eigenvalue of $S(y)$ and has associated left eigenvector $\boldsymbol{\ell}_i$ and so parts 3 and 4 are proved.

Finally, the fact that $y > z \geq \lambda_i$ is sufficient to show that y is the maximal eigenvalue of $S(y)$ and hence part 1 is proved.

Now let's consider the non-diagonalisable case. Here there is not a full set of eigenvectors and so we have to consider generalized eigenvectors. The major effect this has on the above is that a generalized eigenvector may not obey equation (3.5). Part 2 follows as before, but in the proof of part 3 the argument must stop at equation (3.9) as equation (3.10) may not apply.

Define the $M-1\times M$ matrix L whose rows consist of the left eigenvectors (and generalized eigenvectors) of $R_1(z)$ corresponding to the (not necessarily distinct) eigenvalues λ_i, $i = 1, 2, \ldots, M-1$. General theory (see for example Gantmacher [2] or Noble [11]) tells us, subject to a permutation of rows, that

$$\begin{bmatrix} L \\ \boldsymbol{u}(z) \end{bmatrix} R_1(z) \begin{bmatrix} L \\ \boldsymbol{u}(z) \end{bmatrix}^{-1} = \begin{bmatrix} J' & 0 \\ 0 & z \end{bmatrix}, \tag{3.11}$$

where the matrix $J = \begin{bmatrix} J' & 0 \\ 0 & z \end{bmatrix}$ is known as the Jordon-canonical form. (Recall that this matrix has the eigenvalues listed on the diagonal, the super-diagonal may have **some** elements that are one and all other entries are zero.)

Equations (3.8), (3.9) and (3.11) imply that

$$\begin{bmatrix} L \\ \boldsymbol{u}(y) \end{bmatrix} S(y) = \begin{bmatrix} J' & 0 \\ 0 & y \end{bmatrix} \begin{bmatrix} L \\ \boldsymbol{u}(y) \end{bmatrix}. \tag{3.12}$$

Therefore, if we can show that the matrix $\begin{bmatrix} L \\ \boldsymbol{u}(y) \end{bmatrix}$ is non-singular then this is sufficient to directly conclude parts 3 and 4 and then also part 1.

To show that $\begin{bmatrix} L \\ \boldsymbol{u}(y) \end{bmatrix}$ is non-singular it is easiest to show that all the rows are linearly independent. Since L is part of a Jordan decomposition of the matrix $R_1(z)$ we know that its rows are linearly independent. Therefore,

$$\boldsymbol{c}L = \boldsymbol{0} \quad \text{iff} \quad \boldsymbol{c} = \boldsymbol{0}. \tag{3.13}$$

Consequently, it just remains to show that the vector $\boldsymbol{u}(y)$ is linearly independent of the rows of the matrix L.

Recall that $y > \lambda_i$ for all $i = 1, 2, \ldots, M-1$. Consider the equation

$$\begin{aligned} \boldsymbol{0} &= \boldsymbol{d}LS(y) - y\boldsymbol{d}L, \\ &= \boldsymbol{d}J'L - y\boldsymbol{d}L, \\ &= \sum_{i=1}^{M-1} d_i\lambda_i\boldsymbol{\ell}_i + d_i\delta_i\boldsymbol{\ell}_{i+1} - d_iy\boldsymbol{\ell}_i, \end{aligned}$$

$$= \sum_{i=1}^{M-1} d_i(\lambda_i - y)\boldsymbol{\ell}_i + \sum_{i=1}^{M-1} d_i\delta_i\boldsymbol{\ell}_{i+1},$$

$$= \sum_{i=1}^{M-1} \left[d_i(\lambda_i - y) + d_{i-1}\delta_{i-1}\right]\boldsymbol{\ell}_i.$$

where $\boldsymbol{\ell}_i$ is the ith row of L and δ_i is the $(i, i+1)$ entry of the matrix J' and is either 0 or 1. Further, we have $\delta_{M-1} = \delta_0 = 0$.

Now equation (3.13) implies that

$$d_i(\lambda_i - y) + d_{i-1}\delta_{i-1} = 0, \quad \text{for each } i = 1, 2, \ldots, M-1. \tag{3.14}$$

Since $y > \lambda_i$ for all $i = 1, 2, \ldots, M-1$ we can proceed as follows. Consider equation (3.14) with $i = 1$. The fact that $\delta_0 = 0$ implies that $d_1 = 0$, which in turn (when $i = 2$) implies that $d_2 = 0$. We can continue this argument until we get that $d_{M-1} = 0$. Therefore the only vector $\boldsymbol{d}$ that is a solution to

$$\boldsymbol{d}LS(y) = y\boldsymbol{d}L, \tag{3.15}$$

is the vector $\mathbf{0}$.

Now, equation (3.8) says that $\boldsymbol{u}(y)$ is such that

$$\boldsymbol{u}(y)S(y) = y\boldsymbol{u}(y).$$

Since $\boldsymbol{u}(y)$ is non-zero, the fact that $\boldsymbol{d} = \mathbf{0}$ is the only solution to equation (3.15) is sufficient to conclude that $\boldsymbol{u}(y)$ cannot be of the from $\boldsymbol{d}L$ and hence $\begin{bmatrix} L \\ \boldsymbol{u}(y) \end{bmatrix}$ is non-singular. This completes the proof. ■

Lemma 2 *For all* $j \geq 1$,

$$S(y)^j = R_1(z)^j + \boldsymbol{w}(z)\boldsymbol{u}(y)\left[y^j I - R_1(z)^j\right]. \tag{3.16}$$

Proof: The proof follows a simple mathematical induction argument using the fact that $\boldsymbol{w}(z)$ is the right eigenvector of $R_1(z)$ with eigenvalue z and the normalization condition for $\boldsymbol{u}(y)$. ■

Theorem 3 *A nonnegative solution to equation (2.4) for* $\beta = \beta(y) = \beta(z)$ *with maximal eigenvalue* y *is given by*

$$R_2(y) = S(y) = R_1(z) + \boldsymbol{w}(z)\boldsymbol{u}(y)\left[yI - R_1(z)\right]. \tag{3.17}$$

Proof: First we shall prove that $S(y)$ obeys equation (2.4). Consider the right-hand side of equation (2.4).

$$\begin{aligned}
&\beta(z)\left[A_0 + S(y)A_1 + S(y)^2A_2\right] \\
&= \beta(z)\left[A_0 + R_1(z)A_1 + R_1(z)^2A_2\right] \\
&\quad + \beta(z)\boldsymbol{w}(z)\boldsymbol{u}(y)\left[\left(yI - R_1(z)\right)A_1 + \left(y^2I - R_1(z)^2\right)A_2\right], \\
&= R_1(z) + \\
&\quad \boldsymbol{w}(z)\boldsymbol{u}(y)\beta(z)\left[\left(A_0 + yA_1 + y^2A_2\right) - \left(A_0 + R_1(z)A_1 + R_1(z)^2A_2\right)\right], \\
&= R_1(z) + \boldsymbol{w}(z)\boldsymbol{u}(y)\left[yI - R_1(z)\right], \\
&= S(y),
\end{aligned}$$

since $\beta(z) = \beta(y)$, $\chi(y) = y/\beta(y)$ and $R_1(z)$ obeys equation (2.4) when $\beta = \beta(z)$.

Next, we shall prove that $\boldsymbol{u}(y)(yI - R_1(z))$ is a nonnegative vector. To prove this result we follow a very similar argument to that presented in the proofs of Theorems 2 and 4 in Bean *et al.* [1], as follows. Let $W_0(\beta(z))$ be the zero matrix and define for $N \geq 0$

$$W_{N+1}(\beta(z)) = \beta(z)\left[A_0 + W_N(\beta(z))A_1 + W_N(\beta(z))^2A_2\right].$$

It is easy to show by induction that the sequence $\{W_N(\beta(z))\}$ is nondecreasing. It is also possible to show, again by induction, that

$$\boldsymbol{u}(y)W_N(\beta(z)) \leq y\boldsymbol{u}(y), \tag{3.18}$$

because $\beta(z) = \beta(y)$ and $\boldsymbol{u}(y)$ is the strictly positive left eigenvector of $A(y)$ with eigenvalue $y/\beta(y)$. Since Theorem 2 of Bean *et al.* [1] shows that the sequence $\{W_N(\beta(z))\}$ converges monotonically to $R_1(z)$, we can conclude that

$$\boldsymbol{u}(y)R_1(z) \leq y\boldsymbol{u}(y), \tag{3.19}$$

and hence that $S(y) \geq R_1(z) \geq 0$ (where all inequalities are treated elementwise).

Therefore, $S(y)$ is a nonnegative solution to equation (2.4) and by Lemma 1 part 1 has maximal eigenvalue y. We take $R_2(y) = S(y)$. ■

3.2 THE QUASISTATIONARY DISTRIBUTION FOR EACH $\beta < \alpha$

Theorem 4 *If $\beta < \alpha$, then for any $\boldsymbol{x}$ such that $\boldsymbol{x}\boldsymbol{w}(z_1(\beta)) \neq 0$, the distribution $\boldsymbol{m}(\beta) = (\boldsymbol{m}_1(\beta), \ldots)$ given by*

$$\boldsymbol{m}_j(\beta) = c\boldsymbol{x}\left[R_2(z_2(\beta))^j - R_1(z_1(\beta))^j\right], \tag{3.20}$$

with c a normalising constant, is a β-invariant measure, and hence quasistationary distribution, for (X_n). Moreover, $\boldsymbol{m}_j(\beta)$ has the more compact form

$$\boldsymbol{m}_j(\beta) = k\boldsymbol{u}(z_2(\beta)) \left[z_2(\beta)^j I - R_1^j(z_1(\beta))\right], \tag{3.21}$$

with $k = c\boldsymbol{x}\boldsymbol{w}(z_1(\beta))$.

Proof: We first show that $\boldsymbol{m}(\beta)$ obeys equations (2.1) and (2.2). Consider the right-hand side of equation (2.2) for $k \geq 2$.

$$\begin{aligned}
&\beta\left[\boldsymbol{m}_{k-1}(\beta)A_0 + \boldsymbol{m}_k(\beta)A_1 + \boldsymbol{m}_{k+1}(\beta)A_2\right] \\
&\quad = \boldsymbol{x}R_2(z_2(\beta))^{k-1}\beta\left[A_0 + R_2(z_2(\beta))A_1 + R_2(z_2(\beta))^2A_2\right] \\
&\qquad -\boldsymbol{x}R_1(z_1(\beta))^{k-1}\beta\left[A_0 + R_1(z_1(\beta))A_1 + R_1(z_1(\beta))^2A_2\right], \\
&\quad = \boldsymbol{x}\left(R_2(z_2(\beta))^k - R_1(z_1(\beta))^k\right), \\
&\quad = \boldsymbol{m}_k(\beta),
\end{aligned}$$

since $R_2(z_2(\beta))$ and $R_1(z_1(\beta))$ obey equation (2.4). Consider now the right-hand side of equation (2.1).

$$\begin{aligned}
&\beta\left[\boldsymbol{m}_1(\beta)A_1 + \boldsymbol{m}_2(\beta)A_2\right] \\
&= \boldsymbol{x}\beta\left[R_2(z_2(\beta))A_1 + R_2(z_2(\beta))^2A_2\right] - \boldsymbol{x}\beta\left[R_1(z_1(\beta))A_1 + R_1(z_1(\beta))^2A_2\right], \\
&= \boldsymbol{x}\beta\left[A_0 + R_2(z_2(\beta))A_1 + R_2(z_2(\beta))^2A_2\right] \\
&\quad -\boldsymbol{x}\beta\left[A_0 + R_1(z_1(\beta))A_1 + R_1(z_1(\beta))^2A_2\right], \\
&= \boldsymbol{x}\left(R_2(z_2(\beta)) - R_1(z_1(\beta))\right), \\
&= \boldsymbol{m}_1(\beta),
\end{aligned}$$

since $R_2(z_2(\beta))$ and $R_1(z_1(\beta))$ obey equation (2.4).

The only requirement on the choice of the vector $\boldsymbol{x}$ is that $\boldsymbol{m}_j(\beta)$ should be nonnegative for all $j \geq 1$. This raises the question of how many distinct solutions to equations (2.1) and (2.2) of the form given in equation (3.20) there are for each value of β?

It follows from equation (3.19) that $R_2(z_2)^j - R_1(z_1)^j$ which can also be written as $\boldsymbol{w}(z_1)\boldsymbol{u}(z_2)\left(z_2^j I - R_1^j(z_1)\right)$ is nonnegative for all $j \geq 1$. It is also a rank one matrix. Therefore, there is only one solution of the form (3.20) for each value of β. Any choice of vector $\boldsymbol{x}$ such that $\boldsymbol{x}\boldsymbol{w}(z_1(\beta)) \neq 0$ will realise this β-invariant measure on appropriate normalisation. Equation (3.20) reduces to equation (3.21) by applying equation (3.16) and letting $k = c\boldsymbol{x}\boldsymbol{w}(z_1(\beta))$. ■

3.3 CALCULATION OF THE DERIVATIVE OF THE MATRIX $R_1(z)$

When $\beta = \alpha$, we cannot use the form of the β-invariant measures that we found earlier because $z_1(\alpha) = z_2(\alpha)$ and so $R_2(z_2)$ is identical to $R_1(z_1)$. In order to determine the α-invariant measure, proposed in equation (3.2), we need to calculate the derivative of $R_1(z)$.

Lemma 5 *The derivative of $R_1(z)$ is a solution in T to*

$$\frac{\beta'(z)}{\beta(z)} R_1(z) + \beta(z)\Big[TA_1 + (TR_1(z) + R_1(z)T)A_2\Big] - T = 0\,, \tag{3.22}$$

where $\beta'(z) = \dfrac{\chi(z) - z\chi'(z)}{\chi^2(z)}$.

Proof: First, $\beta(z)$ is defined in equation (2.11) as $\beta(z) = \frac{z}{\chi(z)}$. Therefore, the form of $\beta'(z)$ follows trivially. Also, $R_1(z)$ is defined to be the minimal nonnegative solution to

$$S = \beta\left[A_0 + SA_1 + S^2A_2\right]. \tag{3.23}$$

Differentiating this equation with respect to z, while remembering the functional dependencies, completes the proof of the lemma. ■

Throughout the remainder of this section we assume that $z = z_0$. Recall that $\beta(z_0) = \alpha$ and that z_0 is defined as the minimal solution to $\chi(z_0) = z_0\chi'(z_0)$. Hence $\beta'(z_0) = 0$. Therefore, equation (3.22) can be rewritten when $z = z_0$ as

$$\alpha\Big[TA_1 + (TR_1(z_0) + R_1(z_0)T)A_2\Big] = T\,. \tag{3.24}$$

Lemma 6 *A solution to equation (3.24) is given by*

$$T = \boldsymbol{w}(z_0)\left(\boldsymbol{u}(z_0) + \boldsymbol{u}'(z_0)\left(z_0I - R_1(z_0)\right)\right), \tag{3.25}$$

where $\boldsymbol{u}'(z_0)$ is the unique solution to

$$\boldsymbol{b}\left[z_0^2A_2 + z_0(A_1 - \frac{1}{\alpha}I) + A_0\right] = -\boldsymbol{u}(z_0)\left[2z_0A_2 + (A_1 - \frac{1}{\alpha}I)\right] \tag{3.26}$$

subject to $\boldsymbol{be} = 0$.

Proof: The existence of the vector $\boldsymbol{u}'(z_0)$ is shown in Lemma 8 of Bean *et al.* [1]. To show that T obeys equation (3.24) we need to use equations (2.4) and (3.26) and the fact that $\boldsymbol{u}(z_0)$ and $\boldsymbol{w}(z_0)$ are the left and right eigenvectors of $R_1(z_0)$ with eigenvalue z_0, as follows.

Consider the left-hand side of equation (3.24) with T given by equation (3.25).

$$
\begin{aligned}
& \alpha\Big[TA_1 + (TR_1(z_0) + R_1(z_0)T)A_2\Big] \\
= \;& \alpha \boldsymbol{w}(z_0)\left(\boldsymbol{u}(z_0) + \boldsymbol{u}'(z_0)\left(z_0 I - R_1(z_0)\right)\right) A_1 + \\
& \alpha \boldsymbol{w}(z_0)\left(\boldsymbol{u}(z_0) + \boldsymbol{u}'(z_0)\left(z_0 I - R_1(z_0)\right)\right) R_1(z_0) A_2 + \\
& \alpha R_1(z_0) \boldsymbol{w}(z_0)\left(\boldsymbol{u}(z_0) + \boldsymbol{u}'(z_0)\left(z_0 I - R_1(z_0)\right)\right) A_2, \\
= \;& \alpha \boldsymbol{w}(z_0)\boldsymbol{u}'(z_0)\left[z_0 A_1 + z_0^2 A_2 - R_1(z_0)A_1 - R_1(z_0)^2 A_2\right] + \\
& \alpha \boldsymbol{w}(z_0)\boldsymbol{u}(z_0)\left[A_1 + 2z_0 A_2\right], \\
= \;& \alpha \boldsymbol{w}(z_0)\left\{\boldsymbol{u}'(z_0)\left[z_0^2 A_2 + z_0 A_1 + A_0\right] + \boldsymbol{u}(z_0)\left[2z_0 A_2 + A_1\right]\right\} + \\
& -\alpha \boldsymbol{w}(z_0)\boldsymbol{u}'(z_0)\left[A_0 + R_1(z_0)A_1 + R_1(z_0)^2 A_2\right], \\
= \;& \boldsymbol{w}(z_0)\left(\boldsymbol{u}(z_0) + z_0\boldsymbol{u}'(z_0)\right) - \boldsymbol{w}(z_0)\boldsymbol{u}'(z_0)R_1(z_0), \\
= \;& T,
\end{aligned}
$$

which is the right-hand side and so the proof is complete. ■

The matrix T is a solution to the equation that the derivative $R_1'(z_0)$ must obey. However, we have not actually shown that $T = R_1'(z_0)$. Since the fact that T obeys equation (3.24) is sufficient for our purposes, we shall abuse notation somewhat and write $T = R_1'(z_0)$.

3.4 THE QUASISTATIONARY DISTRIBUTION FOR $\beta = \alpha$

Theorem 7 *For any $\boldsymbol{x}$ such that $\boldsymbol{x}R_1'(z_0) \neq \mathbf{0}$, the distribution $\boldsymbol{m}(\alpha) = (\boldsymbol{m}_1(\alpha), \ldots)$ given by*

$$\boldsymbol{m}_j(\alpha) = c\boldsymbol{x}\frac{d}{dz}\left[R_1(z)^j\right]_{z=z_0}, \tag{3.27}$$

with c a normalising constant, is the α-invariant measure, and hence quasistationary and limiting-conditional distribution, for (X_n). Moreover, $\boldsymbol{m}_j(\alpha)$ has the more explicit form

$$\boldsymbol{m}_j(\alpha) = k\left(z_0^j \boldsymbol{u}'(z_0) + j z_0^{j-1}\boldsymbol{u}(z_0) - \boldsymbol{u}'(z_0)R_1(z_0)^j\right), \tag{3.28}$$

with $k = c\boldsymbol{x}\boldsymbol{w}(z_0)$.

Proof: Note that by the product rule

$$\frac{d}{dz}\left[R_1(z)^j\right]_{z=z_0} = \sum_{\ell=1}^{j} R_1(z_0)^{\ell-1} R_1'(z_0) R_1(z_0)^{j-\ell}. \tag{3.29}$$

To prove that $\boldsymbol{m}(\alpha)$ is an α-invariant measure, we simply need to show that it obeys equation (2.1) and (2.2) when $\beta = \alpha$ and is nonnegative. Consider the right-hand side of equation (2.2) for $k \geq 2$.

$$
\begin{aligned}
&\alpha\left[\boldsymbol{m}_{k-1}(\alpha)A_0 + \boldsymbol{m}_k(\alpha)A_1 + \boldsymbol{m}_{k+1}(\alpha)A_2\right] \\
&\quad= \boldsymbol{x}\sum_{\ell=1}^{k-1} R_1(z_0)^{\ell-1}R_1'(z_0)R_1(z_0)^{k-1-\ell}\alpha\left[A_0 + R_1(z_0)A_1 + R_1(z_0)^2A_2\right] \\
&\qquad -\boldsymbol{x}R_1(z_0)^{k-1}\alpha\left[R_1'(z_0)A_1 + R_1'(z_0)R_1(z_0)A_2 + R_1(z_0)R_1'(z_0)A_2\right], \\
&\quad= \boldsymbol{x}\sum_{\ell=1}^{k-1} R_1(z_0)^{\ell-1}R_1'(z_0)R_1(z_0)^{k-\ell} + \boldsymbol{x}R_1(z_0)^{k-1}R_1'(z_0), \\
&\quad= \boldsymbol{x}\sum_{\ell=1}^{k} R_1(z_0)^{\ell-1}R_1'(z_0)R_1(z_0)^{k-\ell}, \\
&\quad= \boldsymbol{m}_k(\alpha),
\end{aligned}
$$

since $R_1(z_0)$ obeys equation (2.4) and $R_1'(z_0)$ obeys equation (3.24). Consider now the right-hand side of equation (2.1).

$$
\begin{aligned}
&\beta\left[\boldsymbol{m}_1(\alpha)A_1 + \boldsymbol{m}_2(\alpha)A_2\right] \\
&\qquad= \boldsymbol{x}\beta\left[R_1'(z_0)A_1 + R_1'(z_0)R_1(z_0)A_2 + R_1(z_0)R_1'(z_0)A_2\right], \\
&\qquad= \boldsymbol{x}R_1'(z_0), \\
&\qquad= \boldsymbol{m}_1(\alpha),
\end{aligned}
$$

again since $R_1'(z_0)$ obeys equation (3.24).

It is easy to show that $R_1'(z_0)$ is a rank one matrix. Therefore, there is a unique solution to equations (2.1) and (2.2) of the form given in equation (3.27). This solution will be realised for all $\boldsymbol{x}$ such that $\boldsymbol{x}R_1'(z_0) \neq \mathbf{0}$.

We now show that equation (3.27) reduces to equation (3.28). Simple substitution is sufficient to complete the proof, on recalling that $\boldsymbol{u}(z_0)$ and $\boldsymbol{w}(z_0)$ are the left and right eigenvectors of $R_1(z_0)$ associated with the eigenvalue z_0, as follows.

$$
\begin{aligned}
&\boldsymbol{m}_j(\alpha) \\
&\quad= c\boldsymbol{x}\frac{d}{dz}\left[R_1(z)^j\right]_{z=z_0}, \\
&\quad= c\boldsymbol{x}\sum_{\ell=1}^{j} R_1(z_0)^{\ell-1}R_1'(z_0)R_1(z_0)^{j-\ell},
\end{aligned}
$$

$$
\begin{aligned}
&= c\boldsymbol{x}\sum_{\ell=1}^{j} R_1(z_0)^{\ell-1}\boldsymbol{w}(z_0)\left(\boldsymbol{u}(z_0)+\boldsymbol{u}'(z_0)\left(z_0 I - R_1(z_0)\right)\right)R_1(z_0)^{j-\ell},\\
&= c\boldsymbol{x}\sum_{\ell=1}^{j} z_0^{\ell-1}\boldsymbol{w}(z_0)\left(\boldsymbol{u}(z_0)+\boldsymbol{u}'(z_0)\left(z_0 I - R_1(z_0)\right)\right)R_1(z_0)^{j-\ell},\\
&= c\boldsymbol{x}\sum_{\ell=1}^{j} z_0^{\ell-1}\boldsymbol{w}(z_0)\\
&\quad\times\left(\boldsymbol{u}(z_0)R_1(z_0)^{j-\ell}+\boldsymbol{u}'(z_0)z_0R_1(z_0)^{j-\ell}-\boldsymbol{u}'(z_0)R_1(z_0)^{j-\ell+1}\right),\\
&= c\boldsymbol{x}\boldsymbol{w}(z_0)\left(j\boldsymbol{u}(z_0)z_0^{j-1}+\boldsymbol{u}'(z_0)\sum_{\ell=1}^{j}\left[z_0^{\ell}R_1(z_0)^{j-\ell}-z_0^{\ell-1}R_1(z_0)^{j-\ell+1}\right]\right),\\
&= k\left(j\boldsymbol{u}(z_0)z_0^{j-1}+\boldsymbol{u}'(z_0)z_0^{j}-\boldsymbol{u}'(z_0)R_1(z_0)^{j}\right),
\end{aligned}
$$

by letting $k = c\boldsymbol{x}\boldsymbol{w}(z_0)$.

Part (iii) of the proof of Theorem 9 of Bean *et al.* [1] shows that the expression for $\boldsymbol{m}_j(\alpha)$ given in equation (3.28) is nonnegative for all $j \geq 1$. Hence, the expression for $\boldsymbol{m}_j(\alpha)$ given in equation (3.27) must also be nonnegative for all $j \geq 1$.

Therefore, the solution to equations (2.1) and (2.2) proposed in equation (3.27) is the unique α-invariant measure, and hence quasistationary and limiting-conditional distribution, for (X_n). ■

4 COMPUTATION OF THE DISTRIBUTIONS

In this section we briefly indicate the steps involved in the computation of the quasistationary distributions. We assume that α and z_0 have already been computed according to Theorem 2, for more computational details see Section 6(i) of Bean *et al.* [1].

4.1 THE CASE $\beta < \alpha$

In this situation we need to calculate $\boldsymbol{m}_j(\beta)$ according to equation (3.20). That is, we need to find $R_1(z_1(\beta))$ and $R_2(z_2(\beta))$. First we need to determine both $z_1(\beta)$ and $z_2(\beta)$. These can be evaluated by performing bisection searches on the intervals $[z_1(1), z_0]$ and $[z_0, 1]$, respectively, to determine the two solutions to $\chi(z) = \frac{z}{\beta(z)}$ (this is a similar procedure to that of finding z_0 and α). At the same time it is most efficient also to generate $\boldsymbol{u}(z_2(\beta))$. Then we must find $R_1(z_1(\beta))$. This can be evaluated using the algorithm explained in

Theorem 10 of Bean *et al.* [1]. It is then easy to calculate $\boldsymbol{w}(z_1(\beta))$ by elementary methods and $R_2(z_2(\beta))$ using Theorem 3. Finally, Theorem 4 can be applied to generate the required quasistationary distribution, where it is computationally easier to use equation (3.21) rather than equation (3.20).

4.2 THE CASE $\beta = \alpha$

In this situation we need to calculate $\boldsymbol{m}_j(\alpha)$ according to equation (3.27). That is, we need to find $R_1(z_0)$ and $R_1'(z_0)$. We assume that z_0 and $\boldsymbol{u}(z_0)$ have already been evaluated at the same time as α. Again, $R_1(z_0)$ can be evaluated using the algorithm explained in Theorem 10 of Bean *et al.* [1]. It is then simple to calculate $\boldsymbol{u}'(z_0)$ as in equation (3.26) (more details are given in Section 6(ii) of Bean *et al.* [1]) and $R_1'(z_0)$ using Lemma 6. Finally, Theorem 7 can be applied to generate the required quasistationary (and in fact limiting-conditional) distribution. It happens that this is not the best numerical method for evaluating the limiting-conditional distribution. From a computational point of view, it is better to use the explicit representation given in equation (3.28) instead of that given in equation (3.27).

Acknowledgements

We thank the referees and the editor for valuable comments and suggestions. The financial support of the Australian Research Council through grants A69532348 and A69022762 is gratefully acknowledged.

References

[1] Bean, N.G., Bright, L., Latouche, G., Pearce, C.E.M., Pollett, P.K. and Taylor, P.G., The quasistationary behavior of quasi-birth-and-death processes. Ann. Appl. Probab., 7 (1997), 134–155.

[2] Gantmacher, F.R., *The Theory of Matrices.* Chelsea (1959), New York.

[3] Kijima, M., Quasi-stationary distributions of phase-type queues. Math. Oper. Res., 18 (1993), 423–437.

[4] Kijima, M. and Makimoto, N., Quasi-stationary distributions of Markov chains arising from queueing processes: A survey. Research reports on Mathematical and Computing Sciences, Series B: Operations Research B-304, Department of Mathematical and Computing Sciences, Tokyo Institute of Technology, Tokyo, Japan, August 1995.

[5] Kingman, J.F.C., A convexity property of positive matrices. Quart. J. Math. Oxford, Series 2, 12 (1961), 283–284.

[6] Kyprianou, E.K., On the quasi-stationary distributions of the GI/M/1 queues. J. Appl. Prob. 9 (1972), 117–128.

[7] Kyprianou, E.K., The quasi-stationary distributions of queues in heavy traffic. J. Appl. Prob. 9 (1972), 821–831.

[8] Latouche, G. and Taylor, P.G., Level and phase independence in QBDs and related processes. In Preparation, 1997.

[9] Makimoto, N., Quasi-stationary distributions in a PH/PH/c queue. Stochastic Models 9 (1993), 195–212.

[10] Neuts, M.F., *Matrix Geometric Solutions in Stochastic Models.* The John Hopkins University Press (1981), Baltimore.

[11] Noble, B., *Applied Linear Algebra.* Prentice-Hall, Englewood Cliffs (1969), New Jersey.

[12] Seneta, E., *Non-negative Matrices and Markov Chains.* Springer-Verlag (1981), New York.