Survival in a quasi-death process

Erik A. van Doorn[a] and Philip K. Pollett[b]

[a] Department of Applied Mathematics, University of Twente
P.O. Box 217, 7500 AE Enschede, The Netherlands
E-mail: e.a.vandoorn@utwente.nl

[b] Department of Mathematics, The University of Queensland
Qld 4072, Australia
E-mail: pkp@maths.uq.edu.au

April 1, 2008

Abstract. We consider a Markov chain in continuous time with one absorbing state and a finite set S of transient states. When S is irreducible the limiting distribution of the chain as $t \to \infty$, conditional on survival up to time t, is known to equal the (unique) quasi-stationary distribution of the chain. We address the problem of generalizing this result to a setting in which S may be reducible, and show that it remains valid if the eigenvalue with maximal real part of the generator of the (sub)Markov chain on S has geometric (but not, necessarily, algebraic) multiplicity one. The result is then applied to pure death processes and, more generally, to quasi-death processes. We also show that the result holds true even when the geometric multiplicity is larger than one, provided the irreducible subsets of S satisfy an accessibility constraint. A key role in the analysis is played by some classic results on M-matrices.

Keywords and phrases: absorbing Markov chain, death process, limiting conditional distribution, migration process, M-matrix, quasi-stationary distribution, survival-time distribution

2000 Mathematics Subject Classification: Primary 60J27, Secondary 15A18

1 Introduction

In the interesting papers [2] and [3] Aalen and Gjessing provide a new explanation for the shape of hazard rate functions in survival analysis. They propose to model survival times as sojourn times of stochastic processes in a set S of transient states until they escape from S to an absorbing state. This "process point of view" entails that (in the words of Aalen and Gjessing) "the shape of the hazard rate is created in a balance between two forces: the attraction of the absorbing state and the general diffusion within the transient space". In other words, the shape of the hazard rate is determined by the interaction of the initial distribution and the distribution over S known as the *quasi-stationary distribution* of the process. Similar ideas have been put forward independently by Steinsaltz and Evans [26].

Aalen and Gjessing discuss several examples of relevant stochastic processes, including finite-state Markov chains with an absorbing state, which is the setting of the present paper. A survival-time distribution in this setting is known as a *phase-type distribution* (see, for example, Latouche and Ramaswami [15, Ch. 2], or Aalen [1]). In their analysis and examples Aalen and Gjessing restrict themselves to chains for which the set S of transient states constitutes a single class, arguing that "irreducibility is important when considering quasi-stationary distributions". As we shall see, however, there are no compelling technical reasons for imposing this restriction. Moreover, in [3, Section 8] Aalen and Gjessing allude to a bottleneck phenomenon that may occur when S is reducible, making it even more desirable to investigate what happens in this case. We note that Proposition 1 in [26], while formulated quite generally, is entirely correct only if one assumes S to be irreducible.

From a modelling point of view there is another argument for extending the analysis to reducible sets S. Namely, if the state of an organism before evanescence is represented by the state of a transient Markov chain, it seems reasonable to allow for the possibility that some transitions are irreversible, reflecting the fact that some real-life processes such as *ageing* are irreversible.

The main aim of the present paper is to provide the tools for hazard rate analysis, by characterizing survival-time distributions and identifying limiting conditional distributions and quasi-stationary distributions, in the setting of finite Markov chains with an absorbing state and a set S of transient states that may be reducible. In Section 2 we perform these tasks under the assumption that the eigenvalue with maximal real part of the generator of the Markov chain has geometric (but not, necessarily, algebraic) multiplicity one. The results are applied in Section 3 to pure death processes, and subsequently in Section 4 to *quasi-death processes*, which may be viewed as death processes in which the sojourn time in each state has a phase-type distribution. By way of illustration we discuss a specific example of a quasi-death process in Section 5. Finally, in Section 6, we generalize some of the results of Section 2 to a setting in which the geometric multiplicity of the eigenvalue with maximal real part may be larger than one. In particular, we obtain a necessary and sufficient condition for the finite Markov chain to have a unique quasi-stationary distribution.

2 Absorbing Markov chains

2.1 Preliminaries

Consider a continuous-time Markov chain $\mathcal{X} := \{X(t),\ t \geq 0\}$ on a state space $\{0\} \cup S$ consisting of an absorbing state 0 and a finite set of transient states $S := \{1, 2, \ldots, n\}$. The generator of $\mathcal{X}$ then takes the form

$$\begin{pmatrix} 0 & \mathbf{0} \\ \boldsymbol{q}^T & Q \end{pmatrix}, \tag{1}$$

where

$$\boldsymbol{q} = -\mathbf{1}Q^T \geq \mathbf{0},\ \ \boldsymbol{q} \neq \mathbf{0}. \tag{2}$$

Here $\mathbf{0}$ and $\mathbf{1}$ are row vectors of zeros and ones, respectively, superscript ^{T} denotes transposition, and $\geq$ indicates componentwise inequality. Since all states in S are transient, state 0 is accessible from any state in S. Hence, whichever the initial state, the process will eventually escape from S into the absorbing state 0 with probability one.

We write $\mathbb{P}_i(\cdot)$ for the probability measure of the process when $X(0) = i$, and let $\mathbb{P}_{\boldsymbol{w}}(\cdot) := \sum_i w_i \mathbb{P}_i(\cdot)$ for any vector $\boldsymbol{w} = (w_1, w_2, \dots, w_n)$ representing a distribution over S. Also, $P_{ij}(\cdot) := \mathbb{P}_i(X(\cdot) = j)$. It is easy to verify (see, for example, Kijima [13, Section 4.6]) that the matrix $P(t) := (P_{ij}(t),\ i, j \in S)$ satisfies

$$P(t) = e^{Qt} := \sum_{k=0}^{\infty} \frac{Q^k}{k!} t^k, \quad t \geq 0.$$

By $T := \sup\{t \geq 0 : X(t) \in S\}$ we denote the *survival time* (or *absorption time*) of $\mathcal{X}$, the random variable representing the time at which escape from S occurs. In what follows we are interested in the limiting distribution as $t \to \infty$ of the residual survival time conditional on survival up to time t, that is,

$$\lim_{t\to\infty} \mathbb{P}_{\boldsymbol{w}}(T > t + s \,|\, T > t), \quad s \geq 0, \tag{3}$$

and in the limiting distribution as $t \to \infty$ of $X(t)$ conditional on survival up to time t, that is,

$$\lim_{t\to\infty} \mathbb{P}_{\boldsymbol{w}}(X(t) = j \,|\, T > t), \quad j \in S, \tag{4}$$

where $\boldsymbol{w}$ is any initial distribution over S.

2.2 Irreducible state space

Let us first suppose that S is irreducible, that is, constitutes a single communicating class. In this case Q has a unique eigenvalue with maximal real part, which we denote by $-\alpha$. It is well known (see, for example, Seneta [25, Theorem 2.6]) that α is real and positive, and that the associated left and right eigenvectors $\boldsymbol{u} = (u_1, u_2, \dots, u_n)$ and $\boldsymbol{v}^T = (v_1, v_2, \dots, v_n)^T$ can be chosen strictly positive componentwise. It will also be convenient to normalize $\boldsymbol{u}$ and $\boldsymbol{v}$ such that

$$\boldsymbol{u}\mathbf{1}^T = 1 \ \text{ and } \ \boldsymbol{u}\boldsymbol{v}^T = 1. \tag{5}$$

It then follows (see Mandl [19]) that the transition probabilities $P_{ij}(t)$ satisfy

$$\lim_{t\to\infty} e^{\alpha t} P_{ij}(t) = v_i u_j > 0, \quad i, j \in S, \tag{6}$$

which explains why α is often referred to as the *decay parameter* of $\mathcal{X}$.

Since $\boldsymbol{u}Q = -\alpha\boldsymbol{u}$, we have $\boldsymbol{u}Q^k = (-\alpha)^k\boldsymbol{u}$ for all k, and hence

$$\boldsymbol{u}P(t) = \sum_{k=0}^{\infty} \frac{\boldsymbol{u}Q^k}{k!} t^k = e^{-\alpha t}\boldsymbol{u}, \quad t \geq 0, \tag{7}$$

that is

$$\mathbb{P}_{\boldsymbol{u}}(X(t) = j) = e^{-\alpha t} u_j, \quad j \in S, \ t \geq 0. \tag{8}$$

Considering that $\mathbb{P}_{\boldsymbol{u}}(T > t) = \mathbb{P}_{\boldsymbol{u}}(X(t) \in S) = e^{-\alpha t}$, it follows that for all $t \geq 0$

$$\mathbb{P}_{\boldsymbol{u}}(T > t + s \,|\, T > t) = e^{-\alpha s}, \quad s \geq 0, \tag{9}$$

and, moreover, that $\boldsymbol{u}$ is a *quasi-stationary distribution* of $\mathcal{X}$ in the sense that for all $t \geq 0$

$$\mathbb{P}_{\boldsymbol{u}}(X(t) = j \,|\, T > t) = u_j, \quad j \in S. \tag{10}$$

So, when $\boldsymbol{u}$ is the initial distribution, the distribution of $X(t)$ conditional on absorption not yet having taken place at time t is constant over t, and the survival time has an exponential distribution with parameter α. Darroch and Seneta [8] have shown that similar results hold true in the limit as $t \to \infty$ when the initial distribution differs from $\boldsymbol{u}$. Namely, for any initial distribution $\boldsymbol{w}$ one has

$$\lim_{t\to\infty} \mathbb{P}_{\boldsymbol{w}}(T > t + s \,|\, T > t) = e^{-\alpha s}, \quad s \geq 0, \tag{11}$$

and

$$\lim_{t\to\infty} \mathbb{P}_{\boldsymbol{w}}(X(t) = j \,|\, T > t) = u_j, \quad j \in S. \tag{12}$$

So when all states in S communicate the limits (3) and (4) are determined by the eigenvalue of Q with maximal real part and the corresponding left eigenvector.

This result can be generalized to a setting in which S may consist of more than one class, as we will show next.

2.3 General state space

Suppose now that S consists of communicating classes $S_1, S_2, \ldots, S_L$, and let Q_k be the submatrix of $Q = (q_{ij})$ corresponding to the states in S_k. We define a partial order on $\{S_1, S_2, \ldots, S_L\}$ by writing $S_i \prec S_j$ when S_i is *accessible* from S_j, that is, when there exists a sequence of states $k_0, k_1, \ldots, k_\ell$, such that $k_0 \in S_j$, $k_\ell \in S_i$, and $q_{k_m k_{m+1}} > 0$ for every m. We will assume in what follows that the states are labelled such that Q is in lower block-triangular form, so that we must have

$$S_i \prec S_j \implies i \leq j. \tag{13}$$

Noting that the matrices Q_k reside on the diagonal of Q, it follows easily that the set of eigenvalues of Q is precisely the union of the sets of eigenvalues of the individual Q_k's. So, if we denote the (unique) eigenvalue with maximal real part of Q_k by $-\alpha_k$ (so that α_k is real and positive) and let $\alpha := \min_k \alpha_k$, then $-\alpha$ is the eigenvalue of Q with maximal real part.

Evidently, $-\alpha$ may be a degenerate eigenvalue. Assuming, however, that $-\alpha$ has geometric (but not, necessarily, algebraic) multiplicity one, there exist, up to constant factors, unique left and right eigenvectors $\boldsymbol{u}$ and $\boldsymbol{v}^T$ corresponding to $-\alpha$. Moreover, it follows, for example, from Theorem I* of Debreu and Herstein [9] (by an argument similar to the proof of [25, Theorem 2.6]) that we may choose $\boldsymbol{u} \geq \mathbf{0}$ and $\boldsymbol{v} \geq \mathbf{0}$, but $\boldsymbol{u}$ and $\boldsymbol{v}$ are not necessarily positive componentwise. As before we will assume $\boldsymbol{u}$ to be normalized such that $\boldsymbol{u}\mathbf{1}^T = 1$. In this setting (8), and hence (9) and (10), retain their validity.

We let $I(\alpha) := \{k : \alpha_k = \alpha\}$, so that $\mathrm{card}(I(\alpha))$ is the algebraic multiplicity of the eigenvalue $-\alpha$, and define

$$a(\alpha) := \min I(\alpha) \quad \text{and} \quad b(\alpha) := \max I(\alpha). \tag{14}$$

Maintaining the assumption that $-\alpha$ has geometric multiplicity one, we note that we must have $u_j = 0$ if j is *not* accessible from $S_{a(\alpha)}$. Indeed, $\boldsymbol{u}$ being the unique solution of the system $\boldsymbol{u}Q = -\alpha\boldsymbol{u}$ and $\boldsymbol{u}\mathbf{1}^T = 1$, it is readily seen that we can determine $\boldsymbol{u}$ by first solving the eigenvector problem in the restricted setting of states that are accessible from $S_{a(\alpha)}$, and subsequently putting $u_j = 0$

whenever j is not accessible from $S_{a(\alpha)}$. Next observe that the union of sets $\cup_{k \in I(\alpha)} S_k$ must be accessible from $\boldsymbol{u}$ (that is, accessible from a state i such that $u_i > 0$), for otherwise α cannot feature in (8), $-\alpha$ being an eigenvalue of Q_k only if $k \in I(\alpha)$. But since $u_j = 0$ if $j \in S_k$ with $k > a(\alpha)$, it follows that, actually, $S_{a(\alpha)}$ must be accessible from $\boldsymbol{u}$. Finally, it is well known that $\mathbb{P}_{\boldsymbol{u}}(X(t) = j) > 0$ for all $t > 0$ if and only if j is accessible from $\boldsymbol{u}$, so, by (8), we must have $u_j > 0$ for all states j that are accessible from $\boldsymbol{u}$, and in particular for all states j that are accessible from $S_{a(\alpha)}$. Combining the preceding results we conclude that $u_j > 0$ if and only if state j is accessible from $S_{a(\alpha)}$.

The counterpart of (8) for the right eigenvector $\boldsymbol{v}^T$ is the relation

$$\sum_{j \in S} P_{ij}(t) v_j = e^{-\alpha t} v_i, \quad i \in S, \tag{15}$$

which may be used in a similar way to show that $v_i > 0$ if and only if $S_{b(\alpha)}$ is accessible from i. It follows in particular that both $u_j > 0$ *and* $v_j > 0$ if (and only if) $a(\alpha) = b(\alpha)$ and $j \in S_{a(\alpha)}$. Since we do not want to exclude the possibility $a(\alpha) > b(\alpha)$, we cannot impose the normalization $\boldsymbol{u}\boldsymbol{v}^T = 1$, but will rather assume in what follows that $\boldsymbol{v}$ satisfies $\boldsymbol{v}\mathbf{1}^T = 1$. We summarize our findings in a theorem.

Theorem 1 If $-\alpha$, the eigenvalue of Q with maximal real part, has geometric multiplicity one, then there are unique vectors $\boldsymbol{u} \geq \mathbf{0}$ and $\boldsymbol{v} \geq \mathbf{0}$ satisfying $\boldsymbol{u}Q = -\alpha\boldsymbol{u}$, $Q\boldsymbol{v}^T = -\alpha\boldsymbol{v}^T$, and $\boldsymbol{u}\mathbf{1}^T = \boldsymbol{v}\mathbf{1}^T = 1$. The jth component of $\boldsymbol{u}$ is positive if and only if state j is accessible from $S_{a(\alpha)}$, whereas the jth component of $\boldsymbol{v}$ is positive if and only if $S_{b(\alpha)}$ is accessible from state j.

The above theorem may alternatively be established by an appeal to the theory of *M-matrices*, which are matrices that can be represented as $cI - P$, where P is a nonnegative matrix and $c \geq \rho(P)$, the spectral radius of P (see, for example, Schneider [24]). Indeed, choosing λ so large that the matrix $Q + \lambda I$ is nonnegative, we have $\rho(Q + \lambda I) = \lambda - \alpha$, and

$$-(Q + \alpha I) = (\lambda - \alpha)I - (Q + \lambda I), \tag{16}$$

so that $-(Q+\alpha I)$ (and, similarly, $-(Q^T+\alpha I)$) is an M-matrix. We can obtain the results of Theorem 1 by applying Schneider's [23, Theorem 2] (see also [24, Theorem 3.1]) to $-(Q+\alpha I)$ and to $-(Q^T+\alpha I)$, and subsequently interpreting the result in the current setting. (We will display further-reaching consequences of Schneider's theorem in Section 6.)

The vector $\boldsymbol{u}$ in Theorem 1 does not necessarily constitute the only quasi-stationary distribution of the process $\mathcal{X}$, that is, the only initial distribution satisfying (10) for all $t \geq 0$. However, we can achieve uniqueness if we restrict ourselves to initial distributions from which $S_{a(\alpha)}$ is accessible. To prove this statement we need the following invariance result.

Lemma 2 If the initial distribution $\boldsymbol{w}$ is such that $S_{a(\alpha)}$ is accessible and satisfies $\boldsymbol{w}Q = x\boldsymbol{w}$ for some real $x < 0$, then $x = -\alpha$, so that $\boldsymbol{w} = \boldsymbol{u}$ if the eigenvalue $-\alpha$ has geometric multiplicity one.

Proof When the initial distribution $\boldsymbol{w} = (w_1, w_2, \ldots, w_n)$ is a left eigenvector corresponding to the eigenvalue x, then, by an argument similar to the one leading to (8), we have

$$\mathbb{P}_{\boldsymbol{w}}(X(t) = j) = e^{xt} w_j, \quad j \in S, \ t \geq 0.$$

It follows that $w_j > 0$ for all states j that are accessible from $\boldsymbol{w}$. So, $S_{a(\alpha)}$ being accessible from $\boldsymbol{w}$, we have $w_j > 0$ for all $j \in S_{a(\alpha)}$. Since $\mathbb{P}_{\boldsymbol{w}}(X(t) = j) \geq w_j P_{jj}(t)$, it follows that

$$P_{jj}(t) \leq e^{xt}, \quad j \in S_{a(\alpha)}, \ t \geq 0.$$

Consequently, in view of (6) applied to the process restricted to $S_{a(\alpha)}$, we must have $x = -\alpha$, whence $\boldsymbol{w} = \boldsymbol{u}$ if $-\alpha$ has geometric multiplicity one. □

We can now copy the arguments in [8] (in which a similar invariance result is implicitly used) and conclude the following.

Theorem 3 If $-\alpha$, the eigenvalue of Q with maximal real part, has geometric multiplicity one then $\mathcal{X}$ has a unique quasi-stationary distribution $\boldsymbol{u}$ from which $S_{a(\alpha)}$ is accessible. The vector $\boldsymbol{u}$ is the (unique, nonnegative) solution of the system $\boldsymbol{u}Q = -\alpha\boldsymbol{u}$ and $\boldsymbol{u}\mathbf{1}^T = 1$.

The restriction to quasi-stationary distributions from which $S_{a(\alpha)}$ is accessible is essential. Without it there may be more than one quasi-stationary distribution; an example is given in Section 3.

Before determining the limits (11) and (12) in the setting at hand we establish the following generalization of (6).

Theorem 4 If $-\alpha$, the eigenvalue of Q with maximal real part, has geometric multiplicity one and algebraic multiplicity $m :=\text{card}(I(\alpha)) \geq 1$, then

$$\lim_{t\to\infty} \frac{e^{\alpha t}}{t^{m-1}} P(t) = c\boldsymbol{v}^T \boldsymbol{u}, \tag{17}$$

where $\boldsymbol{u}$ and $\boldsymbol{v}^T$ are the eigenvectors defined in Theorem 1 and c is some positive constant.

Proof With $J = (J_{ij})$ denoting the Jordan canonical form of Q (see, for example, Friedberg et al. [11]), there exists a nonsingular matrix $\Sigma = (\Sigma_{ij})$ such that $Q = \Sigma J \Sigma^{-1}$, and hence

$$P(t) = e^{tQ} = \Sigma e^{tJ} \Sigma^{-1}, \quad t \geq 0. \tag{18}$$

Denoting the kth Jordan block on the diagonal of the matrix J by $J^{(k)}$, the matrix exponential e^{tJ} will be a block diagonal matrix with blocks $e^{tJ^{(k)}}$. Since $-\alpha$ has geometric multiplicity one there is precisely one Jordan block associated with $-\alpha$, which, without loss of generality, we assume to be $J^{(1)}$. Since the algebraic multiplicity of $-\alpha$ is m, block $J^{(1)}$ is an $m \times m$ matrix. We will treat the cases $m = 1$ and $m > 1$ separately.

First, if $m = 1$ then $J^{(1)} = (-\alpha)$, so that $(e^{tJ})_{11} = e^{-\alpha t}$, while $(e^{tJ})_{1j} = (e^{tJ})_{j1} = 0$ if $j > 1$. It follows that

$$P_{ij}(t) = e^{-\alpha t} \Sigma_{i1} (\Sigma^{-1})_{1j} + o\left(e^{-\alpha t}\right) \quad \text{as} \quad t \to \infty, \quad i, j \in S,$$

and hence

$$\lim_{t\to\infty} e^{\alpha t} P(t) = \boldsymbol{s}^T \boldsymbol{t},$$

where $\boldsymbol{s}^T$ denotes the first column of Σ and $\boldsymbol{t}$ the first row of Σ^{-1}. Since $Q\Sigma = \Sigma J$ we must have $Q\boldsymbol{s}^T = -\alpha \boldsymbol{s}^T$, so it is no restriction to assume that $\boldsymbol{s}$ is normalized such that $\boldsymbol{s} = \boldsymbol{v}$. On the other hand, since $\Sigma^{-1}Q = J\Sigma^{-1}$ we have $\boldsymbol{t}Q = -\alpha\boldsymbol{t}$, so that $\boldsymbol{t} = c\boldsymbol{u}$ for some constant $c \neq 0$. Actually, since $\boldsymbol{t}\boldsymbol{s}^T = \boldsymbol{t}\boldsymbol{v}^T = 1$, we must have $c = 1/\boldsymbol{u}\boldsymbol{v}^T > 0$.

Next, if $m > 1$ then the block $J^{(1)}$ is of the form

$$J^{(1)} = \begin{pmatrix} -\alpha & 1 & 0 & \cdots & 0 & 0 \\ 0 & -\alpha & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -\alpha & 1 \\ 0 & 0 & 0 & \cdots & 0 & -\alpha \end{pmatrix},$$

so that we may write $J^{(1)} = -\alpha I + N$, where I is the $m \times m$ identity matrix and N is the $m \times m$ matrix with ones on the super-diagonal and zeros elsewhere. Obviously, N is nilpotent with index m, whence

$$e^{tN} = I + tN + \frac{t^2N^2}{2!} + \ldots + \frac{t^{m-1}N^{m-1}}{(m-1)!}.$$

Since the act of raising N to the power k amounts to pushing up the diagonal of 1's $k-1$ places, it follows that

$$e^{tJ^{(1)}} = e^{-\alpha t}e^{tN} = e^{-\alpha t} \begin{pmatrix} 1 & t & \frac{t^2}{2!} & \cdots & \frac{t^{m-2}}{(m-2)!} & \frac{t^{m-1}}{(m-1)!} \\ 0 & 1 & t & \cdots & \frac{t^{m-3}}{(m-3)!} & \frac{t^{m-2}}{(m-2)!} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & t \\ 0 & 0 & 0 & \cdots & 0 & 1 \end{pmatrix}.$$

By a similar argument it can be shown that for $k > 1$ the elements of $e^{tJ^{(k)}}$, which correspond to eigenvalues smaller than $-\alpha$, will be $o(e^{-\alpha t})$ as $t \to \infty$. Hence, by (18), the dominant term in $P_{ij}(t)$ is determined by $(e^{tJ})_{1m} = (e^{tJ^{(1)}})_{1m}$, namely

$$P_{ij}(t) = \frac{t^{m-1}e^{-\alpha t}}{(m-1)!}\Sigma_{i1}(\Sigma^{-1})_{mj} + o\left(t^{m-1}e^{-\alpha t}\right) \quad \text{as} \quad t \to \infty, \quad i, j \in S.$$

Hence

$$\lim_{t\to\infty} \frac{e^{\alpha t}}{t^{m-1}}P(t) = \frac{1}{(m-1)!}\boldsymbol{s}^T\boldsymbol{t},$$

where $\boldsymbol{s}^T$ is, again, the first column of Σ and $\boldsymbol{t}$ now stands for the mth row of Σ^{-1}. As before, $Q\Sigma = \Sigma J$ implies that we must have $Q\boldsymbol{s}^T = -\alpha \boldsymbol{s}^T$, so it is no restriction to assume that $\boldsymbol{s}$ is normalized such that $\boldsymbol{s} = \boldsymbol{v}$. Also, $\Sigma^{-1}Q = J\Sigma^{-1}$ implies that $\boldsymbol{t}Q = -\alpha\boldsymbol{t}$, so that $\boldsymbol{t} = d\boldsymbol{u}$ for some constant $d \neq 0$. Since the above limit must be nonnegative we actually have $c = d/(m-1)! > 0$. □

We can now conclude the following.

Theorem 5 If $-\alpha$, the eigenvalue of Q with maximal real part, has geometric multiplicity one, and the initial distribution $\boldsymbol{w}$ is such that $S_{a(\alpha)}$ is accessible, then the limits (3) and (4) exist and are given by (11) and (12), respectively, where $\boldsymbol{u}$ is the unique quasi-stationary distribution from which $S_{a(\alpha)}$ is accessible.

Proof Let the algebraic multiplicity of $-\alpha$ be $m \geq 1$ and let $b(\alpha)$ be as in (14). It is no restriction to assume that $S_{b(\alpha)}$ is accessible from $\boldsymbol{w}$, for otherwise we can rephrase the problem in the setting of a smaller state space. Since, by Theorem 4,

$$\lim_{t\to\infty} \frac{e^{\alpha t}}{t^{m-1}} P_{ij}(t) = c v_i u_j, \quad i, j \in S,$$

and $\boldsymbol{u}\boldsymbol{1}^T = 1$, we have

$$\lim_{t\to\infty} \frac{e^{\alpha t}}{t^{m-1}} \sum_{j\in S} P_{ij}(t) = c v_i, \quad i \in S,$$

which implies that

$$\lim_{t\to\infty} \frac{e^{\alpha t}}{t^{m-1}} \sum_{i\in S} w_i \sum_{j\in S} P_{ij}(t) = c \sum_{i\in S} w_i v_i.$$

Since $v_i > 0$ for all states i from which $S_{b(\alpha)}$ is accessible, while $S_{b(\alpha)}$ is accessible from $\boldsymbol{w}$, we must have $w_i v_i > 0$ for at least one $i \in S$. Hence, for all $j \in S$,

$$\lim_{t\to\infty} \mathbb{P}_{\boldsymbol{w}}(X(t) = j \,|\, T > t) = \lim_{t\to\infty} \frac{\sum_{i\in S} w_i P_{ij}(t)}{\sum_{i\in S} w_i \sum_{j\in S} P_{ij}(t)} = u_j,$$

and, for any $s \geq 0$,

$$\lim_{t\to\infty} \mathbb{P}_{\boldsymbol{w}}(T > t + s \,|\, T > t) = \lim_{t\to\infty} \frac{\sum_{i\in S} w_i \sum_{j\in S} P_{ij}(t+s)}{\sum_{i\in S} w_i \sum_{j\in S} P_{ij}(t)} = e^{-\alpha s},$$

as required. □

Remarks *(i)* The fact that the limiting distribution of the residual survival time exists and is exponentially distributed has been observed by Kalpakam [12] and Li and Cao [16] in a more general setting, namely when the Laplace transform of the survival-time distribution is a rational function (cf. [20]).
(ii) We found it elucidating to prove Theorem 5 by means of Theorem 4, which is of independent interest. We shall see in Section 6, however, that a result encompassing Theorem 5 can be established by an appeal to more general results for quasi-stationary and limiting conditional distributions.
(*iii*) The results in [19] and [8] constitute the continuous-time counterparts of results obtained in [18] and [7], respectively, in a discrete-time setting. The latter results have been generalized (in a more abstract, but still discrete, setting) by Lindqvist [17]. A third approach towards proving Theorem 5 (and its generalization) would be to take Lindqvist results (in particular [17, Theorem 5.8]) as a starting point and prove their analogues in a continuous-time setting.

In what follows we are interested in particular in properties of the left eigenvector $\boldsymbol{u}$ that are determined by structural properties of Q. To set the stage we look more closely into the simple multi-class setting of a pure death process in Section 3, and then generalize our results to quasi-death processes in Section 4. But before doing so we address the problem of verifying whether the condition in Theorem 5 is fulfilled.

2.4 When is the geometric multiplicity of $-\alpha$ equal to 1?

It will be useful to have a simple criterion for establishing that $-\alpha$, the eigenvalue of Q with maximal real part, has geometric multiplicity one. To obtain a sufficient condition we can use a result of Cooper's [6, Theorem 3] on nonnegative matrices that was generalized to M-matrices by Richman and Schneider [22, Corollary 5.8] (see also [24, Corollary 8.6]). Applied to the M-matrix $-(Q+\alpha I)$, the result states that if, for each $j \in I(\alpha)$, the set $\{S_k : S_j \prec S_k,\ k \in I(\alpha)\}$ is linearly ordered, that is, $S_i \prec S_j \Longleftrightarrow i \leq j$ for $i, j \in I(\alpha)$, then the dimension of the null space of $Q + \alpha I$, and hence the geometric multiplicity of $-\alpha$, equals the number of minimal elements in $\{S_k,\ k \in I(\alpha)\}$ with respect to the partial order $\prec$. (If, for each $j \in I(\alpha)$, the set $\{S_k : S_k \prec S_j,\ k \in I(\alpha)\}$ happens to be linearly ordered, we can apply Cooper's result to $-(Q^T + \alpha I)$ to find the geometric multiplicity of $-\alpha$.) A simple consequence of this result is that $-\alpha$ has geometric multiplicity one if $\{S_k,\ k \in I(\alpha)\}$ is linearly ordered. The next theorem states that this condition is, in fact, necessary and sufficient.

Theorem 6 The eigenvalue of Q with maximal real part, $-\alpha$, has geometric multiplicity one if and only if $\{S_k,\ k \in I(\alpha)\}$ is linearly ordered.

Proof It remains to prove the necessity, so let the geometric multiplicity of $-\alpha$ be one. Theorems 1 and 4 imply that when $m = \text{card}(I(\alpha))$, the algebraic multiplicity of $-\alpha$,

$$\lim_{t\to\infty} \frac{e^{\alpha t}}{t^{m-1}} P_{ij}(t) > 0 \tag{19}$$

if the states i and j are such that $S_{b(\alpha)}$ is accessible from i and j is accessible from $S_{a(\alpha)}$. On the other hand, it follows from Mandl [19, Theorem 2] that if i and j satisfy these requirements then (19) holds true provided m is the maximum number of classes $S_k,\ k \in I(\alpha)$, that can be traversed on a path from i to j in the directed graph associated with the Markov chain. Since $m = \text{card}(I(\alpha))$ it follows that $\{S_k,\ k \in I(\alpha)\}$ must be linearly ordered. □

Remark Mandl's result referred to above states that $P_{ij}(t)$ behaves asymptotically as $t^{-(m-1)}e^{-\beta t}$, where $-\beta$ is the largest eigenvalue of *any* class that can be visited on a path from i to j, and m is the largest number of classes with eigenvalue $-\beta$ that can be traversed in a path from i to j. Mandl's proof is based on a careful decomposition of $P_{ij}(t)$. Arguments similar to those of Mandl have been used by Buiculescu [4] in the setting of a denumerable state space.

3 Pure death processes

Let us assume that the Markov chain $\mathcal{X} = \{X(t),\ t \geq 0\}$ of the previous section is a pure death process with death rate μ_i in state $i \in S$, so that the matrix Q of (1) is given by

$$Q = \begin{pmatrix} -\mu_1 & 0 & 0 & \cdots & 0 & 0 \\ \mu_2 & -\mu_2 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -\mu_{n-1} & 0 \\ 0 & 0 & 0 & \cdots & \mu_n & -\mu_n \end{pmatrix}. \tag{20}$$

The classes of S now consist of single states, so, maintaining the notation of the previous section, we let $S_k = \{k\}$, and find that $\alpha_k = \mu_k$ and

$$\alpha = \mu := \min_{i \in S} \mu_i.$$

It follows immediately from Theorem 6 that $-\alpha$, the eigenvalue of Q with maximal real part, has geometric multiplicity one, the setting of the previous section. Letting

$$a := \min\{k : \mu_k = \mu\}, \tag{21}$$

it is clear that an initial distribution $\boldsymbol{w}$ satisfies the requirements of Theorem 5 if and only if $\boldsymbol{w}$ has support in the set of states $\{a, a+1, \ldots, n\}$.

Theorem 7 Let $\mathcal{X}$ be a pure death process with death rate μ_i in state $i \in S$, and let a be as in (21). If the initial distribution $\boldsymbol{w}$ is supported by at least one state $i \geq a$, then

$$\lim_{t\to\infty} \mathbb{P}_{\boldsymbol{w}}(T > t + s \,|\, T > t) = e^{-\mu s}, \quad s \geq 0, \tag{22}$$

and

$$\lim_{t\to\infty} \mathbb{P}_{\boldsymbol{w}}(X(t) = j \,|\, T > t) = u_j, \quad j \in S, \tag{23}$$

where $\boldsymbol{u} = (u_1, u_2, \ldots, u_n)$ is the (unique) quasi-stationary distribution of $\mathcal{X}$ from which state a is accessible, and given by

$$u_j = \begin{cases} \dfrac{\mu}{\mu_j} \displaystyle\prod_{i=1}^{j-1} \left(1 - \frac{\mu}{\mu_i}\right), & j \leq a \\ 0, & j > a, \end{cases} \tag{24}$$

where an empty product denotes unity.

Proof By Theorems 3 and 5 we have to show that the vector $\boldsymbol{u}$ satisfies $\boldsymbol{u}Q = -\mu\boldsymbol{u}$ and $\boldsymbol{u}\mathbf{1}^T = 1$. It is a routine exercise to verify these properties. □

The pure death process thus provides us with an example of the phenomenon of a "bottleneck" state (state a above), alluded to by Aalen and Gjessing in [3, Section 8]. We observe in particular that, conditional on survival, the process does not necessarily become concentrated on state 1, the last state to be visited by the process before absorption, as time increases.

Example The quasi-stationary distribution of the death process on $S = \{1, 2\}$ is given by

$$\boldsymbol{u} = (u_1, u_2) = \begin{cases} \left(\dfrac{\mu_2}{\mu_1}, 1 - \dfrac{\mu_2}{\mu_1}\right) & \text{if } \mu_2 < \mu_1 \\ (1, 0) & \text{if } \mu_1 \leq \mu_2. \end{cases} \tag{25}$$

So when $\mu_1 \leq \mu_2$ and whatever the initial distribution, the process will almost surely be in state 1 if, after a long time, absorption has not yet occurred. Note that $(1, 0)$ is also a quasi-stationary distribution if $\mu_2 < \mu_1$, but one from which state 2 is not accessible. Hence it is a limiting conditional distribution only if $\mathbb{P}(X(0) = 2) = 0$. □

As an aside we remark that the survival time in any birth-death process can be represented by the survival time in a pure death process with the same number of states (see, for example, Aalen [1]). Evidently, the quasi-stationary distributions of the two processes will be different in general.

4 Quasi-death processes

The absorbing continuous-time Markov chain $\mathcal{X} := \{X(t),\ t \geq 0\}$ of Section 2 is a *quasi-death process* if $S = \{(\ell, j) \mid \ell = 1, 2, \ldots, L,\ j = 1, 2, \ldots, J_\ell\}$ and Q takes the block-partitioned form

$$Q = \begin{pmatrix} Q_1 & 0 & 0 & \cdots & 0 & 0 \\ M_2 & Q_2 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & Q_{L-1} & 0 \\ 0 & 0 & 0 & \cdots & M_L & Q_L \end{pmatrix}, \tag{26}$$

where Q_ℓ and M_ℓ are nonzero matrices of dimension $J_\ell \times J_\ell$, and $J_\ell \times J_{\ell-1}$, respectively. We write $X(t) = (L(t), J(t))$ and call $L(t)$ the *level* and $J(t)$ the *phase* of the process at time $t < T$. Throughout this section we assume that $S_\ell := \{(\ell, j) \,|\, j = 1, 2, \ldots, J_\ell\}$ is a communicating class for each level ℓ. Moreover, we suppose

$$\mathbf{1}M_\ell^T + \mathbf{1}Q_\ell^T = \mathbf{0}, \quad \ell = 2, 3, \ldots, L, \tag{27}$$

and, to be consistent with (2),

$$\boldsymbol{q}_1 := -\mathbf{1}Q_1^T \geq \mathbf{0},\ \boldsymbol{q}_1 \neq \mathbf{0}. \tag{28}$$

Hence, with probability one and for any initial state (ℓ, i), the function $L(t),\ 0 \leq t < T$, will be a step function with downward jumps of size one, and the process will eventually escape from S, via a state at level 1, to the absorbing state 0. Extending the notation introduced in Section 2 we write

$$\mathbb{P}_{\boldsymbol{w}_\ell}(\cdot) := \sum_{i=1}^{J_\ell} w_{\ell i}\mathbb{P}_{(\ell,i)}(\cdot)$$

for any distribution $\boldsymbol{w}_\ell = (w_{\ell 1}, w_{\ell 2}, \ldots, w_{\ell J_\ell})$ over S_ℓ.

Evidently, if $J_\ell = 1$ for all levels ℓ then we are in the setting of the simple death process of the previous section with death rate $\mu_1 := \boldsymbol{q}_1$ in state 1 and $\mu_\ell := M_\ell$ in state $\ell > 1$. On the other hand, if the initial distribution concentrates all mass on the first level, we are basically dealing with a Markov chain taking values in the set $\{0\} \cup S_1$, with 0 an absorbing state and S_1 a single communicating class, a setting discussed in Section 2.2. Since $S_1 \prec S_2 \prec \ldots \prec S_L$, Theorem 6 implies that in the more general setting at hand the eigenvalue of Q with maximal real part still has geometric multiplicity one. So we can obtain the limits (4) and (3) by applying the Theorems 3 and 5. However, we can reduce the amount of computation by exploiting the structure of Q, as we shall show next.

We denote the (unique) eigenvalue of Q_ℓ with maximal real part by $-\alpha_\ell$, and the associated left and right eigenvectors by $\boldsymbol{x}_\ell = (x_{\ell 1}, x_{\ell 2}, \ldots, x_{\ell J_\ell})$ and $\boldsymbol{y}_\ell = (y_{\ell 1}, y_{\ell 2}, \ldots, y_{\ell J_\ell})$, respectively. As noted before, α_ℓ is real and positive, and $\boldsymbol{x}_\ell$ and $\boldsymbol{y}_\ell$ can be chosen strictly positive componentwise and such that

$$\boldsymbol{x}_\ell \mathbf{1}^T = 1 \quad \text{and} \quad \boldsymbol{x}_\ell \boldsymbol{y}_\ell^T = 1.$$

In analogy to (6) we have

$$\lim_{t\to\infty} e^{\alpha_\ell t} P_{(\ell,i),(\ell,j)}(t) = y_{\ell i} x_{\ell j}, \quad i,j = 1,2,\ldots,J_\ell,$$

for each level ℓ, so we will refer to α_ℓ as the *decay parameter of* $\mathcal{X}$ *in* S_ℓ. Moreover, the vector $\boldsymbol{x}_\ell$ can be interpreted as the *quasi-stationary distribution of* $\mathcal{X}$ *in* S_ℓ, in the sense that

$$\mathbb{P}_{\boldsymbol{x}_\ell}(X(t) = (\ell,j) \,|\, T_\ell > t) = x_{\ell j}, \quad t \geq 0, \; j = 1,2,\ldots,J_\ell,$$

where T_ℓ denotes the sojourn time of $\mathcal{X}$ in S_ℓ, while

$$\mathbb{P}_{\boldsymbol{x}_\ell}(T_\ell > t) = e^{-\alpha_\ell t}, \quad t \geq 0.$$

If the initial distribution concentrates all mass in S_ℓ (and is represented by the vector $\boldsymbol{w}_\ell = (w_{\ell 1}, w_{\ell 2}, \ldots, w_{\ell J_\ell})$, say) but is otherwise arbitrary, then, by the results of Darroch and Seneta [8] mentioned in Section 2,

$$\lim_{t\to\infty} \mathbb{P}_{\boldsymbol{w}_\ell}(X(t) = (\ell,j) \,|\, T_\ell > t) = x_{\ell j}, \quad j = 1,2,\ldots,J_\ell,$$

and

$$\lim_{t\to\infty} \mathbb{P}_{\boldsymbol{w}_\ell}(T_\ell > t+s \,|\, T_\ell > t) = e^{-\alpha_\ell s}, \quad s \geq 0.$$

We now turn to a general initial distribution $\boldsymbol{w} = (\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_L)$, where $\boldsymbol{w}_\ell = (w_{\ell 1}, w_{\ell 2}, \ldots, w_{\ell J_\ell})$ for $\ell = 1,2,\ldots,L$. We let $\alpha = \min_\ell \alpha_\ell$, and recall that $-\alpha$, the eigenvalue of Q with maximal real part, has geometric multiplicity one. Theorem 5 then tells us that the limiting distribution of the residual survival time in $S = \cup_\ell S_\ell$ is exponentially distributed with parameter α. Regarding the limiting distribution of $X(t)$ conditional on survival in S up to time t, we can state the following generalization of Theorem 7.

Theorem 8 Let $\mathcal{X}$ be a quasi-death process for which Q takes the form (26), and satisfies (27) and (28), and let $-\alpha$ be the eigenvalue of Q with maximal real part (which then has geometric multiplicity one). If the initial distribution $\boldsymbol{w}$

is supported by at least one state in the set $\cup_{\ell \geq a} S_\ell$, where $a = \min\{\ell : \alpha_\ell = \alpha\}$, then

$$\lim_{t\to\infty} \mathbb{P}_{\boldsymbol{w}}(X(t) = (\ell, j) \,|\, T > t) = u_{\ell j}, \;\; j = 1, 2, \ldots, J_\ell, \;\; \ell = 1, 2, \ldots, L, \quad (29)$$

where $\boldsymbol{u}_\ell := (u_{\ell 1}, u_{\ell 2}, \ldots, u_{\ell J_\ell})$ satisfies $\boldsymbol{u}_\ell = \boldsymbol{0}$ if $\ell > a$, and $\boldsymbol{u}_\ell = c\boldsymbol{x}_\ell$, with $\boldsymbol{x}_a$ the (unique and strictly positive) solution of

$$\boldsymbol{x}_a Q_a = -\alpha \boldsymbol{x}_a, \quad \boldsymbol{x}_a \mathbf{1}^T = 1; \qquad (30)$$

for $\ell < a$, $\boldsymbol{u}_\ell$ is defined recursively by

$$\boldsymbol{u}_\ell = -\boldsymbol{u}_{\ell+1} M_{\ell+1} (Q_\ell + \alpha I)^{-1}. \qquad (31)$$

Here I is an identity matrix of appropriate dimensions and $c > 0$ is such that $\boldsymbol{u}\mathbf{1}^T = 1$, where $\boldsymbol{u} := (\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_L)$.

Proof Since, for all $\ell < a$, the eigenvalue with maximal real part of the matrix $Q_\ell + \alpha I$ is given by $-(\alpha_\ell - \alpha) < 0$, it follows from [25, Theorem 2.6(g)] that $-(Q_\ell + \alpha I)^{-1}$ exists and has strictly positive components. So, by induction, $\boldsymbol{u}_\ell$ is positive componentwise for $\ell \leq a$. It follows easily that the vector $\boldsymbol{u}$ satisfies the requirements of Theorem 3. □

In the next section we will apply this theorem to a specific example of a quasi-death process.

5 Example: a migration process

The setup is as follows. We have an ensemble of L particles that move independently of one another in a finite set $\mathcal{M} := \{1, 2, \ldots, m\}$, say, before eventually reaching an absorbing state 0 (outside $\mathcal{M}$). Suppose that at time $t = 0$ the particles are assigned to the states according to some rule, and then each moves in continuous time according to an *irreducible* Markov chain with (necessarily non-conservative) q-matrix of transition rates $Q_{\mathcal{M}} = (q_{ij}, \; i, j \in \mathcal{M})$. The transition rates into the absorbing state are $q_{i0}, \; i \in \mathcal{M}$. We record the *number* of particles in the various states (rather than, say, their positions). Let $N_j(t)$ be the number of particles in state j at time t, and let $\boldsymbol{N} = (N_j,\, j \in \mathcal{M})$. The process $(\boldsymbol{N}(t),\, t \geq 0)$ is also a continuous-time Markov chain. It is an example of a *migration process* (Whittle [28]) or, in queueing-theory parlance, a *network of infinite-server queues*. Since we are assuming that particles move independently, the ensemble process can also be viewed as a (non-interacting) particle system, and thus dates back to at least Doob [10]. The ensemble process takes values in $\tilde{S} = \{\mathbf{0}\} \cup S$, where $\mathbf{0} = (0, 0, \ldots, 0)$ and

$$S = \{\boldsymbol{n} = (n_1, n_2, \ldots, n_m) \in \{0, 1, \ldots\}^m : \textstyle\sum_{j \in \mathcal{M}} n_j = L\},$$

and its q-matrix of transition rates $Q = (q(\boldsymbol{n}, \boldsymbol{m}),\ \boldsymbol{n}, \boldsymbol{m} \in S)$ is given by $q(\boldsymbol{n}, \boldsymbol{n} + \boldsymbol{e}_j - \boldsymbol{e}_i) = n_i q_{ij}$ for all states $j \neq i$ in $\mathcal{M}$, where $\boldsymbol{e}_j$ is the unit vector with a 1 as its j-th entry, and $q(\boldsymbol{n}, \boldsymbol{n} - \boldsymbol{e}_i) = n_i q_{i0}$, for all states i in $\mathcal{M}$. Notice that the total rate out of state $\boldsymbol{n} \in S$ is

$$q(\boldsymbol{n}) := \sum_{\boldsymbol{m} \in \tilde{S} \colon \boldsymbol{m} \neq \boldsymbol{n}} q(\boldsymbol{n}, \boldsymbol{m}) = \sum_{i \in \mathcal{M}} n_i q_i,$$

where $q_i := q_{i0} + \sum_{j \neq i} q_{ij}$. The transition rate into (the absorbing state) $\mathbf{0}$ is $q(\boldsymbol{e}_i, \mathbf{0}) = n_i q_{i0},\ i \in \mathcal{M}$.

Since each of the L particles reaches 0 with probability 1 in finite mean time, so too does the ensemble process reach its absorbing state $\mathbf{0}$ in finite mean time. However, for the ensemble process S is *not* irreducible. The process has irreducible classes

$$S_k = \{\boldsymbol{n} \in \{0, 1, \ldots\}^m : \textstyle\sum_{j \in \mathcal{M}} n_j = k\}, \quad k = 0, 1, \ldots, L,$$

corresponding to there being k particles in $\mathcal{M}$, with S_0 having the single member $\mathbf{0}$, and the process moving from S_k to S_{k-1} when one of the k particles that remain in $\mathcal{M}$ reaches 0. The classes are therefore arranged as follows: $S_0 \prec S_1 \prec \cdots \prec S_{L-1} \prec S_L$. Indeed, the ensemble process is an example of a quasi-death process.

In the next theorem we establish that the limiting condition distribution of the ensemble process assigns positive probability only to the states in S_1, being precisely the unit vectors $\boldsymbol{e}_j,\ j \in \mathcal{M}$, corresponding to the single remaining individual being in state j (and hence the $L - 1$ others in state 0).

Theorem 9 The ensemble process admits a limiting conditional distribution $\boldsymbol{u} = (u(\boldsymbol{m}),\ \boldsymbol{m} \in S)$, which does not depend on the initial distribution over states. It assigns all its mass to S_1, with $u(\boldsymbol{e}_j) = \pi_j,\ j \in \mathcal{M}$, where $(\pi_j,\ j \in \mathcal{M})$ is the limiting conditional distribution associated with $Q_{\mathcal{M}}$.

Proof Let Q_k be the restriction to S_k of the transition rate matrix Q of the ensemble process and let $-\alpha_k$ be the eigenvalue of Q_k with maximum real part $(k = 1, 2, \ldots, L)$. Then, $\alpha_k = k\alpha$. To see this, observe that $\alpha_k = \lim_{t\to\infty} -(1/t)\log \Pr(T > t)$, where T is the time to first exit of the process from S_k (see Kingman [14]); the limit does not depend on the initial distribution over states. However, $T = \min\{T_1, T_2, \ldots, T_k\}$, where T_i is the time it takes individual i to reach 0, and, since the particles move independently, $\Pr(T > t) = \prod_{i=1}^{k} \Pr(T_i > t)$. Since each particle moves according to $Q_{\mathcal{M}}$, we have $-(1/t)\log \Pr(T_i > t) \to \alpha$ as $t \to \infty$. Hence, $\alpha_k = k\alpha$. It follows immediately that $-\alpha$ is the eigenvalue of Q with maximum real part, and, moreover, that its algebraic, and hence geometric, multiplicity is equal to 1. We may therefore appeal to Theorem 8, which implies that the limiting conditional distribution of the ensemble process exists provided the initial distribution assigns mass to at least one of $S_1, S_2, \ldots, S_L$. But we have assumed that all L particles are present initially, and so all this mass is assigned to S_L. Furthermore, the limiting conditional distribution is given by $\boldsymbol{u} = (\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_L)$, where $\boldsymbol{u}_j = \boldsymbol{0}$ for $j > a = 1$, and $\boldsymbol{u}_1$ is the unique (positive) solution to $\boldsymbol{u}_1 Q_1 = -\alpha \boldsymbol{u}_1$ and $\boldsymbol{u}_1 \boldsymbol{1}^T = 1$. Since S_1 consists of the unit vectors $\boldsymbol{e}_i,\ i \in \mathcal{M}$, we have $\boldsymbol{u}_1 = (u(\boldsymbol{e}_1), u(\boldsymbol{e}_2), \ldots, u(\boldsymbol{e}_n))$. Subsequently writing out $\boldsymbol{u}_1 Q_1 = -\alpha \boldsymbol{u}_1$, we get

$$\sum_{i\in\mathcal{M},\, i\neq j} u(\boldsymbol{e}_i) q_{ij} = (q_j - \alpha)\, u(\boldsymbol{e}_j), \quad j \in \mathcal{M},$$

so that we must have $u(\boldsymbol{e}_j) = \pi_j,\ j \in \mathcal{M}$, where $\boldsymbol{\pi} = (\pi_j,\ j \in \mathcal{M})$ is the (unique and strictly positive) solution to $\boldsymbol{\pi} Q_{\mathcal{M}} = -\alpha \boldsymbol{\pi}$ with $\boldsymbol{\pi} \boldsymbol{1}^T = 1$, that is, the limiting conditional distribution associated with $Q_{\mathcal{M}}$. □

6 A further generalization

We finally return to the setting of Subsection 2.3, namely that of a Markov chain $\mathcal{X}$ with a general state space S consisting of communicating classes $S_1, S_2, \ldots, S_L$. In addition to the notation and terminology introduced previously, we let $g \geq 1$ be the geometric multiplicity of $-\alpha$, the eigenvalue of Q with maximal real part, so that there are g linearly independent vectors $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_g$ satisfying

$$\boldsymbol{u}_j Q = -\alpha \boldsymbol{u}_j, \quad 1 \leq j \leq g. \tag{32}$$

Class S_k will be called a *minimal class for* α if it is a minimal element in the set $\{S_j,\ j \in I(\alpha)\}$ with respect to the partial order $\prec$, that is, for all $j \neq k$,

$$S_j \prec S_k \implies j \notin I(\alpha).$$

Letting $m(\alpha)$ be the number of minimal classes for α, we have $m(\alpha) \geq 1$, since $S_{a(\alpha)}$ is always a minimal class for α. On the other hand, we shall see shortly that $m(\alpha) \leq g$. Our purpose in this section is to show that the condition $g = 1$ in Theorems 3 and 5 may be replaced by the weaker condition $m(\alpha) = 1$.

It is known (see, for example, [21, Theorem 1]) that a quasi-stationary distribution $\boldsymbol{u}$ must satisfy $\boldsymbol{u}Q = s\boldsymbol{u}$ for some $s < 0$. An argument similar to the proof of Lemma 2 subsequently implies that a quasi-stationary distribution $\boldsymbol{u}$ from which $S_{a(\alpha)}$ is accessible must satisfy $\boldsymbol{u}Q = -\alpha\boldsymbol{u}$. We can obtain a solution to $\boldsymbol{u}Q = -\alpha\boldsymbol{u}$ by solving the system in the restricted setting of states that are accessible from a single minimal class for α and putting $u_j = 0$ whenever j is one of the remaining states. Since, by Theorem 1, this solution has $u_j > 0$ if and only if j is accessible from the minimal class, there are at least $m(\alpha)$ linearly independent, *nonnegative* vectors $\boldsymbol{u}$ satisfying $\boldsymbol{u}Q = -\alpha\boldsymbol{u}$. (This statement is in fact implied by Schneider's [23, Theorem 2] on M-matrices, referred to earlier in connection with Theorem 1.) So, as announced, $g \geq m(\alpha)$, and we may assume that $\boldsymbol{u}_j \geq \boldsymbol{0}$ and $\boldsymbol{u}_j\boldsymbol{1}^T = 1$ for $j = 1, 2, \ldots, m(\alpha)$. Moreover, it follows from a result of Carlson's [5, Theorem 2] on M-matrices (see also [24, Theorem 3.1]), that any vector $\boldsymbol{u}$ representing a probability distribution and satisfying

$\boldsymbol{u}Q = -\alpha\boldsymbol{u}$, must be a linear combination (with nonnegative coefficients) of $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_{m(\alpha)}$. These observations lead us to the following generalization of Theorem 3.

Theorem 10 Let $-\alpha$ be the eigenvalue of Q with maximal real part. Then $\mathcal{X}$ has a unique quasi-stationary distribution $\boldsymbol{u}$ from which $S_{a(\alpha)}$ is accessible if and only if $m(\alpha) = 1$, that is, $S_{a(\alpha)}$ is the only minimal class for α, in which case $\boldsymbol{u}$ is the (unique) nonnegative solution to the system $\boldsymbol{u}Q = -\alpha\boldsymbol{u}$ and $\boldsymbol{u}\mathbf{1}^T = 1$, and has a positive jth component if and only if state j is accessible from $S_{a(\alpha)}$.

From the proof of Theorem 4, following an argument similar to that used in Section 3 of [8], it is not difficult to see that for any initial distribution $\boldsymbol{w}$ over S the limits (4) exist and constitute a proper distribution $\boldsymbol{u}$, say. Moreover, by [27, Theorem 2], such a limiting conditional distribution must be a quasi-stationary distribution. Hence, assuming that the initial distribution $\boldsymbol{w}$ is such that $S_a(\alpha)$ is accessible, we can repeat the first part of the argument preceding Theorem 10 and conclude that $\boldsymbol{u}Q = -\alpha\boldsymbol{u}$. So, in view of the preceding theorem, we can now state the generalization of Theorem 5 that was announced in Section 2.

Theorem 11 Let $-\alpha$ be the eigenvalue of Q with maximal real part. If $m(\alpha) = 1$, that is, $S_{a(\alpha)}$ is the only minimal class for α, and the initial distribution $\boldsymbol{w}$ is such that $S_{a(\alpha)}$ is accessible, then the limits (3) and (4) exist and are given by (11) and (12), respectively, where $\boldsymbol{u}$ is the unique quasi-stationary distribution of $\mathcal{X}$ from which $S_{a(\alpha)}$ is accessible.

Acknowledgement

Part of this work was done during a period when Phil Pollett held a visiting fellowship at Grey College, University of Durham. The hospitality of Grey College and the Department of Mathematical Sciences is acknowledged with thanks. The work of Phil Pollett is supported by the Australian Research Council Centre of Excellence for Mathematics and Statistics of Complex Systems.

References

[1] Aalen, O.O. (1995) Phase type distributions in survival analysis. *Scand. J. Statist.* **22**, 447-463.

[2] Aalen, O.O. and Gjessing, H.K. (2001) Understanding the shape of the hazard rate: a process point of view. *Statist. Sci.* **16**, 1-22.

[3] Aalen, O.O. and Gjessing, H.K. (2003) A look behind survival data: underlying processes and quasi-stationarity. In: Lindqvist, B.H. and Doksum, K.A. (Eds.), *Mathematical and Statistical Methods in Reliability*. World Scientific Publishing, Singapore, pp. 221-234.

[4] Buiculescu, M. (1972) Quasi-stationary distributions for continuous-time Markov processes with a denumerable set of states. *Rev. Roumaine Math. Pures Appl.* **17**, 1013-1023.

[5] Carlson, D. (1963) A note on M-matrix equations. *J. Soc. Indust. Appl. Math.* **11**, 1027-1033.

[6] Cooper, C.D.H. (1973) On the maximum eigenvalue of a reducible nonnegative real matrix. *Math. Z.* **131**, 213-217.

[7] Darroch, J.N. and Seneta, E. (1965) On quasi-stationary distributions in absorbing discrete-time finite Markov chains. *J. Appl. Probab.* **2**, 88-100.

[8] Darroch, J.N. and Seneta, E. (1967) On quasi-stationary distributions in absorbing continuous-time finite Markov chains. *J. Appl. Probab.* **4**, 192-196.

[9] Debreu, G. and Herstein, I.N. (1953) Nonnegative square matrices. *Econometrica* **21**, 597-607.

[10] Doob, J.L. *Stochastic Processes.* John Wiley & Sons Inc., New Yorks, 1953.

[11] Friedberg, S.H., Insel, A.J. and Spence, L.E. *Linear Algebra,* 3rd ed., Prentice Hall Inc., Upper Saddle River, N.J., 1997.

[12] Kapalkam, S. (1993) On the quasi-stationary distribution of the residual lifetime. *IEEE Trans. Reliab.* **42**, 623-624.

[13] Kijima, M. *Markov Processes for Stochastic Modeling*. Chapman & Hall, London, 1997.

[14] Kingman, J.F.C. (1963) The exponential decay of Markov transition probabilities. *Proc. London Math. Soc.* **13**, 337–358.

[15] Latouche, G. and Ramaswami, V. *Introduction to Matrix Analytic Methods in Stochastic Modeling*. ASA-SIAM, Philadelphia, 1999.

[16] Li, W. and Cao, J. (1993) The limiting distributions of the residual lifetimes of a Markov repairable system. *Reliab. Eng. System Safety* **41**, 103-105.

[17] Lindqvist, B.H. (1989) Asymptotic properties of powers of nonnegative matrices, with applications. *Linear Algebra Appl.* **114/115**, 555-588.

[18] Mandl, P. (1959) On the asymptotic behaviour of probabilities within groups of states of a homogeneous Markov chain. *Časopis Pěst. Mat.* **84,** 140-149 (in Russian).

[19] Mandl, P. (1960) On the asymptotic behaviour of probabilities within groups of states of a homogeneous Markov process. *Časopis Pěst. Mat.* **85,** 448-455 (in Russian).

[20] O'Cinneide, C.A. (1990) Characterization of phase type distributions. *Stoch. Models* **6**, 1-57.

[21] Pollett, P.K. and Vere-Jones, D. (1992) A note on evanescent processes. *Austral. J. Statist.* **34**, 531-536.

[22] Richman, D.J. and Schneider, H. (1978) On the singular graph and the Weyr characteristic of an M-matrix. *Aequationes Math.* **17**, 208-234.

[23] Schneider, H. (1956) The elementary divisors associated with 0 of a singular M-matrix. *Proc. Edinburgh Math. Soc.* **10**, 108-122.

[24] Schneider, H. (1986) The influence of the marked reduced graph of a nonnegative matrix on the Jordan form and on related properties: a survey. *Linear Algebra Appl.* **84**, 161-189.

[25] Seneta, E. *Non-negative Matrices and Markov Chains,* rev. ed., Springer, New York, 1981.

[26] Steinsaltz, D. and Evans, S.N. (2004) Markov mortality models: implications of quasistationarity and varying initial distributions. *Theoret. Population Biol.* **65**, 319-337.

[27] Vere-Jones, D. (1969) Some limit theorems for evanescent processes. *Austral. J. Statist.* **11**, 67-78.

[28] Whittle, P. (1967) Nonlinear migration processes. *Bull. Int. Statist. Inst.* **42**, 642-647.