ANALYTICAL AND COMPUTATIONAL METHODS FOR MODELLING THE LONG–TERM BEHAVIOUR OF EVANESCENT RANDOM PROCESSES

P.K. Pollett

Department of Mathematics
The University of Queensland

Abstract. There are a variety of stochastic systems arising in areas as diverse as wildlife management, chemical kinetics and reliability theory, which eventually "die out", yet appear to be stationary over any reasonable time scale. The notion of a *quasistationary distribution* has proved to be a potent tool in modelling this behaviour. In finite-state systems the existence of a quasistationary distribution is guaranteed. However, in the infinite-state case this may not always be so, and the question of whether or not quasistationary distributions exist requires delicate mathematical analysis. The purpose of this paper is twofold: to present simple conditions for the existence of quasistationary distributions for continuous-time Markov chains, and, to describe an efficient computational procedure for evaluating them. The computational method I shall describe is a variant of Arnoldi's algorithm and it is particularly suited to problems where the transition-rate matrix is both large and sparse, but does not exhibit a banded structure which might otherwise be usefully exploited. The analytical and computational methods will be illustrated with reference to a variety of examples, including birth-death processes, the birth-death and catastrophe process, and an epidemic model for which I shall compare the computed quasistationary distribution with an appropriate diffusion approximation.

1. Introduction

Quasistationary distributions have been used in a variety of diverse contexts for modelling the long-term behaviour of stochastic systems which, in some sense, terminate, but appear to be stationary over any reasonable time scale. For example, in the context of modelling chemical reaction kinetics, there are a number of reactions in which one or more species can become depleted, yet these reactions settle down quickly to a stable equilibrium; quasistationary distributions have been used here to model the concentration of the catalyst in reactions in which the catalyst can become exhausted (see, for example, Oppenheim et al. (1977), Turner and Malek-Mansour (1978), Dambrine and Moreau (1981a, 1981b), Parsons and Pollett (1987) and Pollett (1988b)). In the context of reliability theory, one might wish to determine the distribution of the residual lifetime of a system at

1980 *Mathematics Subject Classification* (1985 *Revision*). 60J27; 65F15.

Key words and phrases. Stochastic modelling; Markov chains; quasistationary distributions; Arnoldi Algorithm.

This work was supported by the Australian Research Council .

some arbitrary time t, conditional on the system being functional (see, for example, Kalpakam and Shahul Hameed (1983), Pijnenburg and Ravichandran (1990) and Pijnenburg et al. (1990)); in the case of a two-unit warm-standby redundant system, the limiting form of this conditional distribution, as t becomes large, is always exponential, no matter what the distribution of lifetimes and repair times (Kalpakam and Shahul Hameed (1983)). Yet another example of the use of quasistationary distributions is in the area of wildlife management, where these have proved to be a potent tool in predicting persistence times, and the distribution of the number of individuals, in animal populations which are subject to large-scale mortality or emigration; in spite of the fact that the usual stochastic models predict eventual extinction, these populations can be surprisingly resilient (see, for example, Scheffer (1951), Mech (1966), Klein (1968), Holling (1973), Pakes (1987), Pollett (1987), and Pakes and Pollett (1989)).

The idea of a quasistationary distribution can be traced back to the work of the Russian Mathematician A.M. Yaglom, who showed that the limiting conditional distribution of the number in the n^{th} generation of the Galton-Watson branching process always exists in the subcritical case (see Yaglom (1947)). But, it was not until the early sixties, and largely stimulated by the remarkable work of Vere-Jones (1962), and later Kingman (1963), that a general theory was annunciated. Since then, there have been a number of significant advances on questions concerned with the existence of quasistationary distributions; in the present context of continuous-time Markov chains, see, for example, Darroch and Seneta (1967), Good (1968), Vere-Jones (1969), Flaspohler (1974), Tweedie (1974), Cavender (1978), van Doorn (1991), Kijima and Seneta (1991), Kijima (1992) and van Doorn and Kijima (1992) (a spectacularly clear account of much of this work is also given in the recent text by Anderson (1991)).

In this paper I shall give a unified account of the theory of quasistationary distributions for continuous-time Markov chains. In Section 2 I shall establish simple conditions for the existence of quasistationary distributions and these will be illustrated, in Section 3, with reference to finite-state systems, birth and death processes, and the birth, death and catastrophe process. Section 4 is devoted to the problem of how one should evaluate quasistationary distributions. I shall describe a computational method, developed previously by David Stewart and I (Pollett and Stewart (1994)), which is particularly suited to problems where the transition-rate matrix is both large and sparse. I shall illustrate the method with reference to an epidemic model, for which there is no analytical expression for the quasistationary distribution, and, for which standard widely available numerical methods fail owing to the size and the structure of the problem; the computed quasistationary distribution will be compared with an appropriate diffusion approximation.

2. The existence of quasistationary distributions

We shall suppose that the system in question can be modelled as a time-homogeneous Markov chain, $(X(t),\ t \geq 0)$, taking values in a discrete set S. Let $Q = (q_{ij},\ i,j \in S)$ be the q-matrix of the chain (assumed to be stable and conservative), so that $q_{ij}\,(\geq 0)$, for $j \neq i$, represents the transition rate from state i

to state j and $q_{ii} = -q_i$, where $q_i = \sum_{j \neq i} q_{ij} \, (< \infty)$ represents the transition rate out of state i. In addition, we shall suppose that Q is regular, so that $X(\cdot)$ is the unique chain with these rates. Checking for regularity should be, though apparently seldom is, a part of the routine of modelling. Simple sufficient conditions for the regularity of Q are contained in Pollett and Taylor (1993). The condition $\sup_j q_j < \infty$, which is predominant in, say, the engineering literature, is certainly too strong for practical purposes; for example, it rules out branching and catastrophe processes and random delay systems.

We shall be concerned with evanescent chains, so, for simplicity, let us take 0 to be the sole absorbing state, that is, $q_0 = 0$, and suppose that $S = \{0\} \cup C$, where $C = \{1, 2, \dots\}$ is an irreducible transient class. In order that there be a positive probability of reaching 0, given that the chain starts in C, we shall suppose that $q_{i0} > 0$ for at least one $i \in C$.

The definition of a quasistationary distribution, which I shall use here, is the one introduced by van Doorn (1991). Let $P(\cdot) = (p_{ij}(\cdot),\ i, j \in S)$ be the transition function of the chain, so that $p_{ij}(t) = \Pr(X(t) = j | X(0) = i)$, for $t \geq 0$.

Definition. *Let $m = (m_j,\ j \in C)$ be a probability distribution over C and define $p(\cdot) = (p_j(\cdot),\ j \in S)$ by*

$$p_j(t) = \sum_{i \in C} m_i p_{ij}(t), \qquad j \in S,\ t \geq 0.$$

Then, m is a quasistationary distribution if, for all $t > 0$ and $j \in C$,

$$\frac{p_j(t)}{\sum_{i \in C} p_i(t)} = m_j.$$

That is, if the chain has m as its initial distribution, then m is a quasistationary distribution if the state probabilities at time t, conditional on the chain being in C at t, are the same for all t.

The relationship between quasistationary distributions and the transition probabilities of the chain is made more precise in the following proposition:

Proposition 1 (Nair and Pollett (1992)). *Let $m = (m_j,\ j \in C)$ be a probability distribution over C. Then, m is a quasistationary distribution if and only if, for some $\mu > 0$, m is μ-invariant on C for P, that is*

$$\sum_{i \in C} m_i p_{ij}(t) = e^{-\mu t} m_j, \qquad j \in C,\ t \geq 0. \tag{1}$$

Thus, in a way which mirrors the theory of *stationary distributions*, one can interpret quasistationary distributions as eigenvectors of the transition function. However, the transition function is available explicitly in only a few simple cases, and so one requires a means of determining quasistationary distributions directly from transition rates of the chain. Since q_{ij} is the right-hand derivative of $p_{ij}(\cdot)$ near 0, an obvious first step is to rewrite (1) as

$$\sum_{i \in C:\, i \neq j} m_i p_{ij}(t) = \left((1 - p_{jj}(t)) - (1 - e^{-\mu t}) \right) m_j, \qquad j \in C,\ t \geq 0,$$

and then divide by t and let $t \downarrow 0$. Proceeding formally, we get

$$\sum_{i \in C:\, i \neq j} m_i q_{ij} = (q_j - \mu) m_j, \qquad j \in C, \tag{2}$$

or, equivalently,

$$\sum_{i \in C} m_i q_{ij} = -\mu m_j, \qquad j \in C. \tag{3}$$

Accordingly, we shall say that m is *μ-invariant on C for Q* whenever (3) holds. The above argument can be justified rigorously (see Proposition 2 of Tweedie (1974)), and so, in view of Proposition 1, we have the following result:

Proposition 2. *If m is a quasistationary distribution then, for some $\mu > 0$, m is μ-invariant on C for Q.*

The more interesting question of when a positive solution, m, to (3) is also a solution to (1) was answered in Pollett (1986, 1988a). However, the necessary and sufficient conditions obtained are usually difficult to verify in practice. If one takes into account the fact that, for m to be a quasistationary distribution, one requires $\sum_{j \in C} m_j = 1$, then much simpler conditions obtain, as the next result demonstrates. It can be deduced from Theorems 3.2, 3.4 and 4.1 of Nair and Pollett (1992); the assumption made here, that Q be regular, facilitates the simpler and more direct proof given below.

Proposition 3. *Let $m = (m_j,\ j \in C)$ be a probability distribution over C and suppose that m is μ-invariant on C for Q. Then,*

$$\mu \leq \sum_{j \in C} m_j q_{j0}, \tag{4}$$

with equality if and only if m is a quasistationary distribution.

Proof. First observe that, since m is a probability distribution, there exists a $j \in C$ such that $m_j > 0$. Hence, because m is μ-invariant on C for Q and C is irreducible, we have, from (2), that $\mu \leq \inf_{j \in C} q_j$ and that $m_j > 0$ for *all* $j \in C$.

Define $Q^* = (q^*_{ij},\ i, j \in C)$ by $q^*_{ij} = m_j q_{ji}/m_i$, for $j \neq i$, and $q^*_{ii} = -q^*_i$, where $q^*_i = q_i - \mu$. Clearly $q^*_{ij} \geq 0$ for all $j \neq i$ and $0 \leq q^*_i < \infty$. And, since m is μ-invariant on C for Q, we have that $\sum_{j \in C} q^*_{ij} = 0$. Thus, Q^* is a stable and conservative q-matrix over C (Q^* is called the μ-reverse of Q with respect to m). Let $P^*(\cdot) = (p^*_{ij}(\cdot),\ i, j \in C)$ be the transition function of the minimal process associated with Q^*. Then, by Lemma 3.3 of Pollett (1988a), we have that

$$m_i p_{ij}(t) = e^{-\mu t} m_j p^*_{ji}(t), \qquad i, j \in C. \tag{5}$$

Summing this equation over $j \in C$, and remembering that, because Q is regular, $\sum_{j \in S} p_{ij}(t) = 1$, we get

$$m_i (1 - p_{i0}(t)) = e^{-\mu t} \sum_{j \in C} m_j p^*_{ji}(t), \qquad i \in C.$$

On summing *this* equation over i, and using Fubini's theorem, we find that

$$\sum_{i\in C} m_i p_{i0}(t) = 1 - e^{-\mu t} \sum_{j\in C} m_j \sum_{i\in C} p^*_{ji}(t),$$

or, equivalently,

$$\sum_{i\in C} m_i p_{i0}(t) = 1 - e^{-\mu t} + e^{-\mu t} \sum_{j\in C} m_j d^*_j(t), \tag{6}$$

where $d^*_i(t) = 1 - \sum_{j\in C} p^*_{ij}(t)$. Notice that $d^*_i(t) \geq 0$, since, because P^* is a transition function, we have that $\sum_{j\in C} p^*_{ij}(t) \leq 1$, for all $i \in C$ and $t \geq 0$.

Now, since P satisfies the forward differential equations, we have, in particular, that

$$p'_{i0}(t) = \sum_{j\in C} p_{ij}(t) q_{j0}, \qquad i \in C,\ t > 0,$$

or, equivalently,

$$p_{i0}(t) = \sum_{j\in C} \int_0^t p_{ij}(s) q_{j0}\, ds, \qquad i \in C,\ t > 0.$$

Multiplying by m_i and summing over $i \in C$, and then using (5) once again, we get

$$\begin{aligned}
\sum_{i\in C} m_i p_{i0}(t) &= \int_0^t \sum_{j\in C} \sum_{i\in C} m_i p_{ij}(s) q_{j0}\, ds \\
&= \int_0^t e^{-\mu s} \sum_{j\in C} m_j \sum_{i\in C} p^*_{ji}(s) q_{j0}\, ds \\
&= \int_0^t e^{-\mu s} \sum_{j\in C} m_j (1 - d^*_j(s)) q_{j0}\, ds \\
&= \int_0^t e^{-\mu s} \left(\sum_{j\in C} m_j q_{j0} - \sum_{j\in C} m_j d^*_j(s) q_{j0} \right) ds \\
&= \frac{1}{\mu}(1 - e^{-\mu t}) \sum_{j\in C} m_j q_{j0} - \int_0^t e^{-\mu s} \sum_{j\in C} m_j d^*_j(s) q_{j0}\, ds.
\end{aligned}$$

On combining this equation with (6) we find that

$$\begin{aligned}
\mu e^{-\mu t} \sum_{j\in C} m_j d^*_j(t) + \int_0^t \mu e^{-\mu s} \sum_{j\in C} m_j d^*_j(s) q_{j0}\, ds& \\
= (1 - e^{-\mu t}) \left(\sum_{j\in C} m_j q_{j0} - \mu \right).&
\end{aligned}$$

The left-hand side of this equation is always non-negative, and so we deduce that (4) must hold. It is also clear that d_i^* is identically 0 for each i if and only if $\mu = \sum_{j \in C} m_j q_{j0}$. But, from (5), we have that

$$\sum_{i \in C} m_i p_{ij}(t) = e^{-\mu t} m_j (1 - d_j^*(t)), \qquad j \in C,$$

and so a necessary and sufficient condition for m to be μ-invariant on C for P is that $d_i^*(t) = 0$, for all $i \in C$ and $t \geq 0$. Thus, in view of Proposition 1, we have proved that m is a quasistationary distribution if and only if equality holds in (4).

Proposition 3 corrects Theorem 6 of Vere-Jones (1969) and the first part of Corollary 1 of Pollett (1986), both of which assert falsely that a μ-invariant probability distribution on C for Q is *always* μ-invariant of C for P. The error was pointed out by van Doorn (1991) and the counter example which he presented provides the basis for the arguments used above. In determining where the error occurred in the original proof, Vere-Jones and I were able to identify a simple sufficient condition (see Corollary 2 of Pollett and Vere-Jones (1992)). It is instructive to see how this condition arises in the context of Proposition 3. Consider the following formal argument, based on summing (3) over $j \in C$:

$$(7) \qquad \sum_{i \in C} m_i q_{i0} = -\sum_{i \in C} m_i \sum_{j \in C} q_{ij} = -\sum_{j \in C} \sum_{i \in C} m_i q_{ij} = \mu \sum_{j \in C} m_j = \mu.$$

The interchange of summation is not permitted in general, but can be justified under various conditions (see Section 3.7 of Knopp (1956)). For example, the interchange is permitted if the double sum in (7) is absolutely convergent, and a necessary and sufficient condition for this is $\sum_{j \in C} m_j q_j < \infty$. Thus, we have the following result:

Corollary 1. *Let $m = (m_j,\ j \in C)$ be a probability distribution over C and suppose that m is μ-invariant on C for Q. Then, if $\sum_{j \in C} m_j q_j < \infty$, m is a quasistationary distribution and $\mu = \sum_{j \in C} m_j q_{j0}$.*

3. Examples

Finite-state systems. If S is a finite set, then clearly $\sum_{j \in C} m_j q_j < \infty$ and so every μ-invariant probability distribution on C for Q is a quasistationary distribution. Indeed, classical matrix theory can be used to show that the q-matrix restricted to C has eigenvalues with negative real parts, that $-\mu$ is the dominant eigenvalue (it has maximal real part), that this eigenvalue always has multiplicity 1, and, that both the corresponding left and right eigenvectors have positive entries (see Mandl (1960), and Darroch and Seneta (1967)); the left eigenvector is, of course, the quasistationary distribution. Thus, for example, in Markovian reliability models, the stationary conditional distribution of the number of functioning units (conditional on the system not having failed) can be obtained as the dominant left eigenvector of the transition-rate matrix restricted to the transient states. In most cases one is forced to evaluate the dominant eigenvector numerically. If the number of states is reasonably small, say 100, then one can use any of

the standard methods (inverse iteration, for example) which are widely available as part of matrix packages, such as MATLAB. If the number of states is even moderately large, these methods are ineffective, both in respect of storage and CPU time. For example, if there are 10^4 states, Q requires over 400 Mbytes of storage. If Q is sparse, or if it possesses a banded structure that can be usefully exploited, then moderately large systems can be handled without difficulty. In Section 4 I shall describe a computational algorithm for dealing with sparse q-matrices. This has been used to evaluate the quasistationary distribution, to within a tolerance of 10^{-6}, for a variety of systems with of the order of 10^4 states, in times ranging from 7 to 15 CPU minutes on a Sun SPARC 10. If the number of states is large, say 10^6, then it is frequently the case that deterministic approximations (see, for example, Pollett and Roberts (1990)) or diffusion approximations (see, for example, Parsons and Pollett (1987), Pollett (1990, 1992) and Pollett and Vassallo (1992)) can be used to provide accurate estimates of the quasistationary distribution.

Birth and death processes. These are widely used in modelling stochastic systems which arise in engineering, the information sciences and biology. Van Doorn (1991) has given a complete treatment of questions concerning the existence of quasistationary distributions for absorbing birth and death processes in cases when the probability of absorption is 1. I shall explain how his conditions for the existence of quasistationary distributions arise in the context of Proposition 3. The q-matrix of an absorbing birth and death process is of the form

$$q_{ij} = \begin{cases} \lambda_i, & \text{if } j = i+1, \\ -(\lambda_i + \mu_i), & \text{if } j = i, \\ \mu_i, & \text{if } j = i-1, \\ 0, & \text{otherwise,} \end{cases}$$

where the birth rates, $(\lambda_i,\ i \geq 0)$, and the death rates, $(\mu_i,\ i \geq 0)$, satisfy $\lambda_i,\ \mu_i > 0$, for $i \geq 1$, and $\lambda_0 = \mu_0 = 0$. Thus, 0 is an absorbing state and $C = \{1, 2, \dots\}$ is an irreducible class.

The classical Karlin and McGregor theory of birth and death processes involves the recursive construction of a sequence of orthogonal polynomials (see van Doorn (1991)). Define $(\phi_i(\cdot),\ i \geq 1)$, where $\phi_i : \mathrm{R} \to \mathrm{R}$, by $\phi_1(x) = 1$,

$$\lambda_1\phi_2(x) = \lambda_1 + \mu_1 - x,$$
$$\lambda_i\phi_{i+1}(x) - (\lambda_i + \mu_i)\phi_i(x) + \mu_i\phi_{i-1}(x) = -x\phi_i(x), \qquad i \geq 2.$$

Now define $\pi = (\pi_i,\ i \geq 1)$ by $\pi_1 = 1$ and

$$\pi_i = \prod_{j=2}^{i} \lambda_{j-1}/\mu_j, \qquad i \geq 2,$$

and let $m_i = \pi_i\phi_i(x)$. It can be shown (van Doorn (1991)) that $\phi_i(x) > 0$ for x in the range $0 \leq x \leq \lambda$, where $\lambda\ (\geq 0)$ is the decay parameter of C (see Kingman (1963)). Since π is a subinvariant measure for Q, that is

$$\sum_{i \in S} \pi_i q_{ij} \leq 0, \qquad j \in S,$$

it follows, from Theorem 4.1 b(ii) of Pollett (1988), that, for each fixed x in the above range, $m = (m_i,\ i \geq 1)$ is an x-invariant measure on C for Q, that is, m satisfies (3) with $\mu = x$. Indeed, m is uniquely determined up to constant multiples. Proposition 3 tells us that if m can be normalized to produce a proper distribution on C, that is, if

$$\sum_{i=1}^{\infty} \pi_i \phi_i(x) < \infty,$$

then the normalized m will be a quasistationary distribution if and only if

$$\sum_{i=1}^{\infty} r_i(x) = 1, \tag{8}$$

where $r_i(x) = \mu_1^{-1} \pi_i x \phi_i(x)$, a conclusion reached by van Doorn using direct methods. Van Doorn's Theorem 3.2 can then be used to determine all the values of x for which (8) holds, at least under the condition

$$\sum_{i=1}^{\infty} \frac{1}{\lambda_i \pi_i} = \infty, \tag{9}$$

which ensures, not only that Q is regular, but that absorption occurs with probability 1. If, in addition, the series

$$\sum_{i=1}^{\infty} \frac{1}{\lambda_i \pi_i} \sum_{j=i+1}^{\infty} \pi_j \tag{10}$$

diverges then (8) holds for *all* x in $(0, \lambda]$, while if it converges then (8) holds if and only if $x = \lambda$. Proposition 3 then tells us that, in either case, $r(x) = (r_i(x),\ i \geq 1)$ is a quasistationary distribution. Indeed, because m is uniquely determined for each x, *all* quasistationary distributions have been obtained under (9); if the series (10) converges, then there is only one, namely $r(\lambda)$, while if (10) diverges, $(r(x),\ 0 < x \leq \lambda)$ comprises a one-parameter family of quasistationary distributions.

The birth, death and catastrophe process. The introduction of a catastrophe component allows greater flexibility in modelling. The q-matrix of the birth, death and catastrophe process is given by

$$\begin{aligned}
q_{i,i+1} &= a q_i, && i \geq 0, \\
q_{i,i} &= -q_i, && i \geq 0, \\
q_{i,i-k} &= q_i b_k, && i \geq 2,\ k = 1, 2 \ldots i-1, \\
q_{i,0} &= q_i \sum_{k=i}^{\infty} b_k, && i \geq 1,
\end{aligned}$$

where $q_0 = 0$, $q_i > 0$, for $i \geq 1$, $a > 0$, $b_i > 0$ for at least one value of $i \geq 1$ and

$$a + \sum_{i=1}^{\infty} b_i = 1.$$

Thus, at a jump time, a birth occurs with probability a, or otherwise a catastrophe occurs, the size of which is determined by the probabilities b_i, $i \geq 1$. Clearly, 0 is an absorbing state and $C = \{1, 2, \dots\}$ is an irreducible class. It is usual to set $q_i = \rho i$, where $\rho > 0$, so that jumps occur at a constant "per capita" rate ρ. Notice that if, of the b_i's, only b_1 is positive, then we recover the simple linear birth and death process. It is well known, and easy to prove (see, for example, Pakes (1987)), that the probability of absorption, starting in state i, is 1 if and only if D, given by

$$D = a - \sum_{i \in C} i b_i = 1 - \sum_{i \in C} (i+1) b_i,$$

is less than or equal to 0. D can be thought of as the drift, and, accordingly, the process is said to be subcritical, critical or supercritical according as D is negative, zero or positive. In a way that is analogous to the theory of Markov branching processes (see for example, Athreya and Ney (1972)), an important role is played by the probability generating function, f, given by

$$f(s) = a + \sum_{i \in C} b_i s^{i+1}, \qquad |s| < 1,$$

and the related function, b, given by $b(s) = f(s) - s$. In identifying the quasi-stationary distribution, we shall need the following facts from branching process theory: that f is convex on $[0,1]$, that $b(s) = 0$ has a unique solution, σ, on this interval, that $\sigma = 1$ or $0 < \sigma < 1$ according as $D \geq 0$ or $D < 0$, and, that $b(s) \geq 0$ on $[0, \sigma]$.

On substituting the transition rates in equation (3), we get

$$-(\rho - \mu) m_1 + \sum_{k=2}^{\infty} k \rho b_{k-1} m_k = 0,$$

$$(j-1)\rho a m_{j-1} - (j\rho - \mu) m_j + \sum_{k=j+1}^{\infty} k \rho b_{k-j} m_k = 0, \qquad j \geq 2.$$

If we try a solution of the form $m_j = t^j$, the first equation tells us that $\mu = -\rho b'(t)$, of necessity, and, on substituting both of *these* quantities in the second equation, we find that $b(t) = 0$. Hence, we may set $t = \sigma$, thus providing a positive solution, $m = (m_j,\ j \in C)$, to (3), such that $\sum_{j \in C} m_j < \infty$ whenever $\sigma < 1$. Under this latter condition, we also have

$$\sum_{j \in C} m_j q_j = \sum_{j \in C} \sigma^j j \rho < \infty.$$

Thus, by Corollary 1, we have the following result, which is implicit in the proof of Theorem 5.1 of Pakes (1987):

Proposition 4. *The subcritical linear birth, death and catastrophe process has a geometric quasistationary distribution, $m = (m_j,\ j \in C)$, given by*

$$m_j = (1-\sigma)\sigma^{j-1}, \qquad j \in C,$$

where σ is the unique solution to $b(s) = 0$ on the interval $[0,1]$.

4. Computational methods for evaluating quasistationary distributions.

If the quasistationary distribution cannot be exhibited explicitly, or if the form of the quasistationary distribution is not amenable to numerical evaluation, one is forced to use a direct computational approach based on determining the dominant left eigenvector of the q-matrix restricted to C, the set of transient states. If the state space is infinite, we may truncate the restricted q-matrix to an $n \times n$ matrix, $Q^{(n)}$, and then obtain a sequence, $\{m^{(n)}\}$, such that $m^{(n)}$ is the left eigenvector of $Q^{(n)}$ associated with the eigenvalue with maximum real part. Then, one estimates the quasistationary distribution by taking successively larger truncations until the difference in the normalized eigenvectors is as small as desired; some appropriate computational algorithm should be used to evaluate the members the sequence $\{m^{(n)}\}$. For a detailed account of truncation procedures, see Seneta (1973) and Tweedie (1973).

In what follows, we shall suppose that one is given a sparse $n \times n$ matrix, Q, which, for simplicity of notation, *now* refers to the *restriction* of the q-matrix to the transient class C; if C is infinite, we shall suppose that Q is a specified $n \times n$ truncation of this matrix. The task is to determine the left eigenvector associated with the eigenvalue with smallest (in modulus) real part. Since it is the custom in the computational mathematics literature, I shall describe the algorithm as it relates to the determination of *right* eigenvectors of a sparse q-matrix; the corresponding left eigenvector can be evaluated by applying the method to the transpose of Q. It should be noted, at this point, that right eigenvectors *do* play an important role in the theory, though, for brevity and simplicity of exposition, I did not mentioned this earlier. They arise both in connection with the so-called Type-II quasistationary distribution (see, for example, Flaspohler (1974)), and in determining absorption probabilities, which are of interest in their own right, but are needed in the present context to evaluate quasistationary distributions for "supercritical" models, where absorption occurs with probability less than 1 (again see Flaspohler (1974)).

An iterative Arnoldi algorithm. The method I shall describe was developed previously by David Stewart and I; further details can be found in (Pollett and Stewart (1994)). It is a variant of the Arnoldi Algorithm, which itself is similar to the more familiar Lanczos Method for finding eigenvalues and eigenvectors of large, sparse *symmetric* matrices (see, for example, Golub and van Loan (1989), Pages 475–493); an assumption that the q-matrix is symmetric would clearly be far too restrictive for practical purposes.

On each iteration of the algorithm, the "basic" Arnoldi Method is used. This involves constructing an $m \times m$ matrix, H_m, where m is smaller (usually much

smaller) than n. The eigenvectors of this matrix are then determined by some efficient dense-matrix method and these are used to provide estimates of the extremal eigenvalues of Q. The method starts with an arbitrary "seed" vector, v_1, from which a sequence, $v_1, v_2, \ldots$, of orthonormal vectors is constructed as follows. First, the vector $w_1 = Qv_1$ is computed. Then, the components of w_1 in the direction of v_1 are subtracted to give a "residual", $r_1 = w_1 - v_1^T w_1 \, v_1$. This vector is normalized, using the Euclidean norm, $\|\cdot\|_2$, to form v_2: $v_2 = r_1/\|r_1\|_2$. In the next cycle $w_2 = Qv_2$ is computed and then the components of w_2 in the directions of v_1 and v_2 are subtracted to give the second residual, r_2. This is normalized to give v_3, and so on. The procedure described gives

$$w_j = Qv_j,$$

$$r_j = w_j - \sum_{i=1}^{j} v_i^T w_j \, v_i,$$

and

$$v_{j+1} = r_j/\|r_j\|_2.$$

If the procedure is halted at say $j = m$, then we shall have that

$$Qv_j = \begin{cases} \sum_{i=1}^{j+1} h_{ij} v_i, & \text{for } j < m, \\ \sum_{i=1}^{m} h_{ij} v_i + r_m, & \text{for } j = m, \end{cases}$$

where $H_m = (h_{ij})$ is an $m \times m$ upper-Hessenberg matrix given by

$$h_{ij} = \begin{cases} w_j^T v_i, & \text{for } i = 1, 2, \ldots, j, \\ \|r_j\|_2, & \text{for } i = j+1, \\ 0 & \text{otherwise.} \end{cases}$$

Thus, if we denote by V_m the matrix with columns $v_1, v_2, \ldots, v_m$, we shall have that

$$QV_m = V_m H_m + r_m e_m^T, \tag{11}$$

where e_m is the unit vector with a 1 as its m^{th} entry, and so, since the columns of V_m are orthonormal and r_m is orthogonal to each of them, we may deduce that $V_m^T Q V_m = H_m$. The idea which is the key to the Arnoldi Method is that if z_m is an eigenvector of H_m, then, for m sufficiently large, $V_m z_m$ should be close to an eigenvector of Q. To make this statement more precise, suppose that z_m satisfies

$$H_m z_m = \hat{\lambda}_m z_m,$$

for some $\hat{\lambda}_m$, and let $x_m = V_m z_m$. Then, on multiplying (11) by z_m, we obtain

$$\begin{aligned} Qx_m &= V_m(H_m z_m) + r_m(z_m)_m \\ &= V_m(\hat{\lambda}_m z_m) + r_m(z_m)_m \\ &= \hat{\lambda}_m x_m + r_m(z_m)_m. \end{aligned}$$

Thus

$$(Q - E_m)x_m = \hat{\lambda}_m x_m,$$

where E_m is given by

$$E_m = \frac{r_m (z_m)_m x_m^T}{\|x_m\|_2^2}.$$

It follows, from standard sensitivity analysis (see, for example, Section 7.2 of Golub and van Loan (1989)), that the error in the eigenvalue can be estimated by $\|r_m\|_2 \, |(z_m)_m| / \|x_m\|_2$. Hence, if the residual vector, r_m, is small or $|(z_m)_m|$ is small, then the approximation will be good. The eigenvalues of H_m can be computed by means of a Schur decomposition and then inverse iteration, since H_m is a (relatively) small matrix. In this way approximations to eigenvalues and eigenvectors of large, sparse matrices can be found in a reasonable time. A more detailed discussion of the basic Arnoldi Method can be found on Pages 501–502 of Golub and van Loan (1989).

One problem with the basic method is that, in order to guarantee an accurate estimate of the required eigenvector of Q, namely the one corresponding to the eigenvalue with smallest (in modulus) real part, m might need to be reasonably large. Stewart and I address this problem by fixing m to be small and adopting an iterative approach, at each stage of which, the basic Arnoldi Method is used. Since m is small, the evaluation of the eigenvectors of H_m at each step is relatively inexpensive. The iteration begins with v_1 specified as an initial estimate of the required eigenvector. The matrices H_m and V_m are computed as above and the eigenvalues of H_m are determined. Then, $\hat{\lambda}$ is set to be the eigenvalue of smallest real part, if it is real, and zero otherwise. The system $(H_m - \hat{\lambda} I)u_1 = z$ is solved for u_1 with z chosen at random for each cycle and this vector is used to produce a new estimate, v_1, of the required eigenvector, in preparation for the next cycle: $v_1 = V_m u_1 / \|V_m u_1\|_2$. The cycles continue until $\|Qv_1 - \hat{\lambda} v_1\|_2$ is sufficiently small. In this way, $\hat{\lambda}$ is at least an eigenvalue of a slightly perturbed matrix, with perturbation of the same size as $\|Qv_1 - \hat{\lambda} v_1\|_2$.

It may appear that solving $(H_m - \hat{\lambda} I)u_1 = z$ will result in a large numerical error in u_1 as the matrix $H_m - \hat{\lambda} I$ will be ill-conditioned if $\hat{\lambda}$ is an accurate estimate of an eigenvalue of H_m. Indeed, there *will be* a large numerical error in u_1. Nonetheless, the normalized vector $u_1 / \|u_1\|_2$ will be close to a corresponding eigenvector of H_m provided the equations are solved using a standard factorize-and-solve method such as Gaussian elimination. The reason is that the computed $\hat{u}_1$ is an exact solution of a perturbed system,

$$(H_m + E - \hat{\lambda} I)\hat{u}_1 = z,$$

where $\|E\|_2 \approx c_m \mathbf{u} \|H_m - \hat{\lambda} I\|_2$, $\{c_m\}$ is a sequence of constants that grows slowly and $\mathbf{u}$ is the "machine epsilon" or "unit roundoff" for the arithmetic used; see Pages 104–107 and Pages 123–124 of Golub and van Loan (1989). Since $\hat{\lambda}$ is close to an eigenvalue of $H_m + E$ (the error in $\hat{\lambda}$ is $O(c_m \mathbf{u} \|H_m\|_2)$ for a simple eigenvalue) it follows that $u_1 / \|u_1\|_2$ is close to an eigenvector of H_m.

Another part of the algorithm that may be vulnerable to roundoff error is the inner Arnoldi iteration when $\|r_{j+1}\|$ is very small. This can occur, for example,

when v_1 is close to an eigenvector of Q. If $\|r_{j+1}\| \approx \mathbf{u}$ then v_{j+1} is largely error and the Arnoldi Method essentially restarts. Note that the v_j are no longer nearly orthogonal. However, since H_m is Hessenberg and $h_{j,j+1} = \|r_{j+1}\| \approx \mathbf{u}$, H_m decouples, and, apart from the multiplicity of the eigenvalues, the eigenvalues are not perturbed by more than approximately one machine epsilon; similar behaviour has been observed by Cullum and Willoughby (1978) in connection with the Lanczos Algorithm for symmetric matrices.

One has a certain amount of freedom in prescribing the value, m, which determines the size of the Hessenberg matrix H_m. This value is perhaps best determined by experimentation. If m is chosen too large or too small, then the algorithm will be slow. If too large, the time taken to evaluate the eigenvectors of H_m will be predominant, while, if too small, the number of outer iterations might be prohibitively large.

Implementation and results. I shall illustrate the algorithm by evaluating the quasistationary distributions of model which is an example of the general epidemic model studied by Ridler-Rowe (1967). The state at time t consists of two entries, $S(t)$ and $I(t)$, being, respectively, the numbers of susceptibles and infectives at time t. The transition rates are given as follows: if $i = (x, y)$, then

$$q_{ij} = \begin{cases} \beta xy & \text{if } j = (x-1, y+1), \\ \gamma y & \text{if } j = (x, y-1), \\ \alpha & \text{if } j = (x+1, y), \\ 0 & \text{otherwise.} \end{cases}$$

The model thus allows for the *immigration* of new susceptibles, at rate $\alpha > 0$, the *removal* of infectives, at per-capita rate $\gamma > 0$, and the *spread of the infection*, the per-encounter infection rate being $\beta > 0$. The state space is

$$S = \{(x, y) : x, y = 0, 1, 2, \dots\},$$

and it is clear, from the form of the transition rates, that C, given by

$$C = \{(x, y) : x = 0, 1, \dots; y = 1, 2, \dots\},$$

is an irreducible transient class. Moreover, once the process leaves C it becomes trapped on the x-axis; all infectives have been removed from the population and it grows without bound at rate α. The state diagram for the Ridler-Rowe (1967) epidemic model is illustrated in Figure 1.

Ridler-Rowe showed that the process is regular (non-explosive) and that eventual departure from C occurs with probability 1. This and related models have been studied extensively; see, for example, Bartlett (1956, 1958), Kendall (1956) and the other references contained in Ridler-Rowe (1967). Various methods have been proposed for studying the long-term behaviour of the population, and, in particular, to explain why over any reasonable time scale the epidemic appears not to die out. For example, one can write down a diffusion approximation which is asymptotically valid in the limit as α becomes large. This provides an approximate analytical formula for the joint distribution of the numbers of infectives and

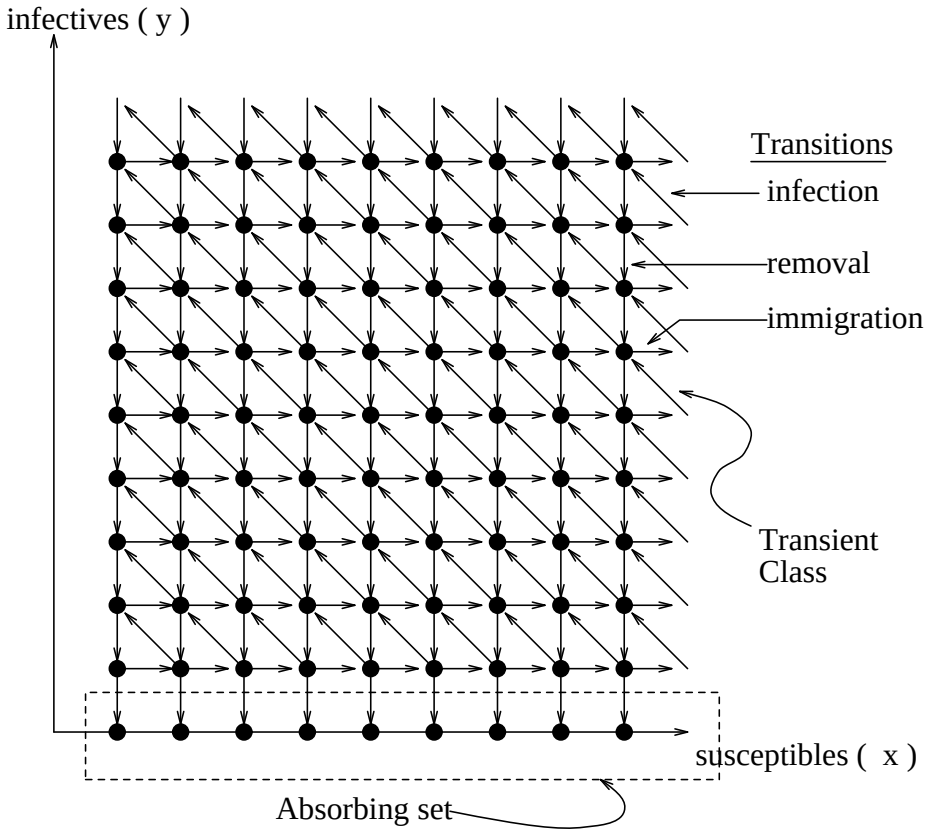

FIG. 1. State diagram for the Ridler-Rowe epidemic model.

susceptibles. It is a bivariate normal distribution centred at the stable equilibrium point of the analogous deterministic model. In contrast, no analytical expression is available for the quasistationary distribution and, although one would expect this to be more accurate than the diffusion approximation, apparently it has not previously been examined.

Suppose that we truncate C to

$$C_N = \{(x, y) : x = 0, 1, \ldots, N-1; y = 1, 2, \ldots, N\},$$

thus placing an upper limit on the numbers of infectives and susceptibles. One can then employ the transformation $i = x + N(y-1)$ to convert the truncated q-matrix into an $n \times n$ matrix, $Q = (q_{ij},\ i, j = 0, 1, \ldots, n-1)$, where $n = N^2$. For N reasonably large, this matrix is large and sparse However, although Q is banded, its bandwidth is of order $\sqrt{n}\,(= N)$ and so direct banded-matrix methods are impracticable.

The Iterative Arnoldi Method was applied to Q^T with m set to 20, and terminated according to the following convergence criterion:

$$\frac{\|Qv_1 - \hat{\lambda} v_1\|_\infty}{\|v_1\|_\infty} \leq tol,$$

where *tol*, the prescribed tolerance, was 10^{-6}. Starting with a random v_1, where $(v_1)_i$ was chosen at random between zero and one, and $n = N^2 \approx 100^2$, the method took around 120 to 170 outer iterations and 7 to 15 CPU minutes on a Sun SPARC 10 to achieve acceptance of the solution at the prescribed tolerance. Note that the size of entries of the q-matrix for the epidemic model with $n = N^2 \approx 100^2$ may be as large as $\beta N^2 \approx 100^2$. This means that the error relative to $\|Q\|_\infty$ is about 10^{-10}.

Of course, the computed eigenvector, when suitably normalized, can be interpreted as a quasistationary distribution. Each entry should therefore have the same sign. There is nothing in the numerical algorithm, bar the error tolerance, to guarantee this. Nevertheless, the method seems to provide very good results: for $n \approx 100^2$, $\min_i(v_1)_i$ was less than zero for the computed eigenvector v_1, but $|\min_i(v_1)_i| < 10^{-19}\|v_1\|_\infty$. This is remarkable when it is realized that, in the double precision arithmetic used, $\mathbf{u} \approx 10^{-16}$ and that some of the positive entries were of size $\approx 10^{-35}$. Why it should work so well is not clearly understood; however, see below. In addition, some numerical experiments were performed in order to determine, roughly, the optimum value of m, the number of Arnoldi steps per major cycle. It seems that $m \approx 20$ is about the best, at least for $n \approx 100^2$. Setting $m = 10$ for $n = 100^2$ reduced the time per major cycle by about half, but the number of major cycles required jumped to over 450. For $m = 30$ the number of major cycles required was reduced to only about 140, with an increase of time per major cycle of about 50%.

Figure 2 shows a mesh plot of the quasistationary distribution of the epidemic process with reflection at the truncation boundary, while Figure 3 shows the corresponding contour plot, together with contours of the stationary distribution of the approximating diffusion. The latter was obtained as follows, using the methods described in Section 3 of Pollett (1990): Take N to be the index ("size of the system") and set $\alpha' = \alpha/N$, $\beta' = \beta N$ and $\gamma' = \gamma$ to obtain "size-independent" quantities (in the numerical experiments these were taken to be $\alpha' = 1$, $\beta' = 4$ and $\gamma' = 2$). In this way the epidemic model can be regarded as being *asymptotically density dependent* in the limit as α becomes large. If we then set

$$X_N(t) = \frac{1}{N}(S(t), I(t)),$$

we can show that the corresponding deterministic model for X_N is

$$\frac{dX}{dt} = F(X),$$

where

$$F(x_1, x_2) = \begin{pmatrix} -\beta' x_1 x_2 + \alpha' \\ \beta' x_1 x_2 - \gamma' x_2 \end{pmatrix}.$$

This model has a unique equilibrium point,

$$x_0 = (\gamma'/\beta', \alpha'/\gamma').$$

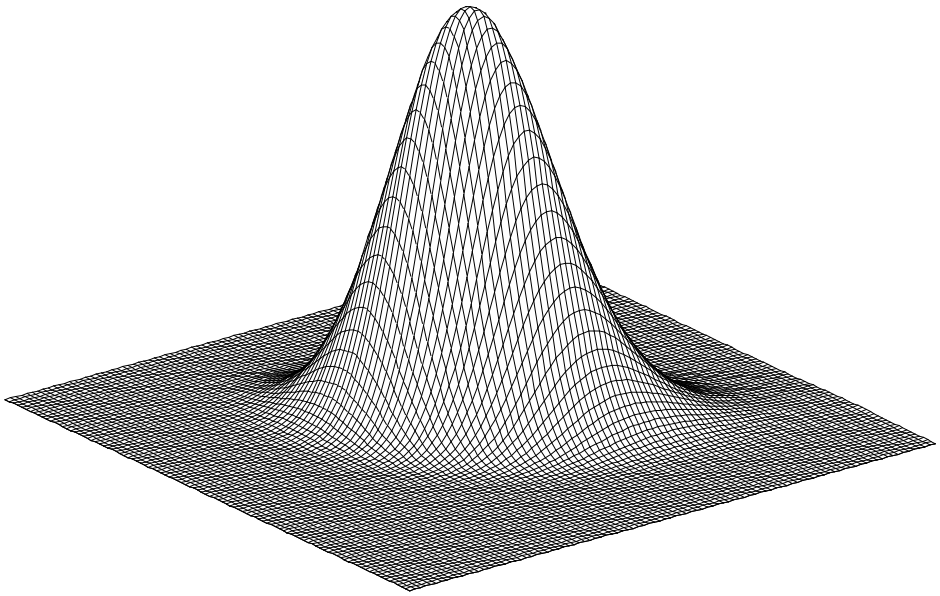

FIG. 2. View of the quasistationary distribution of the (truncated) epidemic model with reflection at the truncation boundary and with parameters $N = 100$, $\alpha = 100$, $\beta = 0.04$ and $\gamma = 2$. The origin is the nearest corner, the x-axis is to the right of the origin and the y-axis is to its left.

It follows, from Theorem 3.2 of Pollett (1990), that $\sqrt{N}(X_N(\cdot) - x_0)$ converges weakly in the space of all sample paths on any given finite time interval to a stable Ornstein-Uhlenbeck process with local drift matrix, B, given by

$$B = \begin{pmatrix} -\alpha'\beta'/\gamma' & -\gamma' \\ \alpha'\beta'/\gamma' & 0 \end{pmatrix},$$

and with local covariance matrix, G, given by

$$G = \begin{pmatrix} 2\alpha' & -\alpha' \\ -\alpha' & 2\alpha' \end{pmatrix}.$$

It has, therefore, a bivariate normal stationary distribution with zero mean and covariance matrix, Σ, which satisfies

$$B\Sigma + \Sigma B^T = -G.$$

It can be shown that

$$\Sigma = \begin{pmatrix} \frac{\gamma'}{\beta'}\left(1 + \frac{(\gamma')^2}{\alpha'\beta'}\right) & -\gamma'/\beta' \\ -\gamma'/\beta' & \frac{\gamma'}{\beta'} + \frac{\alpha'}{\gamma'}. \end{pmatrix}.$$

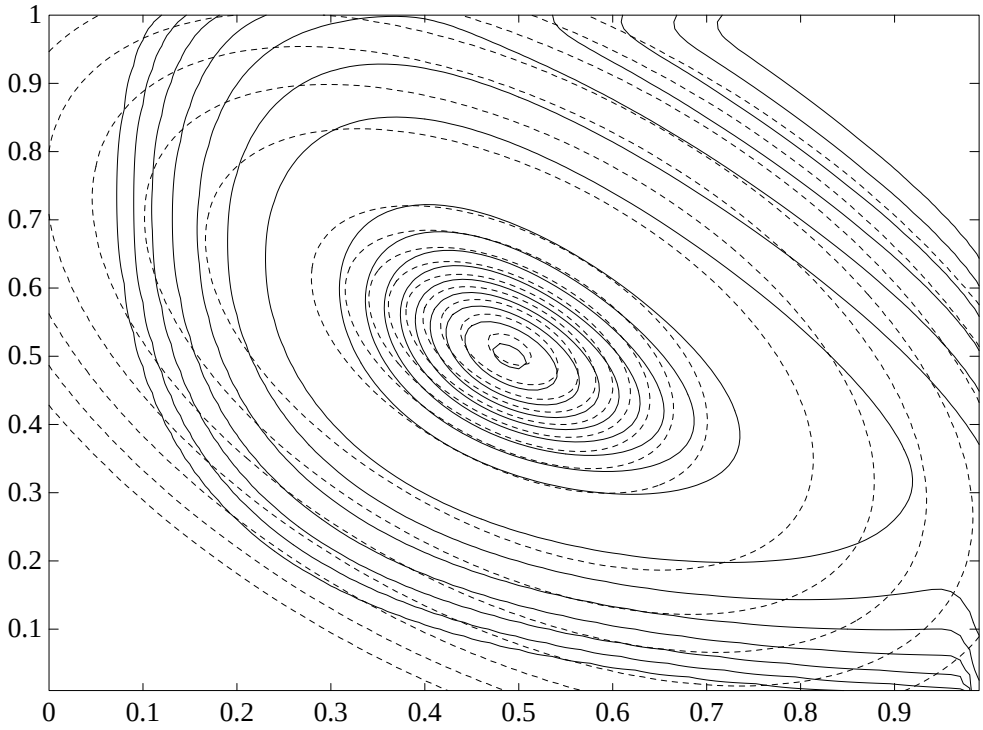

FIG. 3. Contours of the quasistationary distribution (solid) of the epidemic model with reflection at the truncation boundary and with parameters $N = 100$, $\alpha = 100$, $\beta = 0.04$ and $\gamma = 2$, together with those of the diffusion approximation (dashed).

It is useful to consider the contours in Figure 2 as forming two groups. The first, comprising levels 1.78×10^{-3}, 1.6×10^{-3}, 1.4×10^{-3}, 1.2×10^{-3}, 10^{-3}, 8×10^{-4}, 6×10^{-4}, 4×10^{-4} and 2×10^{-4}, account for most of the mass of the distribution. The second group of outer contours, at levels 10^{-5}, 10^{-6}, 10^{-7}, 10^{-8}, 10^{-9}, 10^{-10}, 10^{-11}, account for the extremal behaviour of the epidemic process. Here one would not expect the diffusion approximation to be accurate.

The convergence of the Iterative Arnoldi Method. Figure 4 shows a plot of the value of the standardized residual norm, $\|Qv_1 - \hat{\lambda}v_1\|_\infty / \|v_1\|_\infty$, against the number of outer iterations. Notice that, after each iteration, there may be an increase in this norm, but that, overall, there is a decrease which is approximately log-linear in the number of outer iterations. Observe, also, that the convergence rate per outer iteration increases with m. This is to be expected, because the larger the value of m, the more that the "basic" Arnoldi Method dominates. However, it should be emphasized that, for large m, the "basic" method is slow; as explained above, $m = 20$ appears to provide the fastest overall convergence.

Short of convincing empirical evidence exemplified in Figure 4, it remains an open problem to demonstrate, analytically, why the method works so well. Cer-

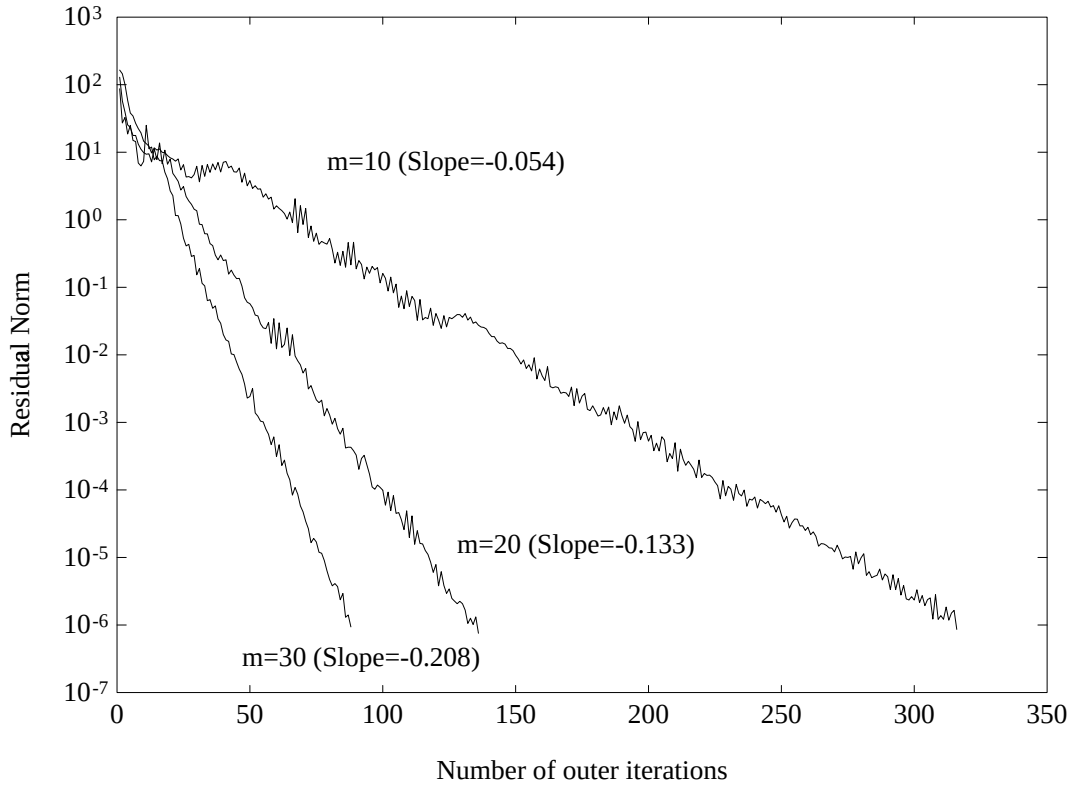

FIG. 4. Plot of the residual norm against the number of outer iterations for Iterative Arnoldi Method applied to the epidemic model with parameters $N = 100$, $\alpha = 100$, $\beta = 0.04$ and $\gamma = 2$. The parameter m is the size of the matrix to which the "basic" Arnoldi Method is applied.

tainly neither Stewart nor I have been able to do this. However, we can at least provide some positive rationale based on variational theory which is valid in the case when Q is symmetric. Suppose, then, that Q is symmetric. Its smallest eigenvalue is the minimum value of $q(v) = v^T Qv/v^T v$ over $v \neq 0$ and the minimum eigenvalue of H_m is the minimum eigenvalue of $q(v)$ over non-zero v in span(V_m), the span of the *columns* of V_m. If V_m is generated using v_1, then, since the $v_1 \in$ span(V_m), the minimum eigenvalue of H_m must be less than or equal to $q(v_1)$. If $\hat{v}_1$ is the vector computed by one iteration of the basic Arnoldi Method, then $\hat{v}_1$ minimizes $q(v)$ over v in span(V_m). So, clearly, $q(\hat{v}_1) \leq q(v_1)$, with equality only if v_1 is already an eigenvector. However, one must be certain that the successive estimates, v_1, don't converge to the wrong eigenvector. This can't happen if the original (starting) v_1 has a component in that direction. To see this, put $v_1 = \sum_{j=1}^{n} \alpha_j x_j$, where $Qx_j = \lambda_j x_j$ and $\|x_j\| = 1$ for $j = 1, 2, \ldots, n$; note that the x_j's are orthogonal as Q is assumed to be symmetric. Then, every vector, v, in span(V_m) has the form $\sum_j \alpha_j p(\lambda_j) x_j$ for some polynomial, p, of degree $\leq m$.

Thus,

$$q(v) = \sum_j \alpha_j^2 p(\lambda_j)^2 \lambda_j / \sum_j \alpha_j^2 p(\lambda_j)^2.$$

The minimizing polynomial will have to satisfy $p(\lambda_1) \geq 1 \geq p(\lambda_j)$, for all j, and so the eigenvector computed will become closer to a true eigenvector.

In addition, we can offer some further evidence based on the work of Saad (1980). If u is the eigenvector corresponding to the smallest eigenvalue, λ, of Q (now, not necessarily symmetric), then Theorem 2.1 of Saad (1980) implies that

$$\|(Q_m - \lambda I)u\| \leq (|\lambda| + \|Q\|)\|(I - \pi_m)u\|,$$

where π_m is the orthogonal projection operator onto the space span(V_m), and $Q_m = \pi_m Q \pi_m$. Proposition 2.1 of Saad (1980) implies that if v_1 can be represented as $v_1 = \sum_i \alpha_i x_i$, where, as above, $x_1, x_2, \ldots$ are the eigenvectors of Q, normalized to have unit length, then

$$\|(I - \pi_m)u\| \leq \left(\sum_{j \neq 1} |\alpha_j| / |\alpha_1| \right) \min_p \max_{j \neq 1} |p(\lambda_j)|,$$

where the minimization is taken over all polynomials, p, of degree $< m$ with $p(\lambda_1) = 1$; note that $x_1 = u$ and $\lambda_1 = \lambda$. This result strongly suggests that, for m sufficiently large, the computed $\hat{\alpha}_1$ should be less than α_1. This would establish the required convergence.

Other methods and extensions. The Arnoldi Method was chosen over other methods (such as inverse iteration) which are appropriate for finding a single eigenvalue/eigenvector pair. One reason is that the matrices involved here are very large. Inverse iteration would require Q to be factorized. Sparse matrix techniques would be needed since, as mentioned earlier, Q, stored as a dense matrix, would require over 400 Mbytes of storage. Even using a Markowitz type sparse matrix factorization results in memory overflow at this size of matrix. It may be possible to implement inverse iteration using a "nested dissection" ordering of the rows and columns of Q. If this is possible, it may be more worthwhile to apply the iterative Arnoldi Method to Q^{-1} to find its largest real eigenvalue, than to use inverse iteration. The main problem with the current implementation, however, is not time, but memory. Specifically, the (dense) matrix V_m, which certainly needs to be stored, can be prohibitively large.

The Iterative Arnoldi Method can be modified to provide not only the smallest eigenvalue and its eigenvector, but also a collection of eigenvalues and eigenvectors, say the three most dominant eigenvalues and their eigenvectors. The eigenvalues could then be used to estimate the rate of convergence of the process to its quasistationary state; see Pollett and Roberts (1990). To obtain a collection of eigenvalues and eigenvectors one would first compute a value of v_1 which is a linear combination of the corresponding eigenvectors of H_m. Provided the subspace of eigenvectors sought has dimension much less than m, the method should be successful.

Finally, research continues, both at the University of Queensland and the Australian National University, on developing methods for evaluating stationary and quasistationary distributions for Markovian models. Presently, researchers at both institutions are evaluating *multigrid methods* and these are currently being implemented on a variety of parallel architectures, including the University of Queensland's MASPAR MP-1, a 4096 (64×64)-processor SIMD machine.

Acknowledgments. I should like to thank Mark Bebbington, Ingemar Nasell and David Stewart for their interest and helpful comments on an earlier draft of the paper. Also, I should like to acknowledge David Stewart's art work; he produced the postscript for Figure 1.

References

1. Anderson, W.J. (1991), *Continuous-time Markov chains: an applications oriented approach*, Springer-Verlag, New York.
2. Athreya, K.B. and Ney, P.E. (1972), *Branching Processes*, Springer-Verlag, Berlin.
3. Bartlett, M.S. (1956), *Deterministic and stochastic models for recurrent epidemics*, Proc. Third Berkeley Symposium on Mathematical Statistics and Probability **4**, 81–109, University of California Press.
4. Bartlett, M.S. (1960), *Some stochastic models in ecology and epidemiology*, Contributions to Probability and Statistics, A Volume dedicated to Harold Hotelling, Stanford University Press, Stanford, pp. 89–96.
5. Cullum, J. and Willoughby, R. (1978), *Lanczos and the computation in specified intervals of the spectrum of large, sparse real symmetric matrices*, Sparse Matrix Proceedings, pp. 220–255.
6. Cavender, J.A. (1978), *Quasistationary distributions for birth-death processes*, Adv. Appl. Probab. **10**, 570–586.
7. Dambrine, S. and Moreau, M. (1981a), *Note on the stochastic theory of a self-catalytic chemical reaction, I*, Physica **106A**, 559–573.
8. Dambrine, S. and Moreau, M. (1981b), *Note on the stochastic theory of a self-catalytic chemical reaction, II*, Physica **106A**, 574–588.
9. Darroch, J.N. and Seneta, E. (1967), *On quasi-stationary distributions in absorbing continuous-time finite Markov chains*, J. Appl. Probab. **4**, 192–196.
10. van Doorn, E.A. (1991), *Quasi-stationary distributions and convergence to quasi-stationarity of birth-death processes*, Adv. Appl. Probab. **23**, 683–700.
11. van Doorn, E.A. and Kijima, M. (1992), *Asymptotics for birth-death polynomials and quasi-limiting distributions for birth-death processes*, Preprint, University of Twente.
12. Flaspohler, D.C. (1974), *Quasi-stationary distributions for absorbing continuous-time denumerable Markov chains*, Ann. Inst. Statist. Math. **26**, 351–356.
13. Golub, G.H. and van Loan, C. (1989), *Matrix Computations*, 2nd Edition, John Hopkins Press.
14. Good, P. (1968), *The limiting behaviour of transient birth and death processes conditioned on survival*, J. Austral. Math. Soc. **8**, 716–722.
15. Holling, C.S. (1973), *Resilience and stability of ecological systems*, Ann. Rev. Ecol. Systematics **4**, 1–23.
16. Kalpakam, S. and Shahul Hameed, M.A. (1983), *Quasi-stationary distribution of a two-unit warm-standby redundant system*, J. Appl. Probab. **20**, 429–435.
17. Kendall, D.G. (1956), *Deterministic and stochastic epidemics in closed populations*, Proc. Third Berkeley Symposium on Mathematical Statistics and Probability, U. of California Press **4**, pp. 149–165.
18. Kijima, M. (1992), *Quasi-limiting distributions of skip-free to the left Markov chains in continuous-time*, Preprint, University of Tsukuba, Tokyo.
19. Kijima, M. and Seneta, E. (1991), *Some results for quasi-stationary distributions of birth-death processes*, Adv. Appl. Probab. **28**, 503–511.

20. Kingman, J.F.C. (1963), *The exponential decay of Markov transition probabilities*, Proc. London Math. Soc. **13**, 337-358.
21. Klein, D.R. (1968), *The introduction, increase, and crash of Reindeer on St. Matthew Island*, J. Wildlife Man. **32**, 351–367.
22. Knopp, K., *Infinite Sequences and Series*, Dover, New York, 1956.
23. Mandl, P. (1960), *On the asymptotic behaviour of probabilities within groups of states of a homogeneous Markov process*, Cas. pest. mat. **85**, 448–456.
24. Mech, L.D. (1966), *The wolves of Ilse Royale*, Fauna of the National Parks: U.S. Fauna Series **7**, U.S. Govt. Printing Office, Washington D.C.
25. Nair, M.G. and Pollett, P.K. (1992), *On the relationship between μ-invariant measures and quasistationary distributions for continuous-time Markov chains*, Adv. Appl. Probab. **24** (to appear).
26. Oppenheim, I. Schuler, K.K. and Weiss, G.H. (1977), *Stochastic theory of nonlinear rate processes with multiple stationary states*, Physica **88A**, 191-214.
27. Pakes, A.G. (1987), *Limit theorems for the population size of a birth and death process allowing catastrophes*, J. Math. Biol. **25**, 307–325.
28. Pakes, A.G. and Pollett, P.K. (1989), *The supercritical birth, death and catastrophe process: limit theorems on the set of extinction*, Stochastic Process. Appl. **32**, 161–170.
29. Parsons, R.W. and Pollett, P.K. (1987), *Quasistationary distributions for some autocatalytic reactions*, J. of Statist. Phys. **46**, 249–254.
30. Pijnenburg, M. and Ravichandran, N. (1990), *Quasi-stationary distribution of a two-unit dependent parallel system*, Preprint, Eindhoven University of Technology.
31. Pijnenburg, M., Ravichandran, N. and Regterschot, G. (1990), *Stochastic analysis of a dependent parallel system*, Preprint, Eindhoven University of Technology.
32. Pollett, P.K. (1986), *On the equivalence of μ–invariant measures for the minimal process and its q-matrix*, Stochastic Process. Appl. **22**, 203–221.
33. Pollett, P.K. (1987), *On the long-term behaviour of a population that is subject to large–scale mortality or emigration*, Ed. S. Kumar, Proceedings of the 8th National Conference of the Australian Society for Operations Research, pp. 196–207.
34. Pollett, P.K. (1988a), *Reversibility, invariance and μ–invariance*, Advances in Applied Probability **20**, 600–621.
35. Pollett, P.K. (1988b), *On the problem of evaluating quasistationary distributions for open reaction schemes*, J. Statist. Phys. **53**, 1207–1215.
36. Pollett, P.K. (1990), *On a model for interference between searching insect parasites*, J. Austral. Math. Soc. Ser. B **31**, 133–150.
37. Pollett, P.K. (1992), *Diffusion approximations for a circuit switching network with random alternative routing*, Austral. Telecom. Res. **25**, 45–52.
38. Pollett, P.K. and Roberts, A.J. (1990), *A description of the long-term behaviour of absorbing continuous-time Markov chains using a centre manifold*, Adv. Appl. Probab. **22**, 111–128.
39. Pollett, P.K. and Stewart, D.E. (1994), *A computational procedure for modelling the long-term behaviour of evanescent random processes*, J. Appl. Probab. **31** (to appear).
40. Pollett, P.K. and Taylor, P.G. (1993), *On the problem of establishing the existence of stationary distributions for continuous-time Markov chains*, Probab. Eng. Informat. Sci. (to appear).
41. Pollett, P.K. and Vassallo, A. (1992), *Diffusion approximations for some simple chemical reaction schemes*, Adv. Appl. Probab. **24**, 875–893.
42. Pollett, P.K. and Vere-Jones, D. (1992), *A note on evanescent processes*, Austral. J. Statist. **34**, 531–536.
43. Ridler-Rowe, C.J. (1967), *On a stochastic model of an epidemic*, J. Appl. Probab. **4**, 19–33.
44. Saad, Y. (1980), *Variations on Arnoldi's method for computing the eigenelements of large unsymmetric matrices*, Linear Algebra Appl. **34**, 269–295.
45. Scheffer, V.B. (1951), *The rise and fall of a reindeer herd*, Sci. Monthly **73**, 356–362.
46. Seneta, E. (1973), *Non-negative matrices and Markov chains*, Springer-Verlag, New York.
47. Turner, J.W. and Malek-Mansour, M. (1978), *On the absorbing zero boundary problem in birth and death processes*, Physica **93A**, 517–525.

48. Tweedie, R.L. (1973), *The calculation of limit probabilities for denumerable Markov processes from infinitesimal properties*, J. Appl. Probab. **10**, 84–99.
49. Tweedie, R.L. (1974), *Some ergodic properties of the Feller minimal process*, Quart. J. Math. Oxford **25**, 485–495.
50. Vere-Jones, D. (1962), *Geometric ergodicity in denumerable Markov chains*, Quart. J. Math. Oxford, Ser. 2 **13**, 7–28.
51. Vere-Jones, D. (1969), *Some limit theorems for evanescent processes*, Austral. J. Statist. **11**, 67–78.
52. Yaglom, A.M. (1947), *Certain limit theorems of the theory of branching processes*, Dokl. Acad. Nauk SSSR **56**, 795–798.

P.K. Pollett, Department of Mathematics, The University of Queensland, Queensland 4072, AUSTRALIA.

E-mail: pkp@markov.maths.uq.oz.au