DES and RES Processes and their Explicit Solutions

Michael N. Katehakis
Dept. of Management Science and Information Systems, Rutgers Business School - Newark and New Brunswick, 1 Washington Park Newark, NJ 07102, mnk@rutgers.edu

Laurens Smit
Mathematisch Instituut, Universiteit Leiden Niels Bohrweg 1, 2333 CA, The Netherlands, & Dept. of Management Science and Information Systems, Rutgers Business School - Newark and New Brunswick, 1 Washington Park Newark, NJ 07102, laurens.smit@rutgers.edu

Flora Spieksma
Mathematisch Instituut, Universiteit Leiden Niels Bohrweg 1, 2333 CA, The Netherlands, spieksma@math.leidenuniv.nl

This article defines and studies the *down entrance state* (DES) and the *restart entrance state* (RES) classes of quasi skip free (QSF) processes specified in terms of the nonzero structure of the elements of their transition rate matrix Q. A QSF process is a Markov chain with states that can be specified by tuples of the form (m, i), where $m \in \mathbb{Z}$ represents the "current" level of the state and $i \in \mathbb{Z}^+$ the current phase of the state, and its transition probability matrix Q does not permit one step transitions to states that are two or more levels away from the current state in one direction of the level variable m. A QSF process is a DES process if and only if *one step "down" transitions from a level m can only reach a single state in level $m-1$, for all m.* A QSF process is a RES process if and only if *one step "up" transitions from a level m can only reach a single set of states in the highest level M_2, largest of all m.*

We derive explicit solutions and simple truncation bounds for the steady state probabilities of both DES and RES processes, when in addition Q insures ergodicity. DES and RES processes have applications in many areas of applied probability comprising computer science, queueing theory, inventory theory, reliability and the theory of branching processes. To motivate their applicability we present explicit solutions for the well-known open problem of the $M/Er/n$ queue with batch arrivals, an inventory model, and a reliability model.

Key words: Successively Lumpable, Stochastic Processes, Queueing, Inventory, Steady State Analysis

1. Introduction

Consider a Markov chain with states that can be specified by tuples of the form (m, i), where $m \in \mathbb{Z}$ represents the "current" level of the state and $i \in \mathbb{Z}^+$ the current phase of the state. The process is called *quasi skip free* (QSF) *to the left* (down) when its transition rate matrix Q (cf. Eq. (1)) does not permit one step transitions to states that are two or more levels away from the current state in the downwards direction of the level variable m. QSF processes generalize the well studied class of *quasi birth and death* processes (QBD) for which one step transitions to states that are two or more levels away from the current state in either direction of the level variable m are not allowed. *QSF to the right* (up) process can be defined with an apparent modification of the definition of the transition rate matrix Q. In the sequel for simplicity we will refer to a QSF process when the skip free direction is apparent (and without loss of generality taken to be the "down" direction).

This article defines and studies the *down entrance state* (DES) and the *restart entrance state* (RES) classes of QSF processes specified in terms of the nonzero structure of the elements of the matrix Q. Explicitly: a) a QSF process satisfies the (DES) property if and only if *one step "down" transitions from a level m can only reach a single state in level $m-1$, for all m* and b) a QSF process satisfies the (RES) property if and only if *one step "up" transitions from a level m can only reach a single set of states in the highest level M_2, largest of all m.*

The main contributions of this paper are as follows. First, we derive explicit solutions for the steady state probabilities of level dependent DES and RES processes, cf. Eqs. (10), (13), (26) and (29). Second, we use state truncation to derive tight bounds for the steady state probabilities. Third, we show that the DES property is satisfied by many models that arise in practice and we obtain explicit solutions for the well-known open problem of the $M/Er/n$ queue with batch arrivals. We note that to our knowledge there are no explicit expressions in the literature for the rate matrix set (and hence for the invariant measure) of this model. This is due to the QSF structure. We also demonstrate the applicability of DES processes to an inventory model with random yield. Finally we discuss the restart hypercube model as an example of a RES process.

Although our methodology applies to infinite values for the number of levels in the process in the downward and/or in the upward direction as well as for the number of phases, we take up explicitly only the case of finite values so that we can employ finite matrices in the analysis. In this way the basic features of the theory will not be obscured by additional formalism. However, we want to emphasize that all the results generalize to infinite values in a natural way using truncation, cf. Section 3.2 herein and Vere-Jones (1967), Tweedie (1973) and Seneta (1980). We will allow the lowest level value to be negative in order to have a natural state description for e.g. inventory models where shortage is allowed. We also note that the results regarding the explicit expressions for the rate matrix set hold for the case that one "boundary-side" (up or down side) of the state space is a transient class of states. However, for simplicity we will not consider this case.

The smaller class of level homogenous QBD (LHQBD) processes cf. Neuts (1981), has been used to model systems in many areas including queueing theory, cf. Riska and Smirni (2002), retrial queues, cf. Artalejo et al. (2010). For algorithmic usages of QBDs we refer to the anti-plagiarism scanner software of Viper. We note that most of the early literature devoted to level homogeneous QBDs typically follows the approach presented in Chapter 6, p. 129, in Latouche and Ramaswami (1999), whereby the computation of the steady state distribution is based on computing a rate matrix "R". This matrix is specified as one of the solutions of the, not easy to solve, matrix quadratic equation: $R^2D + RW + U = 0$, where we use our current (cf. Eq. (1)) notation: D, W, U for the matrices A_2 A_1, A_0 in Latouche and Ramaswami (1999). Numerical methods for computing R involve cyclic reduction Bini and Meini (1996) and logarithmic reduction Pérez and van Houdt (2011), Latouche and Ramaswami (1993). For instance R is expressed in terms of a matrix G (such that $R = U(I - W - UG)^{-1}$) which is the solution of matrix quadratic equation: $UG^2 + WG + D = 0$. For recent work on numerical methods for computing the matrix G for QBD processes of special structure we refer to van Houdt and van Leeuwaarden (2011), Etessami et al. (2010), Bini et al. (2005) and references therein. A general approach to compute G, exists only in special cases when the down transition matrix has the form $c \cdot r$ with c a column vector and r a row vector normalized to one. We note that the DES property is implied by the $c \cdot r$ structure but in our present setting we deal with the much more general class of QSF processes.

In Latouche and Ramaswami (1999), Chapter 13, p. 268-270, a method is given to analyze level homogeneous QSF (LHQSF) processes by considering them as embedded processes in suitably defined QBDs. However, as stated therein, the success of this approach has been limited. We note that the matrix analytic method has been used in Ramaswami (1988) to derive a recursive solution for the $M/G/1$ queue which is a QSF process, using matrix 'G' described above.

In Bright and Taylor (1995) recursive algorithms are given to compute the rate matrices for level dependent QBD processes (LDQBD). We note that when this model satisfies the DES property then our Eq. (8) and Eq. (9) provide a rate matrix set explicitly.

The classes of DES and RES processes have the capacity to model many complex problems, such as Lin et al. (2009), Guillemin et al. (2004), Zwart and Boxma (2000), Gaver et al. (2006) and Baer et al. (2013). Indeed in this paper we obtain explicit solutions for a well-known queueing problem. In addition, RES processes have been used to represent restart systems, cf. Katehakis and Veinott Jr. (1987), Tong et al. (2006) and Sonin (2011), chains that represent inventory systems with random yield or lead times, reliability, cf. Kapodistria (2011), and computer science and the theory of branching processes, cf. Brown et al. (2010).

The rest of the article is organized as follows. In Section 2 we formally define a QSF process and show that the DES property implies a successive lumpable property for a QSF process, cf. Katehakis and Smit (2012). Then, in Theorem 2 we provide explicit expressions for the a rate matrix set, that provides a recursive relation for the steady state probabilities between a level of the state space and its sub-levels. These expressions readily yield the state state probabilities as shown in Theorem 3. In Section 3.2 we show how the state space can be truncated, both in the downwards as in the upward directions. In Section 5 we show how this methodology can be applied to the $M/Er/n$ queue with batch arrivals, and to an inventory model with random yield. We note that the analysis of these models readily extends to the case of the Phase type distributions, we consider the Er distributions for simplicity of the exposition. In the Appendix we give a proof of Theorem 1.

2. Basic Notation for DES and RES Processes

A QSF to the left (or "down") process is a continuous time Markov process $X(t)$ on state space $\mathcal{X}$ that can be expressed as $\mathcal{X} = \bigcup_{m=M_1}^{M_2} \{(m,1),(m,2),\ldots,(m,\ell_m)\}$, where ℓ_m, M_1, M_2, are some fixed finite integers, with $1 \le \ell_m < \infty$, $-\infty < M_1 < M_2 < \infty$, and $M_1 \le m \le M_2$. A state (m,i) specifies its "current" level m and its within the level state i, with $i = 1,\ldots,\ell_m$.

For a QSF process, the transition rate matrix has the form:

$$Q = \begin{bmatrix} W^{M_1} & U^{M_1,M_1+1} & \cdots & U^{M_1,m} & U^{M_1,m+1} & \cdots & U^{M_1,M_2-1} & U^{M_1,M_2} \\ D^{M_1-1} & W^{M_1+1} & \cdots & U^{M_1+1,m} & U^{M_1+1,m+1} & \cdots & U^{M_1+1,M_2-1} & U^{M_1+1,M_2} \\ \vdots & \vdots & \cdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & \cdots & W^m & U^{m,m+1} & \cdots & U^{m-1,M_2-1} & U^{m-1,M_2} \\ 0 & 0 & \cdots & D^{m+1} & W^{m+1} & \cdots & U^{m,M_2-1} & U^{m,M_2} \\ \vdots & \vdots & \cdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & W^{M_2-1} & U^{M_2-1,M_2} \\ 0 & 0 & \cdots & 0 & 0 & \cdots & D^{M_2} & W^{M_2} \end{bmatrix}, \qquad (1)$$

where in the above specification of Q we use the notation D^m, W^m, and $U^{m,k}$ to describe respectively the "down" (to level $m-1$), "within" (level m), and "up" (to level $k = m+1, m+2, \ldots, M_2$) transition rate sub-matrices in relation to the current level m of a state (m,i). The dimensions of these matrices are respectively $\ell_m \times \ell_{m-1}$, $\ell_m \times \ell_m$ and $\ell_m \times \ell_k$; the 0 sub-matrices of Q have position dependent dimensions so that Q is a well defined transition rate matrix.

Note that in the special case that there exist matrices D, W, U^k such that $D^m = D$, $W^m = W$ and $U^{m,k} = U^k$, for all m, the process is called (level) homogeneous. When $U^{m,m+k}$ are all 0 matrices for all $k \ge 2$ and all m, the QSF process reduces to the well studied QBD process, cf. Artalejo and Gómez-Corral (2008).

In the sequel we will assume that the QSF processes discussed are ergodic. Explicit sufficient conditions for the elements of Q can be derived to support this claim, cf. Remark 3; such conditions on ergodicity for multi dimensional Markov Chains can be found in Szpankowski (1988) and various criteria in Tweedie (1981), in Hordijk and Spieksma (1992) and in Spieksma and Tweedie (1994).

We consider Markov process $X(t)$, introduced above. Clearly the state space $\mathcal{X}$ of this process can be partitioned into a (possibly infinite) sequence of mutually exclusive and exhaustive *level sets*: $L_m = \{(m,1),(m,2),\ldots,(m,\ell_m)\}$.

Definition 1. For any fixed m we define the ***sub-level set*** of L_m to be the set of states $\underset{\sim}{L}_m = \cup_{k=M_1}^{m} L_k$ while the set $\widetilde{L}_m = \cup_{k=m}^{M_2} L_k$ is the ***super-level set*** of L_m.

We let $\pi(m,i)$ denote the steady state probability of state (m,i). The vectors $\pi^m := [\pi(m,1),\ldots,\pi(m,\ell_m)]$ and $\underset{\sim}{\pi}^m := [\pi^{M_1},\ldots,\pi^m]$, will denote respectively the steady state probabilities of states in level L_m and sub-level $\underset{\sim}{L}_m$. The vector of the steady state probabilities over all states will be denoted by $\pi := [\pi^{M_1},\ldots,\pi^{M_2}] = \underset{\sim}{\pi}^{M_2}$.

Using the QSF structure of Eq. (1) of the rate matrix Q, the (potentially non-zero) elements of the matrices D^m, W^m, $U^{m,k}$ will be denoted respectively by $d(m-1,j\,|\,m,i)$, (a "down" rate), $w(m,j\,|\,m,i)$ (a "within" rate) and $u(k,j\,|\,m,i)$, (an "up" rate) for $k > m$. Note that the diagonal elements of W^m are the negative sum of all other elements in the m row of Q.

The DES and RES processes we consider in the sequel are are level dependent and ergodic QSF processes with a transition rate matrix Q that satisfies one of the two properties stated in Definition 2. These conditions follow from the structure of the nonzero elements of Q.

Definition 2.

i) The down entrance state (DES) property: one step "down" transitions from a level m can only reach a single state in level $m-1$, for all levels m.

ii) The restart entrance state (RES) property: one step "up" transitions from any level m can only reach a single state in level M_2 (M_2 is the highest level) for all levels m.

3. The Class of DES processes

The following lemmas show that the simple algebraic characterization of the DES property is a sufficient condition for a QSF process to be successively lumpable. Following Katehakis and Smit (2012), a state is an *entrance state* of a subset $\mathcal{X}_0$ of the state space $\mathcal{X}$ if all one step transitions from outside this set $\mathcal{X}_0$ into $\mathcal{X}_0$ can only occur via a transition to the entrance state.

We start with the following Lemma which characterizes an entrance state of a sub-level set L_m of a QSF process in terms of an algebraic property of the "down" transition sub-matrix D^m of its transition rate matrix Q.

LEMMA 1. *For a QSF process $X(t)$ and for a fixed $m \in \{M_1,\ldots,M_2\}$, a state $(m,\boldsymbol{\varepsilon}(L_m)) \in L_m$ is an* **entrance state** *for $\underset{\sim}{L}_m$* if and only if *the following is true for all $(m+1,i) \in L_{m+1}$:*

$$d(m,j\,|\,m+1,i) = 0, \quad \text{if } (m,j) \neq (m,\boldsymbol{\varepsilon}(L_m)). \tag{2}$$

Proof. The structure of the rate matrix Q implies that "down" transitions leaving the set $\widetilde{L}_{m+1} = L_{m+1} \cup L_{m+2} \cup \ldots$ can only come from states in L_{m+1}. Further, by Eq. (2) the latter type of transitions are possible only when they lead into the same state $(m,\boldsymbol{\varepsilon}(L_m)) \in L_m$. □

It is easy to see that Eq. (2) of Lemma 1 is equivalent to the statement that the $\ell_m \times \ell_{m-1}$ matrix D^m has a single nonzero column.

For any fixed $n \in \{M_1,\ldots,M_2\}$, let $\mathcal{D}_n$ denote the partition $\{\underset{\sim}{L}_n, L_{n+1},\ldots,L_{M_2}\}$ of $\mathcal{X}$. For a fixed n, the next lemma establishes that when D^m has a single non-zero column for all $m \geq n+1$, i.e., Q has the DES property, then the QSF process is successively lumpable with respect to the partition $\mathcal{D}_n$.

LEMMA 2. *A QSF process is successively lumpable with respect to a partition $\mathcal{D}_{M_1}$ if D^m contains a single non-zero column vector for all $m = M_1+1,\ldots,M_2$.*

Proof. It is a direct consequence of Lemma 1, that for a QSF process, a sub-level set $\underset{\sim}{L}_m$ has an entrance state $(m,\boldsymbol{\varepsilon}(L_m))$ if D^{m+1} contains a single non-zero column vector. This is true for all $m \geq M_1$. Since $\underset{\sim}{L}_m = \underset{\sim}{L}_{m-1} \cup L_m$ it follows from the definition that the chain is successively lumpable: in the notation of Katehakis and Smit (2012), "D_0" corresponds to $\underset{\sim}{L}_{M_1}$ and for $m > M_1$: "D_m" corresponds to L_{m-M_1}. □

Note that when a QSF process is successively lumpable with respect to a partition $\mathcal{D}_n$ then it is successively lumpable with respect to partition $\mathcal{D}_m$ for all $m > M_1$. A

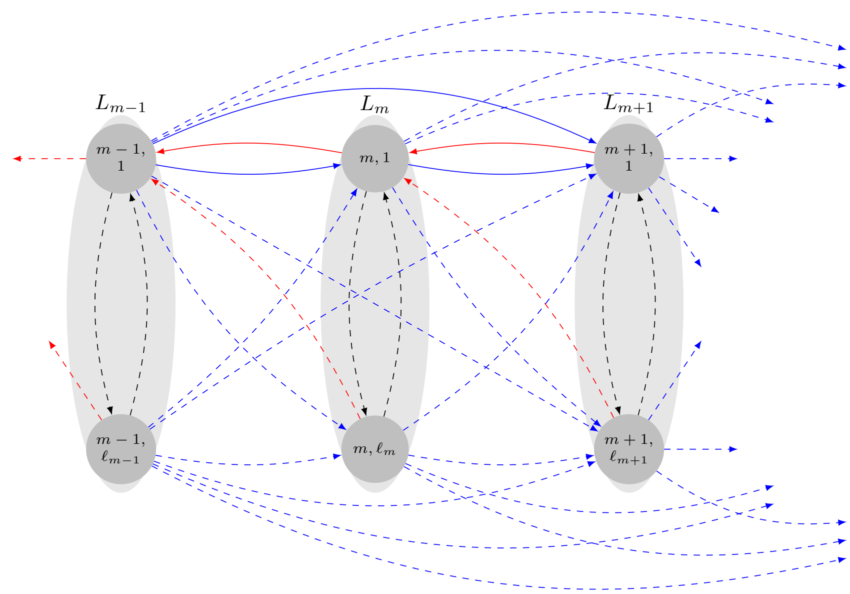

Figure 1 Graphical representation of a Successively Lumpable QSF process

graphical representation of the transitions that are allowed in a successively lumpable QSF process can be found in Figure 3.

We next state the following assumption that will be used in the sequel, where for notational simplicity we let the entrance state of a set $\underset{\sim}{L}_m$ be state $(m,1)$, for all m without loss of generality.

Assumption 1. *The DES process has a transition rate matrix Q with the following properties:*

A1. For all $m \in \{M_1+1,\ldots,M_2\}$, only the first column *of sub-matrix D^m contains non-zero elements, i.e., $d(m-1,1\,|\,m,i) > 0$ for at least one $(m,i) \in L_m$ and all other columns of D^m are equal to zero.*

A2. The QSF process is ergodic (irreducible).

A3. The QSF process has bounded rates.

Remark 1. Part (A3) of Assumption 1 is given to make the proofs applicable for an infinite state space and is not strictly necessary. It can be relaxed without further conditions; it has been added to the assumption to make some of the proofs easier to state.

3.1. Explicit Solutions

We will express the steady state distribution of states in L_m in that of $\underset{\sim}{L}_m$. The presence of entrance states guarantees that sojourn times in $\underset{\sim}{L}_m$ are independent of the rates from states in $\widetilde{L}_m$.

We will use the the following notation. We first define the scalar $\underset{\sim}{\ell}_m := \sum_{k=M_1}^{m} \ell_k$ and use the symbols I_m and $\underset{\sim}{I}_m$ for the identity matrices of dimension $\ell_m \times \ell_m$ and $\underset{\sim}{\ell}_m \times \underset{\sim}{\ell}_m$, respectively. Second we define the (row)vectors of dimension ℓ_m 0_m, 1_m and δ_m to represent a vector identically equal to 0, a vector identically equal to 1, a vector with 1 as its first coordinate and 0 elsewhere respectively. Next we define the (row)vectors of dimension $\underset{\sim}{\ell}_m$: $\underset{\sim}{0}_m$ and $\underset{\sim}{1}_m$ to be the vectors will all its coordinates equal to 0 and 1, respectively. Finally, we define the matrix $\widetilde{U}^{m,n}$ of dimension $\ell_m \times \ell_m$ by:

$$\widetilde{U}^{m,n} = \sum_{k=n}^{M_2} U^{m,k} 1_k' \delta_m, \tag{3}$$

where the elements of $\widetilde{U}^{m,n}$ will be denoted by $\tilde{u}(m,i)$, thus,

$$\tilde{u}(m,i) := \sum_{k=n}^{M_2} \sum_{j=1}^{\ell_k} u(k,j \,|\, m,j).$$

We next define the rate sub-matrices:

$$A_m = \begin{bmatrix} \widetilde{U}^{M_1,m+1} + U^{M_1,m} \\ \vdots \\ \widetilde{U}^{m-1,m+1} + U^{m-1,m} \end{bmatrix}, \tag{4}$$

$$B_m = \widetilde{U}^{m,m+1} + W^m. \tag{5}$$

and

$$\Gamma_m := \begin{bmatrix} A_m \\ B_m \end{bmatrix}.$$

Note that:

i) The matrix A_m contains all rates of Q corresponding to transitions from states in $\underset{\sim}{L}_{m-1}$, *into* states of L_m plus rates corresponding to transitions under Q from states in $\underset{\sim}{L}_{m-1}$ into states of $\widetilde{L}_{m+1}$, which under A_m have been re-directed to transitions into the entrance state $(m,1)$.

ii) The $\ell_m \times \ell_m$ matrix B_m contains all rates of Q corresponding to transitions from states in L_m, *into* states of L_m plus rates corresponding to transitions under Q from states in L_m into states of $\widetilde{L}_{m+1}$, which under B_m have been re-directed to transitions into the entrance state $(m,1)$. Thus, for $m > M_1$ since the construction of B_m excludes all down transitions it a transient transition rate matrix. However, by the definition of $\widetilde{U}^{M_1,M_1+1}$ and by its construction the matrix B_{M_1} is an $\ell_{M_1} \times \ell_{M_1}$ conservative transition rate matrix.

We next state and prove a proposition regarding basic properties of B_m.

Proposition 1. *The following are true:*

i) The matrix B_{M_1} is irreducible.

ii) The matrices B_m are non-singular, for all $m > M_1$.

iii) The inverse of B_m has non-positive elements, for all $m > M_1$.

Proof. To prove i), we will show that every state (M_1, j) in level L_{M_1} that is not the entrance state $(M_1, 1)$ communicates with state $(M_1, 1)$. First, recall that in the construction of B_{M_1}, "up"-transitions are redirected to the entrance state. If there are no "up"-transitions from the communicating class containing (M_1, j), this class would be a closed class in Q which would not contain state $(M_1, 1)$, a contradiction to the irreducibility assumption of Q. So the entrance state is reachable from (M_1, j) under B_{M_1}. Second, we show that (M_1, j) is reachable from state $(M_1, 1)$ under B_{M_1}. Indeed, the only way to reach (M_1, j) from a state in $\widetilde{L}_{M_1+1}$ is via the entrance state, and such a path has to exist by irreducibility of Q. Since the "within" L_{M_1} transition rates under Q are all preserved under B_{M_1}, state (M_1, j) is reachable from state $(M_1, 1)$ under B_{M_1}. Thus every state communicates with the entrance state, and we conclude that B_{M_1} is irreducible.

For ii), we will call a matrix *diagonally dominant* if the absolute value of a diagonal elements are greater or equal than the sum of the absolute values of the off diagonal elements in that row, and strict inequality holds for at least one row. An irreducible diagonally dominant matrix is non singular by the well-known Levy-Desplanques theorem, cf. Varga (1963) (p. 85) or Varga (1976).

The construction of B_m excludes all down transitions and that makes it a diagonally dominant matrix. So when B_m is irreducible, the claim of the lemma is true.

However, in general it is possible that the matrix B_m is not irreducible (even though the matrix Q is irreducible). In this case we will show that the construction of B_m with the irreducibility of Q implies the non-singular property of B_m. Indeed, suppose B_m contains two or more communicating classes of states, say $C_1,\ldots,C_{k_m}$, where C_e contains the entrance state $(m,1)$. We will relabel the states such that states in the same class have adjacent indices and such that if a state in a class C_i has a transition to a state in a class C_j then $i<j$. It is clear that this relabeling is feasible and we can write B_m as:

$$B_m = \begin{bmatrix} Z_{11} & Z_{12} & Z_{13} & \ldots \\ 0 & Z_{22} & Z_{23} & \ldots \\ 0 & 0 & Z_{33} & \ldots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix},$$

where Z_{ii} is a matrix of size $|C_i|\times|C_i|$ containing transition rates within C_i and the matrix Z_{ij} is of size $|C_i|\times|C_j|$ containing transition rates from C_i to C_j (and is possible to have some nonzero elements). We next show that the determinants of Z_{ii} are all non zero. This is sufficient to show that B_m is non-singular, since its determinant is the product of the determinants of the Z_{ii}'s.

First, at least one state in C_e has to have a transition "down" or to another class C_j within B_m, since otherwise C_e would be part of a closed absorbing class under Q. This class does not contain any states in $\underset{\sim}{L}_m$ and this is a contradiction to the irreducibility of Q. Thus Z_{ee} is diagonally dominant and therefore non-singular.

Furthermore, any other class C_j $(j\neq e)$ has to have a state that has a transition rate leaving this class (i.e., down, up or to another class $C_{j'}$ $(j'\neq j)$) otherwise C_j would be a closed class under Q. Therefore, the sum of the off diagonal elements of Z_{jj} in at least one of its row is strictly less than the absolute value of the corresponding diagonal element, since at least one transition leads to a state out of C_j. Thus Z_{jj} is diagonally dominant for all j and the proof of part ii) is complete.

For iii), let τ_i denote the maximum of the absolute values of the diagonal elements of Z_{ii} and let

$$\Gamma_i = \tau_i I + Z_{ii}.$$

Since τ_i is finite and positive and Γ_i is non-negative with row sum less or equal to τ_i, (and strictly less than τ_i for at least one row), the row sum of $\tau_i^{-1}\Gamma_i$ is smaller or equal to one for each row. This implies that $\tau_i^{-1}\Gamma_i$ is a transient transition matrix with spectral radius smaller than 1, and thus all elements of $\sum_{j=0}^{\infty}(\tau_i^{-1}\Gamma_i)^j$ are non-negative and finite, see for example Seneta (1981), Theorem 4.3. Note also that

$$(Z_{ii})^{-1} = (\tau_i(\tau_i^{-1}\Gamma_i - I))^{-1} = \tau_i^{-1}(\tau_i^{-1}\Gamma_i - I)^{-1} = \tau_i^{-1}\sum_{j=0}^{\infty} -(\tau_i^{-1}\Gamma_i)^j.$$

The above implies that all elements of $(Z_{ii})^{-1}$ are non-positive.

By Woodbury's identity, cf. Woodbury (1950) we know:

$$\begin{bmatrix} Z_{11} & Z_{12} \\ 0 & Z_{22} \end{bmatrix}^{-1} = \begin{bmatrix} Z_{11}^{-1} & -Z_{11}^{-1}Z_{12}Z_{22}^{-1} \\ 0 & Z_{22}^{-1} \end{bmatrix},$$

which is a non-positive matrix by the above; an induction argument can establish the same for $(B_m)^{-1}$. □

We next state the following theorem for successively lumpable Markov chains within the context and notation of the DES model. This theorem is a consequence of Theorem 2 of Katehakis and Smit (2012); to avoid the introduction of the notation used in the latter paper, we provide a separate proof in the appendix.

Theorem 1. *Under Assumption 1, the following equality is true for the steady state probabilities* $\underset{\sim}{\pi}^m$ *of* $X(t)$ *for every* m :

$$\underset{\sim}{\pi}^m \Gamma_m = 0_m. \tag{6}$$

We next introduce the idea of a *rate matrix set* for Q as a sequence of matrices $\mathcal{R} = \{\mathcal{R}_m^k\}_{m,k}$ such that $\mathcal{R}_m^k$ satisfy Eq. (7), for all $k = 1, \ldots, m - M_1$ and $m = M_1 + 1, \ldots, M_2$; cf. Latouche and Ramaswami (1999) and references therein.

$$\pi^m = \underset{\sim}{\pi}^{m-k}\mathcal{R}_m^k. \tag{7}$$

Note that there are multiple rate matrix sets, for a given Q. To see this note that for any fixed k, m and known vectors $\pi^m = [\pi(m,1), \ldots, \pi(m,\ell_m)]$ and $\underset{\sim}{\pi}^{m-k} = [\pi^{M_1}, \ldots, \pi^{m-k}]$ Eq. (7) is essentially a system of ℓ_m equations with $\underset{\sim}{\ell}_{m-k} \times \ell_m$ unknowns, the elements of the matrix $\mathcal{R}_m^k$. These equations have many solutions.

In Theorem 2 we show that the specific set $\mathcal{R}_0 := \{R_m^k\}_{m,k}$ obtained recursively using Eq. (8) starting with Eq. (9), is a rate matrix set for Q. For all $m = M_1 + 1, \ldots M_2$ with $k = 2, \ldots m - M_1$ we define:

$$R_m^k := \left[\underset{\sim}{I}_{m-k} \,|\, R^1_{m-(k-1)} \right] R_m^{k-1}, \tag{8}$$

where:

$$R_m^1 := -A_m (B_m)^{-1}. \tag{9}$$

By virtue of Proposition 1ii we know that B_m is non sinular.

THEOREM 2. *The set $\mathcal{R}_0$ defined by Eqs. (8) and (9) above is a rate matrix set for Q.*

Proof. The proof is by induction. For $k=1$ we know by Theorem 1 that

$$\underset{\sim}{\pi}^m \Gamma_m = 0_m.$$

We can rewrite this as:

$$[\underset{\sim}{\pi}^{m-1} | \pi^m] \begin{bmatrix} A_m \\ B_m \end{bmatrix} = 0_m,$$

and thus:

$$\pi^m = -\underset{\sim}{\pi}^{m-1} A_m (B_m)^{-1} = \pi^{m-1} R_m^1.$$

Suppose the statement is true for any m and for $k-1$. We next show that the statement holds for k:

$$\begin{aligned}
\pi^m &= \underset{\sim}{\pi}^{m-(k-1)} R_m^{k-1} \\
&= [\underset{\sim}{\pi}^{m-k} | \pi^{m-(k-1)}] R_m^{k-1} \\
&= [\underset{\sim}{\pi}^{m-k} | \underset{\sim}{\pi}^{m-k} R^1_{m-(k-1)}] R_m^{k-1} \\
&= \underset{\sim}{\pi}^{m-k} \left[\underset{\sim}{I}_{m-k} | R^1_{m-(k-1)} \right] R_m^{k-1} \\
&= \underset{\sim}{\pi}^{m-k} R_m^k.
\end{aligned}$$

Thus the statement is true for $k = 1, \ldots, m - M_1$ and therefore:

$$\pi^m = \underset{\sim}{\pi}^{m-k} R_m^k. \qquad \square$$

Note that the above implies that we can express all vectors π^m in terms of the steady state distribution of level M_1, since M_1 is finite, By the irreducibility assumption all vectors are strictly larger than 0. Therefore we state:

$$\pi^m = \pi^{M_1} R_m^{m-M_1} > 0_m, \tag{10}$$

For any $m_1, m_2 \in \{M_1, \ldots, M_2\}$, with $m_1 < m_2$, we define the column vector $S_{m_1}^{m_2}$ of length ℓ_{m_1} by Eq. (11) below.

$$S_{m_1}^{m_2} = \left[1'_{m_1} + \sum_{m=m_1+1}^{m_2} R_m^{m-m_1} 1'_m \right]. \tag{11}$$

Remark 2.
The elements of R_m^k are non-negative $\forall\ k, m$ where $M_1 + 1 \le m \le M_2$, $1 \le k \le m - M_1$. To check this claim for R_m^1, it suffices to note that $A_m \ge 0$ by definition and $-B_m^{-1} > 0$, by Proposition 1. Alternatively, the (i, j)-th element of R_m^1 can be given an expected first passage time interpretation as is described for QBD processes in Latouche and Ramaswami (1999), chapter 6. The claim for R_m^k, with $k \ge 2$, follows using Eq. (8).

The lemma below establishes the relation between π^{M_1} and $S_{M_1}^{M_2}$.

Lemma 3. *The following relation holds for π^{M_1} and $S_{M_1}^{M_2}$:*

$$\pi^{M_1} S_{M_1}^{M_2} = 1. \tag{12}$$

Proof. Since the chain is ergodic we have $\pi^{M_1} 1'_{M_1} + \sum_{m=M_1+1}^{M_2} \pi^m 1'_m = 1$, thus Eq. (10) implies:

$$\pi^{M_1} \left[1'_{M_1} + \sum_{m=M_1+1}^{M_2} R_m^{m-M_1} 1'_m \right] = 1.$$

Substituting $\left[1'_{M_1} + \sum_{m=M_1+1}^{M_2} R_m^{m-M_1} 1'_m \right]$ by $S_{M_1}^{M_2}$ in the above gives Eq. (12). □

We now state and prove the following theorem.

Theorem 3. *Under Assumption 1, the following is true:*

$$\pi^{M_1} = \delta_{M_1} \left[S_{M_1}^{M_2} \delta_{M_1} - B_{M_1} \right]^{-1}. \tag{13}$$

Proof. Since B_{M_1} is an irreducible rate matrix (Proposition 1 i), it has rank $(\ell_{M_1} - 1)$ by basic linear algebra theory, see for example Seneta (1981).

Furthermore, at the boundary M_1 we have $\Gamma_{M_1} = B_{M_1}$ and $\underset{\sim}{\pi}^{M_1} = \pi^{M_1}$. Thus Theorem 1 implies that Eq. (6) can be written as Eq. (14) below.

$$\pi^{M_1} B_{M_1} = 0_{M_1}. \tag{14}$$

Thus it follows that the vector $S_{M_1}^{M_2}$ is not an element of the linear space spanned by the columns of B_{M_1}, since by Lemma 3 $\pi^{M_1} S_{M_1}^{M_2} = 1$.

The above imply that $-B_{M_1}$ has rank $(\ell_{M_1}-1)$ and when we add to its first column the vector $S_{M_1}^{M_2}$ the resulting matrix $[S_{M_1}^{M_2}\delta_{M_1}-B_{M_1}]$ has full rank and it is invertible.

We use Lemma 3 to state

$$\left[\pi^{M_1}S_{M_1}^{M_2}\right]\delta_{M_1}=\delta_{M_1},$$

and via

$$\pi^{M_1}\left[S_{M_1}^{M_2}\delta_{M_1}-B_{M_1}\right]=\delta_{M_1}-0_{M_1}=\delta_{M_1},$$

we conclude

$$\pi^{M_1}=\delta_{M_1}\left[S_{M_1}^{M_2}\delta_{M_1}-B_{M_1}\right]^{-1}.$$

□

The results above justify the following algorithm to find the steady state distribution of a DES-QSF process.

Algorithm 1 *(DES-QSF)*

- Calculate R_m^1 with Eq. (9) for all $m=M_1+1,\dots,M_2$.
- Compute R_m^k recursively via Eq. (8) for $m=M_1+1\dots,M_2$ and $k=2,\dots,m-M_1$.
- Calculate $S_{M_1}^{M_2}$ via Eq. (11).
- Calculate π^{M_1} via Eq. (13).
- Calculate π^m via Eq. (10) for all $m=M_1+1,\dots,M_2$.

3.2. State Space Truncations

In this section we show how to truncate the state space in the upward direction, in order to obtain upper bounds for the steady state probabilities $\pi(m,i)$ of states in $\underset{\sim}{L}_{m_2}$ where $m_2\in\{M_1,M_1+1,\dots,M_2-1\}$. To this end we first define a process $X_{m_2}(t)$ with truncated state space $\mathcal{X}_{m_2}=\underset{\sim}{L}_{m_2}$ and transition rate matrix:

$$Q_{X_{m_2}}=\begin{bmatrix}\ddots & \vdots & \vdots & \vdots & \vdots \\ \cdots & D^{m_2-2} & W^{m_2-2} & U^{m_2-2,m_2-1} & U^{m_2-2,m_2}+\widetilde{U}^{m_2-2,m_2} \\ \cdots & 0 & D^{m_2-1} & W^{m_2-1} & U^{m_2-1,m_2}+\widetilde{U}^{m_2-1,m_2} \\ \cdots & 0 & 0 & D^{m_2} & W^{m_2}+\widetilde{U}^{m_2,m_2}\end{bmatrix}, \tag{15}$$

where the elements of the last column are given by Eq. (3). We denote the steady state distribution of this process as the row vector $\pi_{X_{m_2}}=[\pi_{X_{m_2}}^{M_1},\dots,\pi_{X_{m_2}}^{m_2}]$ of size:

$\sum_{m=M_1}^{m_2} \ell_m$, where its m^{th} component contains the steady state probabilities for level m of the truncated process.

We next state the following. We emphasize that this proposition clearly holds for $M_2 = \infty$ under the ergodicity assumption.

PROPOSITION 2. *For all finite $m_2 \geq M_1$, and any level $m = M_1, M_1+1, \ldots, m_2$, the following are true:*

i)
$$\pi^m_{X_{m_2}} = \pi^{M_1}_{X_{m_2}} R_m^{m-M_1}, \tag{16}$$
$$\pi^{M_1}_{X_{m_2}} = \delta_{M_1} \left[S^{m_2}_{M_1} \delta_{M_1} - B_{M_1} \right]^{-1}. \tag{17}$$

ii)
$$\pi(m,i) < \pi_{X_{m_2}}(m,i).$$

iii) For all states (m,i), $\pi_{X_\nu}(m,i)$ is a strict decreasing function in $\nu = m_2, m_2+1, \ldots$

Proof. By its construction, the process $X_{m_2}(t)$ is a QSF process which satisfies Assumption $A1$. Further, by its definition the matrix $Q_{X_{m_2}}$ gives rise to the same rate matrices R^k_m as the matrix Q of the original process $X(t)$. This follows from the fact that this specific truncation ensures that the matrices $A_{m,X_{m_2}}$, $B_{m,X_{m_2}}$ of the truncated process corresponding to the matrices A_m B_m of the original process are identical and this proves Eq. (16). The proof of Eq. (17) follows as the proof of Theorem 3, if we replace M_2 with m_2.

For the proof of part ii), using Proposition 1 of Katehakis and Smit (2012), (where $\pi_{X_{m_2}}(m,i) = v_{\underset{\sim}{L}m}(m,i)$) we obtain that Eq. (18) below is valid for all $m \leq \nu$:

$$\pi_{X_\nu}(m,i) = \frac{\pi(m,i)}{\sum_{(k,j)\in \underset{\sim}{L}_\nu} \pi(k,j)} \; for\ all\ \nu = m_2, m_2+1, \ldots. \tag{18}$$

Since $\sum_{(k,j)\in \underset{\sim}{L}_\nu} \pi(k,j) < 1$, for all finite ν, it follows from the above that $\pi(m,i) < \pi_{X_{m_2}}(m,i)$.

For the proof of part iii), note that since $\underset{\sim}{L}_\nu \subset \underset{\sim}{L}_{\nu+1}$, we have that $\sum_{(k,j)\in \underset{\sim}{L}_\nu} \pi(k,j) < \sum_{(k,j)\in \underset{\sim}{L}_{\nu+1}} \pi(k,j)$. Thus, we conclude that $\pi_{X_{\nu+1}}(m,i) < \pi_{X_\nu}(m,i)$. We can repeat this argument for $\underset{\sim}{L}_{\nu+2}, \underset{\sim}{L}_{\nu+3}, \ldots$ and the proof is complete. □

Note that Proposition 2 is closely related the results of Bright and Taylor (1995) (pp. 499-500), derived for LDQBDs. Specifically, Eq. (16)-(17) are the QSF process extensions of Eqs. (1.7)-(1.8) of that paper, with k and m reversed and the change of notation K^*, x_k, R_{k+1}, R_0 in place of our m_2, π^m, R_m R_1.

Note that the matrix $R_m^{m-M_1}$ is finite even when the the QSF process is not ergodic; such a non-ergodic case exists for instance when there is a drift to "up" direction. We can however state the following:

Remark 3. The successively lumpable QSF process is ergodic if $\sum_{m=M_1}^{M_2} R_m^{m-M_1} < \infty$, since then Theorem 3 and Eq. (10) show that there exist positive steady state probabilities for all states. Similarly, it follows that for QSF processes to be ergodic it is sufficient that $S_{M_1}^{M_2} < \infty$.

Remark 4. One can also construct truncations with respect to M_1 or to any ℓ_m separately. This is especially important when some of these constants are infinite. There are various truncations methods possible to truncate the matrix Q of infinite size, some are described in Vere-Jones (1967), Seneta (1980) and Tweedie (1973). Most of these truncations will preserve the successively lumpable property. Such as truncation to $m_1 \geq M_1 = -\infty$ we provide below, following Seneta (1980).

Define a process $X_{m_1}(t)$ with state space $\mathcal{X}_{m_1} = \widetilde{L}_{m_1}$ and transition rate matrix:

$$Q_{X_{m_1}} = \begin{bmatrix} \overline{W}^{m_1} & U^{m_1,m_1+1} & U^{m_1,m_1+2} & U^{m_1,m_1+3} & \dots \\ D^{m_1+1} & W^{m_1+1} & U^{m_1+1,m_1+2} & U^{m_1+1,m+3} & \dots \\ 0 & D^{m_1+2} & W^{m_1+2} & U^{m_1+2,m_1+3} & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}, \tag{19}$$

where $\overline{w}(m_1, j \,|\, m_1, i) = w(m_1, j \,|\, m_1, i)$ if $j \neq 1$ and otherwise we have $\overline{w}(m_1, 1 \,|\, m_1, i) = w(m_1, 1 \,|\, m_1, i) + d(m_1 - 1, 1 \,|\, m_1, i)$. Let $\pi_{X_{m_1}}$ denote the steady state distribution of $X_{m_1}(t)$. For this truncation it can be shown as in Seneta (1980) (Theorem on p. 262) that for all (m, i):

$$\pi(m, i) = \lim_{m_1 \to -\infty} \pi_{X_{m_1}}(m, i).$$

4. The Class of RES processes

In this section we consider QSF processes of the form described in Definition 2ii. Because the current section has the same structure as Section 3.1, most of the theorems and lemmas can be given without proof. We will refer to the corresponding statement in the previous section and add some clarification when necessary.

We start with the following lemmas that show that the simple algebraic characterization of the RES property is a sufficient condition for a QSF process to be successively lumpable.

We start with the following lemma that characterizes an entrance state of the super-level set $\widetilde{L}_m$ of a QSF process in terms of an algebraic property of the "up" transition sub-matrices $U^{m,k}$ of its transition rate matrix Q.

LEMMA 4. *For a QSF process $X(t)$ and for all levels $m \in \{M_1, \ldots, M_2\}$, the state $(M_2, \varepsilon(L_m)) \in L_{M_2}$ is an* **entrance state** *for the set $\widetilde{L}_m$ if the following is true for all states $(n,i) \in \underset{\sim}{L}_{m-1}$:*

$$u(k,j \,|\, n,i) = 0, \quad \text{if } (k,j) \neq (M_2, \varepsilon(L_m)). \tag{20}$$

Proof. Similar to statement made in the proof of Lemma 1 that regards DES processes. □

It is easy to see that Eq. (20) of Lemma 4 is equivalent to the statement that the matrices U^{m,M_2} have a single nonzero column and that $U^{m,k} = 0$ for all $k \in \{m + 1, \ldots, M_2 - 1\}$. This is equivalent to the RES property.

For any fixed $n \in \{M_1, \ldots, M_2\}$, let $\mathcal{D}_n$ denote the partition $\{L_{M_1}, \ldots, L_{n-1}, \widetilde{L}_n\}$ of $\mathcal{X}$. For a fixed n, the next lemma establishes that when Q has the RES property (cf. 2 ii), then the QSF process is successively lumpable with respect to the partition $\mathcal{D}_n$.

LEMMA 5. *A QSF process is successively lumpable with respect to a partition $\mathcal{D}_{M_2}$ if the matrices U^{m,M_2} have a single nonzero column and that $U^{m,k} = 0$ for all $k \in \{m + 1, \ldots, M_2 - 1\}$.*

Proof. Similar to the proof of Lemma 2 in the previous section. □

A graphical representation of the transitions that are allowed in a RES QSF process can be found in Figure 2.

We now state the following assumption that will be used in the sequel of this section, where for notational simplicity we let the entrance state of a set $\widetilde{L}_m$ be state $(M_2, 1)$, for all m without loss of generality.

ASSUMPTION 2. *The RES process under consideration has a transition rate matrix Q with the following properties:*

A1. The process is ergodic (irreducible);

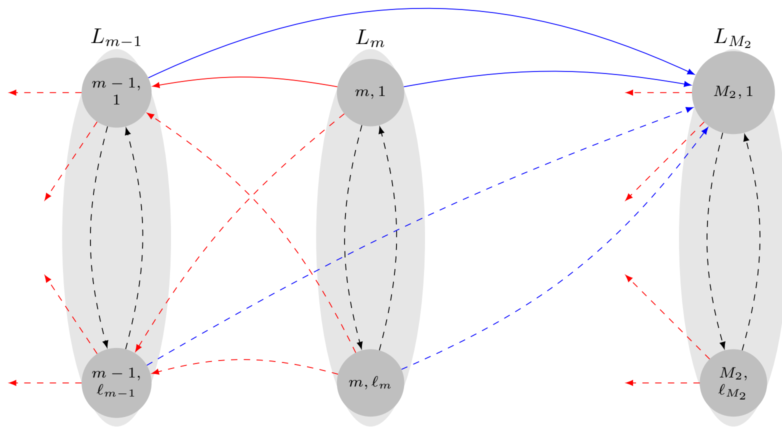

Figure 2 Graphical representation of a RES QSF process

A2. For all $m \in \{M_1, \ldots, M_2 - 1\}$*, only the* first column *of sub-matrix* U^{m,M_2} *can contain non-zero elements.*

A3. The RES process has bounded rates.

Since the process is positive recurrent, there has to be a transition back to the entrance state from a state in $\underset{\sim}{L}_m$, i.e. $u(M_2, 1 \,|\, n, i) > 0$ for at least one $(n, i) \in \underset{\sim}{L}_m$ and all other columns of U^{m,M_2} are equal to zero. In addition, $U^{m,k} = 0$ for all $k \in \{m, \ldots, M_2 - 1\}$.

4.1. Explicit Solutions

We will use the notation introduced in the previous section. In addition we define the matrix $\widetilde{D}^m$ of dimension $\ell_m \times \ell_m$ by:

$$\widetilde{D}^m = D^m 1'_m \delta_m. \tag{21}$$

We next state and prove a proposition regarding basic properties of W_m and $\widetilde{D}^{M_2}$.

Proposition 3. *The following are true:*

i) The matrix $W^{M_2} + \widetilde{D}^{M_2}$ *is irreducible.*

ii) The matrices W^m *are non-singular, for all* $m \in \{M_1, \ldots, M_2\}$*.*

iii) The elements of the inverse of W^m *are non positive, for all* $m < M_2$*.*

Proof. The proves are analogous to that of Proposition 1 in the previous section, that states similar results for a DES process.

We can now state the following theorem for RES process, within the context and notation of the present section.

THEOREM 4. *Under Assumption 2, the following equality is true for the steady state probabilities π^m of $X(t)$ for every $m \in \{M_1, \dots, M_2 - 1\}$:*

$$\pi^m W^m + \pi^{m+1} D^{m+1} = 0_m. \tag{22}$$

$$\pi^{M_2}(W^{M_2} + \widetilde{D}^{M_2}) = 0_{M_2}. \tag{23}$$

Proof. The proof is along the same lines as Theorem 1, since the RES process is successively lumpableas well. We can complete the proof by only considering the states that have possibly positive transitions out of set $\widetilde{L}_{m+1}$. □

For a RES process, we define a *rate matrix set* for Q as a sequence of matrices $\mathcal{R} = \{\mathcal{R}_m\}_m$ such that $\mathcal{R}_m$ satisfy Eq. (24), for all $m = M_1, \dots, M_2 - 1$. Note that this is a slightly different definition than the one introduced for DES processes.

$$\pi^m = \pi^{m+1} \mathcal{R}_m. \tag{24}$$

In Theorem 5 we show that the specific set $\mathcal{R}_0 := \{R_m\}_m$ given by Eq. (25) below, is a rate matrix set for Q. For all $m = M_1, \dots, M_2 - 1$ we define:

$$R_m := -D^{m+1}(W^m)^{-1}. \tag{25}$$

By virtue of Proposition 3ii we know that W^m is non singular.

THEOREM 5. *The set $\mathcal{R}_0$ defined by Eq. (25) above is a rate matrix set for Q.*

Proof. Follows directly from Theorem 4. □

Note that the above implies that we can express all vectors π^m in terms of the steady state distribution of elements in level M_2, since M_2 is finite. By the irreducibility assumption all vectors are strictly larger than 0. Therefore we state:

$$\pi^m = \pi^{M_2} \prod_{k=0}^{M_2-1-m} R_{M_2-1-k} > 0_m, \tag{26}$$

For any $m_1 \in \{M_1, \ldots, M_2\}$, with $m_1 < M_2$, we define the column vector $T_{m_1}^{M_2}$ of length ℓ_{M_2} by Eq. (27) below.

$$T_{m_1}^{M_2} = \left[1'_{M_2} + \sum_{m=m_1}^{M_2-1} \prod_{k=0}^{M_2-1-m} R_{M_2-1-k} 1'_m \right]. \tag{27}$$

Note that R_m is non-negative for all m.

The lemma below establishes the relation between π^{M_2} and $T_{M_1}^{M_2}$.

LEMMA 6. *The following relation holds for π^{M_2} and $T_{M_1}^{M_2}$:*

$$\pi^{M_2} T_{M_1}^{M_2} = 1. \tag{28}$$

Proof. Analogous to Lemma 6. □

We now state and prove the following theorem.

THEOREM 6. *Under Assumption 2, the following is true:*

$$\pi^{M_2} = \delta_{M_2} \left[T_{M_1}^{M_2} \delta_{M_2} - W^{M_2} - \widetilde{D}^{M_2} \right]^{-1}. \tag{29}$$

Proof. Since $W^{M_2} - \widetilde{D}^{M_2}$ is an irreducible rate matrix, Proposition 3 i, it has rank $(\ell_{M_2} - 1)$ by basic linear algebra theory, see for example Seneta (1981). Furthermore, we know that $\pi^{M_2}(W^{M_2} - \widetilde{D}^{M_2}) = 0_{M_2}$ and $\pi^{M_2} T_{M_1}^{M_2} = 1$, thus that the vector $T_{M_1}^{M_2}$ is not an element of the linear space spanned by the columns of $W^{M_2} - \widetilde{D}^{M_2}$. Therefore $[T_{M_1}^{M_2} \delta_{M_1} - W^{M_2} - \widetilde{D}^{M_2}]$ has full rank and is invertible. The remainder of the proof is similar to the proof of Theorem 3. □

The results above justify the following algorithm to find the steady state distribution of a RES QSF process.

Algorithm 2 [RES-QSF]

- Compute R_m via Eq. (25) for $m = M_1 \ldots, M_2 - 1$.
- Calculate $T_{M_1}^{M_2}$ via Eq. (27).
- Calculate π^{M_2} via Eq. (29).
- Calculate π^m via Eq. (26) for all $m = M_1, \ldots, M_2 - 1$.

4.2. State Space Truncations

In this section we show how to truncate the state space of a RES QSF process in the downward direction in order to obtain upper bounds for the steady state probabilities $\pi(m,i)$ of states in $\widetilde{L}_{m_1}$ where $m_1 \in \{M_1+1, M_1+2, \ldots, M_2\}$. To this end we define a process $X_{m_1}(t)$ with truncated state space $\mathcal{X}_{m_1} = \widetilde{L}_{m_1}$ and transition rate matrix:

$$Q_{X_{m_1}} = \begin{bmatrix} W^{m_1} & 0 & \cdots & 0 & 0 & 0 & \widetilde{D}^{m_1} + U^{m_1,M_2} \\ D^{m_1+1} & W^{m_1+1} & \cdots & 0 & 0 & 0 & U^{m_1+1,M_2} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & D^{M_2-2} & W^{M_2-2} & 0 & U^{M_2-2,M_2} \\ 0 & 0 & \cdots & 0 & D^{M_2-1} & W^{M_2-1} & U^{M_2-1,M_2} \\ 0 & 0 & \cdots & 0 & 0 & D^{M_2} & W^{M_2} \end{bmatrix}. \tag{30}$$

We denote the steady state distribution of this process as the row vector $\pi_{X_{m_1}} = [\pi^{m_1}_{X_{m_1}}, \ldots, \pi^{M_2}_{X_{m_1}}]$ of size: $\sum_{m=m_1}^{M_2} \ell_m$, where its m^{th} component contains the steady state probabilities for level m of the truncated process.

We next state the following. We emphasize that this proposition clearly holds for $M_1 = \infty$ under the ergodicity assumption.

Proposition 4. *For all finite $m_1 \leq M_2$, and any level $m = m_1, m_1+1, \ldots, M_2$, the following are true:*

i)
$$\pi^{m}_{X_{m_1}} = \pi^{M_2}_{X_{m_1}} \prod_{k=0}^{M_2-1-m} R_{M_2-1-k} \tag{31}$$

$$\pi^{M_2}_{X_{m_1}} = \delta_{M_2} \left[T^{M_2}_{m_1} \delta_{M_2} - W^{M_2} - \widetilde{D}^{M_2} \right]^{-1}. \tag{32}$$

ii)
$$\pi(m,i) < \pi_{X_{m_1}}(m,i).$$

iii) For all states (m,i), $\pi_{X_\nu}(m,i)$ is a strict decreasing function in $\nu = m_1, m_1-1, \ldots$

Proof. Similar to the proof of Proposition 2: the RES property remains intact, the rate matrices do not change. The entrance state will never be removed from the state space. □

Remark 5. It is not useful to truncate the process in the upward direction. Since we consider a ergodic process with a restart entrance state, returns will go to the highest level. Removing this set would effect the structure of the process too much to give any bounds of interest.

5. Applications

To illustrate the application of the results we provide explicit solutions and approximations to well known open problems of queueing, cf. Adan et al. (2013), and to a stochastic inventory theory problem, cf. Veinott (1965). The queueing models under consideration are the $M/Er/n$ queueing model with batch arrivals and the $Er/M/n$ queueing model. In the two subsequent sections, we take the number of phases to be constant, i.e. $\ell_m = \ell$ for all m, this is done solely for presentation simplicity. The analysis is easy to extend when the number of phases of the corresponding distribution is a function of "m" - the state of the queue, the number of customers in line. The steady state distribution of the $M/Er/n$ is known for the case of Poisson arrivals, as for example discussed in Latouche and Ramaswami (1999). As far as we known this is the first time the steady state distribution of the $M/Er/n$ model with batch arrivals is obtained. We note that the same book gives an exact solution procedure for QBD processes only when M_1 is finite. Below we show that our direct method works for the $Er/M/n$ queueing system, i.e., we provide explicit formulas for the rate matrix set, even when $M_1 = -\infty$. For the inventory model we show that is has the same structure as the $M/Er/n$ and it can be handled similarly.

The construction in the following remark can be used to extend the applicability of the methods described in the previous section to models that are QSF and successively lumpable in the 'upward' direction.

Remark 6. Consider a process with a transition rate matrix Q that has the block form shown in Eq. (33), where its elements are labeled by $(m,i) \in \mathcal{X}$, the states of the underlying process.

$$Q = \begin{bmatrix} \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \iddots \\ \cdots & D^{m-1,m-2} & W^{m-1} & U^{m-1} & 0 & 0 & \cdots \\ \cdots & D^{m,m-2} & D^{m,m-1} & W^{m} & U^{m} & 0 & \cdots \\ \cdots & D^{m+1,m-2} & D^{m+1,m-1} & D^{m+1,m} & W^{m+1} & U^{m+1} & \cdots \\ \iddots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}. \tag{33}$$

Then, we can construct a transition rate matrix $\hat{Q}$ of the form of Eq. (1) by relabeling the states so that a new state $(-m,i)$ corresponds to the original $(m,i) \in \mathcal{X}$ by redefining the down, within and up sub-matrices of $\hat{Q}$ as follows: $\hat{D}^{-m} = U^m$, $\hat{U}^{-m,-k} = D^{m,k}$, and $\hat{W}^{-m} = W^m$. The steady state probabilities of the Q-process can be readily obtained from those of the $\hat{Q}$-process.

5.1. The Classic M/Er/n Model with Batch Arrivals.

In a $M/Er/n$ queueing system with batch arrivals the service of a customer occurs in ℓ phases, each exponentially distributed with parameter μ_i for the i-th phase of the service. For notational simplicity of the exposition we describe in detail the case in which a batch may contain either 1 or 2 customers. Batches with a single customer arrive according to a Poisson process with rate $p\lambda_{m,i}$ when there are m customers in the system and the served customer has gone through the first i phases of services. Similarly, batches of 2 customers arrive with rate $(1-p)\lambda_{m,i}$ with $p \in [0,1]$. The service of a customer has to be completed before another customer can start his first phase.

In order to have state notation that is consistent with that of Section 2, we use the following state description. For $i < \ell$, state (m,i) denotes the event that there are m customers in the waiting line of the system and a customer in service that has gone through i phases of the service. State (m,ℓ) denotes the event that there are m customers in the waiting line and a service completion has just occurred, so that one of the waiting customers is starting service. Note that with this awkward but convenient notation the empty state of the system is state $(0,\ell)$. Then, it is easy to see that this system can be modeled as a QSF process. Its state space is $\mathcal{X} = \{L_0, L_1, \ldots, L_{M_2}\}$ with $L_m = \{(m,1), \ldots, (m,\ell)\}$ and $M_2 \le \infty$. The Q matrix is defined by Eq. (1), with U^m, W^m and D^m (all of size $\ell \times \ell$) as given below.

$$W^0 = \begin{bmatrix} -\lambda_{0,1}-\mu_1 & \mu_1 & 0 & \cdots & 0 \\ 0 & -\lambda_{0,2}-\mu_2 & \mu_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & -\lambda_{0,\ell-1}-\mu_{\ell-1} & \mu_{\ell-1} \\ 0 & \cdots & 0 & 0 & -\lambda_{0,\ell} \end{bmatrix},$$

and for $m \ge 1$:

$$W^m = \begin{bmatrix} -\lambda_{m,1}-\mu_1 & \mu_1 & 0 & \cdots & 0 \\ 0 & -\lambda_{m,2}-\mu_2 & \mu_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & -\lambda_{m,\ell-1}-\mu_{\ell-1} & \mu_{\ell-1} \\ 0 & \cdots & 0 & 0 & -\lambda_{m,\ell}-\mu_\ell \end{bmatrix}, \quad D^m = \begin{bmatrix} 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mu_\ell & 0 & \cdots & 0 & 0 \end{bmatrix}.$$

For $m = 0, 1, \ldots$:

$$U^{m,m+1} = p \begin{bmatrix} \lambda_{m,1} & 0 & 0 & \cdots & 0 \\ 0 & \lambda_{m,2} & 0 & \cdots & 0 \\ 0 & 0 & \lambda_{m,3} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_{m,\ell} \end{bmatrix}, \quad U^{m,m+2} = (1-p) \begin{bmatrix} \lambda_{m,1} & 0 & 0 & \cdots & 0 \\ 0 & \lambda_{m,2} & 0 & \cdots & 0 \\ 0 & 0 & \lambda_{m,3} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_{m,\ell} \\ 0, & & & & \end{bmatrix}.$$

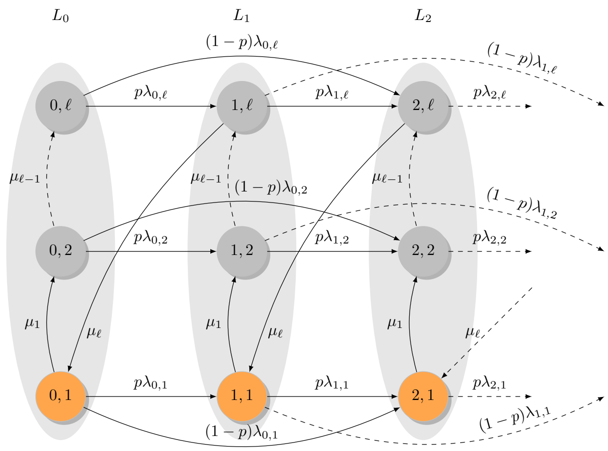

Figure 3 An $M/Er/n$ queueing process with batch arrivals

Note that $(m,1)$ is the entrance state of the set $\utilde{L}_m$, because D_m has a single nonzero column. The matrices A_m and B_m, described in Section 2, are:

$$A_m = \left[\begin{array}{ccccc} \multicolumn{5}{c}{0_{m-3,m}} \\ \hline (1-p)\lambda_{m-2,1} & 0 & 0 & \dots & 0 \\ 0 & (1-p)\lambda_{m-2,2} & 0 & \dots & 0 \\ 0 & 0 & (1-p)\lambda_{m-2,3} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & (1-p)\lambda_{m-2,\ell} \\ \lambda_{m-1,1} & 0 & 0 & \dots & 0 \\ (1-p)\lambda_{m-1,2} & p\lambda_{m-1,2} & 0 & \dots & 0 \\ (1-p)\lambda_{m-1,3} & 0 & p\lambda_{m-1,3} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ (1-p)\lambda_{m-1,\ell} & 0 & 0 & \dots & p\lambda_{m-1,\ell} \end{array}\right]$$

where $0_{m-3,m}$ is a matrix of size $\utilde{\ell}_{m-3} \times \ell_m$ with a 0 at every entry. For $m = 1, 2, \ldots$:

$$B_m = \begin{bmatrix} -\mu_1 & \mu_1 & 0 & \cdots & 0 \\ \lambda_{m,2} & -\lambda_{m,2}-\mu_2 & \mu_2 & \cdots & 0 \\ \lambda_{m,3} & 0 & -\lambda_{m,3}-\mu_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \lambda_{m,\ell} & 0 & 0 & \cdots & -\lambda_{m,\ell}-\mu_\ell \end{bmatrix}$$

and

$$B_0 = \begin{bmatrix} -\mu_1 & \mu_2 & 0 & \cdots & 0 \\ \lambda_{0,2} & -\lambda_{0,2}-\mu_2 & \mu_2 & \cdots & 0 \\ \lambda_{0,3} & 0 & -\lambda_{0,3}-\mu_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \lambda_{0,\ell} & 0 & 0 & \cdots & -\lambda_{0,\ell} \end{bmatrix}.$$

Now we can calculate R^1_m using Eq. (9): $R^1_m = -A_m(B_m)^{-1}$, where $\pi^m = \utilde{\pi}^{m-1} R^1_m$.

Since the first $\underset{\sim}{\ell}_{m-3}$ rows of R^1_m are zero (due to multiplication of $(B_m)^{-1}$ with the $0_{m-3,m}$ sub-matrix of A_m) this expression reduces to the following

$$\pi^m = [\pi^{m-2}|\pi^{m-1}]R^{*1}_m,$$

where R^{*1}_m denotes the nonzero rows of R^1_m.

We can construct the sequence of rate matrices R^k_m using Eq. (8) and the notation R^{*k}_m for the sub-matrix of the nonzero rows of R^k_m we obtain:

$$\pi^m = [\pi^{m-k-1}|\pi^{m-k}]R^{*k}_m$$

and $\pi^m = \pi^0 R^{*m}_m$.

When M_2 is finite (i.e. there is a finite buffer for the number of customers allowed in the system), then Theorem 3 readily provides the solution: $\pi^0 = \delta_0 \left[S_0^{M_2}\delta_0 - B_0\right]^{-1}$, $\pi^m = \pi^0 R^{*m}_m$.

When M_2 is infinite, using Proposition 2, we can construct upper bounds for $\pi(m,i)$ via the process $X_{m_2}(t)$ described therein. This result is stated in the next theorem.

Theorem 7. *The following is true for the $M/Er/n$ model with batch arrivals:*

$$\pi^0_{X_{m_2}} = \delta_0 \left[S_0^{m_2}\delta_0 - B_0\right]^{-1} \tag{34}$$

where $S_0^{m_2} = [1'_0 + \sum_{m=1}^{m_2} R^{*m}_m 1'_m]$ *and*

$$\pi^m \le \pi^m_{X_{m_2}} = \pi^0_{X_{m_2}} R^{*m}_m.$$

Proof. Directly from Theorem 3 (for Eq. (34)) and Section 2 (Proposition 2) for the second claim. □

5.2. An Inventory Model with Random Yield

In this section we consider an inventory model with random yield. Specifically, we investigate a system where customers arrive with rate λ and a batch of products arrives according to an exponential distribution with rate μ. Random yield is possible in this model, i.e. the size of the batch is $n\ell$ with probability p_n, where ℓ is a fixed positive constant, cf. Veinott (1965). We model this inventory model process as a QSF process $X(t)$ on state space $\mathcal{X} = \{L_0, L_1, \ldots\}$ with $L_m = \{(m,1), \ldots, (m,\ell)\}$. In state (m,i)

there are $m\ell + i$ products in stock. Figure 5.2 displays the transition diagram of the described model with $\ell = 3$ and where the size of the batch is 3 with probability p and 6 with probability $1 - p$. This model is a successively lumpable QSF process, where states $(m, \ell - 1)$ are the entrance states of sub-sets $\underset{\sim}{L}_m$. In this QSF process, $U^{m,k}$, W^m and D^m are of size $\ell \times \ell$. It has the same structure as the queueing model described in Section 5.1 and it can be solved analogously.

We note that we can easily obtain explicit formulas for the steady state probabilities even in the case that both λ and μ may depend on the state. For example, when there is too much (little) inventory a discount (premium) price may be used for the product and this may change the arrival rate of the customers, i.e., $\lambda = \lambda(m, i)$. Also, when there is a high level of inventory one may decide not to order. This can be easily incorporated in product arrival rate, i.e., $\mu = \mu(m, i)$. Finally we note that just as easily one can handle the extension where the batch size $n\ell$ is replaced by $n\ell_m$ to represent dependency on the inventory level m. We omitted all these dependencies in this exposition only to simplify the notation.

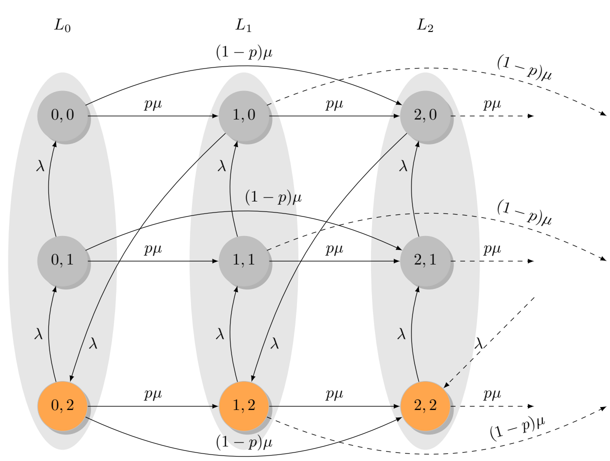

Figure 4 An Inventory Model with random yield.

5.3. Restart Hypercube Models

In this section we illustrate the methods of Section 4.1 using a simplified version of this classical model for the special case of a three dimensional hypercube. The extension to the arbitrary dimension case will become apparent. This type of systems have many applications in diverse fields and have been studied by many authors, cf. the papers we next list and the references therein. Larson (1974), Hooghiemstra and Koole (2000), Katehakis and Melolidakis (1995), Righter (1996), Koole and Spieksma (2001) and Ungureanu et al. (2006).

A basic version of the model is as follows, cf. Derman et al. (1980), Weber (1982), Nash and Weber (1982), Katehakis and Derman (1984), Katehakis (1985), Frostig (1999). A system of known structure is composed of M components and it operates continuously. The time to failure of component $i = 1, \ldots, M$ is exponentially distributed with rate μ_i and it is independent of the state of the other components.

In this section we make the modelling assumption that when the system fails it is restored (or replaced) to a state "as good as new" and the time it takes for this restoration is exponentially distributed with rate λ. It follows under these assumptions that at any point in time the state of the system can be identified by a boolean M-vector $x = (x_1, \ldots, x_M)$, with $x_i = 1$ if the $i-$th component is working, else $x_i = 0$. Hence $\mathcal{X} = \{0,1\}^M$ is the set of all possible states. Under these conditions the time evolution of the state of the system can be described by a continuous time Markov chain. The structure of the system is specified by a binary function ϕ defined on $\mathcal{X}$. Let $G = \{x : \phi(x) = 1\}$ denote the set of all operational (good) states of the system and let $B = \{x : \phi(x) = 0\}$ denote all failed states of the system. For such a system it is important to compute measures of performance such as the availability of the system defined as $\alpha_\phi = \sum_{x \in G} \pi(x)$. Regardless of the choice of the structure ϕ it is easy to see that the corresponding chain is successively lumpable. For example, for the parallel system we have $B = \{(0, \ldots, 0)\}$.

Figure 5 illustrates the transition diagram of the corresponding Markov process for the parallel system when $M = 3$. It is clear that this process is a RES QSF process with respect to partition $\mathcal{D} = \{L_0, L_1, L_2, L_3\}$ of size $M + 1$ and $\forall x \in \mathcal{X}$:

$$x \in L_m \text{ iff } \sum_i x_i = m.$$

It is apparent that in this RES process, $M_1 = 0$ and $M_2 = 3$. The states are ordered as is shown in Figure 5, e.g., $(0,0,1)$ is the first state of level L_1, $(0,1,0)$ is the second state of level $L_1, \ldots$, $(1,1,0)$ is the third state of level L_2, etc. We derive that D^m, W^m

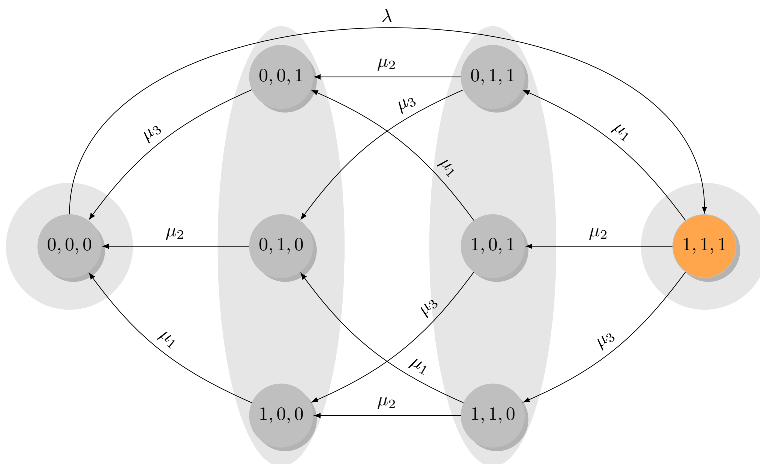

Figure 5 Transition diagram for parallel system with $M = 3$ servers.

and U^{m,M_2} have the from given below:

$$W^0 = -\lambda, \quad U^{0,3} = \lambda,$$

$$D^1 = \begin{bmatrix} \mu_3 \\ \mu_2 \\ \mu_1 \end{bmatrix}, \quad W^1 = \begin{bmatrix} -\mu_3 & 0 & 0 \\ 0 & -\mu_2 & 0 \\ 0 & 0 & -\mu_1 \end{bmatrix}, \quad U^{1,3} = 0'_1,$$

$$D^2 = \begin{bmatrix} \mu_2 & \mu_3 & 0 \\ \mu_1 & 0 & \mu_3 \\ 0 & \mu_1 & \mu_2 \end{bmatrix}, \quad W^2 = \begin{bmatrix} -\mu_2-\mu_3 & 0 & 0 \\ 0 & -\mu_1-\mu_3 & 0 \\ 0 & 0 & -\mu_1-\mu_2 \end{bmatrix}, \quad U^{2,3} = 0'_2,$$

and

$$D^3 = \begin{bmatrix} \mu_1 & \mu_2 & \mu_3 \end{bmatrix}, \quad W^3 = -(\mu_1+\mu_2+\mu_3).$$

The following rate matrices readily follow from Eq. (25), or equivalently the first step of Algorithm 2.

$$R_0 = -D^1(W^0)^{-1} = 1/\lambda \begin{bmatrix} \mu_3 \\ \mu_2 \\ \mu_1 \end{bmatrix}, \quad R_1 = -D^2(W^1)^{-1} = \begin{bmatrix} \mu_2/\mu_3 & \mu_3/\mu_2 & 0 \\ \mu_1/\mu_3 & 0 & \mu_3/\mu_1 \\ 0 & \mu_1/\mu_2 & \mu_2/\mu_1 \end{bmatrix}$$

$$R_2 = -D^3(W^2)^{-1} = \begin{bmatrix} \mu_1/(\mu_2+\mu_3) & \mu_2/(\mu_1+\mu_3) & \mu_3/(\mu_1+\mu_2) \end{bmatrix}.$$

Thus, we can find the steady state distribution using Eqs. (27), (29), (26) or equivalently the remaining steps of Algorithm 2. Note that in this case T_0^3 is a scalar and $\delta_3 = 1$.

$T_0^3 = 1 + \sum_{m=0}^{2} \prod_{k=0}^{2-m} R_{2-k} 1'_m = 1 + R_2 R_1 R_0 1'_0 + R_2 R_1 1'_1 + R_2 1'_2,$

$$\pi^3 = \pi(1,1,1) = \delta_3[T_0^3\delta_3 - W^3 - \widetilde{D}^3]^{-1} = [T_0^3]^{-1},$$
$$\pi^2 = \pi^3 R_2,\ \pi^1 = \pi^2 R_1,\ \pi^0 = \pi^1 R_0.$$

It is important to note the above approach that exploits the RES property of the process results in the following computational gains: instead of solving a system of size 2^M we only need to solve M systems the largest of which is of size $\binom{M}{\lfloor M/2 \rfloor} + 1$.

Appendix. Proof of Theorem 1

We start this appendix with an observation.

REMARK 7. We let $\Delta_0 = \underset{\sim}{L}_m = \{(M_1,1),\ldots,(M_1,\ell_{M_1}),\ldots,(m,1),\ldots,(m,\ell_m)\}$, where m is any fixed integer $m = M_1,\ldots,M_2$. The lumped process on Δ_0 has a rate matrix "U_{Δ_0}" (defined in Katehakis and Smit (2012)) of size $\underset{\sim}{\ell}_m \times \underset{\sim}{\ell}_m$ that can be written as:

$$[\Lambda_m \,|\, \Gamma_m]$$

where Λ_m contains the rates of transitions into states of the set $\underset{\sim}{L}_{m-1}$ (i.e., it is a matrix of dimension $\underset{\sim}{\ell}_m \times \underset{\sim}{\ell}_{m-1}$ and the construction of the $\underset{\sim}{\ell}_m \times \ell_m$, matrix Γ_m is done above following Katehakis and Smit (2012)). Note that we do not need to explicitly define the elements of the matrices Λ_m as they are not explicitly used in the sequel.

For the proof of Theorem 1, consider the partition $\mathcal{D} = \{\underset{\sim}{L}_m, L_{m+1},\ldots,L_{M_2}\}$ of the state space $\mathcal{X}$ of the chain, for any fixed m. We note that the sets Δ_m of Katehakis and Smit (2012) within the present context, are given by: $\Delta_k = \{(k\,0)\} \cup L_k$; where $(k\,0)$ represents the "lumped state".

By Lemma 2 we know that $X(t)$ is successively lumpable with respect to $\mathcal{D}$. Let $v_{\Delta_0} = v_{\underset{\sim}{L}_m}$ denote the steady state probability vector of the lumped process on $\Delta_0 = \underset{\sim}{L}_m$, cf. Remark 7. By Proposition 1 of Katehakis and Smit (2012) (with $(k,i) \in \underset{\sim}{L}_m$ in place of $(0,i) \in \Delta_0$) we know that for all $k \le m$:

$$\pi(k,i) = \sum_{(k',j)\in \underset{\sim}{L}_m} \pi(k'\,j) v_{\underset{\sim}{L}_m}(k,i). \tag{35}$$

Further, since $v_{\underset{\sim}{L}_m}$ is a steady state probability vector of the lumped process on $\underset{\sim}{L}_m$(="Δ_0"), it is the normalized to 1 solution of the equation below (see Remark 7):

$$v_{\underset{\sim}{L}_m}[\Lambda_{m-1} \,|\, \Gamma_m] = \underset{\sim}{0}_m. \tag{36}$$

If we let $c = \sum_{(k' j) \in \underset{\sim}{L}_m} \pi(k', j) \geq 0$, then from Eq. (35) we know $\underset{\sim}{\pi}^m = c\, v_{\underset{\sim}{L}_m}$. Eq. (37) below then follows by multiplying both sides of Eq. (36) by c:

$$\underset{\sim}{\pi}^m [\Lambda_{m-1} \,|\, \Gamma_m] = [\,\underset{\sim}{\pi}^m \Lambda_{m-1} \,|\, \underset{\sim}{\pi}^m \Gamma_m] = \underset{\sim}{0}_m = [\underset{\sim}{0}_{m-1} \,|\, 0_m], \tag{37}$$

and the proof of Theorem 1 is complete. □

Acknowledgments

We would like to express our gratitude to Professor Jerome D. Williams, PhD Program Director of Rutgers Business School, for his support of Laurens C. Smit's studies at Rutgers. This Research has been partially supported by the Rutgers University RBS - Research Resources Committee.

References

Adan, I.J.B.F., S. Kapodistria, J.S.H. van Leeuwaarden. 2013. Erlang arrivals joining the shorter queue. *Queueing Systems* **74**(2-3) 273–302.

Artalejo, J.R., A. Economou, M.J. Lopez-Herrero. 2010. The maximum number of infected individuals in SIS epidemic models: Computational techniques and quasi-stationary distributions. *Journal of computational and applied mathematics* **233**(10) 2563–2574.

Artalejo, J.R., A. Gómez-Corral. 2008. *Retrial queueing systems: a computational approach*. Springer Verlag.

Baer, N., R.J. Boucherie, J.C.W. van Ommeren. 2013. The PH/PH/1 multi-threshold queue. Department of Applied Mathematics, University of Twente, Enschede, The Netherlands.

Bini, D., G. Latouche, B. Meini. 2005. *Numerical methods for structured Markov chains*. Oxford University Press.

Bini, D., B. Meini. 1996. On the solution of a nonlinear matrix equation arising in queueing problems. *SIAM Journal on Matrix Analysis and Applications* **17**(4) 906–926.

Bright, L., P.G. Taylor. 1995. Calculating the equilibrium distribution in level dependent quasi-birth-and-death processes. *Stochastic Models* **11**(3) 497–525.

Brown, M., E.A. Peköz, S.M. Ross. 2010. Some results for skip-free random walk. *Probability in the Engineering and Informational Sciences* **24**(04) 491–507.

Derman, C., G.J. Lieberman, S.M. Ross. 1980. On the optimal assignment of servers and a repairman. *Journal of Applied Probability* 577–581.

Etessami, K., D. Wojtczak, M. Yannakakis. 2010. Quasi-birth-death processes, tree-like qbds, probabilistic 1-counter automata, and pushdown systems. *Performance Evaluation* **67**(9) 837–857.

Frostig, E. 1999. Jointly optimal allocation of a repairman and optimal control of service rate for machine repairman problem. *European Journal of Operational Research* **116**(2) 274–280.

Gaver, D.P., P.A. Jacobs, G. Samorodnitsky, K.D. Glazebrook. 2006. Modeling and analysis of uncertain time-critical tasking problems. *Naval Research Logistics (NRL)* **53**(6) 588–599.

Guillemin, F., P. Robert, A.P. Zwart. 2004. Aimd algorithms and exponential functionals. *The Annals of Applied Probability* **14**(1) 90–117.

Hooghiemstra, G., G. Koole. 2000. On the convergence of the power series algorithm. *Performance Evaluation* **42**(1) 21–39.

Hordijk, A., F. Spieksma. 1992. On ergodicity and recurrence properties of a Markov chain with an application to an open Jackson network. *Advances in Applied Probability* **24**(2) 343–376.

Kapodistria, S. 2011. The M/M/1 queue with synchronized abandonments. *Queueing Systems* **68**(1) 79–109.

Katehakis, M.N. 1985. A note on the hypercube model. *Operations Research Letters* **3**(6) 319–322.

Katehakis, M.N., C. Derman. 1984. Optimal repair allocation in a series system. *Mathematics of Operations Research* **9**(4) 615–623.

Katehakis, M.N., A.F. Veinott Jr. 1987. The multi-armed bandit problem: decomposition and computation. *Mathematics of Operations Research* **12**(2) 262–268.

Katehakis, M.N., C. Melolidakis. 1995. On the optimal maintenance of systems and control of arrivals in queues. *Stochastic Analysis and Applications* **13**(2) 137–164.

Katehakis, M.N., L.C. Smit. 2012. A successive lumping procedure for a class of Markov chains. *Probability in the Engineering and Informational Sciences* **26**(4) 483–508.

Koole, G.M., F.M. Spieksma. 2001. On deviation matrices for birth–death processes. *Probability in the Engineering and Informational Sciences* **15**(2) 239–258.

Larson, R.C. 1974. A hypercube queuing model for facility location and redistricting in urban emergency services. *Computers & Operations Research* **1**(1) 67–95.

Latouche, G, V. Ramaswami. 1993. A logarithmic reduction algorithm for quasi-birth-death processes. *Journal of Applied Probability* 650–674.

Latouche, G., V. Ramaswami. 1999. *Introduction to Matrix Analytic Methods in Stochastic Modeling*, vol. 5. ASA-SIAM Series on Statistics and Applied Probability. SIAM, Philadelphia, PA.

Lin, K., M. Kress, R. Szechtman. 2009. Scheduling policies for an antiterrorist surveillance system. *Naval Research Logistics* **56**(2) 113–126.

Nash, P., R.R. Weber. 1982. Dominant strategies in stochastic allocation and scheduling problems. *Deterministic and Stochastic Scheduling*. Springer, 343–353.

Neuts, M.F. 1981. *Matrix-Geometric Solutions in Stochastic Models - An Algorithmic Approach*. Dover Publications, Inc., New York.

Pérez, J.F., B. van Houdt. 2011. Quasi-birth-and-death processes with restricted transitions and its applications. *Performance Evaluation* **68**(2) 126–141.

Ramaswami, V. 1988. A stable recursion for the steady state vector in Markov chains of M/G/1 type. *Communications in Statistics. Stochastic Models* **4** 183–188.

Righter, R. 1996. Optimal policies for scheduling repairs and allocating heterogeneous servers. *Journal of applied probability* 436–547.

Riska, A., E. Smirni. 2002. M/G/1-type Markov processes: A tutorial. *Performance Evaluation of Complex Systems: Techniques and Tools* **2459** 36–63.

Seneta, E. 1980. Computing the stationary distribution for infinite Markov chains. *Linear Algebra and Its Applications* **34** 259–267.

Seneta, E. 1981. *Non-negative matrices and Markov chains*. Springer, New York.

Sonin, I.M. 2011. Optimal stopping of Markov chains and three abstract optimization problems. *An International Journal of Probability and Stochastic Processes* **83**(4-6) 405–414.

Spieksma, F.M., R.L. Tweedie. 1994. Strengthening ergodicity to geometric ergodicity for Markov chains. *Stochastic Models* **10**(1) 45–74.

Szpankowski, W. 1988. Stability conditions for multidimensional queueing systems with computer applications. *Operations Research* **36**(6) 944–957.

Tong, H., C. Faloutsos, J.Y. Pan. 2006. Fast random walk with restart and its applications. *Data Mining, ICDM '06: Sixth International Conference on Data Mining* 613–622.

Tweedie, R.L. 1973. The calculation of limit probabilities for denumerable Markov processes from infinitesimal properties. *Journal of Applied Probability* 84–99.

Tweedie, R.L. 1981. Criteria for ergodicity, exponential ergodicity and strong ergodicity of Markov processes. *Journal for Applied Probability* **18**(1) 122–130.

Ungureanu, V., B. Melamed, M.N. Katehakis, P.G. Bradford. 2006. Deferred assignment scheduling in cluster-based servers. *Cluster Computing* **9**(2) 57–65.

van Houdt, B., J.S.H. van Leeuwaarden. 2011. Triangular M/G/1-type and tree-like quasi-birth-death Markov chains. *INFORMS Journal on Computing* **23**(1) 165–171.

Varga, R.S. 1963. *Matrix iterative analysis*. Prentice-Hall, Englewood Cliffs, New Jersey.

Varga, R.S. 1976. On recurring theorems on diagonal dominance. *Linear Algebra and its Applications* **13**(1) 1–9.

Veinott, A.F. 1965. The optimal inventory policy for batch ordering. *Operations Research* **13** 424–432.

Vere-Jones, D. 1967. Ergodic properties of nonnegative matrices I. *Pacific Journal of Mathematics* **22**(2) 361–386.

Weber, R.R. 1982. Scheduling jobs with stochastic processing requirements on parallel machines to minimize makespan or flowtime. *Journal of Applied Probability* **19**(1) 167–182.

Woodbury, M.A. 1950. Inverting modified matrices. *Memorandum Report, Statistical Research Group* **42**.

Zwart, A.P., O.J. Boxma. 2000. Sojourn time asymptotics in the M/G/1 processor sharing queue. *Queueing systems* **35**(1-4) 141–166.