Level product form QSF processes and an analysis of queues with Coxian inter-arrival distribution

D. Ertiningsih(1)(2) M.N. Katehakis(3) L.C. Smit(1)(3) F.M. Spieksma(1)
Leiden University, The Netherlands (1) Universitas Gadjah Mada, Indonesia (2)
Rutgers Business School, USA (3)

July 24, 2015

Abstract

In this paper we study a class of Quasi-Skipfree (QSF) processes where the transition rate submatrices in the skipfree direction have a column times row structure. Under homogeneity and irreducibility assumptions we show that the stationary distributions of these processes have a product form as a function of the level. For an application, we will discuss the $Cox(k)/M^Y/1$-queue, that can be modelled as a QSF process on a two-dimensional state space. In addition we study the properties of the stationary distribution and derive monotonicity of the mean number of the customers in the queue, their mean sojourn time and the variance as a function of $k$ for fixed mean arrival rate.

Keywords: QSF process, successively lumpable, Coxian inter-arrival times.

AMS Subject Classification

1 Introduction

A Quasi-Skipfree (QSF) process is a continuous time Markov process $X = \{X_t\}_{t \geq 0}$, on a two-dimensional state space $S = \{(m, i) | m \in \mathbb{Z}, i \in \{0, \ldots, \ell_m\}\}$, where $m$ denotes the ‘level’ of the state and $i$ denotes the ‘phase’ within the level. Additionally, the jump rates are not allowed to cross more than one level in one direction, i.e. either the downward direction or upward direction. This framework is the natural extension of the embedded GI/M/1 and M/G/1-queues, and it has interesting structural properties in common with these processes.

Neuts [12] investigates the embedded GI/M/1-queue as a skip-free to the right process. The matrix-geometric method for computing the stationary distribution of homogeneous Quasi-Birth-Death (QBD) processes has been applied in [12] and [11]. A homogeneous QBD process is a special case of homogeneous QSF processes, where the transition probabilities are not allowed to cross more than one level in both directions. In his book, Neuts discusses some examples of homogeneous QBD processes, e.g., the $M/PH/1$-queue and the $PH/M/c$-queue. He shows that the stationary distribution of these queues can be expressed in terms of a rate matrix.

Recently, Katehakis and Smit [8] have introduced a new procedure to compute the invariant measure for a class of Markov chains. This procedure is called the successive lumping method. Further, an explicit solution and bounds for the steady state probabilities for the class of ergodic QSF processes
that possess the successive lumpability property have been derived in the paper of Katehakis, Smit, and Spieksma \[10\]; an analysis of the applicability requirements and numerical complexity of the successive lumping method is given in \[9\]. Ramaswami and Latouche \[13\] discuss QBD processes with a special structure, namely where the upward or the downward transitions rates form a row times column matrix. In the latter case the rate matrix can be computed explicitly. In the former case they show that the stationary distribution has a level product form, as is the case with the embedded $GI/M/1$-queue. Regarding rate matrix analysis, Ramaswami and Lucantoni, \[14\], exploit the structural properties of the $G|PH|1$-queue, and extend the numerical feasibility of the matrix geometric approach to a wide set of problems by developing efficient schemes to compute the rate matrix.

One of the main assumptions of \[13\] is that the transition rates are bounded as a function of the states. In the particular case of a Quasi Birth and Death (QBD) process, where the upward transition rates form a matrix with one non-zero row, the rate matrix cannot be computed explicitly. However, we will show how the stationary distribution of a higher level can be explicitly expressed in terms of the stationary distribution of lower levels, under an additional invertibility condition. Moreover, we will show that for this specific type of non-homogenous QSF processes it is possible to derive a level product form solution.

We further study a phase-type inter-arrival, batch service queue, denoted $PH/M^{r}/1$-queue, as a particular example of a QSF process, where the upward transition rates form a matrix with one non-zero row. We will consider a special case of a phase type distribution, a Coxian distribution (cf. \[6\]). Herein customers go through a maximum of $k$ exponentially distributed phases, and after each phase the customer can enter the system. We will denote a Coxian distribution with $k$ phases by $Cox(k)$. Since the Coxian distributions are dense in the set of nonnegative distribution functions, we can approximate those by a Coxian distribution by using for example an expectation-maximization algorithm (see \[4\]). Therefore the results presented in this paper can be useful for various queues with different inter-arrival distributions.

We show that the stationary distribution of the phases within a level has a product form as well, if the phase-type inter-arrival distribution is a Coxian distribution. In \[10\] the successive lumpable structure is specified for QSF processes, but we did not investigate the level product form that is derived in this paper for QSF problems with an unbounded number of levels to the left. The parameter of the product form is the solution to a fixpoint equation, that can be numerically approximated efficiently.

We conjecture that allowing for $c$ servers will yield a product form solution based on $c$ factors, similarly to the results presented by Adan et al \[1\] Theorem 4.1. Actually, we can also handle the $Cox(k)/E_{l}/1$ by taking the expected workload instead of the number of customers as the batch service distribution.

This paper is organised as follows. In Section 2 we show that the stationary distribution of the levels has a product form when choosing the downward matrix $D$ as a multiplication of a column vector $c$ and a row vector $r$. In Section 3 we assume that there exists an exit state per level. This is equivalent to the existence of an entrance state per level for a revised level partition of the state space that keeps the QSF property intact. This means that the process is successively lumpable with respect to the new partition. This is used to derive an expression for the rate matrix with respect to the original partition in Section 3.1. This derivation is justified by the invertibility of the generator of a transient Markov chain in the Appendix. Unfortunately, no explicit expression for the rate matrix can be derived. However, we can explicitly determine a matrix that expresses the stationary distribution of a lower level in terms of the stationary distribution of the higher level.
In Section 4 we analyze the non-homogenous $\text{Cox}(k)/M^Y/1$-queue, and find the parameter of the product form, as well as ergodicity properties. In the remainder of this section we specialize this queue to a queue with homogenous rates and an infinite number of phases. In Section 3 we derive monotonicity properties of the stationary mean number of customers in the queue and their mean sojourn time for fixed mean inter-arrival times. To conclude, we show that the stationary distribution of the $E_k/M/1$-queue (a special case of the $\text{Cox}(k)/M^Y/1$ where the number of phases is fixed and the batch size of service is 1) converges monotonically to the stationary distribution of the $D/M/1$-queue.

2 The model and basic properties

In this section we introduce the notation and derive two initial properties for the stationary distribution of homogeneous QSF processes in Lemma 2.1 and Theorem 2.3. Note that by the homogeneity assumption the number of phases per level of the QSF process is constant, say equal to $\ell + 1$, for $\ell \geq 0$. Without loss of generality, we may assume the QSF process to be skip-free to the left. We will also assume that the levels are bounded to the right, since otherwise the stationary distribution cannot exist, as is shown later in this section. Without loss of generality we may then assume that they are non-negative, i.e. $m \leq 0$. Then the infinitesimal generator $Q$ or $q$-matrix (cf. 2 and 10) takes the form:

$$Q = \begin{pmatrix}
\ddots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\ldots & D & W & U^1 & U^2 & U^3 & U^4 \\
\ldots & 0 & D & W & U^1 & U^2 & U^3' \\
\ldots & 0 & 0 & D & W & U^1 & U^2' \\
\ldots & 0 & 0 & 0 & D & W & U^1' \\
\ldots & 0 & 0 & 0 & 0 & D & W' \\
\end{pmatrix}, \quad (2.1)$$

where the $(\ell + 1) \times (\ell + 1)$ sub-matrices $U^s$ $(s = 1, 2, \ldots)$, $W$, and $D$ represent the transition rates to the $s$-th higher level, the same level, and the next lower level respectively. Elements of these matrices are $u^s_{ij}$, $w_{ij}$, and $d_{ij}$ where $u^s_{ij}$ is the $((m, i), (m + s, j))$ element from the matrix $U^s$, $w_{ij}$ is the $((m, i), (m, j))$ element from the matrix $W$, and $d_{ij}$ is the $((m, i), (m - 1, j))$ element from the matrix $D$, for any $i, j \in \{0, \ldots, \ell\}$. The matrices $U^k'$ are such that the row sum of $Q$ is zero. For example these matrices can be $U^k = \sum_{i=k}^{\infty} U^i$ and $W' = W + \sum_{i=1}^{\infty} U^i$.

Throughout the paper we assume that the $q$-matrix $Q$ is irreducible, conservative, stable, and non-explosive. Additionally, we assume that the jump rates are allowed to be unbounded as a function of the state and that $X$ is the minimal process. It is convenient to denote the levels by $L_m$, so that $L_m = \{(m, i) | i \in \{0, \ldots, \ell\}\}$ and $S = \cup_m L_m$, where $-\infty \leq m \leq 0$. In view of the structure of $Q$ given in Eq. (2.1), jumps can only take place to levels $L_k$ for $k \geq m - 1$, given that the current level is $L_m$.

Since the levels are mutually exclusive, they form a partition of the state space. We will introduce some more notation. In accordance with [10], for any fixed $m$ we define the sub-level set of $L_m$ to be the set of states $L_m = \cup_{k \leq m} L_k$, while the set $L_m = \cup_{k \geq m} L_k$ is the super-level set of $L_m$. Then clearly $\{L_{m-1}, L_m\}$ is a partition of $S$ for each $m$.

Suppose that the QSF process is positive recurrent. Let $\pi$ denote the (unique) stationary distribution. We denote the $m$-level sub-vector of $\pi$ by $\pi_m$. The stationary distributions corresponding to $L_{m-1}$ and $L_m$ will be denoted by $\pi_{m-1}$ and $\pi_m$ respectively. Then by a standard taboo decomposition, as
discussed in [10], we can express \( \tilde{\pi}_{m-1} \) in terms of \( \tilde{\pi}_m \) and the expected amount of time spent in the states of \( L_{m-1} \).

We denote by \( m\tau(k,i)(l,j) \) the expected amount of time spent in state \((l,j)\) without passing through states in the super-level set \( L_m \), given that the system starts in state \((k,i)\), where \((k,i),(l,j) \in L_{m-1}\).

Further, denote by \( m\mathcal{T} \) the \( |L_{m-1}| \times |L_{m-1}| \) matrix with elements \( m\tau(k,i)(l,j) \). Write:

\[
D = \begin{bmatrix}
0 & \ldots & 0 & 0 \\
\vdots & \ddots & \vdots & \vdots \\
0 & \ldots & 0 & 0 \\
0 & \ldots & 0 & D
\end{bmatrix},
\]

where 0 stands for an \((\ell + 1) \times (\ell + 1)\) zero matrix and \( D \) has dimension \( |L_m| \times |L_{m-1}| \). Then by the skip free property to the left

\[
\pi_{m-1} = \tilde{\pi}_m D m\mathcal{T}.
\]

We will next express \( m\mathcal{T} \) in terms of \( Q \). The elements of the \( q \)-matrix \( Q \) are denoted by \( q(k,i)(l,j) \), where \((k,i),(l,j) \in S\). Denote \( q(k,i) = -q(k,i)(k,i) \), the parameter of the exponential sojourn time in state \((k,i)\). Further denote by \( mQ \) the taboo-generator with taboo set \( L_m \), restricted to the states of \( L_{m-1} \), with elements

\[
mq(k,i)(l,j) = \begin{cases} q(k,i)(l,j), & (k,i),(l,j) \in L_{m-1}, \\ 0, & \text{otherwise,} \end{cases}
\]

and \( mq(k,i) = -mq(k,i)(k,i) \). The taboo is therefore imposed at the time of the first jump out of a state of the QSF process on \( L_{m-1} \). By assumed irreducibility, \( mQ \) is the \( q \)-matrix of a transient, non-conservative and minimal Markov process on the state space \( L_{m-1} \). Since the number of levels is unbounded to the left, \( mQ \) is of infinite dimension.

Denoting the corresponding (minimal) transition function by \( mP_t = mP_t(k,i)(l,j) \) where all elements \((k,i),(l,j) \in L_{m-1}\). It follows that

\[
m\tau(k,i)(l,j) = \int_0^\infty mP_t(k,i)(l,j) dt < \infty.
\]

We can state the following lemma.

**Lemma 2.1.** The inverse matrix \( mQ^{-1} \) of \( mQ \) exists and \( mQ^{-1} = -m\mathcal{T} \). Further,

\[
m\mathcal{T} mQ = mQ m\mathcal{T} = -I_m,
\]

where \( I_m \) is the identity matrix on \( L_{m-1} \).

**Proof.** By virtue of Anderson [2] (Theorem 2.2.2) the minimal transition function \( mP_t \) satisfies the Kolmogorov forward and backward equations and it is the unique solution on \( S \).

Define a scalar \( a > 0 \). Then \( mR(a) = \int_0^\infty e^{-at} mP_t dt \) is the associated resolvent with elements \( m\tau(k,i)(l,j)(a) \), where \((k,i),(l,j) \in L_{m-1}\). By virtue of Anderson [2], Propositions (2.1.1) and (2.1.2), it holds that

\[
a mR(a) = I_m + mQ mR(a) = I_m + mR(a) mQ.
\]
Hence, by virtue of the backward equation, for \((k, i), (l, j) \in \tilde{L}_{m-1}\) it holds that
\[
a_{m^{\mathcal{F}}(k,i)(l,j)}(a) = \delta_{(k,i)(l,j)} + m q_{(k,i)(k,i)} m^{\mathcal{F}}_{(k,i)(l,j)}(a) + \sum_{(k',i') \neq (k,i)} m q_{(k,i)(k',i')} m^{\mathcal{F}}_{(k',i')(l,j)}(a), \tag{2.5}
\]
where Kronecker delta \(\delta_{(k,i)(l,j)} = \begin{cases} 1, & (k, i) = (l, j) \\ 0, & \text{otherwise} \end{cases}\).

Observe that \(m^{\mathcal{F}}_{(k,i)(l,j)}(a) \uparrow m^{\mathcal{F}}_{(k,i)(l,j)}\) for \(a \downarrow 0\). Using that the summation in Eq. (2.5) contains only non-negative terms, we can now take the limit \(a \downarrow 0\) in Eq. (2.5) and use the monotone convergence theorem to obtain
\[
\delta_{(k,i)(l,j)} + m q_{(k,i)(k,i)} m^{\mathcal{F}}_{(k,i)(l,j)} + \sum_{(k',i') \neq (k,i)} m q_{(k,i)(k',i')} m^{\mathcal{F}}_{(k',i')(l,j)} = 0. \tag{2.6}
\]
Eq. (2.6) is equivalent to
\[
l_{m,(k,i)(l,j)} + (m Q m^T)_{(k,i)(l,j)} = 0.
\]
This yields that \(m Q m^T = -I_m\).

The relation \(m^T m Q = -I_m\) is analogously proved, by using the forward equation for the resolvent.

As a consequence of the result above we obtain:
\[
\pi_{m-1} = -\pi_{m} D m Q^{-1}. \tag{2.7}
\]
Define \(T\) to be the \((\ell + 1) \times (\ell + 1)\) subblock of \((-m Q)^{-1}\) corresponding to the states in \(L_{m-1}\). Note that without further restrictions on \(D\) it is unfortunately not possible to calculate it independently of the rates of states in \(L_{m-1}\). From Eq. (2.3) it follows that
\[
\pi_{m-1} = \pi_{m} B, \quad B = DT,
\]
where \(B_{ij}\) represents the expected local times spent in \((m - 1, j)\), before absorption into \(\tilde{L}_m\), given that the process starts in \((m, i)\).

By homogeneity, we can recursively show that
\[
\pi_m = \pi_0 B^{-m}, \quad \text{for} \quad m \leq 0. \tag{2.8}
\]
Now the existence of the stationary distribution implies that
\[
\sum_{m \leq 0} B^{-m} 1_{\ell+1} = (I - B)^{-1} 1_{\ell+1} < \infty,
\]
with \(I\) the \((\ell + 1) \times (\ell + 1)\) identity matrix and \(1_{\ell+1}\) the \(\ell + 1\)-dimensional vector consisting of ones.

The next lemma shows that it is not possible to have an unbounded state space to the right and a stationary distribution for a homogeneous QSF process.

**Lemma 2.2.** Consider a homogeneous QSF process where the number of levels is unbounded in the negative direction. In order for the process to be positive recurrent, the number of levels has to be bounded in the positive direction.
Proof. Suppose that the levels are not bounded to the right. Were the stationary distribution to exist, the same reasoning as in the above (Eq. (2.8)) would yield that \( \pi_m = \pi_n B^{-m+n}, m \leq n \). Analogously, by homogeneity it would follow that \( \pi_n = \pi_{(n,0)} a \), with \( a \) an \((\ell+1)\)-dimensional positive column vector, independent of \( n \) (see also the proof of Theorem 2.3 below). Then

\[
\sum_{m \leq n, j} \pi_{(m,j)} = \pi_{(n,0)} a^T (I - B)^{-1} 1_{\ell+1}.
\]

Taking the limit \( n \to \infty \), on the one hand yields that \( \sum_{m \leq n, j} \pi_{(m,j)} \to 1 \). On the other hand, the right-hand side of the above equation is the product of two scalars, \( \pi_{(n,0)} \) and \( a^T (I - B)^{-1} 1_{\ell+1} \). Since \( \lim_{n \to \infty} \pi_{(n,0)} = 0 \), the product equals 0 as well, in the limit \( n \to \infty \). This is a contradiction. \( \square \)

2.1 Special choice of \( D \)

Imposing extra structure on \( D \) implies a result on the structure of the stationary distribution, that is described below. This extra structure covers the case when \( D = c \cdot r \), where \( c \) is a column vector and \( r \) is a row vector, and \( \cdot \) denotes matrix product (of potentially non-square matrices). In other words, the rows of \( D \) are proportional to each other. Let the \( i \)th column of \( T \) be denoted by the vector \( t_i = (t_{0i}, t_{1i}, \ldots, t_{\ell i})^T \), for \( i \in \{0, \ldots, \ell\} \).

Then \( B \) has the following form:

\[
B = DT = \begin{bmatrix}
c_0 r \cdot t_0 & c_0 r \cdot t_1 & \cdots & c_0 r \cdot t_\ell \\
c_1 r \cdot t_0 & c_1 r \cdot t_1 & \cdots & c_1 r \cdot t_\ell \\
\vdots & \vdots & \ddots & \vdots \\
c_\ell r \cdot t_0 & c_\ell r \cdot t_1 & \cdots & c_\ell r \cdot t_\ell
\end{bmatrix} = c \cdot \hat{t},
\]

where \( \hat{t} = (r \cdot t_0, r \cdot t_1, \ldots, r \cdot t_\ell) \) and

\[
B^n = (c \cdot \hat{t})^n = c \cdot ((\hat{t} \cdot c)^{n-1} \hat{t}) = \left( \sum_{i=0}^{\ell} c_i r \cdot t_i \right)^{n-1} c \cdot \hat{t}, \quad n \geq 1.
\]

Notice that matrix \( B \) has only one nonzero eigenvalue \( \gamma = \hat{t} \cdot c \) with left eigenvector \( \hat{t} \) of multiplicity 1, provided that \( \hat{t} \cdot c \) is finite. Combination with Eq. (2.8) implies that the stationary distribution of the levels has a product form. The following theorem holds.

**Theorem 2.3.** Assume that \( D \) has a column times row structure as is described above.
If \( \gamma < 1 \) and \( \ell < \infty \), then the Markov process \( X \) is positive recurrent.
If the Markov process \( X \) is positive recurrent, then \( \gamma < 1 \). In either case:

\[
\pi_{(m,j)} = \gamma^{-(m+1)} \sum_{i=0}^{\ell} \pi_{(0,j)} c_i \hat{t}_i.
\]  

(2.9)

Proof. Suppose that \( X \) is positive recurrent. By Eq. (2.8), the stationary distribution necessarily has the form of Eq. (2.9). Taking the summation over the levels \( m \) in Eq. (2.9) yields a finite expression. This implies that \( \sum_{m \leq M} \gamma^{-(m+1)} < \infty \) for any level \( M \). Hence, \( \gamma < 1 \) necessarily.

By separating state \((0,0)\) from level 0, one can compute the expected sojourn time in the states of \( S \setminus \{(0,0)\} \) by inverting the (negative of the) finite rate matrix restricted to this set, analogously to
the above derivations. Define \( x_{(0,i)} \) to be equal to an unknown constant \( x_{(0,0)} \) times the expected sojourn time in state \((0,i)\) given a start in state \((0,0)\) before absorption into \((0,0)\), for all \( i > 0 \). Then define \( x_{(m,j)} \) from Eq. (2.9) by substituting \( \pi_{(0,i)} \) by \( x_{(0,i)} \) in the right-hand side. Note that \( x_{(0,0)} \) is the only unknown constant. Since \( \gamma < 1 \), they can be normalised to yield a (unique) probability distribution on \( S \).

It is still hard to compute \( B \) directly, therefore we introduce even more structure on \( D \) in the next section. More precisely, we assume that \( c \) has only one non-zero component. Without loss of generality, we may assume that \( c_{0} \) is the only non-zero element of \( c \).

3 Exit states and successive lumpability

Our main assumption concerns the presence of exit states, which are defined below.

**Definition 3.1.** Let \( M \subset S \). Then \((m,i) \in M\) is an exit state for \( M \), if

i) \( \sum_{(k',j') \notin M} q_{(k,j)(k',j')} = 0 \) for all \((k,j) \in M\) with \((k,j) \neq (m,i)\).

ii) \( \sum_{(k',j') \notin M} q_{(m,i)(k',j')} > 0 \).

For the remainder of this section we make the following assumption, which implies that for all \( m \), level \( m - 1 \) can only be reached directly from level \( m \) through state \((m,0)\). In other words, \((m,0)\) is an exit state to level \( m - 1 \).

**Assumption 3.2.** State \((m,0)\) is an exit state for superset \( \tilde{L}_{m} \), for all \( m = \ldots, -1, 0 \).

3.1 Stationary distribution

Under assumption 3.2 the sub-matrix \( D \) contains only one non-zero row. This implies that matrix \( B \) has only one non-zero row (the first) as well. Then by virtue of Theorem 2.3:

\[
\pi_{(m-1,j)} = \pi_{(m,0)} b_{0j},
\]

where \( b_{0j} = c_{0} r \cdot t_{j} \), for \( j = 0, \ldots, \ell \).

Let \( \beta_{j} = b_{0j}/b_{00} \). We know: \( \gamma = c_{0} t_{0} = c_{0} r \cdot t_{0} = b_{00} < 1 \) and thus:

\[
\pi_{(m,j)} = \gamma^{-m} \beta_{j} \pi_{(0,0)},
\]

for \( j = 0, \ldots, \ell \).

As has been mentioned at the end of section 2, explicit computation of the matrix \( B \) is in general hard, because it can be of infinite dimension and it can have a complicated structure. In the presence of an exit state however, it turns out to be simpler to express \( \pi_{m} \) in terms of \( \pi_{m-1} \), cf. Eq. (3.3).

By tabooing on the set \( L_{m-1} \), there exists an \( |L_{m-1}| \times (\ell + 1) \) matrix \( R_{m} \), such that

\[
\pi_{m} = \pi_{m-1} R_{m}.
\]

By our homogeneity assumptions, \( R_{m} \) is in fact independent of \( m \), and so we suppress the dependance on \( m \) in our further notation, whenever possible. We will next define an embedded q-matrix \( \hat{Q} \) on \( L_{m} \) and an \( L_{m-1} \times L_{m} \) matrix \( \hat{A} \) by removing the entrance state \((m+1,0)\), such that the mean (local) sojourn times in the state of \( L_{m} \) do not change, given a start in \( L_{m} \). We next state the following theorem using this shift.
Theorem 3.3. Consider a level homogenous process with an exit state in each level. Then:

\[ \pi_m = \pi_{m-1} R, \]

where

\[ R := -\tilde{A}\tilde{Q}^{-1}, \]

where \( \tilde{A} \) and \( \tilde{Q} \) are \(|L_{m-1}| \times (\ell + 1) \) and \((\ell + 1) \times (\ell + 1) \) matrices with elements

\[ \tilde{q}_{ij} = w_{ij} + \sum_{s=1}^{\infty} \sum_{r=0}^{\ell} u_{ir}^s \frac{d_{0j}}{\sum_{v=0}^{\ell} d_{0v}}, \]  

\[ \tilde{a}(k,i)(m,j) = u_{ij}^{m-k} + \sum_{s=m+1-k}^{\infty} \sum_{r=0}^{\ell} u_{ir}^s \frac{d_{0j}}{\sum_{v=0}^{\ell} d_{0v}}, \quad \text{for } k < m, 0 \leq i, j \leq \ell, \]  

respectively.

Notice that \( d_{0j}/\sum_{r=0}^{\ell} d_{0r} \) is the probability that level set \( L_m \) is entered at state \((m,j)\), given that a downward transition (to set \( L_m \)) occurs starting in state \((m+1,0)\).

Proof. We will explicitly compute \( R \), containing the expected sojourn times spent in the states of level \( L_m \), before absorption into \( L_{m-1} \), given that the process starts in \( L_m \).

To this end, note that since \((m+1,0)\) is an exit state for level \( L_{m+1} \), it is an entrance state for the set \( L'_m \), defined below:

\[ L'_m = L_m \cup \{(m+1,0)\}. \]

This implies that \( L'_m \) can only be reached from states in \( S \setminus L'_m \) through \((m+1,0)\). The technique (successive lumping) developed in Katehakis and Smit [8] can be used to compute the expected sojourn time spent in the extended level \( L'_m = L_m \cup \{(m+1,0)\} \) before absorption into \( L_{m-1} \), given that the process starts at this extended level \( L'_m \).

We use this technique to compute \( \tilde{Q} \), the transition rate matrix of size \((\ell + 2) \times (\ell + 2)\) embedded on level \( L'_m \) for any \( m \). Then \( \tilde{Q} \) has elements:

\[ \tilde{q}(k,i)(l,j) = \begin{cases} 
  w_{ij}, & (k,i),(l,j) \in L_m, \\
  \sum_{s=1}^{\infty} \sum_{r=0}^{\ell} u_{ir}^s \frac{d_{0j}}{\sum_{v=0}^{\ell} d_{0v}}, & (k,i) \in L_m, (l,j) = (m+1,0), \\
  -\sum_{j=0}^{\ell} d_{0j}, & (k,i) = (m+1,0), (l,j) \in L_m, \\
  0, & \text{otherwise}.
\end{cases} \]  

Note that the transitions leading to states in superlevel \( L_{m+1} \setminus \{(m+1,0)\} \) are mapped to the entrance state \((m+1,0)\). By virtue of Lemma 2.1, the expected sojourn time spent in level \( L'_m \) before absorption into the sub-level set \( L_{m-1} \) is obtained by inverting \(-\tilde{Q}\). Katehakis, Smit and Spieksma [10], Eq. (9), show that

\[ \pi'_m = -\pi_{m-1} \tilde{A}\tilde{Q}^{-1}, \]

where \( \pi'_m = (\pi_m, \pi_{(m+1,0)}) \) and the elements of \( \tilde{A} \) are given below with \( k < m \):

\[ \tilde{a}(k,i)(l,j) = \begin{cases} 
  u_{ij}^{m-k}, & (l,j) = (m,j) \text{ and } i,j \in \{0, \ldots, \ell\} \\
  \sum_{s=m+1-k}^{\infty} \sum_{r=0}^{\ell} u_{ir}^s, & (l,j) = (m+1,0).
\end{cases} \]

The validity of this exit state lumping construction of Eqs. (3.4) and (3.5) is guaranteed by Lemma A.2 in appendix A. \( \square \)
Restriction to QBD processes  For QBD processes, where $U^1 = U$ and $U^s = 0$ for $s \geq 2$, the above derivations impy that:

$$\pi_m = \pi_{m-1}R,$$

(3.9)

where $R = -UQ^{-1}$. If $U$ is an invertible matrix, then the result above is equivalent to

$$\pi_{m-1} = -\pi_m\hat{Q}U^{-1}.$$  

(3.10)

In this equation we have expressed $\pi_{m-1}$ in terms of $\pi_m$, similarly to Eq. (2.8). However, $\hat{Q}U^{-1}$ is explicitly computable, provided that $U$ is invertible, whereas $B$ may not be invertible.

From Eqns. (3.2) and (3.9) we have the following result

$$\frac{1}{\gamma} \beta = \beta R.$$  

(3.11)

Particularly if $U$ is an invertible matrix, then

$$\beta(-\hat{Q})U^{-1} = \gamma \beta.$$  

(3.12)

**Lemma 3.4.** If $\ell$ is finite, then $1/\gamma$ is the maximum eigenvalue of $R$ in absolute value.

**Proof.** Let $R$ be an irreducible non-negative matrix. Then by the Perron-Frobenius theorem (for example in Seneta [16]) the maximum eigenvalue is positive and real. We denote this eigenvalue by $r$. Further, there exist unique strictly positive left and right eigenvectors. Let $x$ denote this positive right eigenvector. Then for all $i = 0, \ldots, \ell$:

$$rx_i = \sum_j r_{ij}x_j.$$  

(3.13)

From Eq. (3.11) it is clear that $1/\gamma$ is an eigenvalue of $R$ with $\beta$, (where $\beta > 0$) as the left eigenvector corresponding to $1/\gamma$, i.e.:

$$\frac{1}{\gamma} \beta_j = \sum_i \beta_i r_{ij}.$$  

(3.14)

Next, we will show that $1/\gamma$ is the maximum eigenvalue of $R$, thus that $1/\gamma = r$.

From Eq. (3.13) we get

$$\sum_i r_{xi} \beta_i = \sum_j \sum_i \beta_i r_{ij}x_j.$$  

Combining with Eq. (3.14) yields

$$r(x \cdot \beta) = \frac{1}{\gamma}(x \cdot \beta).$$

Since $x \cdot \beta > 0$, it follows that $r = 1/\gamma$. In other words, $1/\gamma$ is the maximum eigenvalue of $R$. 

In the next section we will discuss a single server queueing model with Coxian arrivals and batch services.
Queueing application: analysis of the $\text{Cox}(k)/M^Y/1$ queue

We consider a queueing model in which customers arrive within a maximum of $k$ exponentially distributed phases. The inter-arrival distribution is a Coxian distribution of order $k$, see e.g. [6]. In a Coxian distribution of order $k$, phase $i \in \{0, \ldots, k-1\}$ lasts an exponentially distributed amount of time with parameter $\lambda_i$, at the end of which either a new customer arrives with probability $1-q_i$, or a new phase starts with probability $q_i$, where $q_{k-1} = 0$. Upon arrival of a new customer, a new inter-arrival distribution starts. To avoid trivialities, we assume that $q_i > 0$ for $i \leq k-2$. Hence the inter-arrival distribution is a $\text{Cox}(k)$ distribution with parameters $(\lambda_0, \ldots, \lambda_{k-1}, q_0, \ldots, q_{k-1})$.

We use this naming of the phases to have a more natural state space description: here, state $(m, i)$ denotes the state of the system when there are $m$ customers in the system and the $m+1$ arriving customer has completed $i$ arrival phases. We will use the random variable $C \sim \text{Cox}(k)$ to denote the inter-arrival time.

The $\text{Cox}(k)/M^Y/1$ queueing system can be formulated as a QSF process $X(t)$ on the state space $S = \{L_0, L_1, \ldots\}$, where $L_m = \{(m,0), (m,1), \ldots, (m,k-1)\}$ for all $m \geq 0$. Service occurs according to the distribution induced by $Y$; in batches of size $j$, with $1 \leq j \leq b$, each with probability $p_j$. We will denote the probability-generating function of $Y$ as $\phi_Y(x)$. Note that $\phi_Y(x) = \sum_{j=1}^{b} p_j x^j$.

We will first analyse the general $\text{Cox}(k)/M^Y/1$-queue with finitely many phases. Then we will discuss some special subcases, where the exponential phases all have the same parameter, or the probabilities of a new inter-arrival phase are all equal up to the order. We will briefly discuss the extension to the case of a Coxian distribution of infinite order.

**Analysis of the $\text{Cox}(k)/M^Y/1$-queue** The transition rate matrix $Q$ for this model takes the form:

$$ Q = \begin{bmatrix} W_0 & U & 0 & 0 & 0 & \cdots \\ D_1 & W & U & 0 & 0 & \cdots \\ D_2 & D_1 & W & U & 0 & \cdots \\ D_3 & D_2 & D_1 & W & U & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \\ \end{bmatrix}, $$(4.1)

with $k \times k$ sub-matrices $D_i = p_i \cdot \mu I$, $D'_k = \sum_{i=k}^{y} D_i$, where $I$ is the identity of size $b \times b$ and

$$ W_0 = \begin{bmatrix} -\lambda_0 & q_0 \lambda_0 & \cdots & \cdots & 0 \\ 0 & -\lambda_1 & q_1 \lambda_1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & -\lambda_{k-2} & q_{k-2} \lambda_{k-2} \\ 0 & 0 & \cdots & 0 & -\lambda_{k-1} \\ \end{bmatrix}, \quad U = \begin{bmatrix} (1-q_0)\lambda_0 & \cdots & 0 & 0 \\ (1-q_1)\lambda_1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ (1-q_{k-2})\lambda_{k-2} & \cdots & 0 & 0 \\ \lambda_{k-1} & \cdots & 0 & 0 \\ \end{bmatrix}. $$

and $W = W_0 + \mu I$. Note that $U$ has a $c \cdot r$ structure, with

$$ c = \begin{bmatrix} (1-q_0)\lambda_0 \\ \vdots \\ (1-q_{k-2})\lambda_{k-2} \\ \lambda_{k-1} \end{bmatrix} \quad \text{and} \quad r = (1,0,\ldots,0), $$

10
and therefore fits the analysis of Section 2. Note further that the structure of Eq. (4.1) is the transpose of the matrix introduced in Eq. (2.1). We have to interchange the roles of the matrices $U$ and $D$ in the formulæ of the previous sections to fit the model under consideration. Another possible way to look at the problem is to use the original negative numbering of the levels. Then we do not need the interchanging described above, but the numbering is not induced by the model anymore.

We assume that the QSF process is ergodic. We shall prove below that for ergodicity it is necessary and sufficient to require that the mean number of arrivals per unit time is smaller than the mean number of service completions, which is specified in the relation below:

$$\frac{1}{\sum_{i=0}^{k-1} \prod_{l=0}^{i} q_l \lambda_i^{-1}} = \frac{1}{EC} < \mu \sum_{j=1}^{b} j p_j = \mu EY.$$  

(4.2)

When ergodicity holds, we can use the results of Theorem 2.3, and the stationary distribution takes the form: (again, note the change of notation, since we are considering positive levels now)

$$\pi(m,j) = \gamma^{m-1} \sum_{i=0}^{k-1} \pi(0,i) \hat{c}_i \hat{t}_j = \gamma^{m-1} \sum_{i=0}^{k-1} \pi(0,i) \lambda_i (1 - q_i) \hat{t}_j,$$

where for notational convenience, we use $q_{k-1} = 0$. The factor $\gamma$ and the vector $\hat{t}$ denote the distribution of the phase and are implicitly given as largest positive eigenvalue and corresponding left eigenvector of the rate matrix $R$ for a fixed level. Note that once $\gamma$ and $\hat{t}$ are known, the distribution of level 0 can be deduced.

An alternative procedure is done by observing that the state $(m,0)$ is an entrance state for $\tilde{L}_m$ (and an exit state for the shifted partition $\tilde{L}_{m-1} \cup \{(m,0)\}$). This allows us to use the results from Section 3 for computing the rate matrix in the reverse (downward) direction. However, we prefer not to use the shifted partition, because the emerging stationary distribution in both phases and levels will then have a notationally less amenable form, as we shall see. Therefore we will stay with the current level partitions.

In consideration of the statements above, the following relation holds (cf. Eq. (3.7)) for levels $m \geq 1$:

$$\pi_m = \pi_{m+1} R = -\pi_{m+1} \hat{A} \hat{Q}^{-1},$$

(4.3)

where $\hat{A}$ and $\hat{Q}$ are given in (3.8) and (3.6). We refer to [10], where a similar formula is also provided, but not the results regarding the product form. For this model, the matrix $\hat{Q}$ takes the form:

$$\hat{Q} = \begin{bmatrix}
-\lambda_0 & q_0 \lambda_0 & \ldots & 0 & 0 \\
\mu & - (\lambda_1 + \mu) & q_1 \lambda_1 & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
\mu & 0 & \ldots & - (\lambda_{k-2} + \mu) & q_{k-2} \lambda_{k-2} \\
\mu & 0 & \ldots & 0 & - (\lambda_{k-1} + \mu)
\end{bmatrix}.$$  

(4.4)

Note that $\hat{Q}$ is independent of the batch size distribution. The matrix $\hat{A}$ has dimension $k \times bk$, and can therefore be written as:

$$\hat{A} = (\hat{A}_1 \hat{A}_2 \ldots \hat{A}_b)^T,$$
where \( \tilde{A}_i \) is the following \( k \times k \) matrix:

\[
\tilde{A}_i = \begin{bmatrix}
\sum_{j \geq i} p_j \mu & 0 & \cdots & 0 \\
\sum_{j > i} p_j \mu & p_i \mu & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
\sum_{j > i} p_j \mu & 0 & \cdots & p_i \mu
\end{bmatrix},
\]

(4.5)

for \( 1 < i < b \) and \( \tilde{A}_b = p_b \cdot \mu I \).

Since the rate matrix \( R \) expresses the stationary solution of lower levels in terms of that of higher ones, we cannot recursively compute the stationary distribution in the infinite capacity case. We still need to find \( \gamma \) and \( \hat{t} \) to do so.

Instead we will derive equations for \( \gamma \) and the unknown (conditional) stationary probabilities of the phases. To this end we consider the expression:

\[
-\pi_m \tilde{Q} = \tilde{\pi}_{m+1} \tilde{A}.
\]

Because of the level product form solution (cf. Eq. (3.2)) we can rewrite this equation as:

\[
-\pi_m \tilde{Q} = \sum_{i=1}^{b} \pi_{m+i} \tilde{A}_i = \pi_m \sum_{i=1}^{b} \gamma^i \tilde{A}_i.
\]

(4.6)

Using Eq. (3.1) (i.e. \( \pi_{(m,i)} = \hat{t}_i \pi_{(m,0)} \)) to calculate the components of the vectors with \( 1 \leq i \leq k - 1 \) on both sides of the above equality Eq. (4.6) yields:

\[
(\lambda_i + \mu) \hat{t}_i - q_{i-1} \lambda_{i-1} \cdot \hat{t}_{i-1} = \sum_{j=1}^{b} \gamma^j p_j \mu \cdot \hat{t}_i = \mu \phi_Y(\gamma) \cdot \hat{t}_i.
\]

(4.7)

Hence

\[
\frac{\hat{t}_i}{\hat{t}_{i-1}} = \frac{q_{i-1} \lambda_{i-1}}{\lambda_i + \mu - \mu \phi_Y(\gamma)}.
\]

(4.8)

In other words, writing \( \beta_i = q_{i-1} \alpha_i = \hat{t}_i / \hat{t}_{i-1} \), we get that \( \hat{t}_i = \prod_{l=1}^{i} q_{l-1} \alpha_l \cdot \hat{t}_0 \). Therefore the stationary distribution has the following form for \( m \geq 1 \):

\[
\pi_{(m,i)} = \pi_{(1,0)} \gamma^{m-1} \prod_{l=1}^{i} q_{l-1} \alpha_l,
\]

(4.9)

where \( \alpha_i \) the following function of \( \gamma \) derived from (4.8):

\[
\alpha_i = \frac{\lambda_{i-1}}{\lambda_i + \mu - \mu \phi_Y(\gamma)}.
\]

(4.10)

Note that \( (\alpha_1, \ldots, \alpha_{k-1}) \) only depend on \( (q_0, \ldots, q_{k-2}) \) through \( \gamma \).

Using the balance equation for state \((m,0) \) \( (m \geq 1) \) and Eq. (4.10) we obtain the following expression for \( \gamma \) as a function of \( (\alpha_1, \ldots, \alpha_{k-1}) \):

\[
\gamma = \frac{\alpha_1 (1 - q_0) \lambda_0 + \alpha_1 \sum_{i=1}^{k-1} \prod_{l=1}^{i} \alpha_l q_{l-1} (1 - q_i) \lambda_i}{(\lambda_0 - \lambda_1) \alpha_1 + \lambda_0},
\]

(4.11)
where we have substituted \( \phi_Y(\gamma) \mu = \lambda_1 + \mu - \lambda_0/\alpha_1 \). Equations (4.10) and (4.11) provide a system of \( k \) equations in the unknowns \( \gamma \) and \( \alpha_i, i = 1, \ldots, k - 1 \). As is easily checked, \( \alpha_i = \lambda_{i-1}/\lambda_i, i = 1, \ldots, k - 1, \gamma = 1 \) form a solution of this system.

Using the same balance equation, we obtain a fixpoint equation for \( \gamma \):

\[
\gamma = F(\gamma),
\]

with \( F : \mathbb{R} \to \mathbb{R} \) given by

\[
F(\gamma) = \frac{\lambda_0 (1 - \alpha_0) + \sum_{i=1}^{k-1} \prod_{l=1}^{i} q_{l-1} \lambda_{i-1} \gamma_i}{\lambda_0 + \mu} \frac{\lambda_0 (1 - \alpha_0) \gamma_i}{\lambda_0 + \mu} + \gamma \phi_Y(\gamma).
\]

Again, \( \gamma = 1 \) is a fixpoint of \( F \). The function \( F \) is tediously but easily checked to be a convex function of \( \gamma \) on the interval \([0, 1 + \epsilon]\) for some \( \epsilon > 0 \). It is positive on this interval, and hence larger than the left hand side with \( \gamma = 0 \), i.e. the value of \( F(0) \). Computing the derivative at \( \gamma = 1 \) yields, that this is strictly larger than 1 if and only if condition (4.2) holds. Hence, \( F \) has one fixpoint \( \gamma < 1 \) if and only if (4.2) holds, showing that (4.2) is necessary and sufficient for ergodicity. This is a standard argument used for deriving ergodicity conditions by means of a probability generating function approach, see e.g. [15] Section 4.5.

Hence solutions \( \gamma \) and \( \{\alpha_i\}_{i=1}^{k} \) can be determined (i) by determining the unique fixpoint \( \gamma < 1 \) of \( F \) and then (ii) insert \( \gamma \) into Eq. (4.10).

Next we will express the steady state probabilities of the first level in terms of the steady state probability in state \((0,0)\). For sake of presentation, we omitted the derivation and present the results immediately:

\[
\pi(1,0) = \pi(0,0) \frac{\lambda_0}{\mu} \frac{1 - \gamma}{1 - \phi_Y(\gamma)},
\]

\[
\pi(0,i) = \pi(0,0) \frac{\lambda_0}{\chi_i} \prod_{s=0}^{i-1} q_s \left( 1 + \frac{1 - \gamma}{\gamma(1 - \phi_Y(\gamma))} \sum_{j=1}^{i} \prod_{l=1}^{j} \alpha_l \right), \quad \text{for } i \geq 1,
\]

and

\[
\pi(0,0) = \left( 1 + \frac{\lambda_0}{\mu(1 - \phi_Y(\gamma))} \right) \left( 1 + \sum_{i=1}^{k-1} \prod_{j=0}^{i-1} \frac{\lambda_0}{\lambda_i} q_j \right)^{-1}.
\]

Then by using Eqns. (4.9), (4.14), (4.15), and (4.16), for \( m \geq 1 \) we have

\[
\pi(m,i) = \pi(0,0) \frac{\lambda_0 (1 - \gamma) \gamma^{m-1}}{\mu(1 - \phi_Y(\gamma))} \prod_{l=1}^{i} q_{l-1} \alpha_l.
\]

We summarise all our findings above in the next theorem.

**Theorem 4.1.** The \( \text{Cox}(k)/M^Y/1 \) queue is ergodic in and only if Eq. (4.2) is satisfied. If this is the case, the stationary distribution of the \( \text{Cox}(k)/M^Y/1 \) queue on levels \( \geq 1 \) is given by (4.17). The factors \( \alpha_i, i = 1, \ldots, k - 1 \) and \( \gamma \) can be calculated from Eq. (4.12) and Eq. (4.10). The boundary level can by found by Eq. (4.15) and (4.16).
Remark 1 (Finite capacity queues). In case of a finite capacity queue of size $S$, the stationary distribution can be computed recursively, in terms of $\pi_S$. By using the fact that level $S$ has an entrance state $(S,0)$ from the right, the analysis described in the previous section can be used and yields that

$$\pi_{(S,i)} = -Q_S^{-1}\pi_{(S,0)},$$

where now

$$Q_S = \begin{bmatrix}
-\lambda_0 & \lambda_0 & \cdots & 0 & 0 \\
\mu & -(\lambda_1 + \mu) & \lambda_1 & \cdots & 0 & 0 \\
\vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\
\mu & 0 & \cdots & -(\lambda_{k-2} + \mu) & \lambda_{k-2} \\
\mu & 0 & \cdots & 0 & -\mu
\end{bmatrix}. $$

Then using (4.3) with the matrices $\tilde{Q}_m$ and $\tilde{A}_m$ associated with level $m < S$, one may compute the lower level stationary probabilities (up to a constant). Final renormalisation yields the correct stationary distribution.

Remark 2 (Variable service rates). Suppose that service rates $\mu$ and batch probabilities $p_j$ are equal to $\mu^m$ and $p_j^m$ for level $m \leq S$, $j = 1,\ldots,b$, respectively, for some $S < \infty$. Then the stationary distribution of Cox($k$)/$M^Y$/1 queue on the levels $S + k$, $k \geq 1$, can be computed in exactly the same manner as in the above, yielding the stationary probabilities of these levels up to a constant. Meanwhile, the stationary distribution for levels $m \leq S$ can be computed from $\tilde{\pi}_S$ by using (4.3) with the matrices $\tilde{Q}_m$ and $\tilde{A}_m$ associated with level $m$, as described in Remark 1. Again, final renormalisation yields the correct stationary distribution.

4.1 A Cox($k$) inter-arrival distribution with homogeneous parameters

In this section we consider several subcases for the model described in the previous section.

Homogeneous rates First let us assume that the rates in the exponential distribution are all equal: $\lambda_i = \lambda$, $i = 0,\ldots,k - 1$. Then from Eq. (4.10) we get

$$\alpha_i = \frac{\lambda}{\lambda + \mu - \mu \phi_Y(\gamma)} =: \alpha, \quad i = 1,\ldots,k - 1,$$

in other words, the phase factors $\alpha_i$ have become independent of the phase. Then Eq. (4.11) reduces to

$$\gamma = \alpha(1 - q_0) + \sum_{i=1}^{k-1} \left( \prod_{l=0}^{i-1} q_l \right) \alpha^{i+1}(1 - q_i).$$

(4.19)

The function $F$ can be simplified, but we can also directly insert Eq. (4.18) for $\alpha$ in Eq. (4.19) to obtain that the solution $\gamma$ is a fixpoint of the equation $\gamma = F^h(\gamma)$ with

$$F^h(\gamma) = \frac{\lambda(1 - q_0)}{\lambda + \mu - \mu \cdot \phi_Y(\gamma)} + \sum_{i=1}^{k-1} \left( \prod_{l=0}^{i-1} q_l \right) \left( \frac{\lambda}{\lambda + \mu - \mu \cdot \phi_Y(\gamma)} \right)^{i+1}(1 - q_i).$$

In the same manner as for the function $F$, we can deduce that the ergodicity condition (4.2) is necessary and sufficient for $F^h$ to have a fixpoint $\gamma < 1$. 

14
Homogeneous rates and probabilistic phase transitions  Next we assume that additionally the probabilities of a new phase are all equal, that is, \( q_i = q, i = 0, \ldots, k - 2 \). This has no further impact on \( \alpha \). However, Eq. (4.11) reduces further to

\[
\gamma = \sum_{i=0}^{k-2} q^i \alpha^{i+1} (1 - q) + q^{k-1} \alpha^k. \tag{4.20}
\]

The function \( F^h \) becomes

\[
F^h(\gamma) = \sum_{i=0}^{k-2} q^i \left( \frac{\lambda}{\lambda + \mu - \mu \cdot \phi_Y(\gamma)} \right)^{i+1} (1 - q) + q^{k-1} \left( \frac{\lambda}{\lambda + \mu - \mu \cdot \phi_Y(\gamma)} \right)^k.
\]

Further, the stationary distribution of the levels \( m \geq 1 \) has a product form

\[
\pi(m,i) = \pi(0,0) \frac{\lambda(1 - \gamma) \gamma^{m-1} (q \alpha)^i}{\mu(1 - \phi_Y(\gamma))}, \quad m \geq 1,
\tag{4.21}
\]

and

\[
\pi(0,i) = \pi(0,0) q^i \left( 1 + \frac{(1 - \gamma) \phi_Y(\gamma)}{\gamma(1 - \phi_Y(\gamma))} \sum_{j=1}^{i} \alpha^j \right), \quad \text{for } i = 1, \ldots, k - 1,
\tag{4.22}
\]

where

\[
\pi(0,0) = \frac{(1 - q) \mu(1 - \phi_Y(\gamma))}{(1 - q^k)(\lambda + \mu(1 - \phi_Y(\gamma)))} = \frac{(1 - q)(1 - \alpha)}{1 - q^k}.
\]

Homogeneous rates and deterministic phase transitions: the \( E_k/M^Y/1 \)-queue  Taking \( q_i = q = 1, i = 0, \ldots, k - 2 \), finally yields an Erlang \((k, \lambda)\) inter-arrival distribution. Then Eq. (4.11) takes the very simple form

\[
\gamma = \alpha^k. \tag{4.23}
\]

The function \( F^h \) is simply given by

\[
F^h(\gamma) = \left( \frac{\lambda}{\lambda + \mu - \mu \cdot \phi_Y(\gamma)} \right)^k.
\]

Finally, the stationary distribution has a product form (except for level 0) determined by only one factor

\[
\pi(m,i) = \pi(0,0) \frac{\lambda(1 - \gamma) \alpha^{i+k(m-1)}}{\mu(1 - \phi_Y(\gamma))}, \quad m \geq 1,
\tag{4.24}
\]

and

\[
\pi(0,i) = \pi(0,0) \left( 1 + \frac{(1 - \gamma) \phi_Y(\gamma)}{\gamma(1 - \phi_Y(\gamma))} \sum_{j=1}^{i} \alpha^j \right), \quad \text{for } i = 1, \ldots, k - 1,
\tag{4.25}
\]

where

\[
\pi(0,0) = \frac{1 - \alpha}{k}.
\]

Notice that this is not surprising. The process is then essentially a one-dimensional process with a \( GI/M/1 \)-structure. This is known to have a product form stationary distribution (cf. Asmussen [3]).

We further note that the \( Cox(k)/M^Y/1 \)-queue with constant rates and probabilistic phase transitions is not of the (pure) \( GI/M/1 \)-type. Indeed, the probabilities are only constant with respect to the phases with index smaller than \( k - 1 \).
Distribution of the number of customers in the queue  By taking the summation over the
phases per level we obtain the following modified geometric distribution for the number of customers
in $E_k/M^Y/1$-queue

\[ \bar{\pi}_k^m = \begin{cases} \frac{\lambda(1 - \gamma)^2}{k \mu \cdot (1 - \phi_Y(\gamma))} \gamma^{m-1}, & m > 0, \\ 1 - \frac{\lambda(1 - \gamma)}{k \mu \cdot (1 - \phi_Y(\gamma))}, & m = 0, \end{cases} \]  

(4.26)

where $\bar{\pi}_m^k = \sum_{i=0}^{k-1} \pi_{(m,i)}, m \geq 0$.

4.2 The Cox($\infty$)/$M^Y/1$-queue

Allowing infinitely many phases still fits our framework. We get the natural extensions of formulas
Eq. (4.11) and Eq. (4.13) with $k = \infty$. Clearly the expression for $\alpha_i$ as a function of $\gamma$ is not affected
by the amount of phases.

Restricting to the case of homogeneous rates and probabilities, so that $\lambda_i = \lambda$, $q_i = q$, $i = 0, \ldots$, we get the following results.

The queueing system is ergodic if and only if

\[ \lambda \left( \frac{1 - q}{q} \right) \prec \mu \sum_{j=1}^b j p_j. \]  

(4.27)

Then Eq. (4.11) becomes

\[ \gamma = \frac{\alpha(1 - q)}{1 - q \alpha}, \]  

(4.28)

and so again we obtain

\[ F^h(\gamma) = \frac{\lambda(1 - q)}{\lambda(1 - q) + \mu(1 - \phi_Y(\gamma))}. \]

We summarise our results in the next theorem.

**Theorem 4.2.** Under the ergodicity condition (4.27), the stationary distribution of the Cox($\infty$)/$M^Y/1$ queue on levels $\geq 1$ is given by (4.17) where $\alpha$ and $\gamma$ can be calculated from Eq. (4.18) and Eq. (4.28). The boundary level can by found by Eq. (4.15), where $\pi_{(0,0)} = (1 - q)(1 - \alpha)$.

4.3 Numerical Analysis

Assume an ergodic Cox($k$)/$M^Y/1$ queue of finite order. In order to calculate the solutions \{\(\alpha_i, i = 1, \ldots, k - 1; \gamma\}\}, one can solve (4.12) directly, e.g. by the Newton-Raphson method, and then determine $\alpha_i, i = 1, \ldots, k - 1$, from (4.10).

As $\gamma$ is a fixpoint of the map $F$ associated with (4.12), another possibility is to approximate the
fixpoint $\gamma < 1$ by selecting $\gamma_0$ and by iteratively computing $\gamma_{n+1} = F(\gamma_n)$. It is simply checked that
the fixpoint $\gamma = 1$ is not stable, but that the fixpoint of interest, smaller than 1, is a stable fixpoint . Hence we can use the following scheme to approximate the desired values $\alpha_i, i = 1, \ldots, k - 1, \gamma$.

**Approximation scheme 1**
1. Choose $\gamma < 1$;

2. iteratively put $\gamma := F(\gamma)$, till desired convergence; compute $\alpha_i$ from (4.10), $i = 1, \ldots, k - 1$.

In the case of homogeneous rates, we have shown that the level factor $\gamma$ is the unique fixpoint smaller than 1 of the function $F^h$. This leads to the following adapted scheme, that we only formulate for the case of homogeneous rates.

**Approximation scheme 2 for the case $\lambda_i = \lambda, i \leq k - 1$ (where $k = \infty$ is allowed)**

1. Choose $\gamma < 1$;

2. iteratively put $\gamma := F^h(\gamma)$, until desired convergence, and compute $\alpha$ from (4.18).

It is outside the scope of the paper to discuss the rate of convergence of the scheme, as well as a detailed stopping criterion. Table 1 below shows the non-surprising property that $\gamma$ and $\alpha$ are non-increasing in $q$. Further note that even for a high value of the continuation probability $q$, $\gamma$ is already approximately constant starting from around 20 phases.

<table>
<thead>
<tr>
<th>$q$</th>
<th>$k = 2$</th>
<th>$k = 5$</th>
<th>$k = 20$</th>
<th>$k = 1000$</th>
<th>$k = \infty$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\gamma$</td>
<td>$\alpha$</td>
<td>$\gamma$</td>
<td>$\alpha$</td>
<td>$\gamma$</td>
</tr>
<tr>
<td>0.1</td>
<td>0.4168</td>
<td>0.4414</td>
<td>0.4153</td>
<td>0.4411</td>
<td>0.4153</td>
</tr>
<tr>
<td>0.2</td>
<td>0.3847</td>
<td>0.4339</td>
<td>0.3788</td>
<td>0.4325</td>
<td>0.3788</td>
</tr>
<tr>
<td>0.3</td>
<td>0.3538</td>
<td>0.4272</td>
<td>0.3406</td>
<td>0.4246</td>
<td>0.3406</td>
</tr>
<tr>
<td>0.4</td>
<td>0.3239</td>
<td>0.4214</td>
<td>0.3066</td>
<td>0.4173</td>
<td>0.3065</td>
</tr>
<tr>
<td>0.5</td>
<td>0.2948</td>
<td>0.4163</td>
<td>0.2585</td>
<td>0.4105</td>
<td>0.2582</td>
</tr>
<tr>
<td>0.6</td>
<td>0.2663</td>
<td>0.4117</td>
<td>0.2141</td>
<td>0.4042</td>
<td>0.2134</td>
</tr>
<tr>
<td>0.7</td>
<td>0.2385</td>
<td>0.4076</td>
<td>0.1673</td>
<td>0.3986</td>
<td>0.1658</td>
</tr>
<tr>
<td>0.8</td>
<td>0.2113</td>
<td>0.4039</td>
<td>0.1176</td>
<td>0.3935</td>
<td>0.1147</td>
</tr>
<tr>
<td>0.9</td>
<td>0.1845</td>
<td>0.4006</td>
<td>0.0648</td>
<td>0.3890</td>
<td>0.0659</td>
</tr>
<tr>
<td>1</td>
<td>0.1581</td>
<td>0.3976</td>
<td>0.0085</td>
<td>0.3851</td>
<td>0.0080</td>
</tr>
</tbody>
</table>

Table 1. ($\lambda, \mu, \gamma, p_1, p_2, p_3$) = (0.5, 0.8, 0.35, 0.25, 0.5, 0.25).

5 Monotonicity properties and relation with the $D/M^Y/1$-queue

5.1 Monotonicity properties

Next we will study monotonicity properties for the homogeneous $Cox(k)/M^Y/1$ queues, in the following sense. If $\lambda_i = \lambda$ and $q_i = q$ for $i = 0, \ldots, k - 2$, then we denote the corresponding $k$-order Coxian distribution by $Cox(k, \lambda, q)$.

In the remainder of the paper we assume that the ergodicity condition Eq. (4.2) is satisfied. It is the aim to compare the stationary distribution of the number of customers in the system for $Cox(k, \lambda, q)/M^Y/1$ queues, with different inter-arrival distributions with the same mean inter-arrival times and associated probabilities $q$, but with a different amount of phases.
Let $\lambda^*$ and phase probability $q$ and the number of phases $k$ be given. Then, in order for the mean inter-arrival time to equal $1/\lambda^*$, the parameter $\lambda_k$ of the homogeneous Coxian distribution has to be equal to:

$$\frac{1}{\lambda^*} = \frac{1}{\lambda_k} \sum_{l=0}^{k-1} q^l = \frac{1 - q^k}{\lambda_k (1 - q)}.$$  

that is

$$\lambda_k = \frac{\lambda^*(1 - q^k)}{1 - q}.$$  

Denote the corresponding factors in the stationary distribution (see previous section) of the corresponding QSF process by $\alpha_k$ and $\gamma_k$ respectively, and denote the stationary distribution of the number of customers in the system by $\bar{\pi}_k$.

It does not seem possible to stochastically compare these queueing systems for a different number of phases: Table 2 below shows that there is a lack of monotonicity in the parameter $\gamma_k$, especially for high values of the continuation parameter $q$.

<table>
<thead>
<tr>
<th>$q$</th>
<th>$k = 2$</th>
<th>$k = 5$</th>
<th>$k = 10$</th>
<th>$k = 20$</th>
<th>$k = 50$</th>
<th>$k = 1000$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\gamma$</td>
<td>$\alpha$</td>
<td>$\gamma$</td>
<td>$\alpha$</td>
<td>$\gamma$</td>
<td>$\alpha$</td>
</tr>
<tr>
<td>0.1</td>
<td>0.4485</td>
<td>0.4734</td>
<td>0.4502</td>
<td>0.4764</td>
<td>0.4502</td>
<td>0.4764</td>
</tr>
<tr>
<td>0.2</td>
<td>0.4393</td>
<td>0.4939</td>
<td>0.4501</td>
<td>0.5057</td>
<td>0.4502</td>
<td>0.5058</td>
</tr>
<tr>
<td>0.3</td>
<td>0.4121</td>
<td>0.5120</td>
<td>0.4494</td>
<td>0.5383</td>
<td>0.4502</td>
<td>0.5391</td>
</tr>
<tr>
<td>0.4</td>
<td>0.4286</td>
<td>0.5283</td>
<td>0.4472</td>
<td>0.5738</td>
<td>0.4502</td>
<td>0.5771</td>
</tr>
<tr>
<td>0.5</td>
<td>0.4189</td>
<td>0.5429</td>
<td>0.4415</td>
<td>0.6111</td>
<td>0.4499</td>
<td>0.6206</td>
</tr>
<tr>
<td>0.6</td>
<td>0.4082</td>
<td>0.5563</td>
<td>0.4300</td>
<td>0.6488</td>
<td>0.4482</td>
<td>0.6670</td>
</tr>
<tr>
<td>0.7</td>
<td>0.3968</td>
<td>0.5685</td>
<td>0.4105</td>
<td>0.6853</td>
<td>0.4413</td>
<td>0.7243</td>
</tr>
<tr>
<td>0.8</td>
<td>0.3848</td>
<td>0.5798</td>
<td>0.3811</td>
<td>0.7197</td>
<td>0.4195</td>
<td>0.7796</td>
</tr>
<tr>
<td>0.9</td>
<td>0.3725</td>
<td>0.5902</td>
<td>0.3407</td>
<td>0.7514</td>
<td>0.3658</td>
<td>0.8307</td>
</tr>
</tbody>
</table>

Table 2. $(\lambda^*, \mu, \gamma, p_1, p_2, p_3) = (0.5, 0.8, 0.35, 0.25, 0.5, 0.25)$.

However, in the case of deterministic number phase transitions, an Erlang queue, there exists more structure regarding the change of $\gamma$ with respect to an increasing number of phases $k$.

**Monotonicity properties for the $E_k/M^Y/1$-queue** Let us next restrict to the case $q = 1$, in other words, the case of an Erlang inter-arrival distribution. Then the arrival rate in the $k$-phase system is given by $\lambda_k = \lambda^* k$. Write $\rho = \lambda^*/\mu$.

The following results hold.

**Theorem 5.1. a)** The sequence $\gamma_k$ is strictly decreasing in $k$. It has a limit $\gamma^* = \lim_{k \to \infty} \gamma_k$, which is the unique solution $\xi$ smaller than 1 of the equation below.

$$\xi = e^{-(1-\phi_Y(\xi))/\rho}.$$  

**b)** The map $k \mapsto \bar{\pi}_0^k = 1 - \rho (1 - \gamma_k)/(1 - \phi_Y(\gamma_k))$ is a strictly decreasing function if and only if $P\{Y = 1\} < 1$. If $P\{Y = 1\} = 1$, i.e. the batch size equals 1 with probability 1, then $\bar{\pi}_0^k = 1 - \rho$, for $k = 1, 2, \ldots$.
Proof. By combination of (4.18) and (4.23) we obtain:

\[
1/\gamma_k = \left(1 + \frac{(1 - \phi_Y(\gamma_k))}{k\rho}\right)^k.
\]  

(5.2)

Define \( g_k(x) := \left(1 + \frac{x}{k}\right)^k, \quad k = 0, 1, \ldots \)

Clearly this function is increasing in \( x \). To show that \( g_k \) is also increasing in \( k \), we expand the expression using the binomial formula:

\[
g_{k+1}(x) = \sum_{i=0}^{k+1} \binom{k+1}{i} \left(\frac{x}{\rho(k+1)}\right)^i
= 1 + \sum_{i=1}^{k} \frac{(x/\rho)^i}{i!} \cdot \frac{k}{k+1} \cdots \frac{k+2-i}{k+1} + \frac{(x/\rho)^{k+1}}{(k+1)^{k+1}}.
\]

(5.3)

On the other hand,

\[
g_k(x) = \sum_{i=0}^{k} \binom{k}{i} \left(\frac{x}{\rho(k)}\right)^i
= 1 + \sum_{i=1}^{k} \frac{(x/\rho)^i}{i!} \cdot \frac{k-1}{k} \cdots \frac{k+1-i}{k}.
\]

(5.4)

Eq. (5.3) and Eq. (5.4) yield via a term by term comparison that

\[ g_{k+1}(x) > g_k(x), \quad \text{for all } x > 0. \]  

(5.5)

In other words, \( g_{k+1}(1 - \phi_Y(\gamma)) > g_k(1 - \phi_Y(\gamma)) \), with \( g_{k+1}(1 - \phi_Y(1)) = g_k(1 - \phi_Y(1)) = 1 \) and \( g_{k+1}(1 - \phi_Y(0)) = (1 + \rho^{-1})^{k+1} > (1 + \rho^{-1})^k \). Recall that \( \gamma_i \) is the unique fixpoint of the equation \( 1/\gamma = g_i(1 - \phi_Y(\gamma)) \), with \( \gamma_i \in (0, 1), i = k, k+1 \). For \( x < \gamma_i, 1/x > g_i(1 - \phi_Y(x)) \) and for \( x \in (\gamma_i, 1) \), necessarily \( 1/x < g_i(1 - \phi_Y(x)) \), \( i = k, k+1 \). Hence \( \gamma_{k+1} < \gamma_k \).

Since \( \gamma_k \) are non-increasing, and bounded below, the sequence has a limit, \( \gamma^* \) say, with \( \gamma^* < 1 \). This limit solves (5.1) by the standard limiting argument that \( \lim_{k \to \infty} (1 + x/k)^k = e^x \). The function \( \gamma \mapsto e^{-(1-\phi_Y(\gamma))} \) is a convex function, with fixpoint \( s \gamma = 1 \) and \( \gamma^* < 1 \), derivative larger than 1 at \( \gamma = 1 \), and positive value at \( \gamma = 0 \). The result then follows in a standard manner, thus completing the proof of a).

Part b) follows from the fact that

\[
\frac{1 - \gamma}{1 - \phi_Y(\gamma)} = \frac{1}{\sum_{i=1}^{b} p_i \sum_{j=0}^{i-1} \gamma^j},
\]

(5.6)

which is strictly increasing in \( \gamma < 1 \) if and only if \( P\{Y = 1\} < 1 \).

Clearly, with increasing \( k \), the variance of the inter-arrival time decreases. Hence, the average server utilisation strictly improves with decreasing variance, in the case of batch sizes \( Y \), with \( P\{Y = 1\} < 1 \). Whereas, if the batch size equals 1 with probability 1, the average server utilisation is equal to \( \rho \) and thus constant.

We can say a little more. We define \( L_k = \sum_m m \cdot \bar{\pi}_m^k \) to be the mean number of customers in the system under the stationary distribution. Further, we denote by \( W_k \) the expected sojourn time and by \( V_k \) the variance. The following comparison result holds.
Theorem 5.2. The following are true:

\[ L_{k+1} \leq L_k \quad W_{k+1} \leq W_k, \quad \text{and} \quad V_{k+1} \leq V_k, \quad \text{for} \quad k = 1, 2, \ldots. \quad (5.7) \]

Proof. The expectations with respect to \( \bar{\pi}_{k+1} \) and \( \bar{\pi}_k \) are equal to \( L_{k+1} \) and \( L_k \) respectively.

It is easy to check that

\[ L_k = \frac{\rho}{1 - \phi_Y(\gamma_k)}, \quad \text{for} \quad k = 1, 2, \ldots. \quad (5.8) \]

Since \( \gamma_k \) are non-increasing, then

\[ 1 - \phi_Y(\gamma_k) \leq 1 - \phi_Y(\gamma_{k+1}). \]

It follows that \( L_{k+1} \leq L_k \). Application of Little’s formula yields \( W_{k+1} \leq W_k \).

From (4.26) it is clear that the number of customers in the \( E_k/M/1 \)-queue is (almost) geometrically distributed. This implies that

\[ V_k = \rho^2 \gamma_k / (1 - \gamma_k)^2. \]

It is easy to check that this yields \( V_{k+1} \leq V_k \).

Next, we need to check the variance for a general batch size distribution:

\[ V_k = \left( \frac{\rho}{1 - \phi_Y(\gamma_k)} \right)^2 \sum_{m \geq 1} m^2 \gamma_k^{m-1} - L_k^2. \quad (5.9) \]

Applying geometric series sum yields

\[ \sum_{m \geq 1} m^2 \gamma_k^{m-1} = \gamma_k \sum_{m \geq 1} m(m-1)\gamma_k^{m-2} + \sum_{m \geq 1} m \gamma_k^{m-1} \]

\[ = \frac{2\gamma_k}{(1 - \gamma_k)^3} + \frac{1}{(1 - \gamma_k)^2}. \quad (5.10) \]

By combination of (5.8), (5.9), and (5.10)

\[ V_k = \frac{\rho(1 - \gamma_k)^2}{1 - \phi_Y(\gamma_k)} \left( \frac{2\gamma_k}{(1 - \gamma_k)^3} + \frac{1}{(1 - \gamma_k)^2} \right) - \frac{\rho^2}{(1 - \phi_Y(\gamma_k))^2} \]

\[ = \frac{\rho}{1 - \phi_Y(\gamma_k)} \left( \frac{1 + \gamma_k}{1 - \gamma_k} - \frac{\rho}{1 - \phi_Y(\gamma_k)} \right). \quad (5.11) \]

We define \( h(\gamma) := \frac{1 + \gamma}{1 - \gamma} - \frac{\rho}{1 - \phi_Y(\gamma)}. \)

Since \( \gamma_k \geq \gamma^* \) and \( \gamma_k \downarrow \gamma^* \) by virtue of Theorem 5.1 (a), it is sufficient to show that \( h'(\gamma) = \frac{dh(\gamma)}{d\gamma} \geq 0 \) for \( \gamma \in [\gamma^*, 1) \).

\[ h'(\gamma) = \frac{1}{(1 - \gamma)^2} \left( 2 - \rho \left( \frac{1 - \gamma}{1 - \phi_Y(\gamma)} \right)^2 \phi_Y'(\gamma) \right). \]

Applying Eq. (5.6) yields

\[ h'(\gamma) = \frac{1}{(1 - \gamma)^2} \left( 2 - \rho \left( \frac{1}{\sum_{j=1}^b \sum_{i=0}^{j-1} \gamma_i} \right)^2 \sum_{j=1}^b j p_j \gamma^j \right). \quad (5.12) \]

20
Since \( j \gamma^{j-1} \leq \sum_{i=0}^{j-1} \gamma^i \) for \( \gamma \in [\gamma^*, 1) \), it follows that
\[
h'(\gamma) \geq \frac{1}{(1 - \gamma)^2} \left( 2 - \frac{\rho}{\sum_{j=1}^{b} p_j \sum_{i=0}^{j-1} \gamma^i} \right).
\]

For \( \gamma \in [\gamma^*, 1) \),
\[
\frac{\rho}{\sum_{j=1}^{b} p_j \sum_{i=0}^{j-1} \gamma^i} \leq \frac{\rho}{\sum_{j=1}^{b} p_j \sum_{i=0}^{j-1} (\gamma^*)^i}.
\]

So, to show that \( h'(\gamma \geq 0) \) for \( \gamma \in [\gamma^*, 1) \), it is sufficient to show that
\[
2 \geq \left( \frac{\rho}{\sum_{j=1}^{b} p_j \sum_{i=0}^{j-1} (\gamma^*)^i} \right) \left( 1 - \frac{1 - (\gamma^*)^i}{1 - \gamma^*} \right),
\]
or
\[
2(1 - \phi_Y(\gamma^*))/\rho \geq 1 - \gamma^*.
\]

By virtue of Theorem 5.1 (a), \( \gamma^* \geq 1 - (1 - \phi_Y(\gamma^*))/\rho \) (since \( e^{-x} \geq 1 - x \), for \( x \geq 0 \)).

This implies that
\[
(1 - \phi_Y(\gamma^*))/\rho \geq 1 - \gamma^*.
\]

So that (5.13) follows.

Then
\[
\frac{\rho}{1 - \phi_Y(\gamma_k)} h(\gamma_k) \geq \frac{\rho}{1 - \phi_Y(\gamma_{k+1})} h(\gamma_{k+1}).
\]
In other words, we have proved that \( V_k \geq V_{k+1} \).

Interestingly enough, for general batch size distribution, the stationary distribution does not stochastically decrease with increasing \( k \). This follows immediately from the fact that \( \sum_{m \geq 1} \pi_m^k \) is strictly increasing in \( k \), whereas \( \gamma_k \) is strictly decreasing. However, if the batch size is identically equal to 1, then the stationary distribution has a stochastically monotonic behaviour as a function of \( k \).

**Corollary 5.3.** Suppose that \( P\{Y = 1\} = 1 \). The following is true for \( k = 1, 2, \ldots \):
\[
\bar{\pi}_{k+1}^{st} \preceq \bar{\pi}_k,
\]
or equivalently: \( \sum_{m \geq M} \pi_m^{k+1} \leq \sum_{m \geq M} \pi_m^k \), for all \( M = 0, 1, \ldots \).

In this particular case we can also prove that \( \alpha_k \) is strictly increasing in \( k \).

**Lemma 5.4.** Suppose that \( P\{Y = 1\} = 1 \). The sequence of parameters \( \alpha_k \) is strictly increasing in \( k \).

**Proof.** From Eq. (4.18) we have
\[
\alpha_k = \frac{\lambda_k}{\lambda_k + \mu - \mu \gamma_k}.
\]
We define
\[
f_k(x) = \frac{\mu}{\lambda_k} x^{k+1} - (1 + \frac{\mu}{\lambda_k}) x + 1.
\]
Taking the first derivative of $f_k$ yields
\[ f'_k(x) = \frac{\mu(k+1)}{\lambda_k} x^k - (1 + \frac{\mu}{\lambda_k}). \] (5.16)

Then $\alpha^*_k = \left(\frac{1 + \frac{\mu}{\lambda_k}}{(k+1)}\right)^{1/k}$ is the point where the polynomial $f_k(x)$ has a unique minimum.

Thus
\[ \alpha_k \in \left(\frac{1}{1 + \frac{\mu}{\lambda_k}}, \alpha^*_k\right). \] (5.17)

For any $k \geq 1$, we have $\alpha^*_k = \left(\frac{1 + \frac{\mu}{\lambda_k}}{(k+1)}\right)^{1/k} = \left(1 - \frac{k}{k+1}(1 - \rho)\right)^{1/k}.$

We define
\[ g(k) = \left(1 - \frac{k}{k+1}(1 - \rho)\right)^{1/k}. \] (5.18)

We will show that $g(k)$ is strictly increasing in $k$.

\[ g'(k) = g(k) \left(-\frac{1}{k^2} \log \left(1 - \frac{k}{k+1}(1 - \rho)\right) - \frac{1 - \rho}{k(k+1)(1+k\rho)}\right). \] (5.19)

Since $g(k) = \alpha^*_k > 0$, it is sufficient to show that
\[ -\frac{1}{k^2} \log \left(1 - \frac{k}{k+1}(1 - \rho)\right) - \frac{1 - \rho}{k(k+1)(1+k\rho)} > 0. \] (5.20)

We know that: $\frac{1}{k+1} > \frac{1}{(k+1)(1+k\rho)}$, for $k \geq 1$.

This implies
\[ e^{-\frac{1 - \rho}{k(k+1)(1+k\rho)}} > e^{-\frac{1 - \rho}{k^2}} > \left(1 - \frac{k}{k+1}(1 - \rho)\right)^{1/k} \]
\[ -\frac{1}{k^2} \log \left(1 - \frac{k}{k+1}(1 - \rho)\right) > \frac{1 - \rho}{k(k+1)(1+k\rho)}. \] (5.21)

By combining (5.20) and (5.21) it is clear that $g(k)$ is strictly increasing in $k$. This means that $\alpha^*_k < \alpha^*_{k+1}$ and so $\alpha^*_{k+1} \uparrow 1$, for $k \to \infty$.

Next, for any positive $k$ and $x \in (0, 1)$, the functions $f_k$ and $f_{k+1}$ have 2 intersection points: at 0 and 1. This follows from the relation:
\[ f_{k+1}(x) - f_k(x) = \frac{1}{k(k+1)\rho} x (1-x)^2 \sum_{i=1}^{k} ix^{i-1}. \] (5.22)

This also implies that $f_{k+1}(x) > f_k(x)$ for all $x \in (0,1)$. Hence $\alpha_k < \alpha_{k+1} < 1$.

In other words we can say that $\alpha_k$ is strictly increasing in $k$ and the proof is complete. □
5.2 Comparison of batch service queues with Erlang arrivals versus a deterministic inter-arrival time

Let us assume for the moment that the batch size is identically equal to 1, i.e. \( P\{Y = 1\} \). Consider the \( D/M/1 \)-queue with mean inter-arrival time \( 1/\lambda^* \) and mean service time \( 1/\mu \), where \( \lambda^* = \lambda_k/k \).

In this case it is quite well-known that the stationary distribution of the \( E_k/M/1 \)-queue, with the same mean inter-arrival and mean service time distribution, converges setwise and weakly to the stationary distribution of the \( D/M/1 \)-queue, as \( k \to \infty \). We will generalise these results when there are batch services. By virtue of Asmussen [3] and Bhat [5] the stationary distribution \( \bar{\pi}_D \) of the \( D/M/1 \)-queue is given by

\[
\bar{\pi}_D^m = \begin{cases} 
(1 - \sigma)\rho\sigma^{m-1}, & m > 0 \\
1 - \rho, & m = 0, 
\end{cases}
\]

where \( \sigma \) is the unique root smaller than 1 of

\[
\sigma = e^{-(1-\sigma)/\rho}, \quad (5.23)
\]

when \( \rho < 1 \). By virtue of Theorem 5.1, this means that \( \sigma = \gamma^* = \lim_{k \to \infty} \gamma_k \), and in particular \( \sigma < \gamma_k \). It follows directly that:

\[
\bar{\pi}_D \preceq \bar{\pi}_k \preceq \bar{\pi}_{k+1}, \quad k = 1, 2, \ldots \quad (5.24)
\]

Define \( L_D, W_D, \) and \( V_D \) as the mean number of customers, the expected sojourn time, and the variance of customers of the \( D/M/1 \)-queue respectively, under the stationary distribution. We now derive the result below.

**Corollary 5.5.** The following are true:

i) Eq. (5.24) holds for all \( k = 1, \ldots, \) and

ii) \( L_k \downarrow L_D, \quad W_k \downarrow W_D, \) and \( V_k \downarrow V_D, \) as \( k \to \infty, \) monotonically.

Intuitively it seems clear that a similar result holds as well for the \( E_k/M^Y/1 \)- and \( D/M^Y/1 \)-queues, with the same mean inter-arrival times and the same batch service distributions. So far, we have not been able to find any result on (setwise and weak) convergence of the stationary distribution of the \( E_k/M^Y/1 \)-queue to the stationary distribution of the \( D/M^Y/1 \)-queue, although it should be completely similar to convergence results in the case of batch size equal to 1.

**General batch size**  Suppose now again that \( P\{Y = 1\} < 1 \).

**Theorem 5.6.** For the stationary distribution \( \bar{\pi}_D \) of the \( D/M^Y/1 \)-queue the following holds:

i) \( \bar{\pi}_m^k \to \bar{\pi}_m^D, \) for \( m = 0, \ldots \) and

ii) \[
\bar{\pi}_D^m = \begin{cases} 
\frac{\rho(1 - \sigma)^2}{1 - \phi_Y(\sigma)}\sigma^{m-1}, & m > 0, \\
\frac{\rho(1 - \sigma)}{1 - \phi_Y(\sigma)}, & m = 0, 
\end{cases}
\]

(5.25)
where $\sigma$ is the unique root smaller than 1 of

$$
\sigma = e^{-(1-\phi_Y(\sigma))/\rho}, \quad \phi_Y(\sigma) = \sum_{j=1}^{b} p_j \sigma^j. \quad (5.26)
$$

Proof. Since the proof is quite standard, we only provide a sketch of the proof.

First, notice that the embedded processes on arrival instants are of the GI/M/1-type considered in [7, pp. 82–86]. As has been derived there, it follows for the corresponding stationary distribution $\bar{\pi}_D^A$ of the $D/M^Y/1$-queue that

$$
\bar{\pi}_D^{A,m} = (1 - \sigma)\sigma^m,
$$

with $\sigma$ the unique root smaller than 1 of Eq. (5.26).

Secondly, by using semi-regeneration, cf. Asmussen [3, Ch. VII.5 and p. 283], we then obtain formula Eq. (5.25) for the stationary distribution of the non-embedded $D/M^Y/1$-queue. The desired convergence properties then follow from the explicit formulae for the respective stationary distributions.

It directly follows from Theorem 5.6(i) and Theorem 5.2 that the result of Corollary 5.5 holds in the batch service case as well.

**Corollary 5.7.** The monotonicity result in Corollary 5.5 (ii) holds in the case of batch service.

As a consequence, the mean number of customers, expected sojourn time and variance are minimized by deterministic inter-arrival times. This is a well-known result for non-batch systems (cf. Asmussen [3, pp. 336-339]). Results on other performance measures are given in this section as well.

**Acknowledgement**

This research was supported by Directorate General of Human Resource for Science, Technology and Higher Education of the Ministry of National Education of Indonesia.

**A Appendix**

**Lemma A.1.** Let $X$ be a possibly non-conservative, non-explosive, transient, stable Markov process on state space $S$ with q-matrix $Q$. Then $Q^{-1}$ exists and $Q^{-1} = T$ where the $(i,j)$-th element of $T$ is defined as follows: $\tau_{i,j} = \int_0^\infty p_{t,(i,j)} dt < \infty$, where $p_{t,(i,j)}$ are the elements of the transition function $P_t$.

Proof. A generalization of Lemma 2.1 and the proof is analogous. \qed

**Lemma A.2.** Let $X$ be a possibly non-conservative, non-explosive, transient, stable Markov process on state space $S$ with q-matrix $Q$. Let $s \in S$ be given. Consider the transition rate matrix $\tilde{Q}$ on $X \backslash \{s\}$ where the elements are given by:

$$
\tilde{q}_{ij} = q_{ij} + q_{is} \frac{q_{sj}}{q_s},
$$
where \( q_s = -q_{ss} \).

Let \( \tilde{q}^{-1}_{ij} \) denote the \((i,j)\)-th elements of matrices \( Q^{-1} \) and \( \tilde{Q}^{-1} \), respectively. Then the following are true for \( i, j \neq s \):

- \( \tilde{q}^{-1}_{ij} = q^{-1}_{ij} \),
- \( q^{-1}_{sj} = \sum_{r \neq s} \frac{q_{sr}}{q_s} \tilde{q}^{-1}_{rj} \).

**Proof.** By Lemma A.1 the inverse matrices \( Q^{-1} \) and \( \tilde{Q}^{-1} \) exist, and the entries are equal to the expected sojourn time spent in each state of \( S \) and \( S_s = S \setminus \{s\} \) respectively. It is convenient to use a representation based on the jump chain on \( S \) with transition matrix \( P^J \). The entries of \( P^J \) are given by:

\[
p^J_{i,j} = \frac{q_{i,j}}{q_i},
\]

where \( q_i = -q_{i,i} \), for \( i, j \in S \).

Clearly \( P^J \) is a sub-stochastic matrix, since \( Q \) is transient. We denote its \( n \)-th iterate by \( P^{J,n} \). The 0-th iterate is the identity. It follows from Anderson [2], Proposition (4.1.1) (see also Eq. (2.4) of this paper) that:

\[
\tau_{i,j} = \sum_{n \geq 0} P^{J,n}_{i,j} \frac{1}{q_j}.
\]

The jump transition probability matrix \( \tilde{P}^J \) associated with \( \tilde{Q} \) on \( S_s \), is given by

\[
\tilde{P}^J_{i,j} = p^J_{i,j} + p^J_{i,s} p^J_{s,j}.
\]

This is precisely the transition matrix of the jump chain associated with \( Q \), embedded on \( S_s \). By using Eq. (A.2) and the fact that the \((i,j)\)-th element of \(-Q^{-1}\) represents the expected amount of time spent in state \( j \), given a start in state \( i \) before absorption outside \( S \), we directly obtain that \(-\tilde{Q}^{-1}\) is the expected amount of time spent in \( S_s \) before absorption outside this set.

This proves the first statement. For the second statement, we invoke (A.2). Then for \( j \neq s \)

\[
\tau_{s,j} = \sum_{r \neq s} P^{J}_{s,r} \tau_{r,j} = \sum_{r \neq s} \frac{q_{sr}}{q_s} \tau_{r,s}.
\]

The result follows.

**References**


