# Remaining Loads in a PH/M/*c* Queue with Impatient Customers

**Anders Rygh Swensen[1]**

## Abstract

There are three contributions to the literature in this note. Firstly, we point out that under some weak conditions the result in Swensen (1986) on the remaining loads in a GI/M/*c* queue with impatient customers, derived under the assumption that the distribution of inter-arrival times is Coxian, is also valid for the more general phase type distribution. From this result the distributions of the virtual and the actual waiting time can easily be obtained. Secondly, the relation to an alternative expression for the distribution of the virtual waiting time derived by Kawanishi and Takine (2015) is also clarified. Finally, we explain how the results on the remaining loads can be used to find distributions describing particular fixed servers and provide a couple of numerical examples of how this can be done.

## 1 Introduction

A queue where the customers that cannot be served within a fixed time leave the system is usually referred to as a queue with impatient customers or a queue with bounded waiting time. To describe such systems balking and reneging have also been used as denominations, see Ancker and Gafarian (1963).

The motivation for studying such queues has often been problems in telecommunication where the main challenge is to dimension the system so that on the one hand not too many calls are lost and on the other hand not having too many servers, which means that more equipment than necessary is used. An early contribution with the possibility that the customers hang up and leave the system is Palm (1957). Also notice that bounding the waiting time can act as a simple load control.

✉   Anders Rygh Swensen
     swensen@math.uio.no

[1]   Department of Mathematics, University of Oslo, P.O. Box 1053, N-0316 Blindern, Oslo, Norway

Another field where models allowing impatient customers arise is perishable inventory theory. In blood banks the supply of blood portions is random and shelf time is limited, see Perry and Stadje (2022) for some discussion and also the monographs of Nahmias (2012) and Karaesmen et al. (2011). In contrast to the previous situation the impatience time is usually less manageable and must often be considered as fixed.

The operation of call centers is a third area where impatience has to be taken into account. This is stressed in several papers by Mandelbaum and coauthors, see e.g. Mandelbaum and Zeltyn (2009). The crucial question here is to avoid that the center is understaffed, rejecting too many customers, or overstaffed such that the proportion of time the servers are unoccupied is larger than necessary.

In all areas of application mentioned above there is a trade-off between loosing too many customers and having too many idle servers. Knowledge of the distribution of the waiting times and loads is helpful for choosing an appropriate impatience time, when that is possible. For managerial decisions and design of this kind of systems, especially deciding on the number of servers, such insight is crucial.

In this note we will consider a first-come first-served queue with $c$ servers where the service times are independent exponential random variables with expectation $1/\mu$. The arrivals of new customers follow a renewal process where the distribution of the inter-arrival times is of the phase type. Customers whose waiting time exceeds the fixed non-random impatience time $\tau$ leave the system. This system is often denoted as PH/M/c+D.

The main contribution of the present note is that in a PH/M/$c$ queue with impatient customers where the impatient time is deterministic, explicit solutions, up to some matrix inversions, exist for the stationary asymptotic distribution of the remaining loads and consequently for the virtual and the actual waiting time. It turns out that the procedure proposed by Swensen (1986) with a Coxian distribution describing the renewal input process, can be extended, with some adjustments, to the more general case where a phase type distribution is used. The key result for the generalization can be found in Lemma 2 in Appendix B.

Relevant papers dealing with impatient customers are Choi et al. (2004) and Kawanishi and Takine (2016) who considered the more general Markovian arrival process, MAP/M/$c$ queue with bounded waiting or impatience time. He and Wu (2020) treated the situation where the impatience time is not constant. Kawanishi and Takine (2015) derived the stationary distribution for the virtual waiting time in a PH/M/$c$ queues with impatient customers. It is worth noticing that the results in Choi et al. (2004) and Kawanishi and Takine (2015, 2016) are complementary to the results in the present paper. The distribution of the virtual waiting time in those papers is derived directly by solving a second order differential equation. The distributions of the waiting times, virtual and actual, in Swensen (1986) are derived as an implication of the distribution of the loads of the servers. Thus, the results and methods of proof are different and a natural question is how they are related. We will discuss this issue. In particular we compare the density for the virtual waiting time derived by Kawanishi and Takine (2015) with the expression of the present note.

Although the distribution of the virtual waiting time is important, knowledge of the simultaneous distribution of the remaining loads of the servers represents an additional value. For example, one can focus on the simultaneous remaining loads of a particular group of servers, cf. the famous Palm-Jacobæus formula in the telecommunication literature.

It is also worth remarking that the methods for the numerical computations of the distributions in the two cases are quite different. In papers by Choi et al. (2004) and Kawanishi and Takine (2015, 2016) the expressions that are used involve matrix exponentials. On the other hand, the formula for the probability density of the remaining loads in the present

paper is quite easy to evaluate when a procedure for computing matrix eigenvalues and eigenvectors is available.

The note is organized as follows. The distribution of the remaining loads for phase type distributed arrival times is derived in the next section. In section 3 the relation between the implied distribution of the virtual waiting time relying on the distribution of the remaining load and the more direct version in Kawanishi and Takine (2015) is discussed. In section 4 the solution of a linear equation which is important for finding the asymptotic distribution of the remaining loads is considered. In section 5 we discuss how the results on the distribution of the remaining loads can be used to find distributions for fixed particular servers and carry out the actual computations in an exampel. In the Appendices the necessary modifications of the Kolmogorov forward equations for the remaining loads for phase type distributed inter-arrival times are explained. The stationary solution to these equations is found and some lemmas crucial for the derivation are proved.

We use the following notation, as in Swensen (1986): a phase type distribution is described by the pair $(\gamma, T)$ where $\gamma = (\gamma_1, \ldots, \gamma_m)$ is a probability vector, i.e. $\gamma_1, \ldots, \gamma_m \geq 0$, $\sum_{i=1}^m = 1$, and $T$ is a non-singular $m \times m$ matrix with negative diagonal elements and non-negative off-diagonal elements. The Coxian distributions define the subclass of the phase type distributions where $\gamma = (1, 0, \ldots, 0)$ and the only nonnegative elements of the matrix $T = \{t_{ij}\}$ are the elements $t_{i(i+1)}$, $i = 1, \ldots, m - 1$.

To a large extent the arguments used in the treatment of the Coxian case carry over to the phase type case. Therefore there will be some overlap with Swensen (1986) when the arguments are verified in the phase type setup.

In the following we consider the scalar field of complex numbers and matrices whose elements also are complex numbers.

## 2 Distribution of Remaining Loads

Consider the process $(Y(t), V_1(t), \ldots, V_c(t))$ where $V_i(t)$ is the unfinished work, i.e. the remaining load, for sever $i = 1, \ldots, c$ and $Y(t)$ is the phase of the arrival process. Under the assumptions that the service times are independent, exponentially distributed and that the arrival process is a renewal process where the distribution of the inter-arrival times is of the phase type, the process is Markovian.

The phase type distribution of the inter-arrival times satisfies

### Assumption 1

(i)   The m solutions $\{\eta_1, \ldots, \eta_m\}$ of $\det[c\mu\mathbf{e} \cdot \gamma + T - c\eta I_m] = 0$ are distinct and are not equal to any of the eigenvalues of $T/c$.
(ii)  The matrix $T + \mathbf{T} \cdot \gamma$ where $\mathbf{T} = -T\mathbf{e}, \mathbf{e} = (1, \cdots, 1)'$, is irreducible and has distinct eigenvalues.

*Remark 1.* The claim in Assumption 1 that there are no equal eigenvalues of $\mu\mathbf{e} \cdot \gamma + T/c$ and no equal eigenvalues of $T + \mathbf{T} \cdot \gamma$ simplifies the derivations since there is only one eigenvector associated with each eigenvalue. We therefore exclude the situation where the dimension of the eigenspace of an eigenvalue is strictly less than the algebraic multiplicity of the eigenvalue, i.e. where the geometric multiplicity is less than the algebraic multiplicity, see Horn and Johnson (2013).

The $m$ solutions $\{\eta_1, \ldots, \eta_m\}$ are sorted in decreasing order according to size of the moduli. For a conjugate pair the number with positive imaginary part is counted first. Also, define B as the matrix where the rows are the left normalized eigenvectors of $T + \mathbf{T} \cdot \boldsymbol{\gamma}$. Let D be the diagonal matrix with diagonal elements $(c\mu - \kappa_\ell)^{-1}$, $\ell = 1, \ldots, m$ where $\kappa_\ell$, $\ell = 1, \ldots, m$ are the eigenvalues of $T + \mathbf{T} \cdot \boldsymbol{\gamma}$ and where one, $\kappa_1$ say, is equal to 0. Let E be the diagonal matrix with diagonal elements $\exp(-\tau c \eta_\ell)$, $\ell = 1, \ldots, m$.

The following theorem is a description of the density of the asymptotic distribution of the remaining loads.

**Theorem 1** *Consider a first-come first-served GI/M/c queuing model with waiting time bounded by $\tau$, service rate $\mu$ and inter-arrival distribution of the phase type $(\boldsymbol{\gamma}, T)$. Under Assumption* 1 *the stationary asymptotic distribution of the remaining loads $V_1, \ldots, V_c$ has density*

$$\mu^{c-1} \exp(-\mu(v_1 + \cdots + v_c) + c\mu(\tau \wedge \min v_i)) \sum_{k=1}^{m} \delta_k \underline{y}_c^k \mathbf{e} \exp(-c\eta_k(\tau \wedge \min v_i)) \qquad (1)$$

*for $\min V_i > 0$ where $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_m)$ is the solution of*

$$\boldsymbol{\delta} E Y_{c-1}[c\mu \mathbf{T} \cdot \boldsymbol{\gamma} B^{-1} D B - (c-1)\mu I_m + (c-1)\mu \mathbf{e} \cdot \boldsymbol{\gamma} + T] = 0 \qquad (2)$$

*and*

$$\sum_{k=1}^{m} \delta_k \{ \sum_{i=0}^{c-1} \binom{c}{i} \underline{y}_i^k \mathbf{e} + \underline{y}_c^k \mathbf{e}[(\frac{1}{\mu} - \frac{1}{\eta_k}) \exp(-c\eta_k \tau) + \frac{1}{\eta_k}] \} = 1. \qquad (3)$$

*The matrix $Y_{c-1}$ has rows $\underline{y}_{c-1}^1, \ldots, \underline{y}_{c-1}^m$ where $\underline{y}_{c-1}^k$ is the solution of the homogenous linear equation*

$$\underline{y}_{c-1}^k[(c-1)\mu I_m - (c-1)\mu \mathbf{e} \cdot \boldsymbol{\gamma} - T - c\mu R_k] = 0$$

*with $y_{c-11}^k = 1$ and the $m \times m$ matrix $R_k = \mathbf{T} \cdot \boldsymbol{\gamma}[c\eta_k I_m - T]^{-1}$, $k = 1, \ldots, m$.*

Furthermore, when $c = 1$, $\underline{y}_1^k = c\mu \underline{y}_0^k R_k$. When $c > 1$, $\underline{y}_c^k = c\mu \underline{y}_{c-1}^k R_k$ and $\underline{y}_0^k, \ldots, \underline{y}_{c-2}^k$, $k = 1, \ldots, m$ are defined by the recursion

$$\underline{y}_i^k = (c-i)\mu \underline{y}_{i+1}^k (i\mu I_m - i\mu \mathbf{e} \cdot \boldsymbol{\gamma} - T)^{-1}, \quad i = c-2, \ldots, 1$$
$$\underline{y}_0^k = -c\mu \underline{y}_1^k T^{-1}.$$

*Remark 2.* From Lemma 3 it follows that the matrices $[(c-1)\mu I_m - (c-1)\mu \mathbf{e} \cdot \boldsymbol{\gamma} - T - c\mu R_k]$ have rank m-1 such that $\underline{y}_{c-1}^k$, $k = 1, \ldots, m$ are determined up to a normalizing constant. Also it follows frem Lemma 1 in Neuts (1982) that when $c > 1$, the matrices $(i\mu I_m - i\mu \mathbf{e} \cdot \boldsymbol{\gamma} - T)$, $i = 1, \ldots, c-2$ are nonsingular.

The proof of the theorem is an elaboration of Swensen (1986) for the case where the inter-arrival distribution is Coxian. In the first step to prove the main result we show, in Appendix A, that when $t \to \infty$ the functions

$$q_{k0j} = y_{0j}^k, \quad j = 1, \ldots, m, \qquad (4)$$

$$q_{kij}(v_1, \ldots, v_i) = \mu^i y_{ij}^k \exp(-\mu(v_1 + \cdots + v_i)), \; i = 1, \ldots, c-1, \; j = 1, \ldots, m, \qquad (5)$$

$$q_{kcj}(v_1, \ldots, v_c) = \mu^{c-1} y_{cj}^k \exp(-\mu(v_1 + \cdots + v_c) + c(\mu - \eta_k)v_{(1)}), \; j = 1, \ldots, m, \qquad (6)$$

for $k = 1, \ldots, m$, satisfy the differential equation defined by the Kolmogorov forward equations in the region $v_{(1)} = \min v_i \leq \tau$. Since $\eta_1, \ldots, \eta_m$ are distinct the solutions are independent and the general solution in this region therefore has the form $\mathbf{Q} = \sum_{k=1}^m \delta_k \mathbf{Q}_k$.

As a second step, the function $\mathbf{Q}$ is inserted in the Kolmogorov forward equation to obtain a differential equation for values in the region $\min v_i > \tau$. In Appendix A it is shown that the solution has the form

$$K \exp(-\mu(v_1 + \cdots + v_c)) \boldsymbol{\gamma} B^{-1} D B$$

where $K = \sum_{k=1}^m \delta_k c \mu^c y_{c-1}^k \mathbf{T} \exp(c(\mu - \eta_k)\tau)$.

Finally, in the third step the distribution of the remaining loads is determined by requiring that the solutions for $\min v_i \leq \tau$ and $\min v_i > \tau$ are continuous at $\min v_i = \tau$.

## 3 Distribution of the Virtual Waiting Time

Building on the results in Choi et al. (2004) and Kawanishi and Takine (2015) found an explicit expression for the stationary joint distribution of the phase of the arrival process and the virtual waiting time. The density of the stationary virtual waiting time has the form

$$\mathbf{v}(v)\mathbf{e} = (v_1(v), \ldots, v_m(v))\mathbf{e} = p\hat{\mathbf{v}} \exp[(c\mu\mathbf{e} \cdot \boldsymbol{\gamma} + T)(\tau - v)]\mathbf{e}, \; 0 < v < \tau \qquad (7)$$

where $\hat{\mathbf{v}} = (\hat{v}_1, \ldots, \hat{v}_m) = \boldsymbol{\gamma}(c\mu I - T)^{-1}/\boldsymbol{\gamma}(c\mu I - T)^{-1}\mathbf{e}$, and $p$ is the normalizing constant given in Kawanishi and Takine (2015).

However, an expression for the density of the stationary virtual waiting time can also be obtained from the density of the remaining loads. Differentiating the expression for the complimentary distribution function in Swensen (1986) one gets the probability density of virtual waiting time when Assumption 1 is satisfied

$$f(v) = \sum_{k=1}^m \delta_k (y_c^k \mathbf{e}) \, c \exp(-c\eta_k v), \; 0 < v < \tau \qquad (8)$$

where $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_m)$ is defined in Theorem 1. By the same arguments as in the next section it follows that the expression (8) has an imaginary part equal to zero.

The following proposition summarizes the relation between the two expressions of the density.

**Proposition 1** *If Assumption 1 holds the coefficients of (8) are given by*

$$\delta_k (y_c^k \mathbf{e}) = p \exp(c\eta_k \tau) \sum_{j,\ell=1}^m \hat{v}_j f_{j,k} f^{k\ell} / c, \; k = 1, \ldots, m \qquad (9)$$

*where $c\eta_1, \ldots, c\eta_m$ are the eigenvalues of $c\mu\mathbf{e} \cdot \boldsymbol{\gamma} + T$, $F = \{f_{j\ell}\}_{j,\ell=1}^m$ is the matrix where the columns are the right eigenvectors of the matrix $c\mu\mathbf{e} \cdot \boldsymbol{\gamma} + T$ and $F^{-1} = \{f^{j\ell}\}_{j,\ell=1}^m$.*

*Also*

$$p = c \sum_{k=1}^{m} \delta_k(y_c^k \mathbf{e}) \exp(-c\eta_k \tau). \qquad (10)$$

*Proof.* Assumption 1 *i*) implies that $c\mu\mathbf{e} \cdot \boldsymbol{\gamma} + T$ is diagonalizable, that is $c\mu\mathbf{e} \cdot \boldsymbol{\gamma} + T = F \,\mathrm{diag}\,(c\eta_1, \ldots, c\eta_m)F^{-1}$ where the m eigenvalues $\{c\eta_1, \ldots, c\eta_m\}$ are sorted in decreasing order according to the size of the moduli. From elementary properties of the exponential matrix it follows that

$$
\begin{aligned}
p\hat{\mathbf{v}} &\exp\left[(c\mu\mathbf{e} \cdot \boldsymbol{\gamma} + T)(\tau - v)\right]\mathbf{e} \\
&= p\hat{\mathbf{v}}F \exp\left[\,\mathrm{diag}\,(c\eta_1(\tau - v), \ldots, c\eta_m(\tau - v))\right]F^{-1}\mathbf{e} \\
&= p\hat{\mathbf{v}}F \,\mathrm{diag}\,(\exp(c\eta_1(\tau - v)), \ldots, \exp(c\eta_m(\tau - v)))F^{-1}\mathbf{e} \\
&= p \sum_{k=1}^{m}[\exp(c\tau\eta_k) \sum_{j,\ell=1}^{m} \hat{v}_j f_{j,k} f^{k\ell}] \exp(-c\eta_k v),\ 0 < v < \tau.
\end{aligned}
$$

Hence by comparing the coefficients of $\exp(-c\eta_1), \ldots, \exp(-c\eta_m)$,

$$\delta_k(y_c^k \mathbf{e}) = p \exp(c\tau\eta_k) \sum_{j,\ell=1}^{m} \hat{v}_j f_{j,k} f^{k\ell} /c,\ k = 1, \ldots, m$$

which is (9). This shows how the density (8) for the virtual waiting time can be expressed using the results (7) from Kawanishi and Takine (2015).

Furthermore, summing over $k$, since $\sum_{k=1}^{m} f_{j,k} f^{k\ell} = 1$ if $j = \ell$ and 0 if $j \neq \ell$,

$$
\begin{aligned}
\sum_{k=1}^{m} \delta_k(y_c^k \mathbf{e}) \exp(-c\tau\eta_k) &= p \sum_{j,\ell=1}^{m} \hat{v}_j \sum_{k=1}^{m} f_{j,k} f^{k\ell} /c = p \sum_{j=1}^{m} \hat{v}_j /c = p\hat{\mathbf{v}}\mathbf{e}/c \\
&= \frac{1}{c}p\boldsymbol{\gamma}(c\mu I - T)^{-1}\mathbf{e}/\boldsymbol{\gamma}(c\mu I - T)^{-1}\mathbf{e} = p/c
\end{aligned}
$$

which is (10). ∎

*Remark 3.* For numerical computations the representation in (8) yields an alternative to the approach using matrix exponentials suggested by Kawanishi and Takine (2015). With a procedure computing eigenvalues and eigenvectors the quantities $\delta_1, \ldots, \delta_m$ can be found as described in Theorem 1.

## 4 Solution of the Linear Equations (2) and (3)

The linear equations (2) and (3) can be solved directly by writing them on the form $A\boldsymbol{\delta}' = e_{m+1}$ where $A$ is a $(m + 1) \times m$ matrix and $e_{m+1}$ is a $(m + 1)$-dimensional vector with all elements equal to 0 except the last which equals 1. Then, using a $QR$ factorization, as described in Theorem 2.1.14 in Horn and Johnson (2013), there exist a $(m + 1) \times m$ matrix $Q$ with orthonormal columns and an upper triangular $m \times m$ matrix $R$ such that the linear equations may be expressed as $QR\boldsymbol{\delta}' = e_{m+1}$ from which $\boldsymbol{\delta}$ is easily found.

But there is an alternative way to obtain a solution starting with solving

$$\boldsymbol{\phi}[c\mu\mathbf{T}\cdot\boldsymbol{\gamma}B^{-1}DB - (c-1)\mu I_m + (c-1)\mu\mathbf{e}\cdot\boldsymbol{\gamma} + T] = 0 \tag{11}$$

which is determined up to a normalization by Lemma 4. Normalizing the first element of $\boldsymbol{\phi}$ as 1, it follows from Lemma 1 that all elements of $\boldsymbol{\phi}$ are real.

**Lemma 1** *The solutions of Eq.* (11) *have real elements*.

*Proof.* If $\boldsymbol{\kappa}$ is the diagonal matrix with diagonal elements $\kappa_1, \dots, \kappa_m$, $B(T + \mathbf{T}\cdot\boldsymbol{\gamma})B^{-1} = \boldsymbol{\kappa}$. Furthermore, $Bc\mu I_m B^{-1} = c\mu I_m$. Subtracting $B(-T - \mathbf{T}\cdot\boldsymbol{\gamma} + c\mu I_m)$ $B^{-1} = c\mu I_m - \boldsymbol{\kappa}$. Then $B(c\mu I_m - T - \mathbf{T}\cdot\boldsymbol{\gamma})^{-1}B^{-1} = (c\mu I_m - \boldsymbol{\kappa})^{-1}$ and $(c\mu I_m - T - \mathbf{T}\cdot\boldsymbol{\gamma})^{-1}$ $= B^{-1}(c\mu I_m - \boldsymbol{\kappa})^{-1}B = B^{-1}DB$. Thus since the matrix $(c\mu I_m - T - \mathbf{T}\cdot\boldsymbol{\gamma})$ has real elements, so has the matrix $B^{-1}DB$ and the matrix $c\mu\mathbf{T}\cdot\boldsymbol{\gamma}B^{-1}DB - (c-1)\mu I_m + (c-1)\mu\mathbf{e}\cdot\boldsymbol{\gamma} + T$. Therefore also the solutions of (11) must be real. ∎

Remark that $\boldsymbol{\phi} = \delta EY_{c-1}$. The matrix $Y_{c-1}$ defined in Theorem 1 has rows corresponding to $\eta_1, \dots, \eta_m$. Then $\bar{Y}_{c-1}$ is the rows corresponding to $\bar{\eta}_1, \dots, \bar{\eta}_m$ where $\bar{\eta}_i$ is the conjugate of $\eta_i$ if this number is not real. Let $\delta$ and $\tilde{\delta}$ be the solutions of equation (2). Since the elements of $\boldsymbol{\phi}$ are real

$$\boldsymbol{\phi} = \delta Y_{c-1}E = \tilde{\delta}\bar{Y}_{c-1}\bar{E}.$$

The matrices $\bar{Y}_{c-1}$ and $\bar{E}$ are nonsingular, such that $\bar{\delta} = \tilde{\delta}$. This means that in the solutions of equation (3) the complex elements occur in conjugate pairs. But then

$$\sum_{k=1}^m \delta_k\{\sum_{i=0}^{c-1}\binom{c}{i}\underline{y}_i^k\mathbf{e} + \underline{y}_c^k\mathbf{e}[(\frac{1}{\mu} - \frac{1}{\eta_k})\exp(-c\eta_k\tau) + \frac{1}{\eta_k}]\}$$
$$= \sum_{k=1}^m \delta_k\underline{y}_{c-1}^k\{\sum_{i=0}^{c-1}\binom{c}{i}\prod_{\ell=i}^{c-2}(c-\ell)\mu^{c-\ell+1}(\ell\mu I_m - \ell\mu\mathbf{e}\cdot\boldsymbol{\gamma} - T)^{-1} \tag{12}$$
$$+ cI_m + c\mu R_k[(\frac{1}{\mu} - \frac{1}{\eta_k})\exp(-c\eta_k\tau) + \frac{1}{\eta_k}]\}\mathbf{e}$$

must be a real number, since conjugate pairs of $\delta_i$ correspond to conjugate pairs of other terms in the sum (12).

But the sum of a complex number and its conjugate must be real. Therefore (12) is real and can be used for normalizing $\delta$.

## 5 Computation of the Loads for a Particular Group of Servers

Theorem 1 describes the distribution of the remaining load of the $c$ servers on the open set $< 0, \infty >^c$. As pointed out in Remark 1 in Swensen (1986) the distribution on $[0, \infty >^c$ can easily be obtained. This result can be used to find expressions for probabilities and distributions for particular fixed servers. In addition to the distribution of the load we will consider two such probabilities as a function of the threshold $\tau$: the probability that a particular server is occupied and the probability that customers are rejected at the particular server. Denote the load of this server by $V_{p1}$. The main argument for finding the distribution is, for $V_{p2}$ the load at another fixed particular server,

$$P(V_{p1} > v) = P(V_{p1} = \min_{1 \le i \le m} V_i > v) + P(V_{p1} > v, V_{p1} > \min_{1 \le i \le m} V_i)$$

$$= \frac{1}{c} P(\min_{1 \le i \le m} V_i > v) + (c-1) P(V_{p1} > v, V_{p2} = \min_{1 \le i \le m} V_i > v)$$

$$+ P(V_{p1} > v, \min_{1 \le i \le m} V_i = 0).$$

The probabilities on the right-hand side follow from Theorem 1. More details can be found in Lemma A.7 in Swensen (1986).

In Fig. 1 the probability that a particular server is busy, i.e. $P(V_{p1} > 0)$, is shown as a function of the bound $\tau$ for a queue with 4 servers.

The distributions of the inter-arrival distributions which are considered are an exponential distribution and three phase type distributions with intensity matrix

$$\begin{pmatrix} -16.0 & 4.0 & b \\ d & -2.0 & a \\ c & 0 & -0.4 \end{pmatrix}.$$

The case with $a = 0.4, b = c = d = 0.0$ corresponds to the Coxian distribution considered in Swensen (1986).

The probability is plotted in Fig. 1 for four distributions which all have expectation 0.3125. In addition to the exponential distribution there are 3 phase type distributions; PH1: $a = 0.346, b = d = 0.0, c = 0.5$ with coefficient of variation 2.98, PH2: $a = 0.07, b = 0.5, c = 0.8, d = 0.0$ with coefficient of variation 2.84, PH3: $a = 0.05, b = 0.5, c = 0.0, d = 0.8$ with coefficient of variation 2.59. The intensity of the server distribution, $\mu$, is taken equal to 1.
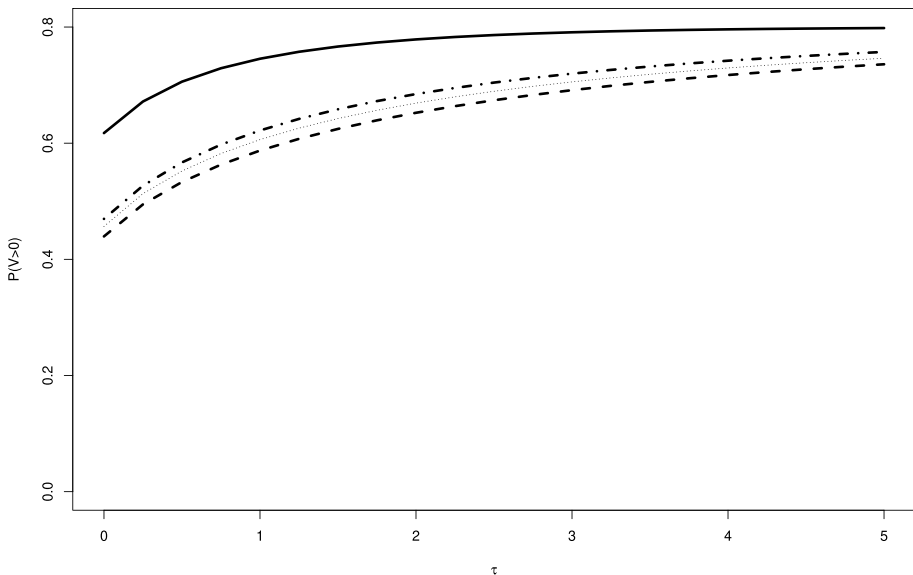


**Fig. 1** Probability of server 1 being occupied as function of $\tau$. Exponential inter-arrivals: solid line, PH1 inter-arrivals: dashed line, PH2 inter-arrivals: dotted line, PH3 inter-arrivals: dot-dashed line
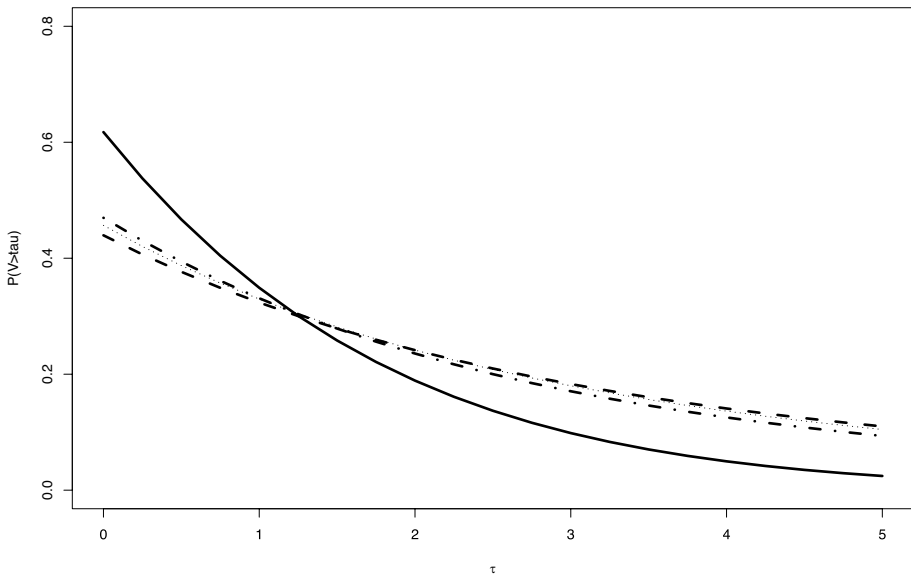
**Fig. 2** Probability of customer being rejected at server 1 as function of $\tau$. Exponential inter-arrivals: solid line, PH1 inter-arrivals: dashed line, PH2 inter-arrivals: dotted line, PH3 inter-arrivals: dot-dashed line

As one can see in Fig. 1, $P(V_{p1} > 0)$, the probability of server 1 being busy, is, not surprisingly, increasing in $\tau$. This follows from the fact that larger value of $\tau$ means that there are more customers in the system. Therefore the probability that the server is idle is less and the probability of being busy is larger.

In Fig. 2 the probability that a particular server has a load such that no new customer is allowed at this server, i.e. $P(V_{p1} > \tau)$, is shown as a function of the bound $\tau$ for the a queue with 4 servers for the same queuing systems as those used in Fig. 1. The decreasing probability is due to the fact that as $\tau$ increases fewer customer are rejected, and the probability of being rejected therefore becomes smaller.

As $\tau \to 0$, $\lim P(V_{p1} > \tau) = P(V_{p1} > 0)$ for all inter-arrival distributions. In particular, in the situation displayed in Figs. 1 and 2 for the exponential distribution, the system reduces to Erlang's loss system when $\tau = 0$. The probability that a fixed server is busy is then given by the Palm-Jacobæus formula which in this case equals 0.617, see page 63 in Asmussen (1987).

The calculations were done using the software package R, R Core Team (2020).

# Appendix

## A. Stationary Solutions of the Kolmogorov Forward Equations

We will use the following notation $T = \{t_{jl}\}_{j,l=1}^m$, $t_{jj} = -\lambda_j$, $t_{jl} = \lambda_j \beta_{jl}$, $j \neq l$ for $j, l = 1 \ldots, m$. Also $\beta_{jj} = 0$ and $\alpha_j = 1 - \sum_{l=1}^m \beta_{jl}$ for $j = 1, \ldots, m$ such that $\mathbf{T} = (\lambda_1 \alpha_1, \ldots, \lambda_m \alpha_m)'$.

The forward Kolmogorov equations are a generalization of those used in Swensen (1986). Due to the restrictions on $\gamma$ and $T$ in the Coxian case some modifications are necessary when a general phase type distribution is used to describe the arrival process.

Let $p_{ij}(t + h, v_1, \dots, v_i)$ be the density of the remaining load for $i$ occupied servers, $i$, $i = 0, \dots, c$, and arrival process in state $j$, $j = 1, \dots, m$ at time $t + h$. Letting first $t \to \infty$ and then $h \to 0$ it follows that the following equations must be satisfied by the stationary solutions

$$[cp_{11}(0), \dots, cp_{1m}(0)] = -[\sum_{l=1}^{m} p_{0l}t_{l1}, \dots, \sum_{l=1}^{m} p_{0l}t_{lm}], \tag{13}$$

$$
\begin{aligned}
-\sum_{l=1}^{i} \frac{dp_{ij}}{dv_l} = {} & \sum_{l=1}^{m} p_{il}(v_1, \dots, v_i)t_{lj} \\
& + \frac{\gamma_j}{(c - i + 1)} \sum_{l=1}^{m} \sum_{k=1}^{i} p_{i-1l}(v_1, \dots, v_{k-1}, v_{k+1}, \dots, v_i)\lambda_l\alpha_l\mu \exp(-\mu v_k) \\
& + (c - i)p_{i+1j}(v_1, \dots, v_i, 0)
\end{aligned}
\tag{14}
$$

for $i = 1, \dots, c - 1$ when $c > 1$ and

$$
\begin{aligned}
-\sum_{l=1}^{c} \frac{dp_{cj}}{dv_l} = {} & \sum_{l=1}^{m} p_{cl}(v_1, \dots, v_c)t_{lj} \\
& + \gamma_j \sum_{l=1}^{m} \sum_{k=1}^{c} p_{c-1l}(v_1, \dots, v_{k-1}, v_{k+1}, \dots, v_c)\lambda_l\alpha_l\mu \exp(-\mu v_k) \\
& + \gamma_j \sum_{l=1}^{m} \sum_{k=1}^{c} \int_0^{\tau \wedge \min v_i} p_{cl}(v_1, \dots, v_{k-1}, v, v_{k+1}, \dots, v_c)\lambda_l\alpha_l\mu \exp(-\mu(v_k - v))dv \\
& + \gamma_j I(\tau < \min v_i) \sum_{l=1}^{m} p_{cl}(v_1, \dots, v_c)\lambda_l\alpha_l
\end{aligned}
\tag{15}
$$

for $j = 1, \dots, m$.

Firstly, we claim that a solution of the previous equations has the form defined by the Eqs. (4)–(6) in the region $\min v_i \leq \tau$. As in Swensen (1986) where the similar situation is treated for the Coxian distributions the verification is by straight forward insertion in (13) and (14) for $i = 1, \dots, c - 1$.

For the case $i = c$ the insertion of Eqs. (5) and (6) in the left- and right-hand side of Eq. (15) yields that the following equations must be satisfied for $k = 1, \dots, m$

$$\frac{\mu}{\mu - \eta_k} \underline{y}^k_c \mathbf{T} = c\mu \underline{y}^k_{c-1} \mathbf{T} \tag{16}$$

and

$$\underline{y}^k_c [c\eta I_m - T] = \frac{\mu[\underline{y}^k_c \mathbf{T}]\boldsymbol{\gamma}}{\mu - \eta_k}. \tag{17}$$

It follows from Lemma 2 that $\eta_k$ is a solution of $(\mu - \eta_k)/\mu = \boldsymbol{\gamma}[c\eta_k I_m - T]^{-1}\mathbf{T}$. Therefore the Eq. (16) may be written

$$\underline{y}^k_c \mathbf{T} = c\mu \underline{y}^k_{c-1} \mathbf{T} \cdot \boldsymbol{\gamma}[c\eta_k I_m - T]^{-1}\mathbf{T} = c\mu \underline{y}^k_{c-1} R_k \mathbf{T}, \tag{18}$$

where the last equality follows from the definition of $R_k$. But the definition of $\underline{y}^k_c$ and $\underline{y}^k_{-c-1}$ implies (18).

To verify (17) note that from the definition of $\underline{y}^k_{-c-1}$ and $R_k$ it follows that

$$\underline{y}^k_{-c}[c\eta_k I_m - T] = c\mu \underline{y}^k_{-c-1} R_k[c\eta_k I_m - T] = c\mu \underline{y}^k_{-c-1} \mathbf{T} \cdot \boldsymbol{\gamma}$$

such that the Eq. (17) may be written, using the definition of $\underline{y}^k_{-c-1}$ and $\underline{y}^k_{-c}$

$$c\mu \underline{y}^k_{-c-1} \mathbf{T} \cdot \boldsymbol{\gamma} = \frac{\mu}{(\mu - \eta_k)} \underline{y}^k_{-c} \mathbf{T} \cdot \boldsymbol{\gamma}. \tag{19}$$

But from the definition of $\underline{y}^k_{-c}$, $\underline{y}^k_{-c-1}$ and Lemma 2 it follows that

$$\underline{y}^k_{-c} \mathbf{T} = c\mu \underline{y}^k_{-c-1} \mathbf{T} \cdot \boldsymbol{\gamma}[c\eta_k I - T]^{-1}\mathbf{T} = c\mu \underline{y}^k_{-c-1} \mathbf{T}\frac{(\mu - \eta_k)}{\mu} \tag{20}$$

which implies (19) and therefore (17). Thus the proposed solutions given by Eqs. (4)–(6) satisfy the requirements from Eqs. (13)–(15) in the area $\min v_i \leq \tau$ for each $\eta_1, \dots, \eta_m$.

Secondly, the general solution will be a linear combination of these solutions. Inserting such a solution in the Eq. (15) one gets when $\min v_i > \tau$ the following equation on vector form

$$(-\sum_{l=1}^c \frac{dp_{c1}}{dv_l}, \dots, -\sum_{l=1}^c \frac{dp_{cm}}{dv_l}) - (p_{c1}, \dots, p_{cm})(T + \mathbf{T} \cdot \boldsymbol{\gamma}) = K\boldsymbol{\gamma} \exp(-\mu(v_1 + \dots + v_c)) \tag{21}$$

where by using (20)

$$K = \sum_{k=1}^m \delta_k \underline{y}^k_{-c} \mathbf{T}\mu^c \exp(c(\mu - \eta_k)\tau))/(\mu - \eta_k) = c\sum_{k=1}^m \delta_k \underline{y}^k_{-c-1} \mathbf{T}\mu^c \exp(c(\mu - \eta_k)\tau)).$$

Let $\kappa_1, \dots, \kappa_m$ be the eigenvalues of $T + \mathbf{T} \cdot \boldsymbol{\gamma}$, $B = (b'_1, \dots, b'_m)'$ be the matrix with the normalized left eigenvectors of $T + \mathbf{T} \cdot \boldsymbol{\gamma}$ as rows and let D be the diagonal matrix with diagonal elements $1/(c\mu - \kappa_1), \dots, 1/(c\mu - \kappa_m)$. Since $T + \mathbf{T} \cdot \boldsymbol{\gamma}$ is a generator the Perron-Frobenius theorem implies that one eigenvalue, $\kappa_1 = 0$ say, is equal to zero and that the other eigenvalues have strictly negative real parts. Also by Assumption 1 *ii*) $\kappa_1, \dots, \kappa_m$ are distinct. Since $\kappa_2, \dots, \kappa_m$ have negative real parts they are not equal to $c\mu > 0$.

As in Swensen (1986) one can show that

$$\psi_j(T + \mathbf{T} \cdot \boldsymbol{\gamma}) = b_j(T + \mathbf{T} \cdot \boldsymbol{\gamma}) \exp(-\kappa_j v_1)\phi_j = \kappa_j b_j \exp(-\kappa_j v_1)\phi_j = -\sum_{l=1}^c \frac{d\psi_j}{dv_l}$$

where $\phi_1, \dots, \phi_m$ are arbitrary differentiable functions in one variable and $\psi_j(v_1 + \dots + v_c) = b_j \exp(-\kappa_j v_1)\phi(-(c-1)v_1 + v_2 + \dots + v_c)$, $j = 1, \dots, m$. Therefore $\psi_1, \dots, \psi_m$ are $m$ independent solutions of the homogenous differential equation corresponding to (21).

Now define,

$$\psi_0(v_1, \dots, v_c) = K \exp(-\mu(v_1 + \dots + v_c))\boldsymbol{\gamma} B^{-1}DB.$$

Then

$$-\sum_{l=1}^{c} \frac{d\psi_0}{dv_l} = c\mu K \exp(-\mu(v_1 + \cdots + v_c))\gamma B^{-1} DB.$$

Thus, if $\kappa$ is the diagonal matrix with diagonal elements $\kappa_1, \ldots, \kappa_m$,

$$-\sum_{l=1}^{c} \frac{d\psi_0}{dv_l} - \psi_0(T + \mathbf{T} \cdot \gamma) = K \exp(-\mu(v_1 + \cdots + v_c))\gamma B^{-1}(c\mu D - D\kappa)B$$

$$= K\gamma \exp(-\mu(v_1 + \cdots + v_c))$$

such that $\psi_0$ is a particular solution of (21) and the general solution is $\psi_0 + \sum_{l=1}^{m} v_l \psi_l$.

Since $\kappa_1 = 0$ and $\kappa_2, \ldots, \kappa_m$ have strictly negative real parts, $v_1 = \cdots = v_m = 0$ because $p_{cj} \to 0$, $j = 1, \ldots m$ as $\sum_{l=1}^{c} v_l^2 \to \infty$.

Finally, from the Kolmogorov forward equations it follows that

$$p_{cj}(v_1, \ldots, v_c) = \lim_{h \to 0} p_{cj}(v_1 + h, \ldots, v_c + h), \ j = 1, \ldots m$$

such that for $(v_1, \ldots, v_c) \in \{(v_1, \ldots, v_c) : \min v_i = \tau\}$

$$\sum_{k=1}^{m} \delta_k c\mu \underline{y}_{-c-1}^{k} \exp(c(\mu - \eta_k)\tau)\mathbf{T} \cdot \gamma B^{-1} DB = \sum_{k=1}^{m} \delta_k \underline{y}_{-c}^{k} \exp(c(\mu - \eta_k)\tau).$$

Using the definition of $\underline{y}_{-c}^{k}$ and letting $Y_{c-1}$ be the matrix with rows $\underline{y}_{-c-1}^{k}$, $k = 1, \ldots, m$ this equals

$$\delta E Y_{c-1}[c\mu \mathbf{T} \cdot \gamma B^{-1} DB - (c-1)\mu I_m + (c-1)\mu \mathbf{e} \cdot \gamma + T] = 0. \tag{22}$$

The matrix $E$ has full rank and so has $Y_{c-1}$ by Lemma 5. Hence it follows from Lemma 4 that Eq. (22) has a solution up to a multiplicative constant, which is found by normalization.

## B. Some Lemmas

**Lemma 2** *Under Assumption 1 i) the set of solutions $\eta_1, \ldots, \eta_m$ of the equation*

$$\det[c\mu \mathbf{e} \cdot \gamma + T - c\eta I_m] = 0 \tag{23}$$

*is contained in the set of solutions of*

$$\frac{\mu - \eta}{\mu} = \gamma[c\eta I_m - T]^{-1}\mathbf{T}.$$

*Proof.* Let $\eta$ be a solution of Eq. (23). From Assumption 1 i) $\eta$ is different from the eigenvalues of $T/c$ so that $(c\eta I_m - T)$ is invertible. Then

$$f(c\eta) = \gamma(c\eta I_m - T)^{-1}(-T)\mathbf{e} = \gamma(c\eta I_m - T)^{-1}(c\eta I_m - T - c\eta I_m)\mathbf{e}$$

$$= \gamma \mathbf{e} - c\eta\gamma(c\eta I_m - T)^{-1}\mathbf{e} = 1 - c\eta\gamma(c\eta I_m - T)^{-1}\mathbf{e}.$$

Using Cauchy's formula for the determinant of a rank-one perturbation, see Horn and Johnson (2013) p. 26, the relation

$$0 = \det[c\mu\mathbf{e} \cdot \boldsymbol{\gamma} + T - c\eta I_m] = \det(T - c\eta I_m)(1 + c\mu\boldsymbol{\gamma}(T - c\eta I_m)^{-1}\mathbf{e})$$

implies that

$$\frac{1}{\mu} = -c\boldsymbol{\gamma}(T - c\eta I_m)^{-1}\mathbf{e}.$$

Hence

$$\frac{\mu - \eta}{\mu} = 1 - \frac{\eta}{\mu} = 1 - c\eta\boldsymbol{\gamma}(c\eta I_m - T)^{-1}\mathbf{e} = f(c\eta) = \boldsymbol{\gamma}(c\eta I_m - T)^{-1}\mathbf{T}. \quad \blacksquare$$

**Lemma 3** *The matrices*

$$(c - 1)\mu I_m - (c - 1)\mu\mathbf{e} \cdot \boldsymbol{\gamma} - T - c\mu R_k, \ k = 1, \dots, m \tag{24}$$

*have rank m-1.*

*Proof.* As pointed out in Swensen ([1986](#)) it suffices to prove that the matrix ([24](#)) is singular. The case where $R_k - I$ is non-singular is proved as in Swensen ([1986](#)). To deal with the case where $R_k - I$ is singular note that from Lemma [2](#) it follows that $R_k\mathbf{T} = \mathbf{T} \cdot \boldsymbol{\gamma}[c\eta_k I_m - T]^{-1}\mathbf{T} = [(\mu - \eta_k)/\mu]\mathbf{T}$ so $\mathbf{T}$ is a right eigenvector of $R_k$ and $(\mu - \eta_k)/\mu$ is the non-zero eigenvalue. When $R_k - I$ is singular, 1 has to be the non-zero eigenvalue of $R_k$ such that $1 = (\mu - \eta_k)/\mu$ and $\eta_k = 0$. Furthermore, for $\eta_k = 0$ and using Cauchy's formula for the determinant of a rank-one perturbation

$$0 = \det(c\mu\mathbf{e} \cdot \boldsymbol{\gamma} + T - c\eta_k I_m) = \det(c\mu\mathbf{e} \cdot \boldsymbol{\gamma} + T) = (\det T)(1 + c\mu\boldsymbol{\gamma}T^{-1}\mathbf{e})$$

such that $c\mu\boldsymbol{\gamma}T^{-1}\mathbf{e} = -1$ and

$$[(c - 1)\mu I_m - (c - 1)\mu\mathbf{e} \cdot \boldsymbol{\gamma} - T - c\mu R_k]\mathbf{e} = (-T + c\mu\mathbf{T} \cdot \boldsymbol{\gamma}T^{-1})\mathbf{e} = (-T\mathbf{e}) - \mathbf{T}) = 0,$$

which shows that the matrix ([24](#)) must be singular also when $R_k - I$ is singular. $\blacksquare$

**Lemma 4** *Let $\kappa_1, \dots, \kappa_m$ be the eigenvalues of $T + \mathbf{T} \cdot \boldsymbol{\gamma}$ and B the matrix with rows as normalized left eigenvectors of $T + \mathbf{T} \cdot \boldsymbol{\gamma}$ and let D be the diagonal matrix with diagonal elements $1/(c\mu - \kappa_1), \dots, 1/(c\mu - \kappa_m)$. Under Assumption [1](#) ii) the matrix*

$$c\mu\mathbf{T} \cdot \boldsymbol{\gamma}B^{-1}DB - (c - 1)\mu I_m + (c - 1)\mu\mathbf{e} \cdot \boldsymbol{\gamma} + T$$

*has rank m-1.*

*Proof.* From Lemma 1 in Neuts ([1982](#)) it follows that the matrix $(c - 1)\mu I_m - (c - 1)\mu\mathbf{e} \cdot \boldsymbol{\gamma} - T$ has rank m. The matrix $c\mu\mathbf{T} \cdot \boldsymbol{\gamma}B^{-1}DB - (c - 1)\mu I_m + (c - 1)\mu\mathbf{e} \cdot \boldsymbol{\gamma} + T$ therefore has rank at least m-1. Hence, it suffices to show that it is singular.

Let $\boldsymbol{\kappa}$ be the diagonal matrix with diagonal elements $\kappa_1, \dots, \kappa_m$. Without loss of generality we can assume $\kappa_1 = 0$. Then

$$B(T + \mathbf{T} \cdot \boldsymbol{\gamma})\mathbf{e} = \boldsymbol{\kappa}B\mathbf{e},$$

$$BT\mathbf{e} + B\mathbf{T} \cdot \boldsymbol{\gamma}\mathbf{e} = \boldsymbol{\kappa}B\mathbf{e},$$

$$0 = -B\mathbf{T} + B\mathbf{T} = \kappa B\mathbf{e}.$$

but $\kappa_i \neq 0$ such that $b_i\mathbf{e} = 0$ for the ith row of B, $b_i$, $i = 2, \ldots, m$ and $B\mathbf{e} = (\mathbf{e}'b_1', 0, \cdots, 0)'$ $= a(1, 0, \cdots, 0)' = a\mathbf{e}_1$. This implies that $\mathbf{e} = B^{-1}B\mathbf{e} = aB^{-1}\mathbf{e}_1$ and

$$1 = \boldsymbol{\gamma}\mathbf{e} = a\boldsymbol{\gamma}B^{-1}\mathbf{e}_1. \tag{25}$$

Also,

$$DB\mathbf{e} = Da\mathbf{e}_1 = \frac{a}{c\mu}\mathbf{e}_1.$$

Hence,

$$c\mu\mathbf{T} \cdot \boldsymbol{\gamma}B^{-1}DB\mathbf{e} = c\mu\mathbf{T} \cdot \boldsymbol{\gamma}B^{-1}\frac{a}{c\mu}\mathbf{e}_1$$

$$= c\mu\mathbf{T} \cdot (a\boldsymbol{\gamma}B^{-1}\mathbf{e}_1)\frac{1}{c\mu} = \mathbf{T} = -T\mathbf{e}$$

by (25), which shows that $c\mu\mathbf{T} \cdot \boldsymbol{\gamma}B^{-1}DB - (c-1)\mu I_m + (c-1)\mu\mathbf{e} \cdot \boldsymbol{\gamma} + T$ is singular. ∎

**Lemma 5** *The vectors $y_{c-1}^k$, $k = 1 \ldots, m$ are linearly independent.*

*Proof.* See Lemma A.6 in Swensen (1986). The argument for the Coxian distributions is also valid for phase type distributions. ∎

**Data Availability**  There are no data in this article.

## Declarations

## References

Ancker CJ, Gafarian AV (1963) Some queuing problems with balking and reneging. Oper Res 11:88–100. https://doi.org/10.1287/opre.11.1.88

Asmussen S (1987) Applied probability and queues. Wiley, New York. https://doi.org/10.1007/b97236

Choi B, Kim B, Zhu D (2004) MAP/M/c queue with constant impatient time. Math Oper Res 29:309–325. https://doi.org/10.1287/moor.1030.0081

He Q-M, Wu H (2020) Multi-layer MMFF processes and the MAP/PH/K+GI queue: Theory and algorithms. Queueing Models Serv Manage 3:37–87

Horn RA, Johnson CR (2013) Matrix analysis, 2nd edn. Cambridge University Press, New York

Karaesmen I, Scheller-Wolf A, Deniz B (2011) Managing perishable and aging inventories: Review and future research directions. In: Kempf K, Keskinocak A, Uzsoy P (eds.) Handbook of Production Planning, vol. 1, pp 393–436. Kluwer, Academic Publishers. https://doi.org/10.1007/978-1-4419-6485-4-15

Kawanishi K, Takine T (2015) A note on the virtual waiting time in the stationary PH/M/c+D queue. J Appl Probab 52:899–903. https://doi.org/10.1239/jap/1445543855

Kawanishi K, Takine T (2016) MAP/M/c and M/PH/c queues with constant impatience times. Queuing Syst 82:381–420. https://doi.org/10.1007/s11134-015-9455-9

Mandelbaum A, Zeltyn S (2009) Staffing many-server queues with impatient customers: Constraint satisfaction in call centers. Oper Res 57:1189–1205. https://doi.org/10.1287/opre.1080.0651

Nahmias S (2012) Perishable inventory systems. International Series in Operations Research, Springer, New York

Neuts MF (1982) Explicit steady-state solutions to some elementary queuing models. Oper Res 30:480–489. https://doi.org/10.1287/opre.30.3.480

Palm RCA (1957) Research on telephone traffic carried by full availability groups. Tele 1

Perry D, Stadje W (2022) Perishable inventories and queues with impatience. Queuing Syst 100:237–239. https://doi.org/10.1007/s11134-022-09787-2

R Core Team (2020) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Swensen AR (1986) On a GI/M/c queue with bounded waiting times. Oper Res 34:895–908. https://doi.org/10.1287/opre.34.6.895