J. Korean Math. Soc. **38** (2001), No. 4, pp. 807–842

A RECENT PROGRESS IN ALGORITHMIC ANALYSIS OF FIFO QUEUES WITH MARKOVIAN ARRIVAL STREAMS

TETSUYA TAKINE

ABSTRACT. This paper summarizes recent developments of analytical and algorithmical results for stationary FIFO queues with multiple Markovian arrival streams, where service time distributions are general and they may differ for different arrival streams. While this kind of queues naturally arises in considering queues with a superposition of independent phase-type arrivals, the conventional approach based on the queue length dynamics (i.e., M/G/1 paradigm) is not applicable to this kind of queues. On the contrary, the workload process has a Markovian property, so that it is analytically tractable. This paper first reviews the results for the stationary distributions of the amount of work-in-system, actual waiting time and sojourn time, all of which were obtained in the last six years by the author. Further this paper shows an alternative approach, recently developed by the author, to analyze the joint queue length distribution based on the waiting time distribution. An emphasis is placed on how to construct a numerically feasible recursion to compute the stationary queue length mass function.

1. Introduction

This paper provides a survey of a recent progress in the algorithmic analysis of work-conserving FIFO single-server queues with (possibly correlated) multiple non-Poissonian arrival streams governed by a continuous-time finite-state Markov chain. A particular feature of these queueing models is that the service time distribution of customers may be different for different streams.

Section 2 describes the arrival process considered in this paper, called Markovian arrival stream (MAS) [8], and provides some properties of

Received October 28, 2000.

2000 Mathematics Subject Classification: 60K25, 60J25, 60J27.

Key words and phrases: FIFO queue, Markovian arrival stream, waiting time, queue length.

MAS. Note that the counting process associated with MAS follows a marked MAP (Markovian arrival process) [17, 18], which is considered as an extension of MAP [26] for a single arrival stream to (possibly correlated) multiple arrival streams. Note that MAP is a class of semi-Markovian arrival processes and includes Markov modulated Poisson processes, phase-type renewal processes and the superposition of those arrival processes as special cases. In MAS input, the service time distributions of customers from different arrival streams can be different.

When service times of customers from all arrival streams are independent and identically distributed (i.i.d.) with a common distribution function, the stochastic behavior of the total number of customers in the system and the state of the underlying Markov chain immediately after departures can be formulated as a bivariate Markov chain of M/G/1 type and algorithmic procedures to compute performance metrics are known as M/G/1 paradigm [28]. This framework is described in section 3, together with an application to the MAP/GI/1 queue.

When we allow different service time distributions for different arrival streams, it causes the dependency of service times on the state of the underlying Markov chain at the time of arrival. Thus the bivariate process of the total number of customers and the state of the underlying Markov chain immediately after departures no longer forms a Markov chain, and therefore the conventional M/G/1 paradigm is not applicable to analyze such a queue. In fact, to construct a Markov chain representing the queue length dynamics, we have to keep track of from which arrival stream each customer in the system arrived. Without doubt, such a Markov chain is too complicated to obtain any analytical results. Note also that even in a special case of i.i.d. service times, M/G/1 paradigm does not enable us to characterize the *joint* queue length distribution.

The author has been tackling this kind of queues for several years. Contrary to the queue length process, the workload process (i.e., the process representing the amount of work-in-system) has a Markovian property. Namely, for a broad class of queues with MAS input, the amount of work-in-system (or some related stochastic process depending on the queueing discipline) is characterized as a bivariate Markov process with the spatial homogeneity and the skip-free to the left property, which is viewed as a continuous analogue of a Markov chain of M/G/1 type, and an algorithmic approach to such a process (i.e., a continuous version of the M/G/1 paradigm) was established in [36, 41]. Further the closed-form formula for the amount of work-in-system was recently found in [45]. Section 4 provides some of those results. Related

studies on work-in-system (or waiting time) in queues with multiple non-Poissonian arrival streams having different service time distributions can be found in [7, 30, 50].

Further the author recently succeeded in obtaining a purely algorithmic procedure to compute the stationary joint queue length distribution in the work-conserving FIFO queue with MAS input [43]. Some of the results in [43] are considered as the relationship between the time-average joint queue length distribution and the customer-average joint queue length distribution at departures, and it was recently proved to hold for a broad class of queues with MAS input [47]. Moreover, for a class of FIFO queues, this relationship is specialized to obtain a distributional form of Little's law, i.e., relationship between the time-average joint queue length distribution and the sojourn time distributions of customers from respective arrival streams [47]. Note that only for stationary FIFO M/G/1 queues, a distributional form of Little's law has already been known in the literature [22]. These results are summarized in section 5.1.

The distributional form of Little's law suggests a new approach to analyze the stationary joint queue length distribution based on the stationary sojourn time distribution. In fact, we can construct a numerically feasible recursion to compute the stationary joint queue length mass function [43, 47]. First, with the uniformization technique [49], each Stieltjes integral of a matrix exponential with respect to the sojourn time distribution function is rewritten to be an infinite series of matrices. By doing so, the problem is reduced to find coefficients vectors of this series. It can be shown that those coefficient vectors are identical to the steady-state solution of a certain Markov chain of M/G/1 type and the latter can numerically be obtained by the conventional M/G/1 paradigm. Section 5.3 summarizes this approach.

2. MAS: Markovian arrival stream [43, 47]

This section describes the arrival process called *Markovian arrival stream* (MAS). There exist K arrival streams, where customers who arrive from the kth ($k = 1, \ldots, K$) arrival stream are called class k customers. Customer arrivals are governed by a time-homogeneous, stationary Markov chain with finite state space $\{1, \ldots, M\}$, which is called the underlying Markov chain hereafter. The underlying Markov chain is assumed to be irreducible.

The underlying Markov chain stays in state i $(i = 1, \ldots, M)$ for an exponential interval of time with finite mean μ_i^{-1}. When the sojourn time in state i has elapsed, the underlying Markov chain changes its state to state j without any arrivals with probability $\sigma_{i,j}^{(0)}$ $(j = 1, \ldots, M,$ $j \neq i)$. Also, with probability $\sigma_{i,j}^{(k)}$ $(j = 1, \ldots, M,\ k = 1, \ldots, K)$, the underlying Markov chain changes its state to state j and a class k customer arrives. It is assumed that for each k $(k = 1, \ldots, K)$, there exist some i and j such that $\sigma_{i,j}^{(k)} > 0$, so that arrivals of class k customers are certain. Note that

$$\sum_{\substack{j=1 \\ j \neq i}}^{M} \sigma_{i,j}^{(0)} + \sum_{k=1}^{K} \sum_{j=1}^{M} \sigma_{i,j}^{(k)} = 1, \qquad i = 1, \ldots, M.$$

Service times of class k customers who arrive with a state transition from state i to state j are i.i.d. according to a distribution function $h_{k,i,j}(x)$ $(k = 1, \ldots, K,\ i, j = 1, \ldots, M)$, with finite mean. Without loss of generality, we assume that $h_{k,i,j}(0) = 0$ for k, i, j such that $\sigma_{i,j}^{(k)} > 0$.

We now introduce a formal representation of MAS. Let $\boldsymbol{C}$ denote an $M \times M$ matrix whose (i, j)th element $C_{i,j}$ is given by

$$C_{i,j} = \begin{cases} \mu_i \sigma_{i,j}^{(0)}, & i \neq j,\ i, j = 1, \ldots, M, \\ -\mu_i, & i = j, i = 1, \ldots, M. \end{cases}$$

Also let $\boldsymbol{D}_k(x)$ $(k = 1, \ldots, K,\ x \geq 0)$ denote an $M \times M$ matrix whose (i, j)th element $D_{k,i,j}(x)$ is given by

$$D_{k,i,j}(x) = \mu_i \sigma_{i,j}^{(k)} h_{k,i,j}(x), \qquad k = 1, \ldots, K,\ i, j = 1, \ldots, M.$$

Thus MAS has representation $(\boldsymbol{C}, \boldsymbol{D}_1(x), \ldots, \boldsymbol{D}_K(x))$.

DEFINITION 2.1. Let $\mathcal{S}$ denote a jointly stationary sequence $\{(T_n, K_n, H_n);\ n = 1, 2, \ldots\}$, where T_n, K_n and H_n denote the length of time interval between the $n-1$st and nth transitions, the class index, and the service time of the nth customer, respectively, generated by a Markov chain with finite state space $\{1, \ldots, M\}$ and an infinitesimal generator $\boldsymbol{C} + \boldsymbol{D}_1 + \cdots + \boldsymbol{D}_K$, and distribution functions $\{h_{k,i,j}(x)\}$ $(k = 1, \ldots, K,$ $i, j = 1, \ldots, M)$ on $(0, \infty)$. Associated with this, we denote the marked point process $\mathcal{N}$ which is defined on $[0, \infty) \times \{1, \ldots, K\} \times (0, \infty)$. MAS with representation $(\boldsymbol{C}, \boldsymbol{D}_1(x), \ldots, \boldsymbol{D}_K(x))$ is then defined as either a stationary sequence $\mathcal{S}$ or its associated point process $\mathcal{N}$.

We now consider the superposition of independent MASs. For example, suppose there exist two independent MASs with representation $(\widetilde{\boldsymbol{C}}_1, \widetilde{\boldsymbol{D}}_{1,1}(x), \ldots, \widetilde{\boldsymbol{D}}_{1,K_1}(x))$ and $(\widetilde{\boldsymbol{C}}_2, \widetilde{\boldsymbol{D}}_{2,1}(x), \ldots, \widetilde{\boldsymbol{D}}_{2,K_2}(x))$, where $\widetilde{\boldsymbol{C}}_j$ and $\widetilde{\boldsymbol{D}}_{j,k_j}(x)$ $(j = 1, 2,\ k_j = 1, \ldots, K_j)$ are $M_j \times M_j$ matrices. Then the superposed arrival process of these independent MASs is again an MAS with representation $(\boldsymbol{C}, \boldsymbol{D}_{1,1}(x), \ldots, \boldsymbol{D}_{1,K_1}(x), \boldsymbol{D}_{2,K}(x), \ldots, \boldsymbol{D}_{2,K_2}(x))$, where with $M_j \times M_j$ identity matrix $\boldsymbol{I}_{M_j}$ $(j = 1, 2)$

$$\begin{aligned} \boldsymbol{C} &= \widetilde{\boldsymbol{C}}_1 \oplus \widetilde{\boldsymbol{C}}_2, \\ \boldsymbol{D}_{1,k_1}(x) &= \widetilde{\boldsymbol{D}}_{1,k_1}(x) \otimes \boldsymbol{I}_{M_2}, \qquad k_1 = 1, \ldots, K_1, \\ \boldsymbol{D}_{2,k_2}(x) &= \boldsymbol{I}_{M_1} \otimes \widetilde{\boldsymbol{D}}_{2,k_2}(x), \qquad k_2 = 1, \ldots, K_2. \end{aligned}$$

Here $\otimes$ and $\oplus$ denote Kronecker product and Kronecker sum, respectively [15]. Thus we have the following result.

PROPOSITION 2.1. *The class of MAS is close under superposition of independent streams.*

The generality of MAS is clearly stated in the following theorem which is considered as a little adaptation of Theorem 1 in [8].

THEOREM 2.1 ([8]). *The class of MAS is weakly dense in the class of stationary marked point processes. That is,*

(a) *for a given $\mathcal{S}$, there exists a sequence $\{\mathcal{S}^{(m)};\ m = 1, 2, \ldots\}$ generated by MASs such that the $\mathcal{S}^{(m)}$ converges to $\mathcal{S}$ in distribution as m goes to infinity, and*

(b) *for a given $\mathcal{N}$, there exists a sequence $\{\mathcal{N}^{(m)};\ m = 1, 2, \ldots\}$ generated by MASs such that the $\mathcal{N}^{(m)}$ converges to $\mathcal{N}$ in distribution as m goes to infinity.*

Theorem 2.1 implies that, for example, for any ergodic and stationary G/G/1 queue with MPP input, there exists a sequence of stationary MASs such that their actual waiting times converges in distribution to the actual waiting time of the G/G/1 queue.

Let $A_k(t)$ and $U_k(t)$ denote the number and the total amount of workload of class k customers arriving during an interval $(0, t]$, and let $S(t)$ denote the state of the underlying Markov chain at time t. Then the (i, j)th element of

$$\exp\left[\left(\boldsymbol{C} + \sum_{k=1}^{K} z_k \boldsymbol{D}_k^*(s)\right) t\right]$$

represents

$$E\left[z_1^{A_1(t)}\cdots z_K^{A_K(t)}e^{-s_1U_1(t)}\cdots e^{-s_KU_k(t)}1(S(t)=j)\,\middle|\,1(S(0)=i)\right],$$

where $1(\chi)$ denotes the indicator function of event χ and $\boldsymbol{D}_k^*(s_k)$ denotes the Laplace-Stieltjes transform of $\boldsymbol{D}_k(x)$:

$$\boldsymbol{D}_k^*(s)=\int_0^\infty \exp(-sx)d\boldsymbol{D}_k(x), \qquad Re(s)>0, k=1,\ldots,K.$$

We now introduce some notations related to MAS with representation $(\boldsymbol{C},\boldsymbol{D}_1(x),\ldots,\boldsymbol{D}_K(x))$. Let $\boldsymbol{D}(x)$ denote $\boldsymbol{D}_1(x)+\cdots+\boldsymbol{D}_K(x)$ and $\boldsymbol{D}^*(s)$ denote the LST of $\boldsymbol{D}(x)$, i.e., $\boldsymbol{D}^*(s)=\boldsymbol{D}_1^*(s)+\cdots+\boldsymbol{D}_K^*(s)$. We denote $\boldsymbol{D}_k(\infty)=\boldsymbol{D}_k^*(0+)$ by $\boldsymbol{D}_k$ and its (i,j)th element by $D_{k,i,j}$. Further we define $\boldsymbol{D}$ as $\boldsymbol{D}(\infty)=\boldsymbol{D}^*(0+)$. The infinitesimal generator of the underlying Markov chain is then given by $\boldsymbol{C}+\boldsymbol{D}$. Let λ_k $(k=1,\ldots,K)$ denote the arrival rate of class k customers, which is given by

$$\lambda_k=\boldsymbol{\pi}\boldsymbol{D}_k\boldsymbol{e}, \qquad k=1,\ldots,K, \tag{1}$$

where $\boldsymbol{e}$ denotes an $M\times 1$ vector whose elements are all equal to one and $\boldsymbol{\pi}$ denotes the stationary probability vector of the underlying Markov chain. Note that $\boldsymbol{\pi}$ satisfies

$$\boldsymbol{\pi}(\boldsymbol{C}+\boldsymbol{D})=\boldsymbol{0}, \qquad \boldsymbol{\pi}\boldsymbol{e}=1.$$

Further let ρ_k $(k=1,\ldots,K)$ denote the utilization factor of class k customers. Note that ρ_k is given by

$$\rho_k=\boldsymbol{\pi}\int_0^\infty x d\boldsymbol{D}_k(x)\boldsymbol{e}.$$

Let ρ be $\rho_1+\cdots+\rho_K$. Note that

$$\rho=\boldsymbol{\pi}\int_0^\infty x d\boldsymbol{D}(x)\boldsymbol{e}. \tag{2}$$

In the rest of this paper, it is assumed that $\rho<1$, which ensures the stability of the queue [24, 36], and the queue is in steady state.

REMARK 2.1. The counting process of MAS is called marked MAP (Markovian arrival process) [17, 18], which is considered as an extension of MAP [26] for a single arrival stream to (possibly correlated) multiple arrival streams. Thus a marked MAP has a representation $(\boldsymbol{C},\boldsymbol{D}_1,\ldots,\boldsymbol{D}_K)$ and a MAP has a representation $(\boldsymbol{C},\boldsymbol{D})$.

3. M/G/1 paradigm and its application to MAP/GI/1 queue

In this section we introduce Markov chains of M/G/1 type and its solution method, known as M/G/1 paradigm. Note that the (imbedded) queue length processes in queues with MAS input *do not* fall into this category unless service times of all customers are i.i.d.. As you will see in section 5.3, however, we still rely on M/G/1 paradigm when we develop a numerically feasible algorithm to compute the joint queue length mass function in a FIFO single-server queue with MAS input.

3.1. Markov chains of M/G/1 type and M/G/1 paradigm [28]

Consider a nonpreemptive, work-conserving single server queue, where Q_n denotes the number of customers in the queue immediately after the departure of the nth customer and A_n denote the number of customers arriving during the service time of the nth customer. We then have

$$Q_{n+1} = \max(Q_n - 1, 0) + A_{n+1}, \quad n = 0, 1, \ldots. \tag{3}$$

In particular, we consider a queue, where customers of class k ($k = 1, \ldots, K$) arrive according to a Poisson process with rate λ_k and service times of class k customers are independent, and identically distributed according to a general distribution function $h_k(x)$ with finite mean h_k. Letting

$$\lambda = \sum_{k=1}^{K} \lambda_k, \qquad h(x) = \sum_{k=1}^{K} \lambda_k h_k(x)/\lambda, \qquad h = \sum_{k=1}^{K} \lambda_k h_k/\lambda,$$

then this queue is considered as an M/GI/1 queue with arrival rate λ and the service time distribution $h(x)$ having finite mean h, so that $\{A_n\}$ is a sequence of i.i.d. random variables. Thus $\{Q_n\}$ forms a Markov chain with a transition probability matrix

$$\begin{pmatrix} a_0 & a_1 & a_2 & a_3 & \ldots \\ a_0 & a_1 & a_2 & a_3 & \ldots \\ 0 & a_0 & a_1 & a_2 & \ldots \\ 0 & 0 & a_0 & a_1 & \ldots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}, \tag{4}$$

where a_k ($k = 0, 1, \ldots$) denotes $\Pr(A_n = k)$, i.e.,

$$a_k = \int_0^\infty e^{-\lambda x} \frac{(\lambda x)^k}{k!} dh(x), \qquad k = 0, 1, \ldots.$$

When $\rho = \lambda h < 1$, this Markov chain has the stationary state probabilities $q_k = \Pr(Q = k)$ $(k = 0, 1, \ldots)$, where Q denotes a generic random variable for the Markov chain $\{Q_n,\ n = 0, 1, \ldots\}$. Namely, $q_0 = 1 - \rho$ and q_k $(k = 1, 2, \ldots)$ is recursively computed by

$$q_k = \left(q_0 \overline{a}_k + \sum_{j=1}^{k-1} q_j \overline{a}_{k-j+1} \right) / (1 - \overline{a}_1),$$

where $\overline{a}_k = \sum_{j=k}^{\infty} a_j$. Note that the q_k is of customer-average, and in the M/GI/1, it is identical to the time-average distribution of the number of customers in the system.

Next we consider (3) under the assumption that customer arrivals follow a MAP with representation $(\boldsymbol{C}, \boldsymbol{D})$ and service times of customers are i.i.d. according to a distribution function $h(x)$. Note that the A_n is not a sequence of i.i.d. random variables. However, A_n is conditionally independent of A_m $(m < n)$ given that Q_{n-1} and the state of the underlying Markov chain immediately after the departure of the $n-1$st customer. Thus in a MAP/GI/1 queue, we define S_n^D as the state of the underlying Markov chain immediately after the departure of the nth customer. Then the $\{(Q_n, S_n^D),\ n = 0, 1, \ldots\}$ forms a bivariate Markov chain whose transition probability matrix takes a form:

$$\begin{pmatrix} \boldsymbol{B}_0 & \boldsymbol{B}_1 & \boldsymbol{B}_2 & \boldsymbol{B}_3 & \cdots \\ \boldsymbol{A}_0 & \boldsymbol{A}_1 & \boldsymbol{A}_2 & \boldsymbol{A}_3 & \cdots \\ \boldsymbol{O} & \boldsymbol{A}_0 & \boldsymbol{A}_1 & \boldsymbol{A}_2 & \cdots \\ \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{A}_0 & \boldsymbol{A}_1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}, \tag{5}$$

where the $\boldsymbol{A}_k$ $(k = 0, 1, \ldots)$ denotes a sequence of $M \times M$ matrices which satisfies

$$\boldsymbol{A}^*(z) = \sum_{k=0}^{\infty} \boldsymbol{A}_k z^k = \int_0^{\infty} \exp\left[(\boldsymbol{C} + z\boldsymbol{D})\, x\right] dh(x),$$

and $\boldsymbol{B}_k$ $(k = 0, 1, \ldots)$ is given by $(-\boldsymbol{C})^{-1}\boldsymbol{D}\boldsymbol{A}_k$. Note that the (i,j)th element of $\boldsymbol{A}_k$ represents $\Pr(A_n = k, S_n^D = j \mid S_{n-1}^D = i, Q_{n-1} > 0)$ and the (i,j)th element of $(-\boldsymbol{C})^{-1}\boldsymbol{D}$ appeared in $\boldsymbol{B}_k$ represents the conditional probability that the underlying Markov chain is in state j immediately after the first arrival after time 0 given the underlying Markov chain being in state i at time 0.

We note that the transition probability matrix (5) is considered as the generalization of (4) for the M/GI/1 queue, where each element in the former is replaced by a matrix.

DEFINITION 3.1. A time-homogeneous, bivariate Markov chain $\{(Q_n, S_n^D);\ n = 0, 1, \ldots\}$ with state space $\{0, 1, \ldots\} \times \{1, \ldots, M\}$ is called a Markov chain of M/G/1 type, if (i) Q_n is skip free to the left, i.e., it can decrease at most by one with a one-step transition, and (ii) transition probabilities are spatially homogeneous except for the boundary, i.e., for every $k = 0, 1, \ldots$ and $i, j = 1, \ldots, M$, transition probabilities

$$\Pr((Q_{n+1}, S_{n+1}^D) = (m - 1 + k, j) \mid (Q_n, S_n^D) = (m, i))$$

do not depend on m for all $m = 1, 2, \ldots$.

Thus a Markov chain of M/G/1 type has a transition probability matrix of the form (5).

Utilizing properties (i) and (ii) in Definition 3.1, we can numerically compute the stationary probabilities of $\{(Q_n, S_n^D);\ n = 0, 1, \ldots\}$, which is known as the matrix-analytic method, and its framework is called M/G/1 paradigm. In what follows, we summarize the main steps to compute the stationary probabilities of $\{(Q_n, S_n^D);\ n = 0, 1, \ldots\}$ whose transition probability matrix is given by (5). See [28] for details.

ALGORITHM 3.1 (Conventional matrix-analytic method [28]).

1. Input: $\boldsymbol{A}_k$ and $\boldsymbol{B}_k$ $(k = 0, 1, \ldots)$.
2. Stability test: Let $\boldsymbol{A}$ denote

$$\boldsymbol{A} = \sum_{k=0}^{\infty} \boldsymbol{A}_k,$$

and $\boldsymbol{\pi}_A$ denote the invariant probability vector of $\boldsymbol{A}$. The necessary and sufficient condition that the Markov chain has the stationary probability is

$$\rho = \boldsymbol{\pi}_A \sum_{k=1}^{\infty} k \boldsymbol{A}_k \boldsymbol{e} < 1,$$

and each element of

$$\boldsymbol{\beta}_B = \sum_{k=1}^{\infty} k \boldsymbol{B}_k \boldsymbol{e}$$

is finite.

3. Computation of fundamental matrix $\boldsymbol{G}$: We define $\boldsymbol{G}$ as an $M \times M$ stochastic matrix which satisfies

$$\boldsymbol{G} = \sum_{k=0}^{\infty} \boldsymbol{A}_k \boldsymbol{G}^k. \tag{6}$$

Note that the ith row of $\boldsymbol{G}$ is considered as the transition probability vector of the underlying Markov chain over the first passage time from states $(k+1, i)$ to state $(k, \cdot)$ for any $k = 0, 1, \ldots$. It is known that $\boldsymbol{G}$ is given as the limit $n \to \infty$ in the recursion

$$\boldsymbol{G}_{n+1} = \sum_{k=0}^{\infty} \boldsymbol{A}_k \boldsymbol{G}_n^k, \qquad n = 0, 1, \ldots, \tag{7}$$

with $\boldsymbol{G}_0 = \boldsymbol{A}_0$. Note that the $\boldsymbol{G}_n$ is an element-wise increasing sequence of matrices. Then compute the invariant probability vector $\boldsymbol{g}$ of $\boldsymbol{G}$.

4. Computation of recurrence matrix $\boldsymbol{K}$: We define $\boldsymbol{K}$ as an $M \times M$ stochastic matrix whose ith row represents the transition probability vector over the recurrence time from state $(0, i)$ to state $(0, \cdot)$. Note that $\boldsymbol{K}$ is given by

$$\boldsymbol{K} = \sum_{k=0}^{\infty} \boldsymbol{B}_k \boldsymbol{G}^k.$$

Then compute the invariant probability vector $\boldsymbol{f}$ of $\boldsymbol{K}$.

5. Computation of the stationary probabilities $\boldsymbol{x}_k$: Let $\boldsymbol{x}_k$ $(k = 0, 1, \ldots)$ denote a $1 \times M$ vector whose jth element represents $\Pr(Q = k, S^D = j)$, where Q and S^D denote generic random variables for the Q_n and the S_n^D, respectively. Then the $\boldsymbol{x}_k$ is recursively computed by [29]

$$\boldsymbol{x}_0 = \boldsymbol{f}/c,$$

and $\boldsymbol{x}_k$ $(k = 1, 2, \ldots)$ is recursively computed by

$$\boldsymbol{x}_k = \left(\boldsymbol{x}_0 \boldsymbol{B}_k^+ + \sum_{j=1}^{k-1} \boldsymbol{x}_j \boldsymbol{A}_{k-j+1}^+ \right) (\boldsymbol{I} - \boldsymbol{A}_1^+)^{-1},$$

where with $\boldsymbol{B} = \sum_{k=0}^{\infty} \boldsymbol{B}_k$ and $\boldsymbol{\beta} = \sum_{k=1}^{\infty} k \boldsymbol{A}_k \boldsymbol{e}$,

$$c = 1 + \frac{\boldsymbol{f}\boldsymbol{\beta}_B}{1-\rho} + \boldsymbol{f}(\boldsymbol{B} - \boldsymbol{I})(\boldsymbol{I} - \boldsymbol{A} + (\boldsymbol{e} - \boldsymbol{\beta})\boldsymbol{g})^{-1}\boldsymbol{e},$$

$$\boldsymbol{A}_k^+ = \sum_{j=k}^{\infty} \boldsymbol{A}_j \boldsymbol{G}^{j-k}, \qquad \boldsymbol{B}_k^+ = \sum_{j=k}^{\infty} \boldsymbol{B}_j \boldsymbol{G}^{j-k}.$$

REMARK 3.1. The most intensive part in Algorithm 3.1 is the computation of $\boldsymbol{G}$ in (7). Several other algorithms to compute $\boldsymbol{G}$ are known [4, 16]. Also, an entirely different solution method (called the invariant subspace approach) for Markov chains of M/G/1 type has also been developed recently [5, 6].

3.2. Application to MAP/GI/1 queue [26, 46]

We revisit the MAP/GI/1 queue. Recall that Q and S_D denote generic random variables representing the number of customers in the system and the state of the underlying Markov chain immediately after departures. As in the preceding subsection, let $\boldsymbol{x}_k$ $(k = 0, 1, \ldots)$ denote a $1 \times M$ vector whose jth element represents $\Pr(Q = k, S_D = j)$. It then follows from (5) and $\boldsymbol{B}_k = (-\boldsymbol{C})^{-1}\boldsymbol{D}\boldsymbol{A}_k$ that the vector PGF (probability generating function)

$$\boldsymbol{x}^*(z) = \sum_{k=0}^{\infty} z^k \boldsymbol{x}_k$$

satisfies

$$\boldsymbol{x}^*(z)\,[z\boldsymbol{I} - \boldsymbol{A}^*(z)] = \boldsymbol{x}_0(-\boldsymbol{C})^{-1}(\boldsymbol{C} + z\boldsymbol{D})\boldsymbol{A}^*(z). \tag{8}$$

We now define Y and S as generic random variables representing the number of customers in the system and the state of the underlying Markov chain at a randomly chosen instant of time in the stationary MAP/GI/1 queue. Let $\boldsymbol{y}_k$ $(k = 0, 1, \ldots)$ denote a $1 \times M$ vector whose jth element represents $\Pr(Y = k, S = j)$ and $\boldsymbol{y}^*(z)$ denote the vector probability generating function of the $\boldsymbol{y}_k$:

$$\boldsymbol{y}^*(z) = \sum_{k=0}^{\infty} z^k \boldsymbol{y}_k.$$

It is known that $\boldsymbol{y}^*(z)$ and $\boldsymbol{x}^*(z)$ are related by [26, 42]

$$\boldsymbol{y}^*(z)(\boldsymbol{C} + z\boldsymbol{D}) = \lambda(z-1)\boldsymbol{x}^*(z), \tag{9}$$

from which and (8) it follows that

$$\boldsymbol{y}^*(z)[z\boldsymbol{I} - \boldsymbol{A}^*(z)] = (z-1)\boldsymbol{y}_0\boldsymbol{A}^*(z). \tag{10}$$

In [46], it is shown that (10) implies that the $\boldsymbol{y}_k$ is identical to the steady-state solution of a Markov chain of M/G/1 type whose transition probability matrix is given by

$$\begin{pmatrix} \boldsymbol{A}_0 & \boldsymbol{A}_1 & \boldsymbol{A}_2 & \boldsymbol{A}_3 & \cdots \\ \boldsymbol{A}_0 & \boldsymbol{A}_1 & \boldsymbol{A}_2 & \boldsymbol{A}_3 & \cdots \\ \boldsymbol{O} & \boldsymbol{A}_0 & \boldsymbol{A}_1 & \boldsymbol{A}_2 & \cdots \\ \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{A}_0 & \boldsymbol{A}_1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

Thus applying Algorithm 3.1, we can numerically obtain the $\boldsymbol{y}_k$. In what follows, we provide some remarks in applying Algorithm 3.1 to the MAP/GI/1 queue. See [37] also, for its implementation.

REMARK 3.2.

1. As the input to Algorithm 3.1, we have to compute the sequence of matrices $\boldsymbol{A}_k$, which is provided in Appendix A.1.
2. The fundamental matrix $\boldsymbol{G}$ in the MAP/GI/1 queue satisfies

$$\boldsymbol{G} = \int_0^\infty \exp[(\boldsymbol{C} + \boldsymbol{D}\boldsymbol{G})x] dh(x), \tag{11}$$

and is given as the limit $n \to \infty$ in the recursion

$$\boldsymbol{G}_{n+1} = \int_0^\infty \exp[(\boldsymbol{C} + \boldsymbol{D}\boldsymbol{G}_n)x] dh(x), \qquad n = 0, 1, \ldots, \tag{12}$$

with $\boldsymbol{G}_0 = \boldsymbol{O}$. Note that the $\boldsymbol{G}_n$ is an element-wise increasing sequence of matrices. It is also known that for MAP, the convergence of the recursion in (12) is faster than that in (7). We provide an algorithm to compute the integral on the right-hand side of (12) in Appendix A.2.

3. The recurrence matrix $\boldsymbol{K}$ in Step 4 is identical to $\boldsymbol{G}$, so that $\boldsymbol{f} = \boldsymbol{g}$.

4. Waiting time distribution in the FIFO queue with MAS input

4.1. Stationary distribution of work-in-system [36, 41]

In this section, we consider the stationary distribution of the amount of work-in-system in a work-conserving single-server queue with MAS input with representation $(\boldsymbol{C}, \boldsymbol{D}_1(x), \ldots, \boldsymbol{D}_K(x))$. Recall that we assume that $\rho < 1$, which ensures that the queue is stable [24, 36].

Let V_t and S_t denote the amount of work-in-system and the state of the underlying Markov chain at time t. We then define τ as

$$\tau = \inf\{t;\ t \geq 0, V_t = 0\},$$

i.e., τ denotes the first passage time to the idle state after time 0. Let $\boldsymbol{P}(x)$ denote an $M \times M$ matrix whose (i,j)th element represents $\Pr(S_\tau = j \mid S_0 = i, V_0 = x)$. Note here that considering the preemptive-resume last-come, first-served service discipline, we have for $x, y > 0$

$$\begin{aligned}&\Pr(S_\tau = j \mid S_0 = i, V_0 = x + y)\\ &= \sum_{k \in \mathcal{M}} \Pr(S_\tau = k \mid S_0 = i, V_0 = x) \Pr(S_\tau = j \mid S_0 = k, V_0 = y),\end{aligned}$$

or equivalently

$$\boldsymbol{P}(x+y) = \boldsymbol{P}(x)\boldsymbol{P}(y), \qquad x, y \geq 0. \tag{13}$$

Thus replacing y in (13) by infinitesimal real $\delta x > 0$ and ignoring terms of order $o(\delta x)$, we have

$$\boldsymbol{P}(x + \delta x) = \boldsymbol{P}(x)\left[\boldsymbol{I} + \boldsymbol{C}\delta x + \int_0^\infty d\boldsymbol{D}(y)\delta x \boldsymbol{P}(y)\right],$$

from which it follows that

$$\frac{d}{dx}\boldsymbol{P}(x) = \boldsymbol{P}(x)\left[\boldsymbol{C} + \int_0^\infty d\boldsymbol{D}(y)\boldsymbol{P}(y)\right].$$

Thus, taking account of $\boldsymbol{P}(0+) = \boldsymbol{I}$, we obtain the following theorem.

THEOREM 4.1. *$\boldsymbol{P}(x)$ is given by*

$$\boldsymbol{P}(x) = \exp(\boldsymbol{Q}x), \tag{14}$$

where $\boldsymbol{Q}$ is an $M \times M$ matrix which satisfies

$$\boldsymbol{Q} = \boldsymbol{C} + \int_0^\infty d\boldsymbol{D}(x)\exp(\boldsymbol{Q}x). \tag{15}$$

REMARK 4.1. The matrix-exponential result in (14) is almost self-evident because $\boldsymbol{P}(x)$ satisfies the semi-group property (13) [10].

REMARK 4.2.

1. Note that $\boldsymbol{Q}$ is considered as a infinitesimal generator of an irreducible Markov chain which is obtained by excising busy periods, i.e., if we observe the underlying Markov chain only when the system is idle, the observed Markov chain has a generator $\boldsymbol{Q}$.

2. When service times of all customers are i.i.d. according to $h(x)$ (i.e, MAP/GI/1 queue), we have $\boldsymbol{Q} = \boldsymbol{C} + \boldsymbol{D}\boldsymbol{G}$, where $\boldsymbol{G}$ is given in (11).

$\boldsymbol{Q}$ in (15) is obtained as the limit $\boldsymbol{Q}_\infty$ in the recursion

$$\boldsymbol{Q}_{n+1} = \boldsymbol{C} + \int_0^\infty d\boldsymbol{D}(x) \exp(\boldsymbol{Q}_n x), \quad n = 0, 1, \ldots$$

with $\boldsymbol{Q}_0 = \boldsymbol{C}$, where the integral on the right hand side of the above recursion can be numerically obtained in a way similar to (12) (see Appendix A.2). Note that the $\boldsymbol{Q}_n$ is an element-wise increasing sequence of matrices. Let $\boldsymbol{\kappa}$ denote a $1 \times M$ vector which satisfies

$$\boldsymbol{\kappa}\boldsymbol{Q} = \boldsymbol{0}, \qquad \boldsymbol{\kappa}\boldsymbol{e} = 1. \tag{16}$$

Let V and S denote generic random variables representing the amount of work-in-system and the state of the underlying Markov chain in steady state. Note that the jth element κ_j of $\boldsymbol{\kappa}$ represents $\kappa_j = \Pr(S = j \mid V = 0)$. We then define $\boldsymbol{v}(x)$ as a $1 \times M$ vector whose jth element represents $\Pr(V \le x, S = j)$. Because $\Pr(V = 0) = 1 - \rho$ (which comes from Little's law) and the jth element of $\boldsymbol{\kappa}$ represents $\Pr(S = j \mid V = 0)$, we obtain the following theorem.

THEOREM 4.2. *$\boldsymbol{v}(0)$ is given by*

$$\boldsymbol{v}(0) = (1 - \rho)\boldsymbol{\kappa}. \tag{17}$$

Now we consider $\boldsymbol{v}(x)$ for $x > 0$. A standard birth-and-death argument leads to

$$\boldsymbol{v}(x) = \boldsymbol{v}(x + \delta t)(\boldsymbol{I} + \boldsymbol{C}\delta t) + \int_0^x \boldsymbol{v}(x - y + \delta t) d\boldsymbol{D}(y)\delta t + o(\delta t),$$

from which it follows that

$$\frac{d}{dx}\boldsymbol{v}(x) + \boldsymbol{v}(x)\boldsymbol{C} + \int_0^x \boldsymbol{v}(x - y) d\boldsymbol{D}(y) = 0. \tag{18}$$

(18) is considered as a generalization of Takács's integro-differential equation for the M/GI/1 queue [35]. Note further that integrating (18) over $(0, x]$ yields

$$\boldsymbol{v}(x) = \boldsymbol{v}(0) - \int_0^x \boldsymbol{v}(y)(\boldsymbol{C} + \boldsymbol{D}(x - y))dy, \tag{19}$$

where $\boldsymbol{v}(0)$ is given in (17). The integral equation (19) for $\boldsymbol{v}(x)$ is considered as a generalization of the Pollaczek-Khinchin's integral equation for the M/GI/1 queue (see [11, 27]).

Recently a closed-form solution of (19) was found in [45]. Let $\boldsymbol{R}(x)$ denote an $M \times M$ matrix which is given by

$$\boldsymbol{R}(x) = \int_0^x dw \int_w^\infty d\boldsymbol{D}(y) \exp(\boldsymbol{Q}(y-w)).$$

THEOREM 4.3 ([45]). *$\boldsymbol{v}(x)$ is given by*

$$\boldsymbol{v}(x) = (1-\rho)\boldsymbol{\kappa} \sum_{n=0}^{\infty} \boldsymbol{R}^{(n)}(x), \qquad x \geq 0, \tag{20}$$

where $\boldsymbol{R}^{(0)}(x) = \boldsymbol{I}$, $\boldsymbol{R}^{(1)}(x) = \boldsymbol{R}(x)$ and $\boldsymbol{R}^{(n)}(x)$ $(n = 2, 3, \ldots)$ denotes the n-fold convolution of $\boldsymbol{R}(x)$ with itself, i.e.,

$$\boldsymbol{R}^{(n)}(x) = \int_0^x \boldsymbol{R}^{(n-1)}(x-y) d\boldsymbol{R}(y).$$

REMARK 4.3. For the M/GI/1 queue with the mean service time h, let $h_e(x)$ denote the equilibrium distribution function

$$h_e(x) = \int_0^x \frac{1-h(y)}{h} dy$$

of the service time distribution $h(x)$ with itself. Then (20) is reduced to Pollaczek-Khinchin's formula:

$$v(x) = (1-\rho) \sum_{n=0}^{\infty} h_e^{(n)}(x),$$

where $h_e^{(0)}(x) = 1$, $h_e^{(1)}(x) = h_e(x)$ and $h_e^{(n)}(x)$ $(n = 2, 3, \ldots)$ denotes the n-fold convolution of the equilibrium distribution function $h_e(x)$ with itself.

Let $\boldsymbol{v}^*(s)$ denote the LST of $\boldsymbol{v}(x)$. From (17) and (19) (or (18)), we have the following theorem.

THEOREM 4.4. *$\boldsymbol{v}^*(s)$ satisfies*

$$\boldsymbol{v}^*(s)[s\boldsymbol{I} + \boldsymbol{C} + \boldsymbol{D}^*(s)] = (1-\rho)s\boldsymbol{\kappa}, \tag{21}$$

where $\boldsymbol{D}^*(s)$ denotes the LST of $\boldsymbol{D}(x)$.

REMARK 4.4. When service times of all customers are i.i.d. according to a distribution function $h(x)$ (i.e, MAP/GI/1 queue), (21) is reduced to

$$\boldsymbol{v}^*(s)[s\boldsymbol{I} + \boldsymbol{C} + h^*(s)\boldsymbol{D}] = (1-\rho)s\boldsymbol{\kappa}, \tag{22}$$

where $h^*(s)$ denotes the LST of $h(x)$. In [26], (22) is obtained by a totally different approach. When arrivals follow a Poisson process, (22) is further reduced to

$$v^*(s) = \frac{(1-\rho)s}{s - \lambda + \lambda h^*(s)}. \tag{23}$$

Besides Theorem 4.4, using (20), we can obtain a closed-form formula for $\boldsymbol{v}^*(s)$. Let $\boldsymbol{R}^*(s)$ denote the LST of $\boldsymbol{R}(x)$, i.e.,

$$\boldsymbol{R}^*(s) = \int_0^\infty \exp(-sx) \int_x^\infty d\boldsymbol{D}(y) \exp(\boldsymbol{Q}(y-x)).$$

THEOREM 4.5. *$\boldsymbol{v}^*(s)$ is given by*

$$\boldsymbol{v}^*(s) = (1-\rho)\boldsymbol{\kappa}(\boldsymbol{I} - \boldsymbol{R}^*(s))^{-1}, \qquad Re(s) > 0, \tag{24}$$

where $\boldsymbol{R}^*(s)$ satisfies

$$\boldsymbol{R}^*(s)\,(\boldsymbol{e}\boldsymbol{\kappa} - s\boldsymbol{I} - \boldsymbol{Q}) = \left(\frac{\boldsymbol{C} + \boldsymbol{D}^*(s)}{s}\right)(s\boldsymbol{I} - \boldsymbol{e}\boldsymbol{\kappa}) - \boldsymbol{Q}. \tag{25}$$

REMARK 4.5. From (21) and (24), we have

$$\boldsymbol{v}^*(s)\left[\boldsymbol{I} + \frac{\boldsymbol{C} + \boldsymbol{D}^*(s)}{s}\right] = \boldsymbol{v}^*(s)\left[\boldsymbol{I} - \boldsymbol{R}^*(s)\right].$$

However, for $M \geq 2$,

$$\boldsymbol{I} + \frac{\boldsymbol{C} + \boldsymbol{D}^*(s)}{s} \neq \boldsymbol{I} - \boldsymbol{R}^*(s).$$

In fact, it can be shown that if the equality held, $\boldsymbol{Q}$ would be $\boldsymbol{O}$.

Taking derivatives of both sides of (25), we obtain a numerically feasible recursion to compute moments of V [45], which is given in Appendix A.3.

REMARK 4.6. Using (21), we can obtain another recursion for the $\boldsymbol{v}^{(n)}$ [36], where an equation for $\boldsymbol{v}^{(n)}$ $(n = 1, 2, \ldots)$ is derived under the assumption that $\boldsymbol{v}^{(n+1)}$ (as well as $\boldsymbol{v}^{(m)}$ $(m = 1, \ldots, n)$) is finite.

4.2. Actual waiting time and sojourn time distributions [43]

We now consider the stationary distribution of the actual waiting time of customers in respective classes. Note that the actual waiting time of a customer is defined as the amount of work-in-system which the customer finds on arrival. Let W_k and S_k^A $(k = 1, \ldots, K)$ denote generic random variables representing the actual waiting time of class k customers and the state of the underlying Markov chain immediately after arrivals of class k customers, respectively. We then define $\boldsymbol{w}_k(x)$ $(k = 1, \ldots, K)$ as a $1 \times M$ vector whose jth element represents $\Pr(W_k \leq x, S_k^A = j)$. Let $\boldsymbol{w}_k^*(s)$ $(k = 1, \ldots, K)$ denote the LST of $\boldsymbol{w}_k(x)$. Applying the conditional PASTA [23], we have

$$\boldsymbol{w}_k(x) = \frac{\boldsymbol{v}(x)\boldsymbol{D}_k}{\lambda_k}, \qquad \boldsymbol{w}_k^*(s) = \frac{\boldsymbol{v}^*(s)\boldsymbol{D}_k}{\lambda_k}, \tag{26}$$

where we use (1). See [20] also.

Let T_k $(k = 1, \ldots, K)$ denote a generic random variable representing the sojourn time (i.e., the sum of actual waiting time and service time) of class k customers. We then define $\boldsymbol{t}_k(x)$ as a $1 \times M$ vector whose jth element represents $\Pr(T_k \leq x, S_k^A = j)$. Let $\boldsymbol{t}_k^*(s)$ $(k = 1, \ldots, K)$ denote the LST of $\boldsymbol{t}_k(x)$. We then have

$$\boldsymbol{t}_k(x) = \frac{1}{\lambda_k} \int_0^x \boldsymbol{v}(x-y) d\boldsymbol{D}_k(y), \qquad \boldsymbol{t}_k^*(s) = \frac{\boldsymbol{v}_k^*(s)\boldsymbol{D}_k^*(s)}{\lambda_k}. \tag{27}$$

4.3. Asymptotic formulas of tail distributions [48]

While the closed-form expression of the distribution function of the amount of work-in-system is obtained in Theorem 4.3, it cannot be used to compute $\Pr(V \leq x)$ for a specific value of x, as in the case of the M/GI/1 queue. Thus in order to compute the distribution function based on the results in the preceding subsection, we have to rely on some numerical inversion technique of Laplace transforms. Therefore in this subsection, we consider the asymptotics of the tail distributions of the amount of work-in-system, actual waiting time and sojourn time. As you will see, under some conditions, the asymptotic tail distributions have simple forms which are numerically tractable.

Let $\overline{\boldsymbol{v}}(x)$, $\overline{\boldsymbol{w}_k}(x)$ and $\overline{\boldsymbol{t}_k}(x)$ $(k = 1, \ldots, K)$ denote the tail distributions of $\boldsymbol{v}(x)$, $\boldsymbol{w}_k(x)$ and $\boldsymbol{t}_k(x)$, respectively, i.e, (see (26) and (27))

$$\overline{\boldsymbol{v}}(x) = \boldsymbol{v}(\infty) - \boldsymbol{v}(x) = \boldsymbol{\pi} - \boldsymbol{v}(x),$$

$$\overline{\boldsymbol{w}_k}(x) = \boldsymbol{w}_k(\infty) - \boldsymbol{w}_k(x) = \frac{\overline{\boldsymbol{v}}(x)\boldsymbol{D}_k}{\lambda_k}, \quad k = 1, \ldots, K,$$

$$\overline{\boldsymbol{t}_k}(x) = \boldsymbol{t}_k(\infty) - \boldsymbol{t}_k(x)$$
$$= \frac{1}{\lambda_k}\left(\boldsymbol{\pi}\boldsymbol{D}_k - \int_0^x \boldsymbol{v}(x-y)d\boldsymbol{D}_k(y)\right), \qquad k = 1, \ldots, K.$$

These tail probabilities often have an exponential form, even though we need some conditions for it [1, 2]. On the other extreme, subexponential asymptotics are also known for various queues [9, 21, 34, 48].

We start with exponential asymptotics. The results presented below are very similar to those in [1] for the batch BMAP/GI/1 queue. Note however that [1] considers the asymptotics of the waiting time distributions in connection with the queue length counterpart. Thus the results in [1] are not transparent in our settings. For this reason, we provide complete proofs of the results for exponential asymptotics.

The exponential asymptotics of the tail distributions heavily relies on the analytical properties of $\boldsymbol{D}^*(s)$ [1]. Let $-A$ $(A \geq 0)$ denote the convergence abscissa of $\boldsymbol{D}^*(s)$.

LEMMA 4.1. *For real $s \in (-A, \infty)$, there exists the dominant eigenvalue $\delta^*(s)$ of $\boldsymbol{C} + \boldsymbol{D}^*(s)$ such that*

(a) *$\delta^*(s)$ is real and there exist strictly positive left and right eigenvectors associated with $\delta^*(s)$, which are unique to constant multiples,*
(b) *$\delta^*(s)$ is a simple root of the characteristic equation of $\boldsymbol{C} + \boldsymbol{D}^*(s)$ and $Re(\phi^*(s)) < \delta^*(s)$ for any other eigenvalue $\phi^*(s)$ of $\boldsymbol{C} + \boldsymbol{D}^*(s)$, and*
(c) *$\delta^*(s)$ is a strictly decreasing convex function in $(-A, \infty)$.*

The proof of Lemma 4.1 is given in Appendix B.1.

Let $\boldsymbol{l}^*(s)$ and $\boldsymbol{r}^*(s)$ $(s \in (-A, \infty))$ denote the left and right eigenvectors associated with the dominant eigenvalue $\delta^*(s)$, respectively, which satisfy the normalizing conditions:

$$\boldsymbol{l}^*(s)\boldsymbol{e} = \boldsymbol{l}^*(s)\boldsymbol{r}^*(s) = 1.$$

LEMMA 4.2. *For real $s \in (-A, 0)$, the equation*

$$-s = \delta^*(s) \tag{28}$$

has at most one root.

The proof of Lemma 4.2 is given in Appendix B.2.

We now introduce the assumption that is required for exponential asymptotics.

ASSUMPTION 4.1. (28) *has a root in* $(-A, 0)$, *i.e.,*

$$\eta = \delta^*(-\eta), \tag{29}$$

for some $\eta \in (0, A)$.

We then have the following theorem.

THEOREM 4.6. *Under Assumption 4.1, we have*

$$\lim_{x\to\infty} \frac{\overline{v}(x)}{\exp(-\eta x)} = \frac{(1-\rho)\boldsymbol{\kappa}\boldsymbol{r}^*(-\eta)}{\boldsymbol{l}^*(-\eta)\left(-\boldsymbol{D}^{(1)}(-\eta)\right)\boldsymbol{r}^*(-\eta) - 1}\boldsymbol{l}^*(-\eta), \tag{30}$$

and for all $k = 1, \ldots, K$,

$$\begin{aligned}\lim_{x\to\infty} \frac{\overline{\boldsymbol{w}_k}(x)}{\exp(-\eta x)} \\ = \frac{(1-\rho)\boldsymbol{\kappa}\boldsymbol{r}^*(-\eta)}{\boldsymbol{l}^*(-\eta)\left(-\boldsymbol{D}^{(1)}(-\eta)\right)\boldsymbol{r}^*(-\eta) - 1} \cdot \frac{\boldsymbol{l}^*(-\eta)\boldsymbol{D}_k}{\lambda_k},\end{aligned} \tag{31}$$

$$\begin{aligned}\lim_{x\to\infty} \frac{\overline{\boldsymbol{t}_k}(x)}{\exp(-\eta x)} \\ = \frac{(1-\rho)\boldsymbol{\kappa}\boldsymbol{r}^*(-\eta)}{\boldsymbol{l}^*(-\eta)\left(-\boldsymbol{D}^{(1)}(-\eta)\right)\boldsymbol{r}^*(-\eta) - 1} \cdot \frac{\boldsymbol{l}^*(-\eta)\boldsymbol{D}_k^*(-\eta)}{\lambda_k}.\end{aligned} \tag{32}$$

The proof of Theorem 4.6 is given in Appendix B.3.

Next we consider subexponential asymptotics. For any distribution function $f(x)$ with finite mean f, let $\overline{f}(x)$ denote the complementary distribution function of $f(x)$, i.e., $\overline{f}(x) = 1 - f(x)$.

DEFINITION 4.1. A distribution function $f(x)$ (and the corresponding random variable F) is called heavy-tailed if $\overline{f}(x) > 0$ for all $x \geq 0$ and

$$\lim_{x\to\infty} \frac{\overline{f}(x+y)}{\overline{f}(x)} = 1, \qquad \text{for all } y \geq 0. \tag{33}$$

It is known that the convergence in (33) is uniform on compact y sets [31]. Let $f^{(n)}(x)$ denote the n-fold convolution of $f(x)$ with itself, i.e.,

$$f^{(n)}(x) = \int_0^x f(x-y)df^{(n-1)}(y), \qquad n = 2, 3, \ldots,$$

with $f^{(1)}(x) = f(x)$, and $\overline{f^{(n)}}(x)$ denote $1 - f^{(n)}(x)$.

DEFINITION 4.2. The heavy-tailed $f(x)$ (and the corresponding random variable F) is called subexponential if

$$\lim_{x\to\infty} \frac{\overline{f^{(2)}}(x)}{\overline{f}(x)} = 2.$$

LEMMA 4.3 ([12]). *Suppose $f(x)$ is subexponential and $g(x)$ is any distribution function such that*

$$\lim_{x\to\infty} \frac{\overline{g}(x)}{\overline{f}(x)} = c < \infty.$$

If $c > 0$, $g(x)$ is subexponential.

Readers are referred to [13, 21, 34] for details.

We now define H as a generic random variable representing a service time of a randomly chosen customer, whose distribution function is denoted by $h(x)$. We then have

$$\overline{h}(x) = 1 - \frac{\boldsymbol{\pi} \boldsymbol{D}(x) \boldsymbol{e}}{\lambda},$$

and the mean $h = E[H]$ is given by (see (2))

$$h = \frac{\rho}{\lambda}.$$

Let $h_e(x)$ denote the equilibrium distribution of $h(x)$:

$$h_e(x) = h^{-1} \int_0^x \overline{h}(y) dy.$$

ASSUMPTION 4.2. *The equilibrium distribution function $h_e(x)$ of the overall (i.e., customer-average) service time distribution $h(x)$ is subexponential.*

REMARK 4.7. Assumption 4.2 includes the cases that the most heaviest-tailed service time distribution among $h_{k,i,j}(x)$ follows Pareto, log-normal, heavy-tailed Weibull or a distribution with regularly varying tails.

THEOREM 4.7 ([48]). *Under Assumption 4.2, we have*

$$\lim_{x\to\infty} \frac{\overline{\boldsymbol{v}}(x)}{\overline{h_e}(x)} = \frac{\rho}{1-\rho} \boldsymbol{\pi},$$

$$\lim_{x\to\infty} \frac{\overline{\boldsymbol{w}_k}(x)}{\overline{h_e}(x)} = \lim_{x\to\infty} \frac{\overline{\boldsymbol{t}_k}(x)}{\overline{h_e}(x)} = \frac{\rho}{1-\rho} \cdot \frac{\boldsymbol{\pi} \boldsymbol{D}_k}{\lambda_k}, \qquad k = 1, \ldots, K.$$

REMARK 4.8. Theorem 4.7 implies that for all $k = 1, \ldots, K$,

$$\lim_{x\to\infty} \frac{\Pr(V > x)}{\overline{h_e}(x)} = \lim_{x\to\infty} \frac{\Pr(W_k > x)}{\overline{h_e}(x)} = \lim_{x\to\infty} \frac{\Pr(T_k > x)}{\overline{h_e}(x)} = \frac{\rho}{1-\rho}. \tag{34}$$

Namely, all random variables V, W_k and T_k are subexponential (see Lemma 4.3) and have the same asymptotic constant given only in terms of the utilization factor ρ, independent of the structure of MAS (i.e., $\boldsymbol{C}$ and $\boldsymbol{D}(x)$).

5. Joint queue length distribution in FIFO queues

In this section we consider the joint queue length distribution. It is widely recognized that the queue length distribution in a FIFO queue with multiple non-Poissonian arrival streams having different service time distributions is very hard to analyze, since we have to keep track of the complete order of customers in the queue to describe the queue length dynamics. It is clear that the structure of any Markov chains, keeping track of the type of customers in the waiting line and the service position, is extremely complicated and therefore it is a formidable task to obtain the stationary distribution from such a formulation. In what follows, we summarize a recent development to analyze the queue length distribution in such circumstances.

5.1. Invariance relationship and distributional form of Little's law [42, 47]

We first show an invariance relationship between the time-average joint queue length distribution and customer-average joint-queue length distribution at departures. When the FIFO service discipline is employed, this relationship leads to a distributional form of Little's law. Even though the original theorems in [47] are stated in a more general arrival process than MAS defined in section 2, we specialize them for MAS below.

Consider a stationary queue with marked MAP arrivals with representation $(\boldsymbol{C}, \boldsymbol{D}_1, \ldots, \boldsymbol{D}_K)$, where we *do not* specify any particular service mechanism, e.g., service discipline, service time process, and the number of servers. Let $L_k(t)$ $(k = 1, \ldots, K,\ t \geq 0)$ denote the number of class k customers in the system at time t, and let $A_k(t)$ and $B_k(t)$ $(k = 1, \ldots, K,\ t \geq 0)$ denote the numbers of arrivals and departures,

respectively, of class k customers during an interval $(0,t]$. We assume that the multivariate queueing process $(L_1(t),\dots L_K(t))$ satisfies

$$\begin{aligned}&(L_1(t),\dots L_K(t))\\ =&(L_1(0),\dots L_K(0))+(A_1(t),\dots A_K(t))-(B_1(t),\dots B_K(t)),\end{aligned}$$

and $(A_1(t),\dots A_K(t))$ follows a marked MAP with representation $(\boldsymbol{C}, \boldsymbol{D}_1, \dots, \boldsymbol{D}_K)$.

Let L_k $(k=1,\dots,K)$ and S denote the number of class k customers in the system and the state of the underlying Markov chain, respectively, in steady state. Also, for each $k=1,\dots,K$, let $Q_k(n)$ $(n=1,\dots,K)$ and S_k^D denote the number of class n customers in the system and the state of the underlying Markov chain, respectively, immediately after departures of class k customers in steady state. We then define $\boldsymbol{l}(\boldsymbol{z})$ and $\boldsymbol{q}_k(\boldsymbol{z})$ $(k=1,\dots,K)$ as $1\times M$ vectors whose jth elements $l_j(\boldsymbol{z})$ and $q_{k,j}(\boldsymbol{z})$ represent

$$l_j(\boldsymbol{z})=E\left[z_1^{L_1}\cdots z_K^{L_K}1_{\{S=j\}}\right],\quad q_{k,j}(\boldsymbol{z})=E\left[z_1^{Q_k(1)}\cdots z_K^{Q_k(K)}1_{\{S_k^D=j\}}\right],$$

respectively.

THEOREM 5.1. *Under some technical assumptions (see Remark 5.1 below), $\boldsymbol{l}(\boldsymbol{z})$ and $\boldsymbol{q}_k(\boldsymbol{z})$ $(k=1,\dots,K)$ in a stationary queue with marked MAP arrivals with representation $(\boldsymbol{C},\boldsymbol{D}_1,\dots,\boldsymbol{D}_K)$ are related by*

$$\boldsymbol{l}(\boldsymbol{z})\left(\boldsymbol{C}+\sum_{k=1}^{K}z_k\boldsymbol{D}_k\right)=\sum_{k=1}^{K}\lambda_k(z_k-1)\boldsymbol{q}_k(\boldsymbol{z}). \tag{35}$$

REMARK 5.1. In applying Theorem 5.1 to a model where a state transition of the underlying Markov chain and departure(s) of customers can occur simultaneously, we need to determine the order of those events to satisfy some technical condition. See [47] for details.

Because we do not assume any particular service mechanism, the relationship (35) holds for a broad class of stationary queues with MAS input. Apparently, (9) is a special case of this relationship. Besides, special cases of this relationship can be found in the literature. When $K=1$ (i.e., MAP arrivals), (35) is reduced to the result in [42]. In [43], it is analytically shown for a FIFO queue with MAS input. This relationship can also be shown to hold for nonpreemptive priority queues with MAS inputs, with the results in [40, 44].

We now restrict our attention to FIFO queues. More precisely, we assume the followings.

ASSUMPTION 5.1. *Customers leave the system in order of arrival, and for any time t, the sojourn time of any customer arriving before time t is independent of the arrival process after time t.*

Let R_k denote the stationary sojourn time of a class k customer and S_k^A denote the state of the underlying Markov chain immediately after arrivals of class k customers. We then define $\boldsymbol{r}_k(x)$ $(k = 1, \ldots, K)$ as a $1 \times M$ vector whose jth element represents $\Pr(R_k \leq x, S_k^A = j)$. We then have the following theorem [47].

THEOREM 5.2. *Under Assumption 5.1, $\boldsymbol{l}^*(\boldsymbol{z})$ and $\boldsymbol{r}_k(x)$ are related by*

$$\boldsymbol{l}^*(\boldsymbol{z})\left(\boldsymbol{C} + \sum_{k=1}^{K} z_k \boldsymbol{D}_k\right) = \sum_{k=1}^{K} \lambda_k (z_k - 1) \int_{0-}^{\infty} d\boldsymbol{r}_k(x) \exp\left[\left(\boldsymbol{C} + \sum_{k=1}^{K} z_k \boldsymbol{D}_k\right) x\right]. \tag{36}$$

Obviously, Theorem 5.2 generalizes a distributional form of Little's law for the number of customers in FIFO queues with Poisson arrivals [22]. We allow more general arrival processes than Poisson arrivals and it gives a relationship between the joint queue length distribution and the sojourn time distributions of customers in respective classes.

5.2. PGFs of joint queue length distribution [47]

Using Theorem 5.2, we show how to develop a numerically feasible algorithm to compute the probability mass function of the joint queue length in the stationary FIFO single-server queue with MAS input. Let X_k $(k = 1, \ldots, K)$ denote the number of class k customers waiting for service in steady state. We then define $\boldsymbol{x}(\boldsymbol{z})$ as a $1 \times M$ vector whose jth element $x_j(\boldsymbol{z})$ represents

$$x_j(\boldsymbol{z}) = E\left[z_1^{X_1} \cdots z_K^{X_K} 1_{\{S=j\}}\right].$$

Similarly, let Y_k denote the number of class k customers in the system in steady state. We then define $\boldsymbol{y}(\boldsymbol{z})$ as a $1 \times M$ vector whose jth element

$y_j(\boldsymbol{z})$ represents

$$y_j(\boldsymbol{z}) = E\left[z_1^{Y_1} \cdots z_K^{Y_K} 1_{\{S=j\}}\right].$$

The following corollary is an immediate consequence of Theorem 5.2.

COROLLARY 5.1. *In the work-conserving FIFO single-server queue with MAS input with representation* $(\boldsymbol{C}, \boldsymbol{D}_1, \ldots, \boldsymbol{D}_K)$, *the vector PGFs* $\boldsymbol{x}(\boldsymbol{z})$ *and* $\boldsymbol{y}(\boldsymbol{z})$ *of the stationary joint probabilities of the number of customers are given by*

$$\boldsymbol{x}^*(\boldsymbol{z})\left(\boldsymbol{C} + \sum_{k=1}^{K} z_k \boldsymbol{D}_k\right) = \sum_{k=1}^{K} \lambda_k (z_k - 1) \int_{0-}^{\infty} d\boldsymbol{w}_k(x) \exp\left[(\boldsymbol{C} + \sum_{i=1}^{K} z_i \boldsymbol{D}_i)x\right], \tag{37}$$

$$\boldsymbol{y}^*(\boldsymbol{z})\left(\boldsymbol{C} + \sum_{k=1}^{K} z_k \boldsymbol{D}_k\right) = \sum_{k=1}^{K} \lambda_k (z_k - 1) \int_{0-}^{\infty} d\boldsymbol{t}_k(x) \exp\left[(\boldsymbol{C} + \sum_{i=1}^{K} z_i \boldsymbol{D}_i)x\right], \tag{38}$$

where $\boldsymbol{w}_k(x)$ *and* $\boldsymbol{t}_k(x)$ $(k = 1, \ldots, K)$ *are given in (26) and (27), respectively.*

5.3. Numerical Algorithm to compute the joint probability mass function [47]

In this subsection we provide a way to develop recursion formulas to compute the joint probability mass function of queue length. With $\boldsymbol{n} = (n_1, \ldots, n_K) \geq \boldsymbol{0}$, let $\boldsymbol{x}(\boldsymbol{n})$ and $\boldsymbol{y}(\boldsymbol{n})$ denote $1 \times M$ vectors whose jth elements $x_j(\boldsymbol{n})$ and $y_j(\boldsymbol{n})$ represent

$$x_j(\boldsymbol{n}) = \Pr(X_1 = n_1, \ldots, X_K = n_K, S = j),$$
$$y_j(\boldsymbol{n}) = \Pr(Y_1 = n_1, \ldots, Y_K = n_K, S = j),$$

respectively. By definition,

$$\boldsymbol{x}^*(\boldsymbol{z}) = \sum_{n_1=0}^{\infty} \cdots \sum_{n_K=0}^{\infty} z_1^{n_1} \cdots z_K^{n_K} \boldsymbol{x}(\boldsymbol{n}),$$

$$\boldsymbol{y}^*(\boldsymbol{z}) = \sum_{n_1=0}^{\infty} \cdots \sum_{n_K=0}^{\infty} z_1^{n_1} \cdots z_K^{n_K} \boldsymbol{y}(\boldsymbol{n}).$$

Let θ denote the maximum absolute value of the diagonal elements of matrix $\boldsymbol{C}$. It then follows from (38) that

$$\begin{aligned}(39)\qquad &\sum_{n_1=0}^{\infty} \cdots \sum_{n_K=0}^{\infty} z_1^{n_1} \cdots z_K^{n_K} \boldsymbol{x}(\boldsymbol{n}) \left(\boldsymbol{C} + \sum_{k=1}^{K} z_k \boldsymbol{D}_k \right) \\ &= \sum_{k=1}^{K} (z_k - 1) \sum_{m=0}^{\infty} \boldsymbol{v}^{(m)}(\theta) \boldsymbol{D}_k \left[\boldsymbol{I} + \theta^{-1} \left(\boldsymbol{C} + \sum_{i=1}^{K} z_i \boldsymbol{D}_i \right) \right]^m,\end{aligned}$$

where

$$\boldsymbol{v}^{(m)}(\theta) = \int_{0-}^{\infty} e^{-\theta x} \frac{(\theta x)^m}{m!} d\boldsymbol{v}(x).$$

We now define $\boldsymbol{F}_m(\boldsymbol{n})$'s ($m = 0, 1, \ldots$, $\boldsymbol{n} \geq \boldsymbol{0}$ and $n_1 + \cdots + n_K \leq m$) as $M \times M$ matrices which satisfy

$$(40)\qquad \left[\boldsymbol{I} + \theta^{-1} \left(\boldsymbol{C} + \sum_{i=1}^{K} z_i \boldsymbol{D}_i \right) \right]^m = \sum_{\substack{n_1 \geq 0 \\ n_1 + \cdots + n_K \leq m}} \cdots \sum_{n_K \geq 0} z_1^{n_1} \cdots z_K^{n_K} \boldsymbol{F}_m(\boldsymbol{n}).$$

It then follows from (39) and (40) that

$$\begin{aligned}&\sum_{n_1=0}^{\infty} \cdots \sum_{n_K=0}^{\infty} z_1^{n_1} \cdots z_K^{n_K} \boldsymbol{x}(\boldsymbol{n}) \left(\boldsymbol{C} + \sum_{k=1}^{K} z_k \boldsymbol{D}_k \right) \\ &= \sum_{k=1}^{K} (z_k - 1) \sum_{m=0}^{\infty} \boldsymbol{v}^{(m)}(\theta) \boldsymbol{D}_k \sum_{\substack{n_1 \geq 0 \\ n_1 + \cdots + n_K \leq m}} \cdots \sum_{n_K \geq 0} z_1^{n_1} \cdots z_K^{n_K} \boldsymbol{F}_m(\boldsymbol{n}) \\ &= \sum_{k=1}^{K} (z_k - 1) \sum_{n_1=0}^{\infty} \cdots \sum_{n_K=0}^{\infty} z_1^{n_1} \cdots z_K^{n_K} \sum_{m=n_1+\cdots+n_K}^{\infty} \boldsymbol{v}^{(m)}(\theta) \boldsymbol{D}_k \boldsymbol{F}_m(\boldsymbol{n}).\end{aligned}$$

Comparing coefficient vectors of $z_1^{n_1} \cdots z_K^{n_K}$ of both sides of the above equation, we obtain the following theorem.

THEOREM 5.3. *The vector joint probabilities* $\boldsymbol{x}(\boldsymbol{n})$ $(\boldsymbol{n} \geq \boldsymbol{0})$ *for the number of waiting customers are recursively obtained by*

$$\boldsymbol{x}(\boldsymbol{0}) = \sum_{m=0}^{\infty} \boldsymbol{v}^{(m)}(\theta)\boldsymbol{D}\boldsymbol{F}_m(\boldsymbol{0})(-\boldsymbol{C})^{-1},$$

and for $n_1 + \cdots + n_K \geq 1$,

$$\boldsymbol{x}(\boldsymbol{n}) = \Bigg[\sum_{\substack{k=1 \\ n_k \geq 1}}^{K} \boldsymbol{x}(\boldsymbol{n} - \boldsymbol{e}_k)\boldsymbol{D}_k + \sum_{m=n_1+\cdots+n_K}^{\infty} \boldsymbol{v}^{(m)}(\theta)\boldsymbol{D}\boldsymbol{F}_m(\boldsymbol{n}) - \sum_{\substack{k=1 \\ n_k \geq 1}}^{K} \sum_{m=n_1+\cdots+n_K-1}^{\infty} \boldsymbol{v}^{(m)}(\theta)\boldsymbol{D}_k\boldsymbol{F}_m(\boldsymbol{n} - \boldsymbol{e}_k) \Bigg] (-\boldsymbol{C})^{-1}.$$

Similarly, we can obtain the recursion for the $\boldsymbol{y}(\boldsymbol{n})$.

THEOREM 5.4. *The vector joint probabilities* $\boldsymbol{y}(\boldsymbol{n})$ $(\boldsymbol{n} \geq \boldsymbol{0})$ *for the number of customers in the system are recursively obtained by*

$$\boldsymbol{y}(\boldsymbol{0}) = (1 - \rho)\boldsymbol{\kappa},$$

and for $n_1 + \cdots + n_K \geq 1$,

$$\boldsymbol{y}(\boldsymbol{n}) = \Bigg[\sum_{\substack{k=1 \\ n_k \geq 1}}^{K} \boldsymbol{y}(\boldsymbol{n} - \boldsymbol{e}_k)\boldsymbol{D}_k + \sum_{m=n_1+\cdots+n_K}^{\infty} \sum_{i=0}^{m} \boldsymbol{v}^{(i)}(\theta)\boldsymbol{D}^{(m-i)}(\theta)\boldsymbol{F}_m(\boldsymbol{n}) - \sum_{\substack{k=1 \\ n_k \geq 1}}^{K} \sum_{m=n_1+\cdots+n_K-1}^{\infty} \sum_{i=0}^{m} \boldsymbol{v}^{(i)}(\theta)\boldsymbol{D}_k^{(m-i)}(\theta)\boldsymbol{F}_m(\boldsymbol{n} - \boldsymbol{e}_k) \Bigg] (-\boldsymbol{C})^{-1},$$

where

$$\boldsymbol{D}^{(n)}(\theta) = \int_{0-}^{\infty} e^{-\theta x} \frac{(\theta x)^n}{n!} d\boldsymbol{D}(x), \qquad n = 0, 1, \ldots.$$

It is clear from Theorems 5.3 and 5.4 that $\boldsymbol{x}(\boldsymbol{n})$ and $\boldsymbol{y}(\boldsymbol{n})$ can be recursively computed if we obtain $\boldsymbol{v}^{(m)}(\theta)$ and $\boldsymbol{F}_m(\boldsymbol{n})$ $(m = 0, 1, \ldots)$.

We first consider the recursion for $\boldsymbol{F}_m(\boldsymbol{n})$. By definition, we have

$$\sum_{\substack{n_1\geq 0 \\ n_1+\cdots+n_K\leq m+1}} \cdots \sum_{n_K\geq 0} z_1^{n_1}\cdots z_K^{n_K}\boldsymbol{F}_{m+1}(\boldsymbol{n})$$

$$= \sum_{\substack{n_1\geq 0 \\ n_1+\cdots+n_K\leq m}} \cdots \sum_{n_K\geq 0} z_1^{n_1}\cdots z_K^{n_K}\boldsymbol{F}_m(\boldsymbol{n})\left[\boldsymbol{I}+\theta^{-1}\left(\boldsymbol{C}+\sum_{i=1}^{K} z_i\boldsymbol{D}_i\right)\right],$$

from which we have the following theorem.

THEOREM 5.5. *The $\boldsymbol{F}_m(\boldsymbol{n})$ ($m = 0,1,\ldots$, $\boldsymbol{n}\geq\boldsymbol{0}$ and $n_1+\cdots+n_K\leq m$) is computed by the following recursion with $\boldsymbol{F}_0(\boldsymbol{0}) = \boldsymbol{I}$: For $m = 0,1,\ldots$,*

$$\boldsymbol{F}_{m+1}(\boldsymbol{0}) = (\boldsymbol{I}+\theta^{-1}\boldsymbol{C})^{m+1},$$

$$\boldsymbol{F}_{m+1}(\boldsymbol{n}) = \boldsymbol{F}_m(\boldsymbol{n})(\boldsymbol{I}+\theta^{-1}\boldsymbol{C}) + \sum_{\substack{k=1 \\ n_k\geq 1}}^{K} \boldsymbol{F}_m(\boldsymbol{n}-\boldsymbol{e}_k)\theta^{-1}\boldsymbol{D}_k$$

$$\text{for} \quad \sum_{k=1}^{K} n_k = 1,2,\ldots,m,$$

$$\boldsymbol{F}_{m+1}(0,\ldots,0,\underset{k\text{th}}{m+1},0,\ldots,0) = (\theta^{-1}\boldsymbol{D}_k)^{m+1}.$$

Next we consider the computation of the $\boldsymbol{v}^{(m)}(\theta)$. We note that

$$\sum_{m=0}^{\infty} \boldsymbol{v}^{(m)}(\theta)z^m = \boldsymbol{v}^*(\theta-\theta z), \qquad |z|\leq 1,$$

and therefore $\sum_{m=0}^{\infty}\boldsymbol{v}^{(m)}(\theta) = \boldsymbol{\pi}$. Roughly speaking, $\boldsymbol{v}^{(m)}(\theta)$'s represent the vector probability mass functions of m arrivals from an independent Poisson stream with rate θ during the stationary virtual waiting time whose vector LST is given by $\boldsymbol{v}^*(s)$.

Substituting s in (21) by $\theta-\theta z$ yields

$$\sum_{m=0}^{\infty} \boldsymbol{v}^{(m)}(\theta)z^m[\theta(1-z)\boldsymbol{I}+\boldsymbol{C}+\boldsymbol{D}^*(\theta-\theta z)] = \theta(1-z)(1-\rho)\boldsymbol{\kappa},$$

from which it follows that

$$\sum_{m=0}^{\infty} \boldsymbol{v}^{(m)}(\theta) z^m \Big[\theta(1-z)\boldsymbol{I} + \boldsymbol{C} + \sum_{n=0}^{\infty} \boldsymbol{D}^{(n)}(\theta) z^n\Big] = \theta(1-z)(1-\rho)\boldsymbol{\kappa}. \tag{41}$$

Comparing coefficient vectors of z^m $(m = 0, 1, \ldots)$ in both sides of (41), we obtain

$$\left(\boldsymbol{v}^{(0)}(\theta), \boldsymbol{v}^{(1)}(\theta), \ldots\right) = \left(\boldsymbol{v}^{(0)}(\theta), \boldsymbol{v}^{(1)}(\theta), \ldots\right) \boldsymbol{\Gamma} \tag{42}$$

where $\boldsymbol{\Gamma}$ is given by

$$\boldsymbol{\Gamma} = \begin{pmatrix} \boldsymbol{B}_0 + \boldsymbol{B}_1 & \boldsymbol{B}_2 & \boldsymbol{B}_3 & \boldsymbol{B}_4 & \cdots \\ \boldsymbol{B}_0 & \boldsymbol{B}_1 & \boldsymbol{B}_2 & \boldsymbol{B}_3 & \cdots \\ \boldsymbol{O} & \boldsymbol{B}_0 & \boldsymbol{B}_1 & \boldsymbol{B}_2 & \cdots \\ \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{B}_0 & \boldsymbol{B}_1 & \cdots \\ \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{B}_0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}, \tag{43}$$

with

$$\boldsymbol{B}_0 = \boldsymbol{I} + \theta^{-1}\left(\boldsymbol{C} + \boldsymbol{D}^{(0)}(\theta)\right), \quad \boldsymbol{B}_m = \theta^{-1}\boldsymbol{D}^{(m)}(\theta), \quad m = 1, 2, \ldots. \tag{44}$$

Note here that $\boldsymbol{B}_m$'s $(m = 0, 1, \ldots)$ are nonnegative $M \times M$ matrices which satisfies $\sum_{m=0}^{\infty} \boldsymbol{B}_m \boldsymbol{e} = \boldsymbol{e}$. Thus matrix $\boldsymbol{\Gamma}$ is considered as the transition probability matrix of a certain Markov chain of M/G/1 type. Those and (42) lead to the following theorem.

THEOREM 5.6. *$\boldsymbol{v}^{(m)}(\theta)$ is identical to the stationary distribution of a Markov chain of $M/G/1$ type, whose transition probability matrix is given in (43).*

Theorem 5.6 implies that $\boldsymbol{v}^{(m)}(\theta)$ $(m = 0, 1, \ldots)$ can numerically be obtained by applying Algorithm 3.1 in section 3.1 to (43).

REMARK 5.2. Owing to a special structure of (43), the following simplifications are made in applying Algorithm 3.1 to (43).

1. $\boldsymbol{G}$ in (6) is given in terms of $\boldsymbol{Q}$:

$$\boldsymbol{G} = \boldsymbol{I} + \theta^{-1}\boldsymbol{Q}.$$

2. $\boldsymbol{v}^{(0)}(\theta)$ is given in terms of $\boldsymbol{f}$ in Step 4:

$$\boldsymbol{v}^{(0)}(\theta) = \frac{1-\rho}{\boldsymbol{f}\boldsymbol{B}_0\boldsymbol{e}} \cdot \boldsymbol{f}.$$

3. $\boldsymbol{\kappa}$ is given in terms of $\boldsymbol{f}$ in Step 4:

$$\frac{\boldsymbol{f}\boldsymbol{B}_0}{\boldsymbol{f}\boldsymbol{B}_0\boldsymbol{e}} = \boldsymbol{\kappa},$$

so that if we compute $\boldsymbol{f}$ first, $\boldsymbol{\kappa}$ is obtained by the above equation without solving a system (16) of linear equations.

6. Concluding remarks

This paper summarizes recent developments by the author, in matrix-analytic methods for work-conserving single-server queues with MAS input. For most of customer-average quantities of interest, we assume the FIFO service discipline. As shown in section 4, the stationary distributions of the amount of work-in-system, actual waiting time and sojourn times are fully characterized. Further utilizing those and the distributional form of Little's law, this paper shows how to obtain a numerically feasible recursion to compute the joint queue length mass function in the FIFO queue. In what follows, we provides some remarks on the results in queues with MAS input, which are not included in the paper.

For a queue with independent MAS and M/GI inputs, the stationary distribution of the amount of work-in-system is decomposed into two factors [38]. Namely consider a work-conserving single-server queue with independent MAS and M/GI inputs, where the MAS has representation $(\boldsymbol{C}, \boldsymbol{D}(x))$ and the M/GI input has rate λ_2 and the service time distribution function $h_2(x)$ whose LST is denoted by $h_2^*(s)$. We then have

$$\boldsymbol{v}^*(s) = v^*_{\mathrm{M/GI/1}}(s)\boldsymbol{v}^*_{\mathrm{MAS}}(s - \lambda_2 + \lambda_2 h_2^*(s)),$$

where $v^*_{\mathrm{M/GI/1}}(s)$ denotes the LST of the stationary distribution of the amount of work-in-system in the M/GI/1 queue (see (23)), and $\boldsymbol{v}^*_{\mathrm{MAS}}(s)$ denotes the LST of the stationary distribution of the amount of work-in-system in a special queue with MAS input having modified service time distributions. See [38] for details.

The time-dependent LST of the distribution of the amount of work-in-system was obtained in [36]. Also, the waiting time and joint queue length distributions in vacation queues with MAS input was studied in [47]. Last-come, first-served queues with MAS input are considered in [19, 36, 39].

Multiple arrival streams are naturally arise in priority queues. The LSTs of the stationary waiting time and sojourn time distributions in

the preemptive-resume priority queue with MAS input were obtained in [36]. For the nonpreemptive case, the waiting time distribution and the marginal queue length distributions were analyzed in [40, 44].

A. Computational algorithms

A.1. Computation of $\boldsymbol{A}_k$ for MAP arrivals [25, 37]

The $\boldsymbol{A}_k$ is given by

$$\boldsymbol{A}_k = \sum_{n=k}^{\infty} \gamma_n \boldsymbol{F}_n(k), \qquad k = 0, 1, \ldots,$$

where γ_n is given in (45) and $\boldsymbol{F}_n(k)$ $(n = 0, 1, \ldots,\ k = 0, \ldots, n)$ is recursively computed by

$$\boldsymbol{F}_{n+1}(k) = \begin{cases} \boldsymbol{F}_n(0)(\boldsymbol{I} + \theta^{-1}\boldsymbol{C}) = (\boldsymbol{I} + \theta^{-1}\boldsymbol{C})^{n+1}, & k = 0, \\ \boldsymbol{F}_n(k)(\boldsymbol{I} + \theta^{-1}\boldsymbol{C}) + \boldsymbol{F}_n(k-1)(\theta^{-1}\boldsymbol{D}), & 1 \le k \le n, \\ \boldsymbol{F}_n(n)(\theta^{-1}\boldsymbol{D}) = (\theta^{-1}\boldsymbol{D})^{n+1}, & k = n+1. \end{cases}$$

A.2. Computation of $\boldsymbol{G}$ for MAP arrivals [26]

We define θ and γ_n $(n = 0, 1, \ldots)$ as

$$\theta = \max_{i \in \mathcal{M}} |C_{i,i}|, \qquad \gamma_n = \int_0^{\infty} e^{-\theta x} \frac{(\theta x)^n}{n!}. \tag{45}$$

Then $\boldsymbol{G}$ is given by the limit $\boldsymbol{G}_\infty$ in the recursion

$$\boldsymbol{G}_{n+1} = \sum_{n=0}^{\infty} \gamma_n (\boldsymbol{I} + \theta^{-1}(\boldsymbol{C} + \boldsymbol{D}\boldsymbol{G}_n))^n, \qquad n = 0, 1, \ldots,$$

with $\boldsymbol{G}_0 = \boldsymbol{O}$.

A.3. Recursion for moments $\boldsymbol{v}^{(n)}$ of work-in-system [45]

Let $\boldsymbol{D}^{(n)}$ $(n = 1, 2, \ldots)$ denote

$$\boldsymbol{D}^{(n)} = \lim_{s \to 0+} (-1)^n \frac{d^n}{ds^n} \boldsymbol{D}^*(s).$$

$\boldsymbol{v}^{(n)}$ $(n = 1, 2, \ldots)$ is then recursively computed by

$$\boldsymbol{v}^{(0)} = \boldsymbol{\pi},$$
$$\boldsymbol{v}^{(n)} = \sum_{m=0}^{n-1} \boldsymbol{v}^{(m)} \boldsymbol{R}^{(n-m)} (\boldsymbol{I} - \boldsymbol{R}^{(0)})^{-1}, \qquad n = 1, 2, \ldots,$$

where $\boldsymbol{R}^{(n)}$'s are $M \times M$ matrices which are determined by

$$\boldsymbol{R}^{(0)} = \boldsymbol{I} + (\boldsymbol{D} - \boldsymbol{I})\boldsymbol{e}\boldsymbol{\kappa} + (\boldsymbol{C} + \boldsymbol{D})(\boldsymbol{e}\boldsymbol{\kappa} - \boldsymbol{Q})^{-1}.$$
$$\boldsymbol{R}^{(n)} = \left(\boldsymbol{D}^{(n)} + \frac{1}{n} \boldsymbol{D}^{(n+1)} \boldsymbol{e}\boldsymbol{\kappa} - n\boldsymbol{R}^{(n-1)} \right) (\boldsymbol{e}\boldsymbol{\kappa} - \boldsymbol{Q})^{-1}, \qquad n = 1, 2, \ldots.$$

B. Proofs

B.1. Proof of Lemma 4.1

For any real $s \in (-A, \infty)$, $\boldsymbol{C} + \boldsymbol{D}^*(s)$ is an irreducible ML-matrix [32], so that Lemma 4.1 (a) and (b) follow from Theorem 2.6 in [32].

For any real $s \in (-A, \infty)$, $\exp(\boldsymbol{C} + \boldsymbol{D}^*(s))$ is an irreducible, nonnegative matrix whose Perron-Frobenius eigenvalue is given by $\exp(\delta^*(s))$. Because the (i, j)th element of $\exp(\boldsymbol{C} + \boldsymbol{D}^*(s))$ is given by

$$\int_0^\infty e^{-sx} dB_{i,j}(x)$$

for some (defective) distribution function $B_{i,j}(x)$, Theorem A.1.1 in [28] implies that $\log(\exp(\delta^*(s))) = \delta^*(s)$ is a strictly decreasing convex function in $s \in (-A, \infty)$, which completes the proof. □

B.2. Proof of Lemma 4.2

Let $\delta^{(1)}(s)$ and $\boldsymbol{D}^{(1)}(s)$ denote the first derivatives of $\delta^*(s)$ and $\boldsymbol{D}(s)$, respectively. Note that $\delta^{(1)}(0) = \boldsymbol{l}^*(0)\boldsymbol{D}^{(1)}(0)\boldsymbol{r}^*(0) = \boldsymbol{\pi}\boldsymbol{D}^{(1)}(0)\boldsymbol{e} = -\rho > -1$. Thus the convexity of $\delta^*(s)$ leads to Lemma 4.2. □

B.3. Proof of Theorem 4.6

Note first that

$$\int_0^\infty \exp(-sx)\overline{\boldsymbol{v}}(x)dx = \frac{\boldsymbol{\pi} - \boldsymbol{v}^*(s)}{s}.$$

It then follows from (21) that

$$\frac{\boldsymbol{\pi} - \boldsymbol{v}^*(s)}{s}[s\boldsymbol{I} + \boldsymbol{C} + \boldsymbol{D}^*(s)] = \frac{\boldsymbol{\pi}}{s}[s\boldsymbol{I} + \boldsymbol{C} + \boldsymbol{D}^*(s)] - (1 - \rho)\boldsymbol{\kappa}. \tag{46}$$

To prove the theorem, we shall apply Theorem 4 (Tauberian Theorem) of Chapter XIII.5 in [14] with $\rho = 1$ and constant $L(x)$. Post-multiplying both sides of (46) by $\boldsymbol{r}^*(s)$, using $(\boldsymbol{C} + \boldsymbol{D}^*(s))\boldsymbol{r}^*(s) = \delta^*(s)\boldsymbol{r}^*(s)$, and rearranging terms, we obtain

$$\frac{\boldsymbol{\pi} - \boldsymbol{v}^*(s)}{s}\boldsymbol{r}^*(s) = \frac{\boldsymbol{\pi}\boldsymbol{r}^*(s)}{s} - \frac{1-\rho}{s+\delta^*(s)}\boldsymbol{\kappa}\boldsymbol{r}^*(s),$$

and therefore

$$\lim_{s\to-\eta}(s+\eta)\frac{\boldsymbol{\pi} - \boldsymbol{v}^*(s)}{s}\boldsymbol{r}^*(s) = -\frac{1-\rho}{1+\delta^{(1)}(-\eta)}\boldsymbol{\kappa}\boldsymbol{r}^*(-\eta). \tag{47}$$

On the other hand, with (46), we have

$$\begin{aligned}&\frac{\boldsymbol{\pi} - \boldsymbol{v}^*(s)}{s}[s\boldsymbol{I} + \boldsymbol{C} + \boldsymbol{D}^*(s) - \delta^*(s)\boldsymbol{r}^*(s)\boldsymbol{l}^*(s)]\\ =&\frac{\boldsymbol{\pi}}{s}[s\boldsymbol{I} + \boldsymbol{C} + \boldsymbol{D}^*(s)] - (1-\rho)\boldsymbol{\kappa} - \frac{\boldsymbol{\pi} - \boldsymbol{v}^*(s)}{s}\delta^*(s)\boldsymbol{r}^*(s)\boldsymbol{l}^*(s),\end{aligned}$$

from which and (29), it follows that

$$\begin{aligned}&\left(\lim_{s\to-\eta}(s+\eta)\frac{\boldsymbol{\pi} - \boldsymbol{v}^*(s)}{s}\right)[-\eta\boldsymbol{I} + \boldsymbol{C} + \boldsymbol{D}^*(-\eta) - \eta\boldsymbol{r}^*(-\eta)\boldsymbol{l}^*(-\eta)]\\ =&-\eta\left(\lim_{s\to-\eta}(s+\eta)\frac{\boldsymbol{\pi} - \boldsymbol{v}^*(s)}{s}\boldsymbol{r}^*(s)\right)\boldsymbol{l}^*(-\eta),\end{aligned}$$

and using (47), we obtain

$$\begin{aligned}&\left(\lim_{s\to-\eta}(s+\eta)\frac{\boldsymbol{\pi} - \boldsymbol{v}^*(s)}{s}\right)[-\eta\boldsymbol{I} + \boldsymbol{C} + \boldsymbol{D}^*(-\eta) - \eta\boldsymbol{r}^*(-\eta)\boldsymbol{l}^*(-\eta)]\\ =&\frac{(1-\rho)\eta\boldsymbol{\kappa}\boldsymbol{r}^*(-\eta)}{1+\delta^{(1)}(-\eta)}\boldsymbol{l}^*(-\eta).\end{aligned} \tag{48}$$

Note here that $\boldsymbol{F}(-\eta) = -\eta\boldsymbol{I} + \boldsymbol{C} + \boldsymbol{D}^*(-\eta) - \eta\boldsymbol{r}^*(-\eta)\boldsymbol{l}^*(-\eta)$ is nonsingular, which is shown in the following way. If $\boldsymbol{F}(-\eta)\boldsymbol{u} = \boldsymbol{0}$ for some column vector $\boldsymbol{u}$, then $\boldsymbol{l}^*(-\eta)\boldsymbol{F}(-\eta)\boldsymbol{u} = -\eta\boldsymbol{l}^*(-\eta)\boldsymbol{u} = \boldsymbol{0}$, so that $[-\eta\boldsymbol{I} + \boldsymbol{C} + \boldsymbol{D}^*(-\eta)]\boldsymbol{u} = \boldsymbol{0}$ and therefore $\boldsymbol{u} = k\boldsymbol{r}^*(-\eta)$ for some k. Thus from $\boldsymbol{F}(-\eta)\boldsymbol{u} = \boldsymbol{0}$, we have $k\boldsymbol{r}^*(-\eta) = \boldsymbol{0}$. Because $\boldsymbol{r}^*(-\eta)$ is strictly positive (Lemma 4.1 (a)), we have $k = 0$. As a result, $\boldsymbol{u} = \boldsymbol{0}$ if $\boldsymbol{F}(-\eta)\boldsymbol{u} = \boldsymbol{0}$, which implies $\boldsymbol{F}(-\eta)$ is nonsingular.

Thus noting $\boldsymbol{l}^*(-\eta)\boldsymbol{F}(-\eta) = -\eta\boldsymbol{l}^*(-\eta)$, we have from (48)

$$\lim_{s\to-\eta}(s+\eta)\frac{\boldsymbol{\pi} - \boldsymbol{v}^*(s)}{s} = -\frac{(1-\rho)\boldsymbol{\kappa}\boldsymbol{r}^*(-\eta)}{1+\delta^{(1)}(-\eta)}\boldsymbol{l}^*(-\eta). \tag{49}$$

Note here that $\delta^{(1)}(-\eta) < -1$ because of Lemma 4.1 (c) and $\delta^{(1)}(0) > -1$. Further noting $\delta^{(1)}(-\eta) = \boldsymbol{l}^*(-\eta)\boldsymbol{D}^{(1)}(-\eta)\boldsymbol{r}^*(-\eta)$ and applying Theorem 2 of Chapter XIII.5 in [14] to (49), we obtain (30).

It is easy to see that (32) comes from the definition of $\overline{\boldsymbol{w}_k}(x)$ and (30). For $\overline{\boldsymbol{t}_k}(x)$, we follow an argument in [3]. Note that

$$\begin{aligned} e^{\eta x}\overline{\boldsymbol{t}_k}(x) &= \frac{1}{\lambda_k}\int_0^x e^{\eta(x-y)}\overline{\boldsymbol{v}}(x-y)e^{\eta y}d\boldsymbol{D}_k(y)+\frac{1}{\lambda_k}e^{\eta x}\boldsymbol{\pi}(\boldsymbol{D}_k-\boldsymbol{D}_k(x)) \\ &= \frac{1}{\lambda_k}\int_0^\infty 1(y\in[0,x])e^{\eta(x-y)}\overline{\boldsymbol{v}}(x-y)e^{\eta y}d\boldsymbol{D}_k(y)+\frac{1}{\lambda_k}e^{\eta x}\boldsymbol{\pi}\overline{\boldsymbol{D}_k}(x), \end{aligned}$$

where $\overline{\boldsymbol{D}_k}(x) = \boldsymbol{D}_k - \boldsymbol{D}_k(x)$. Because $\eta > -A$, $\boldsymbol{D}_k^*(-\eta) < \infty$, so that $\exp(\eta x)\overline{\boldsymbol{D}_k}(x)$ goes to 0 as $x \to \infty$. Thus dominated convergence theorem leads to

$$\begin{aligned} \lim_{x\to\infty} e^{\eta x}\overline{\boldsymbol{t}_k}(x) &= \frac{1}{\lambda_k}\int_0^\infty \left(\lim_{x\to\infty} 1(y\in[0,x])e^{\eta(x-y)}\overline{\boldsymbol{v}}(x-y)\right)e^{\eta y}d\boldsymbol{D}_k(y) \\ &= \frac{1}{\lambda_k}\left(\lim_{x\to\infty} e^{\eta x}\overline{\boldsymbol{v}}(x)\right)\boldsymbol{D}_k^*(-\eta), \end{aligned}$$

from which and (30), (32) follows.

ACKNOWLEDGEMENT. The author is thankful to Bong Dae Choi for his invitation to writing this paper. This research was supported in part by Grant-in-Aid for Scientific Research (C) of Japan Society for the Promotion of Science under Grant No. 12650380.

References

[1] J. Abate, G. L. Choudhury, and W. Whitt, *Asymptotics for steady-state tail probabilities in structured Markov chain models*, Stoch. Model. **10** (1995), 99–143.

[2] ———, *Exponential approximations for tail probabilities in queues, I: Waiting times*, Opns. Res. **43** (1995), 885–901.

[3] ———, *Exponential approximations for tail probabilities in queues, II: Sojourn time and workload*, Opns. Res. **44** (1995), 758–763.

[4] A. S. Alfa, B. Sengupta, and T. Takine, *The use of non-linear programming in matrix analytic methods*, Stoch. Model. **14** (1998), 351–367.

[5] N. Akar and K. Sohraby, *An invariant subspace approach in M/G/1 and G/M/1 type Markov chains*, Stoch. Model. **13** (1997).

[6] N. Akar, N. C. Oguz, and K. Sohraby, *Matrix-geometric solution in finite and infinite M/G/1 type Markov chains: a unifying generalized state-space approach*, IEEE J. Select. Areas Communi. **16** (1998), 626–639.

[7] S. Asmussen, *Ladder heights and the Markov-modulated M/G/1 queue*, Stochastic Process. Appl. **37** (1991), 313–326.

[8] S. Asmussen and G. Koole, *Marked point processes as limits of Markovian arrival streams*, J. Appl. Probab. **30** (1993), 365–372.

[9] S. Asmussen, L. F. Henriksen, and C. Klüppelberg, *Large claims approximations for risk processes in a Markovian environment*, Stochastic Process. Appl. **54** (1994), 29–43.

[10] P. L. Butzer and H. Berens, *Semi-Groups of Operators and Approximation*, Springer-Verlag, New York, 1967.

[11] J. W. Cohen, *The Single Server Queue, Revised Ed.*, North-Holland, Amsterdam, 1982.

[12] P. Embrechts and C. M. Goldie, *On convolution tails*, Stochastic Process. Appl. **13** (1982), 263–278.

[13] P. Embrechts, C. Klüppelberg, and T. Mikosch, *Modelling Extremal Events for Insurance and Finance*, Springer-Verlag, Berlin, 1997.

[14] W. Feller, *An Introduction to Probability Theory and Its Applications, Volume II*, John Wiley and Sons, New York, 1966.

[15] A. Graham, *Kronecker Products and Matrix Calculus with Applications*, Ellis Horwood, Chichester, 1981.

[16] L. Gün, *Experimental results on matrix-analytic solution techniques – extensions and comparisons*, Stoch. Model. **5** (1989), 669–682.

[17] Q.-M. He, *Queues with marked customers*, Adv. in Appl. Probab. **28** (1996), 567–587.

[18] Q.-M. He and M. F. Neuts, *Markov chains with marked transitions*, Stochastic Process. Appl. **74** (1998), 37–52.

[19] Q.-M. He, *Classification of Markov processes of matrix M/G/1 type with tree structure and its applications to the MMAP[K]/G[K]/1 queues*, to appear in Stoch. Model. **16** (2000).

[20] ———, *The versatility of MMAP[K] and the MMAP[K]/G[K]/1 queue*, to appear in QUESTA, 2000.

[21] P. R. Jelenković and A. A. Lazar, *Subexponential asymptotics of a Markov-modulated random walk with queueing applications*, J. Appl. Probab. **35** (1998), 325–347.

[22] J. Keilson and L. D. Servi, *A distributional form of Little's law*, Oper. Res. Lett. **7** (1988), 223–227.

[23] D. König and V. Schmidt, *Extended and conditional versions of the PASTA property*, Adv. in Appl. Probab. **22** (1990), 510–512.

[24] R. M. Loynes, *The stability of a queue with non-independent interarrival and service times*, Proc. Cambridge Philos. Soc. **58** (1962), 497–520.

[25] D. M. Lucantoni and V. Ramaswami, *Efficient algorithms for solving the non-linear matrix equations arising in phase type queues*, Stoch. Model. **1** (1985), 29–52.

[26] D. M. Lucantoni, K. S. Meier-Hellstern, and M. F. Neuts, *A single-server queue with server vacations and a class of non-renewal arrival processes*, Adv. in Appl. Probab. **22** (1990), 676–705.

[27] M. F. Neuts, *Generalizations of the Pollaczek-Khinchin integral equation in the theory of queues*, Adv. in Appl. Probab. **18** (1986), 952–990.

[28] ———, *Structured Stochastic Matrices of M/G/1 Type and Their Applications*, Marcel Dekker, New York, 1989.

[29] V. Ramaswami, *Stable recursion for the steady state vector for Markov chains of M/G/1 type*, Stoch. Model. **4** (1988), 183–188.

[30] G. J. K. Regterschot and J. H. A. de Smit, *The queue* M|G|1 *with Markov modulated arrivals and services*, Math. Oper. Res. **11** (1986), 465–483.

[31] E. Seneta, *Regularly Varying Functions*, Lecture Notes in Math., Springer-Verlag, Berlin **508** (1976).

[32] ———, *Non-negative Matrices and Markov Chain*, Springer-Verlag, New York, 1980.

[33] B. Sengupta, *An invariance relationship for the G/G/1 queue*, Adv. in Appl. Probab. **21** (1989), 956–957.

[34] K. Sigman, *Appendix: A primer on heavy-tailed distribution*, QUESTA **33** (1999), 261–275.

[35] L. Takács, *Investigation of waiting time problems by reduction to Markov processes*, Acta Math. Academ. Sci. Hungar. **6** (1955), 101–129.

[36] T. Takine and T. Hasegawa, *The workload in the MAP/G/1 queue with state-dependent services: Its application to a queue with preemptive resume priority*, Stoch. Model. **10** (1994), 183–204.

[37] T. Takine, Y. Matsumoto, T. Suda, and T. Hasegawa, *Mean waiting times in nonpreemptive priority queues with Markovian arrival and i.i.d. service processes*, Perfor. Eval. **20** (1994), 131–149.

[38] T. Takine, *On the single-server queue with independent MAP/G and M/GI input streams*, Stoch. Model. **11** (1995), 227–234.

[39] T. Takine, B. Sengupta, and R. Yeung, *A generalization of the matrix M/G/1 paradigm for Markov chains with a tree structure*, Stoch. Model. **11** (1995), 411–421.

[40] T. Takine, *A nonpreemptive priority MAP/G/1 queue with two classes of customers*, J. Opns. Res. Soc. Japan **39** (1996), 266–290.

[41] ———, *A continuous version of matrix-analytic methods with the skip-free to the left property*, Stoch. Model. **12** (1996), 673–682.

[42] T. Takine and Y. Takahashi, *On the relationship between queue lengths at a random instant and at a departure in the stationary queue with BMAP arrivals*, Stoch. Model. **14** (1998), 601–610.

[43] T. Takine, *Queue length distribution in a FIFO single-sever queue with multiple arrival streams having different service time distributions*, submitted for publication (Tech. Report #98011, Department of Applied Mathematics and Physics, Kyoto University), 1998.

[44] ———, *The nonpreemptive priority MAP/G/1 queue*, Oper. Res. **47** (1999), 917–927.

[45] ———, *Matrix product-form solution for an LCFS-PR single-server queue with multiple arrival streams governed by a Markov chain*, submitted for publication (Tech. Report #2000-003, Department of Applied Mathematics and Physics, Kyoto University), 2000.

[46] ———, *A new recursion for the queue length distribution in the stationary BMAP/G/1 queue*, Stoch. Model. **16** (2000), 335–341.

[47] ———, *Distributional form of Little's law for FIFO queues with multiple Markovian arrival streams and its application to queues with vacations*, to appear in Queueing Systems, 2001.

[48] ———, *Subexponential asymptotics of the waiting time distribution in a single-server queue with multiple Markovian arrival streams*, to appear in Stoch. Model., 2001.

[49] H. C. Tijms, *Stochastic Models, An Algorithmic Approach*, Wiley, Chichester, 2000.

[50] Y. Zhu and N. U. Prabhu, *Markov-modulated PH/G/1 queueing systems*, QUESTA **9** (1991), 313–322.

Department of Applied Mathematics and Physics
Graduate School of Informatics
Kyoto University
Kyoto 606-8501, Japan
E-mail: takine@amp.i.kyoto-u.ac.jp