

# The stationary tail asymptotics in the $GI/G/1$ type queue with countably many background states\*

Masakiyo Miyazawa,  
Tokyo University of Science

Yiqiang Q. Zhao  
Carleton University

## Abstract

We consider asymptotic behaviors of the stationary tail probabilities in the discrete time  $GI/G/1$  type queue with a countable background state space. These probabilities are presented in a matrix form with respect to the background state space, and shown to be the solution of a Markov renewal equation. Using this fact, we consider their decay rates. Applying the Markov renewal theorem, it is shown that certain reasonable conditions lead to the geometric decay of the tail probabilities as the level goes to infinity. We exemplify this result using a discrete time priority queue with a single server and two types of customers.

**Keywords:**  $GI/G/1$  type queue, infinite background states, decay rate, stationary distribution, Markov additive process, Wiener-Hopf factorization, dual process.

## 1. Introduction

We consider asymptotic behaviors of the stationary tail probabilities in the discrete time  $GI/G/1$  type queue with a countable background state space. For finitely many background states, this model has been introduced in [12] under the name “blocked-structured Markov chains with repeating rows”. A fundamental structure has been found in [12] that is useful in computing stationary distributions. It extends the framework of the  $M/G/1$  and  $GI/M/1$  types of queues developed by Neuts [20, 21]. More results about this model can be found in [28, 29]. In particular, the decay rate of the stationary distribution is considered for the light tailed case by Li and Zhao [13]. However, all these results are only concerned with a finite background state space. If the background states are countably many, finding the decay rate would be much more challenging.

We aim to characterize the decay rate in this countable case. Our interest is in the light-tailed case, specifically in the case that the tail probabilities decay geometrically fast. This study is motivated by the Markov additive paradigm of Miyazawa [17], which proposes the use of one dimensional Markov additive processes in studying the decay rate problems. One of the key issues here is to have sufficient information in background

---

\*This research is supported in part by JSPS under grant No. 13680532, and by a research grant from NSERC.

states to cope with the multidimensional structure in states of queueing networks. In this paper, we limit our attention to discrete state and discrete time models, i.e., the  $GI/G/1$  type queue with countable background states, as a first step of research in this direction. Nevertheless, the method can be applied in a wide range of applications. It should be noted that this  $GI/G/1$  type queue is not a Markov modulated  $GI/G/1$  queue with countable phases (see Remark 2.1 for more details).

We employ the Markov renewal approach developed in Miyazawa [16, 18], combining it with the censoring representation of the stationary distribution of [12, 28] and the Wiener Hopf factorization for a Markov additive process (e.g., see [4, 29]). We also use the characterization on the periodicity of the ladder heights in the Markov additive process due to Alsmeyer [3], in which the Markov additive process is termed as a Markov random walk. These ideas enable us to get conditions under which the tail probabilities decay geometrically (see Theorem 4.1). This result generalizes the corresponding result for the  $M/G/1$  type queue in Miyazawa [18]. We also exemplify it by a discrete time single server priority queue with two types of customers. This example shows where we have to be careful in applications of the main result.

Historically, decay rate problems in queues have been widely studied in literature because of their importance. In particular, the large deviation technique is popular, since it can be applied not only to single queues but also to some queueing networks with feed forward routing as well as multiple queues of single servers such as generalized processor sharing (e.g., see [7, 26]). This technique is particularly useful when arrival times of customers have complicated dependencies. However, it has two main drawbacks. First, it only gives the order of the exponential decay, and provides no information about its prefactor. Secondly, it may be difficult to be applied to the study of the tail of multidimensional distributions in queueing networks because they usually behave differently at the boundary of their state spaces.

To obtain more qualitative results, the change of measure technique together with martingales has also been used (e.g., see [6]). This approach is sometimes powerful to find upper bounds of the tail probabilities, but may be difficult to evaluate the prefactor of the decay rate. Some related work can also be found in Foley and McDonald [11] and McDonald [15], in which the change of measure and harmonic functions are used to get the prefactor. Classical approaches, such as using matrix computations or complex variables, are also useful. These may provide us with more detailed information, particularly when the background state space is finite (see, e.g., [13]). However, they do not seem to work well generally when the space is infinite.

The Markov additive approach used in this paper is rather classical as well (e.g., see [5]). There are many studies on Markov additive processes and their asymptotic behaviors (e.g., see [8, 10, 22] and their references). In particular, Chan and Lai [8] derive Cram el-Lundberg type asymptotics for hitting probabilities of a multi-dimensional additive component under a very general setting. Some of those results are closely related to ours. For example, Theorem 3 of [8] assumes similar but slightly stronger conditions on the Markov additive kernel (see Section 4). However, such asymptotics of the hitting probabilities have been scarcely applied for queueing processes, particularly, of queueing networks, since these have different transition structure at boundaries, e.g., they are not simple functions such as the supremum of the additive components in general as well

as the background process itself may not be Markov (see Section 2). In this respect, the Markov additive approach seems to have not been well developed or explored for queueing applications, particularly when the background states are infinite. This paper intends to fill this gap.

This paper consists of six sections. We introduce the  $GI/G/1$  type queue in Section 2, and describe a Markov additive process associated with it in Section 3. The main result and its proof are given in Section 4. Pilot examples are considered in Section 5. We finally make some remarks in Section 6.

## 2. $GI/G/1$ type queue

Let  $\mathbb{N}$  be the set of all nonnegative integers and let  $S_0$  and  $S_1$  be finite or countable sets. We consider a Markov chain with two dimensional state space  $S = (\{0\} \times S_0) \cup ((\mathbb{N} - \{0\}) \times S_1)$ , and transition probability matrix  $P$  of the following Markov additive structure. The entries of  $P$  are given by

$$\begin{aligned} P((0, i), (0, j)) &= B_0(i, j), & i, j \in S_0, \\ P((0, i), (n, j)) &= B_n(i, j), & n \geq 1, i \in S_0, j \in S_1, \\ P((n, i), (0, j)) &= B_{-n}(i, j), & n \geq 1, i \in S_1, j \in S_0, \\ P((m, i), (n, j)) &= A_{n-m}(i, j), & n, m \geq 1, i, j \in S_1, \end{aligned}$$

where  $B_0, B_n$  for  $n \geq 1, B_{-n}$  for  $n \leq -1$ , and  $A_n$  are arbitrary  $S_0 \times S_0, S_0 \times S_1, S_1 \times S_0$  and  $S_1 \times S_1$  nonnegative matrices, respectively, such that for  $S_0$ - and  $S_1$ -dimensional column vectors  $\mathbf{e}_{(0)}$  and  $\mathbf{e}$  of ones, respectively,

$$B_0 \mathbf{e}_{(0)} + \sum_{\ell=1}^{\infty} B_{\ell} \mathbf{e} = \mathbf{e}_{(0)}, \quad B_{-n} \mathbf{e}_{(0)} + \sum_{\ell=-n+1}^{\infty} A_{\ell} \mathbf{e} = \mathbf{e}, \quad n \geq 1.$$

Obviously, these conditions ensure that  $P$  is a stochastic matrix. Symbolically, the matrix  $P$  can be written as

$$P = \begin{pmatrix} B_0 & B_1 & B_2 & B_3 & \cdots \\ B_{-1} & A_0 & A_1 & A_2 & \cdots \\ B_{-2} & A_{-1} & A_0 & A_1 & \cdots \\ B_{-3} & A_{-2} & A_{-1} & A_0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \cdots \end{pmatrix}.$$

Throughout the paper, we assume that  $P$  is irreducible and aperiodic. To have the irreducibility, we can not accommodate dummy states. This is the reason why we have introduced two sets  $S_0$  and  $S_1$ .

When both  $S_0$  and  $S_1$  are finite, this model has been studied under the name “the  $GI/G/1$  type queue”, which is an extension of the  $GI/M/1$  and  $M/GI/1$  type queues. In these models,  $S_0$  is typically different from  $S_1$ . The first component  $n$  of state  $(n, i) \in S$  is referred to as the level, and the second one as the phase or background state. We refer to this model as the  $GI/G/1$  type queue with countable background states. One may consider this model as a discrete valued Markov random walk reflected at level zero,

but this is incorrect since the background process may depend on the level and therefore it may not be a Markov chain by itself. This situation is crucial for queueing network applications.

Our primary concern is with the stationary distribution of  $P$ , and in particular its asymptotic tail behavior. Unlike the case in which  $S_0$  and  $S_1$  are finite, the stability condition, i.e. conditions for the existence of the stationary distribution, is not obvious (e.g., see Remark 2.1 below and Theorem 3.1 of [18]). However, such a condition is well known for many queueing models. So, we shall just assume the stability of  $P$ .

Although the background process itself may not be a Markov chain, it has Markov structure as long as the level never hits the origin. This off-boundary process will be useful to see the asymptotic tail behavior as well as is related to the stability. Thus, we define  $A$  as

$$A = \sum_{n=-\infty}^{+\infty} A_n. \quad (2.1)$$

This  $A$  is a transition matrix for the off-boundary background process but not for the whole background process. Note that  $A$  is substochastic but may not be stochastic. Even if it is stochastic, it may not be recurrent. Furthermore, the irreducibility of  $P$  may not imply the irreducibility of  $A$ . Models for which this can happen will be discussed in Section 5. However, we need good conditions for  $A$  to consider the asymptotic behavior. For this, we assume the following condition throughout the paper:

- (i)  $A$  is irreducible and aperiodic.

**Remark 2.1** In this paper, we do not assume that  $A$  has the stationary distribution. If this is the case, a typical stability condition requires that, if the boundary is removed, the level drifts to  $-\infty$ , i.e.,

$$\sum_{n=-\infty}^{+\infty} n\pi A_n \mathbf{e} < 0, \quad (2.2)$$

where  $\pi$  is the stationary distribution of  $A$ . But, this condition may not be sufficient for the  $GI/G/1$  type queue since the model may have different state transitions at boundaries, which are needed for describing multi-line queues and queueing networks (e.g., see [17]). This situation is contrasted with a Markov modulated  $GI/G/1$  queue, in which the interarrival and service times are modulated by a given Markov chain. In this case, the background process is this Markov chain. For this class of models, condition (2.2) is well known as the stability condition due to Loynes [14].

### 3. Markov additive process

Because of the random walk structure of  $P$ , except in level 0, it is useful to consider the Markov additive process generated by  $\{A_\ell; \ell \in \mathbb{Z}\}$ , where  $\mathbb{Z}$  is the set of all integers.

Let  $\{(X_n, Y_n); n = 0, 1, \dots\}$  be a sequence of random vectors generated by the following transition function.

$$P(X_{n+1} = j, Y_{n+1} - Y_n = \ell | X_k, Y_k, k = 0, 1, \dots, n) = A_\ell(X_n, j), \quad n \geq 0, j \in S_1, \ell \in \mathbb{Z}.$$

This  $\{(X_n, Y_n)\}$  is said to be a Markov additive process with background state  $X_n$  and additive component  $Y_n$ . We also refer to  $Y_n$  as level. This order of the components are standard for a Markov additive process. The order should not be confused with the state description in the transition matrix  $P$ , which has the level as the first component. This is convenient for matrix description. Furthermore, it should be noticed that the background process of the Markov additive process is different from the one of  $P$  in general. Since  $A$  of (2.1) may be strictly substochastic, this Markov additive process may be terminating in finite time. Throughout the paper, we assume

- (ii) Markov additive process  $\{(X_n, Y_n)\}$  is 1-arithmetic. Since  $Y_n$  is integer valued, this is equivalent to that there is no integer  $d > 1$  and a function  $\delta$  from  $S_1$  to  $\{0, 1, \dots, d-1\}$  such that  $A_n(i, j) > 0$  implies  $n - \delta(i) + \delta(j) \in d\mathbb{Z}$  for all  $i, j \in S_1$  (see [2] and [24]).

We also refer to this condition as  $\{A_\ell; \ell \in \mathbb{Z}\}$  to be 1-arithmetic.

**Remark 3.1** The  $\delta(\cdot)$  is called a shift function. Note that the shift function must be identically 0 for the 1-arithmetic case since the additive component is integer valued. Also note that 1-arithmetic of  $\{A_\ell; \ell \in \mathbb{Z}\}$  is different from the aperiodicity of  $A$ . A sufficient condition for 1-arithmetic is that the greatest common divisor of

$$\{n_1 + n_2 + \dots + n_k; A_{n_1}(i, j_1)A_{n_2}(j_1, j_2) \times \dots \times A_{n_k}(j_{k-1}, i) > 0, n_\ell \in \mathbb{Z}, j_\ell \in S_1\}$$

is one for some  $i \in S_1$ . This is an easy condition to verify 1-arithmetic.

We next introduce two important measures for the Markov additive process. Let

$$\tau_\ell^- = \inf\{n \geq 1; Y_n - Y_0 \leq \ell\}, \quad \ell \in \mathbb{Z},$$

where  $\tau_\ell^- = \infty$  if there is no  $n$  such that  $Y_n \leq \ell$ . That is,  $\tau_\ell^-$  is the hitting time of  $Y_n$  at the set  $(-\infty, \ell]$ . Define  $S_1 \times S_1$  matrices  $R_\ell^+$ ,  $G_\ell^-$  and  $H_0^-$  by

$$\begin{aligned} R_\ell^+(i, j) &= E\left(1(\tau_{\ell-1}^- < \infty) \sum_{n=1}^{\tau_{\ell-1}^- - 1} 1(X_n = j, Y_n - Y_0 = \ell) \middle| X_0 = i\right), \quad \ell \geq 1, \\ G_\ell^-(i, j) &= P(X_{\tau_{\ell-1}^-} = j, Y_{\tau_{\ell-1}^-} - Y_0 = \ell | X_0 = i), \quad \ell \leq -1, \\ H_0^-(i, j) &= P(X_{\tau_0^-} = j, Y_{\tau_0^-} - Y_0 = 0 | X_0 = i), \end{aligned}$$

where  $X_{\tau_{\ell-1}^-} = j$  tacitly assumes that  $\tau_{\ell-1}^- < \infty$ , and similar conventions are made for  $Y_{\tau_{\ell-1}^-}$  and others. Denote the matrix generating functions for the sequences of matrices  $\{A_\ell; \ell \in \mathbb{Z}\}$ ,  $\{R_\ell^+; \ell \geq 1\}$  and  $\{G_\ell^-; \ell \leq -1\}$  by

$$A_*(z) = \sum_{\ell=-\infty}^{\infty} z^\ell A_\ell, \quad R_*^+(z) = \sum_{\ell=1}^{\infty} z^\ell R_\ell^+, \quad G_*^-(z) = \sum_{\ell=-\infty}^{-1} z^\ell G_\ell^-.$$

This notation for matrix generating functions will be used for other matrices without statements below. Then, we have the following Wiener-Hopf factorization (e.g., see Theorem 14 of [29]). This is slightly different from the standard form, which will be mentioned after Lemma 3.2 below.

**Lemma 3.1** For the Markov additive process generated by  $\{A_\ell; \ell \in \mathbb{Z}\}$ ,

$$I - A_*(z) = (I - R_*^+(z))(I - H_0^-)(I - G_*^-(z)). \quad (3.1)$$

This lemma has been used to consider the tail decay rate of the stationary distribution in [13]. In this paper, we also consider the decay rate using it. However, we shall not use the lemma directly, but apply it to a modified Markov additive process in which the corresponding  $A$  is stochastic.

We introduce a dual Markov additive process so as to consider the measure  $\{R_\ell^+(i, j)\}$ . To this end, we tentatively assume that  $A$  has the stationary vector  $\boldsymbol{\pi}$ . Define

$$\tilde{A}_n = \Delta_{\boldsymbol{\pi}}^{-1} A_n^T \Delta_{\boldsymbol{\pi}}, \quad n \in \mathbb{Z},$$

where  $\Delta_{\boldsymbol{\pi}}$  is the diagonal matrix whose  $i$ -th entry is the corresponding entry of vector  $\boldsymbol{\pi}$ . Let  $\{(\tilde{X}_n, \tilde{Y}_n)\}$  be the Markov additive process generated by  $\{\tilde{A}_\ell; \ell \in \mathbb{Z}\}$ . It is not hard to see that  $\{(\tilde{X}_n, \tilde{Y}_n)\}$  is stochastically equivalent to  $\{(X_{-n}, -Y_{-n})\}$  under the background stationary measure  $\boldsymbol{\pi}$  (not necessarily to be a probability vector). So, we call it a dual of the Markov additive process  $\{(X_n, Y_n)\}$ . Let

$$\tilde{\tau}_\ell^+ = \inf\{n \geq 1; \tilde{Y}_n - \tilde{Y}_0 \geq \ell\}, \quad \ell \geq 0.$$

That is,  $\tilde{\tau}_\ell^+$  is the hitting time at  $[\ell, \infty)$ . Define

$$\tilde{G}_\ell^+(i, j) = P(\tilde{X}_{\tilde{\tau}_\ell^+} = j, \tilde{Y}_{\tilde{\tau}_\ell^+} - \tilde{Y}_0 = \ell | \tilde{X}_0 = i), \quad \ell \geq 1.$$

Thus,  $\{\tilde{G}_\ell^+(i, j); j \in S_1, \ell \geq 1\}$  is the ladder height distribution for each  $i \in S_1$ . Similarly, the occupation measure  $\tilde{R}_\ell^-$  and the hitting probability from below  $\tilde{H}_0^+$  are defined. The following results show how the measure  $\{R_\ell^+(i, j)\}$  is related to this distribution.

**Lemma 3.2** If  $A$  has the stationary measure  $\boldsymbol{\pi}$ , then we have

$$R_\ell^+ = \Delta_{\boldsymbol{\pi}}^{-1} (\tilde{G}_\ell^+)^T \Delta_{\boldsymbol{\pi}}, \quad \ell \geq 1, \quad (3.2)$$

$$G_\ell^- = \Delta_{\boldsymbol{\pi}}^{-1} (\tilde{R}_\ell^-)^T \Delta_{\boldsymbol{\pi}}, \quad \ell \leq -1, \quad (3.3)$$

$$H_0^- = \Delta_{\boldsymbol{\pi}}^{-1} (\tilde{H}_0^+)^T \Delta_{\boldsymbol{\pi}}. \quad (3.4)$$

**PROOF.** For finite  $S_1$ , the first and second equations are obtained in Theorem 1 of [27]. They as well as the third equation can be similarly obtained for countable  $S_1$ . However, we here prove them for this paper to be self-contained. Assume that  $\{X_n\}$  is stationary. Then, from the definitions of the dual processes, we have, for  $n, \ell \geq 1$ ,

$$\begin{aligned} & P(X_n = j, Y_k - Y_0 \geq \ell, k = 1, \dots, n-1, Y_n - Y_0 = \ell | X_0 = i) P(X_0 = i) \\ &= P(X_0 = j, Y_{k-n} - Y_{-n} \geq \ell, k = 1, \dots, n-1, Y_0 - Y_{-n} = \ell, X_{-n} = i) \\ &= P(\tilde{X}_0 = j, \tilde{Y}_{n-k} - \tilde{Y}_n \leq -\ell, k = 1, \dots, n-1, \tilde{Y}_n - \tilde{Y}_0 = \ell, \tilde{X}_n = i) \\ &= P(\tilde{X}_0 = j, \tilde{Y}_{n-k} - \tilde{Y}_0 \leq 0, k = 1, \dots, n-1, \tilde{Y}_n - \tilde{Y}_0 = \ell, \tilde{X}_n = i) \\ &= P(\tilde{X}_0 = j, \tilde{Y}_k - \tilde{Y}_0 \leq 0, k = 1, \dots, n-1, \tilde{Y}_n - \tilde{Y}_0 = \ell, \tilde{X}_n = i) \\ &= P(\tilde{X}_0 = j) P(\tilde{Y}_k - \tilde{Y}_0 \leq 0, k = 1, \dots, n-1, \tilde{Y}_n - \tilde{Y}_0 = \ell, \tilde{X}_n = i | \tilde{X}_0 = j). \end{aligned}$$

Since

$$R_\ell^+(i, j) = \sum_{n=1}^{\infty} P(X_n = j, Y_k \geq \ell, k = 1, \dots, \ell - 1, Y_n - Y_0 = \ell | X_0 = i),$$

we get (3.2). Similarly, (3.3) and (3.4) can be proved.  $\square$

From (3.2) of this lemma, (3.1) can be written as

$$I - A_*(z) = (I - \Delta_\pi^{-1}(\tilde{G}_*^+(z))^T \Delta_\pi)(I - H_0^-)(I - G_*^-(z)). \quad (3.5)$$

This is the standard form of the Wiener-Hopf factorization appeared in the literature (see, e.g., Theorem 5.3 of [4] and Theorem 2.12 in Chapter XI of [5]). For our arguments, (3.1) is more convenient than this standard form because (3.1) has no dual term. We also note a dual version of (3.1), which is immediate from Lemmas 3.1 and 3.2.

$$I - \tilde{A}_*(z) = (I - \tilde{R}_*^+(z))(I - \tilde{H}_0^+)(I - \tilde{G}_*^+(z)). \quad (3.6)$$

We now assume

(iii)  $P$  has the unique stationary distribution.

Denote this distribution by row vector  $\boldsymbol{\nu}$ , and partition its components with respect to levels in such a way that  $\boldsymbol{\nu} = (\boldsymbol{\nu}_0, \boldsymbol{\nu}_1, \dots)$ . An important step in our arguments is to express  $\boldsymbol{\nu}$  using  $\{R_\ell^+\}$ . We follow the idea of using censoring processes due to Grassmann and Heyman [12] (see also [28]). Let  $Q_{\ell, m} = A_{m-\ell}$  for  $\ell, m \geq 1$ , and let  $Q$  be a block structured matrix whose  $(\ell, m)$ -th block is  $Q_{\ell, m}$  for  $\ell, m \geq 1$ . Since  $Q$  is strictly substochastic,  $\hat{Q} \equiv (I - Q)^{-1}$  exists. Denote the  $(\ell, m)$ -th block of  $\hat{Q}$  by  $\hat{Q}_{\ell, m}$ , and define the  $S_0 \times S_0$  matrix  $C_0$  by

$$C_0 = B_0 + \sum_{\ell=1}^{\infty} \sum_{m=1}^{\infty} B_\ell \hat{Q}_{\ell, m} B_{-m}.$$

Clearly,  $C_0$  is a transition matrix for the censored process at level 0 of the Markov chain with transition matrix  $P$ . Hence, if the stationary distribution  $\boldsymbol{\nu}$  exists, we must have

$$\boldsymbol{\nu}_0 = \boldsymbol{\nu}_0 C_0. \quad (3.7)$$

For  $n \geq 1$ , define the  $S_0 \times S_1$  matrix  $R_{0, n}^+$  by

$$R_{0, n}^+ = \sum_{k=1}^{\infty} B_{n+k-1} \hat{Q}_{k, 1}.$$

For  $\ell \leq -1$ , define  $S_1 \times S_1$  matrix  $H_\ell^-$  by

$$H_\ell^-(i, j) = P(X_{\tau_\ell^-} = j, Y_{\tau_\ell^-} - Y_0 = \ell | X_0 = i) \quad i, j \in S_1.$$

Since  $\hat{Q}_{k,1}(i, j)$  is the expected number of visits of the additive process  $Y_n$  at level 1 with background state  $j$  before it goes below level 1, provided that it started at level  $k$  with background state  $i$ , we can see  $\hat{Q}_{k,1} = H_{-(k-1)}^-(I - H_0^-)^{-1}$  for  $k \geq 2$ . Hence,

$$R_{0,n}^+ = (B_n + \sum_{k=1}^{\infty} B_{n+k} H_{-k}^-)(I - H_0^-)^{-1}. \quad (3.8)$$

Then, censoring the original Markov chain at level set  $\{0, 1, \dots, n\}$  yields the following equations, which are obtained in Theorems 2.1 and 2.6 of [28] (also, see [12]).

$$\boldsymbol{\nu}_n = \boldsymbol{\nu}_0 R_{0,n}^+ + \sum_{\ell=1}^{n-1} \boldsymbol{\nu}_\ell R_{n-\ell}^+, \quad n \geq 1. \quad (3.9)$$

#### 4. Twisted process and decay rate

We now consider the decay rate of  $\boldsymbol{\nu}_n$  as  $n \rightarrow \infty$ . To this end, we shall require the right and left invariant vectors of the matrix generating function  $A_*(z)$  for an appropriately chosen  $z$ . Note that, for  $z > 0$ ,

$$A_*(z) = \left( E(z^{Y_1 - Y_0} \mathbf{1}(X_1 = j) | X_0 = i) \right)$$

is viewed as an operator for functions from  $S_1$  to  $\mathbb{R}$ . For each fixed  $z > 0$ , let  $\gamma(z)$  be the spectral radius of  $A_*(z)$ . Obviously  $\gamma(1) \leq 1$  since  $A$  is substochastic. Then, finding the invariant vectors requires to find  $z > 1$  such that  $\gamma(z) = 1$ . Denote this  $z$  by  $\alpha$ . If  $S_1$  is finite, it is known that  $\gamma(z)$  is a convex function of  $z$ . So, if  $A_*(z)$  exists for  $z \in (1 - \epsilon_0, 1 + \epsilon_1)$  for  $\epsilon_0, \epsilon_1 > 0$ , and  $\liminf_{z \uparrow \epsilon_1} \gamma(z + 1) > 1$ , then, by the Perron Frobenius theorem (e.g. see [23]), there exist such an  $\alpha > 1$  and unique positive vectors  $\boldsymbol{\mu}(\alpha)$  and  $\boldsymbol{h}(\alpha)$  such that

$$\boldsymbol{\mu}(\alpha) A_*(\alpha) = \boldsymbol{\mu}(\alpha), \quad (4.1)$$

$$A_*(\alpha) \boldsymbol{h}(\alpha) = \boldsymbol{h}(\alpha). \quad (4.2)$$

If  $S_1$  is infinite, this finding problem becomes to be complicated. First, the  $\alpha$  may not be unique as is known (see Section 5 for this example). So, we further request that

$$\boldsymbol{\mu}(\alpha) \boldsymbol{h}(\alpha) < \infty. \quad (4.3)$$

Secondly, there seem no good sufficient conditions for our applications (see Remark 4.1 below). So far, we assume that (4.1), (4.2) and (4.3) hold for some  $\alpha > 1$  and for positive vectors  $\boldsymbol{\mu}(\alpha)$  and  $\boldsymbol{h}(\alpha)$  throughout this section. These conditions are equivalent to that  $A_*(\alpha)$  is 1-positive recurrent (see Theorem 6.4 of [23]). Note that we have to check this positivity in our applications.

**Remark 4.1** When  $A$  is recurrent, finding the  $\alpha$  and corresponding vectors has been studied in a much more general context by Ney and Nummelin [22]. Namely,  $\gamma(z)$  is strictly convex (Corollary 3.3 of [22]), and (4.1), (4.2) and (4.3) are obtained for  $\alpha > 1$



under certain regularity conditions (Lemma 4.1 of [22]). They also provide some specific examples to satisfy the regularity conditions. However, these conditions may not be easy to check in queueing applications. For instance, Example 6.8 of [22] notices that it is hard to check them for the  $A$  of a birth and death process, which is typical in queueing networks.

For this  $\alpha$ , we define

$$\begin{aligned}\boldsymbol{\xi}(\alpha) &= (\mathbf{h}(\alpha))^T (I - (G_*^-(\alpha))^T) (I - (H_0^-)^T) \Delta_{\boldsymbol{\mu}(\alpha)} \\ \beta(\alpha) &= \alpha \boldsymbol{\mu}(\alpha) A_*'(\alpha) \mathbf{h}(\alpha),\end{aligned}$$

as long as these are well defined, where the derivative of a matrix function is taken component-wise. The following theorem is a main result of this paper.

**Theorem 4.1** Under assumptions (i), (ii) and (iii), if (4.1), (4.2) and (4.3) hold true for some  $\alpha > 1$ , then  $\boldsymbol{\xi}(\alpha)$  is nonnegative and nonzero vector satisfying  $\boldsymbol{\xi}(\alpha)\mathbf{e} < \infty$ , and  $\beta(\alpha)$  is positive but may be infinite. In addition to these conditions, if

$$\boldsymbol{\nu}_0 B_*^+(\alpha) \mathbf{h}(\alpha) < \infty, \quad (4.4)$$

where  $B_*^+(z) = \sum_{n=1}^{\infty} z^n B_n$ , then we have

$$\lim_{n \rightarrow \infty} \alpha^n \boldsymbol{\nu}_n = \frac{\boldsymbol{\omega}_0^{(\alpha)} \Delta_{\boldsymbol{\mu}(\alpha)}^{-1} \boldsymbol{\xi}(\alpha)^T}{\beta(\alpha)} \boldsymbol{\mu}(\alpha) < \infty, \quad (4.5)$$

where the limit is taken component-wise, the right-hand side is zero if  $\beta(\alpha) = \infty$ , and

$$\boldsymbol{\omega}_0^{(\alpha)} = \sum_{n=1}^{\infty} \alpha^n \boldsymbol{\nu}_0 R_{0,n}^+.$$

**Remark 4.2** If the additive component  $Y_n$  is bounded from above, then the  $\beta(\alpha)$  is always finite. In most applications,  $A$  is positive recurrent. But, Theorem 4.1 does not require any recurrence condition on  $A$ .

**Remark 4.3** This theorem is an extension of Theorem 5.1 of [18], which relaxes conditions as well as modeling assumptions of [25] (see Theorems 1 and 4 of their paper). Matrix analysis is employed in [25], while the Markov renewal approach is used in [18]. The latter paper also discussed various kind of Markov renewal equations for the  $M/G/1$  type queue with a countable background state space. Some of them could be extended for the  $GI/G/1$  type queue. However,  $M/G/1$  type queues provide more information on the prefactor of the decay rate than the  $GI/G/1$  type queues.

In the rest of this section, we prove Theorem 4.1 using a few of lemmas. The first lemma is concerned with a twisted process by the  $\alpha$ . Define

$$\begin{aligned}A_n^{(\alpha)} &= \alpha^n \Delta_{\mathbf{h}(\alpha)}^{-1} A_n \Delta_{\mathbf{h}(\alpha)}, \\ \boldsymbol{\eta}(\alpha) &= \boldsymbol{\mu}(\alpha) \Delta_{\mathbf{h}(\alpha)}.\end{aligned}$$

From (4.1) and (4.2),  $A^{(\alpha)} \equiv \sum_{n=-\infty}^{+\infty} A_n^{(\alpha)}$  is stochastic and from (4.3), it is positive recurrent with stationary row vector  $\boldsymbol{\eta}(\alpha)$ . Let  $\{(X_n^{(\alpha)}, Y_n^{(\alpha)}); n \geq 0\}$  be the Markov additive process generated by  $\{A_\ell^{(\alpha)}; \ell \in \mathbb{Z}\}$ . This additive process is said to be twisted. Denote the matrices corresponding to  $R_\ell^+$ ,  $G_\ell^-$  and  $H_0^-$  of the original Markov additive process  $\{(X_n, Y_n); n \geq 0\}$  by  $R_\ell^{(\alpha)+}$ ,  $G_\ell^{(\alpha)-}$  and  $H_0^{(\alpha)-}$ , respectively. The matrix generating functions of the first two characteristics are similarly denoted by  $R_*^{(\alpha)+}(z)$  and  $G_*^{(\alpha)-}(z)$ , respectively. Then, we can obtain the following result from the definition of  $A_n^{(\alpha)}$ .

**Lemma 4.1**

$$\begin{aligned} R_\ell^{(\alpha)+} &= \alpha^\ell \Delta_{\mathbf{h}(\alpha)}^{-1} R_\ell^+ \Delta_{\mathbf{h}(\alpha)}, & \ell \geq 1 \\ G_\ell^{(\alpha)-} &= \alpha^\ell \Delta_{\mathbf{h}(\alpha)}^{-1} G_\ell^- \Delta_{\mathbf{h}(\alpha)}, & \ell \leq -1 \\ H_0^{(\alpha)-} &= \Delta_{\mathbf{h}(\alpha)}^{-1} H_0^- \Delta_{\mathbf{h}(\alpha)}. \end{aligned}$$

This observation is a key step in our arguments.

We next take a dual of the twisted process  $\{(X_n^{(\alpha)}, Y_n^{(\alpha)}); n \geq 0\}$  with respect to the stationary vector  $\boldsymbol{\eta}(\alpha)$ . Denote this Markov additive process by  $\{(\tilde{X}_n^{(\alpha)}, \tilde{Y}_n^{(\alpha)}); n \geq 0\}$ . Then, from the definition, this additive process is generated by

$$\tilde{A}_\ell^{(\alpha)} \equiv \alpha^\ell \Delta_{\boldsymbol{\mu}(\alpha)}^{-1} A_\ell^T \Delta_{\boldsymbol{\mu}(\alpha)}. \quad (4.6)$$

Denote the matrices corresponding to  $\tilde{G}_\ell^+$ ,  $\tilde{R}^-$  and  $\tilde{H}_0^+$  of the dual process defined in Section 3 by  $\tilde{G}_\ell^{(\alpha)+}$ ,  $\tilde{R}^{(\alpha)-}$  and  $\tilde{H}_0^{(\alpha)+}$ , respectively. Then, from Lemmas 3.2 and 4.1, we have

$$\tilde{G}_\ell^{(\alpha)+} = \alpha^\ell \Delta_{\boldsymbol{\mu}(\alpha)}^{-1} (R_\ell^+)^T \Delta_{\boldsymbol{\mu}(\alpha)}, \quad \ell \geq 1, \quad (4.7)$$

$$\tilde{R}_\ell^{(\alpha)-} = \alpha^\ell \Delta_{\boldsymbol{\mu}(\alpha)}^{-1} (G_\ell^-)^T \Delta_{\boldsymbol{\mu}(\alpha)}, \quad \ell \leq -1, \quad (4.8)$$

$$\tilde{H}_0^{(\alpha)+} = \Delta_{\boldsymbol{\mu}(\alpha)}^{-1} (H_0^-)^T \Delta_{\boldsymbol{\mu}(\alpha)}, \quad (4.9)$$

and (3.6) becomes

$$I - \tilde{A}_*^{(\alpha)}(z) = (I - \tilde{R}_*^{(\alpha)-}(z))(I - \tilde{H}_0^{(\alpha)+})(I - \tilde{G}_*^{(\alpha)+}(z)). \quad (4.10)$$

Note that this is also a direct consequence of (3.1) by (4.7), (4.8) and (4.9).

Let  $\{(\hat{X}_n^{(\alpha)}, \hat{Z}_n^{(\alpha)})\}$  be the Markov additive process generated by  $\{\tilde{G}_\ell^{(\alpha)+}; \ell \geq 1\}$ . Since  $\hat{Z}_n^{(\alpha)} - \hat{Z}_{n-1}^{(\alpha)} \geq 1$ , this process can also be considered as a Markov renewal process (cf. [9]).

**Lemma 4.2** (a) The background process  $\{\hat{X}_n^{(\alpha)}\}$  is Harris ergodic, i.e., it has one positive recurrent class and all other states can reach this class with positive probabilities, and  $\boldsymbol{\xi}(\alpha)$  is its unique stationary measure up to constant multiplication.

(b) The expectation of  $\hat{Z}_1^{(\alpha)} - \hat{Z}_0^{(\alpha)}$ , i.e, the transition interval of the Markov renewal process, with respect to  $\boldsymbol{\xi}(\alpha)$  is  $\beta(\alpha) > 0$ , which may be infinite.

(c) The Markov additive process  $\{(\hat{X}_n^{(\alpha)}, \hat{Z}_n^{(\alpha)})\}$  is 1-arithmetic.

PROOF. Since  $G_*^-(\alpha)$  with  $\alpha > 1$  and  $H_0^-$  are strictly substochastic,  $I - G_*^-(\alpha)$  and  $I - H_0^-$  are invertible. So,  $I - \tilde{R}_*^{(\alpha)-}(1)$  and  $I - \tilde{H}_0^{(\alpha)+}$  are also invertible. Hence, postmultiplying  $\mathbf{e}$  to (4.10), (4.1) implies that transition matrix  $\tilde{G}_*^{(\alpha)+}(1)$  of the background process is stochastic. Since  $\boldsymbol{\xi}(\alpha)$  can be written as

$$\boldsymbol{\xi}(\alpha) = \boldsymbol{\eta}(\alpha)(I - \tilde{R}_*^{(\alpha)-}(1))(I - \tilde{H}_0^{(\alpha)+}), \quad (4.11)$$

and  $\boldsymbol{\eta}(\alpha)$  is the left invariant vector of  $\tilde{A}_*^{(\alpha)}(1)$ , we have

$$\boldsymbol{\xi}(\alpha) = \boldsymbol{\xi}(\alpha)\tilde{G}_*^{(\alpha)+}(1), \quad (4.12)$$

where  $\boldsymbol{\xi}(\alpha)$  can not be the zero vector since  $\boldsymbol{\eta}(\alpha)$  is positive. For vector  $\mathbf{a} \equiv \{a_i; i \in S_1\}$ , denote the vector whose  $i$ -th component is  $|a_i|$  by  $|\mathbf{a}|$ . From (4.11),

$$|\boldsymbol{\xi}(\alpha)|\mathbf{e} \leq 2\boldsymbol{\eta}(\alpha)(I + \tilde{R}_*^{(\alpha)-}(1))\mathbf{e}.$$

From (4.8) and Lemma 4.1, we have

$$\begin{aligned} \boldsymbol{\eta}(\alpha)\tilde{R}_*^{(\alpha)-}(1)\mathbf{e} &= \boldsymbol{\mu}(\alpha)(G_*^-(\alpha))\mathbf{h}(\alpha) \\ &= \boldsymbol{\mu}(\alpha)\Delta_{\mathbf{h}(\alpha)}G_*^{(\alpha)-}(1)\mathbf{e} \\ &\leq \boldsymbol{\mu}(\alpha)\mathbf{h}(\alpha), \end{aligned}$$

where the last inequality is obtained since  $G_*^{(\alpha)-}(1)$  is substochastic. Thus,  $|\boldsymbol{\xi}(\alpha)|\mathbf{e}$  is finite by (4.3). Using (4.12) repeatedly, we have, for all  $n \geq 1$ ,

$$\boldsymbol{\xi}(\alpha) = \boldsymbol{\xi}(\alpha)(\tilde{G}_*^{(\alpha)+}(1))^n.$$

Note that  $\tilde{G}_*^{(\alpha)+}(1)$  is stochastic, and has at most one recurrent class since  $\tilde{A}_*^{(\alpha)}(1)$  is irreducible. Hence, there exists a row vector  $\mathbf{p}$  such that  $(\tilde{G}_*^{(\alpha)+}(1))^n$  converges to matrix  $\mathbf{e}\mathbf{p}$  componentwise as  $n$  goes to infinity (if necessary, choosing an appropriate lattice for  $n$ ), and the  $\mathbf{p}$  is either zero or a probability vector. Since  $|\boldsymbol{\xi}(\alpha)|\mathbf{e}$  is finite, the bounded convergence theorem yields

$$\boldsymbol{\xi}(\alpha) = (\boldsymbol{\xi}(\alpha)\mathbf{e})\mathbf{p}.$$

This implies that neither  $\boldsymbol{\xi}(\alpha)\mathbf{e}$  nor  $\mathbf{p}$  can be zero, and all the entries of  $\boldsymbol{\xi}(\alpha)$  have the same sign. However,  $\boldsymbol{\xi}(\alpha)$  can not be negative because we have, from (4.11),

$$\boldsymbol{\xi}(\alpha)(I - \tilde{H}_0^{(\alpha)+})^{-1}(I - \tilde{R}_*^{(\alpha)-}(1))^{-1} = \boldsymbol{\eta}(\alpha) > \mathbf{0}.$$

Consequently,  $\boldsymbol{\xi}(\alpha)$  is the stationary and finite measure of  $\tilde{G}_*^{(\alpha)+}(1)$ , so the irreducibility of  $\tilde{A}_*^{(\alpha)}(1)$  implies that  $\{\hat{X}_n^{(\alpha)}\}$  is Harris ergodic. These prove (a).

Since  $\tilde{A}_*^{(\alpha)}(1/\alpha)$  and  $\tilde{A}_*^{(\alpha)}(1)$  are finite nonnegative matrices,  $\tilde{A}_*^{(\alpha)}(z)$  is also finite and nonnegative for  $1/\alpha < z < 1$ . Hence, differentiating (4.8) componentwise for  $z$  satisfying  $1/\alpha < z < 1$  and letting  $z \uparrow 1$ , we have

$$\begin{aligned} \beta(\alpha) &= \boldsymbol{\eta}(\alpha)(\tilde{A}_*^{(\alpha)})'(1)\mathbf{e} \\ &= \boldsymbol{\xi}(\alpha)(\tilde{G}_*^{(\alpha)+})'(1)\mathbf{e} > 0. \end{aligned} \quad (4.13)$$

Hence, we get (b). From the first equality of (4.13),  $\{\tilde{Y}_n^{(\alpha)}\}$  has the mean drift  $\beta(\alpha) > 0$ , so condition (ii) implies (c) by (ii) of Theorem 1 of [3].  $\square$

**Remark 4.4** (c) can be directly proved by (4.10). Let  $\mathbf{f}$  be a column vector such that  $|\mathbf{f}| = \mathbf{e}$ . Since  $I - \tilde{R}_*^{(\alpha)-}(e^{ip})$  and  $I - \tilde{H}_0^{(\alpha)+}$  are invertible for each positive number  $p$ , we have that  $\tilde{A}_*^{(\alpha)}(e^{ip})\mathbf{f} = \mathbf{f}$  if and only if  $\tilde{G}_*^{(\alpha)+}(e^{ip})\mathbf{f} = \mathbf{f}$ . Hence,  $\{\tilde{A}_n^{(\alpha)}; n \in \mathbb{Z}\}$  and  $\{\tilde{G}_n^{(\alpha)+}; n \geq 1\}$  have the same periodicity as Markov additive processes (see [2, 24]).

Let  $\tilde{\mathbf{x}}_\ell^{(\alpha)} = \alpha^\ell \Delta_{\boldsymbol{\mu}(\alpha)}^{-1} \boldsymbol{\nu}_\ell^\top$  and  $\tilde{G}_{0,n}^{(\alpha)+} = \alpha^n \Delta_{\boldsymbol{\mu}(\alpha)}^{-1} (R_{0n}^+)^\top \Delta_{\boldsymbol{\mu}(\alpha)}$ . Then, using (4.7), equation (3.9) is converted to

$$\tilde{\mathbf{x}}_n^{(\alpha)} = \tilde{G}_{0,n}^{(\alpha)+} \tilde{\mathbf{x}}_0^{(\alpha)} + \sum_{\ell=1}^{n-1} \tilde{G}_{n-\ell}^{(\alpha)+} \tilde{\mathbf{x}}_\ell^{(\alpha)}, \quad n \geq 1. \quad (4.14)$$

This is a proper Markov renewal equation since  $\{\tilde{G}_n^{(\alpha)+}; n \geq 1\}$  is a stochastic, positive recurrent and 1-arithmetic kernel by Lemma 4.2. Hence, if

$$c(\alpha) \equiv \boldsymbol{\xi}(\alpha) \sum_{n=1}^{\infty} \tilde{G}_{0,n}^{(\alpha)+} \tilde{\mathbf{x}}_0^{(\alpha)}$$

is finite, then we can apply the Markov renewal theorem to the sequence  $\{\tilde{\mathbf{x}}_n^{(\alpha)}\}$  (e.g., [2, 9, 24]). The following lemma is useful for this.

**Lemma 4.3** We have

$$c(\alpha) \leq \boldsymbol{\nu}_0 B_*^+(\alpha) \mathbf{h}(\alpha). \quad (4.15)$$

PROOF. Let  $R_{0*}^+(\alpha) = \sum_{n=1}^{\infty} \alpha^n R_{0n}^+$ . By substituting the definitions, we have

$$\begin{aligned} c(\alpha) &= \boldsymbol{\xi}(\alpha) \Delta_{\boldsymbol{\mu}(\alpha)}^{-1} (R_{0*}^+(\alpha))^\top \Delta_{\boldsymbol{\mu}(\alpha)} \tilde{\mathbf{x}}_0^{(\alpha)} \\ &= (\mathbf{h}(\alpha))^\top (I - (G_*^-(\alpha))^\top) (I - (H_0^-)^\top) (R_{0*}^+(\alpha))^\top \boldsymbol{\nu}_0^\top \\ &= \boldsymbol{\nu}_0 R_{0*}^+(\alpha) (I - H_0^-) (I - G_*^-(\alpha)) \mathbf{h}(\alpha). \end{aligned} \quad (4.16)$$

We next evaluate  $R_{0*}^+(\alpha)$ . From (3.8),

$$\begin{aligned} R_{0*}^+(\alpha) &= \sum_{n=1}^{\infty} \alpha^n B_n (I + \sum_{k=1}^{n-1} \alpha^{-k} H_{-k}^-) (I - H_0^-)^{-1} \\ &\leq B_*^+(\alpha) (I + H_*^-(\alpha)) (I - H_0^-)^{-1}, \end{aligned}$$

where  $H_*^-(\alpha) = \sum_{n=1}^{\infty} \alpha^{-n} H_{-n}^-$ . Substituting this into (4.16) and using the fact that  $H_*^-(\alpha) = G_*^-(\alpha) (I - G_*^-(\alpha))^{-1}$ , we arrive at (4.15) because  $(I - H_0^-) (I - G_*^-(\alpha)) \mathbf{h}(\alpha)$  is nonnegative by Lemma 4.2.  $\square$

PROOF OF THEOREM 4.1 The first part of the theorem is proved by Lemma 4.2. Condition (4.4) implies that  $c(\alpha)$  is finite by Lemma 4.3, and the mean transition interval of Markov renewal process generated by  $\{G_n^{(\alpha)}\}$  is  $\beta(\alpha)$  by Lemma 4.2. Hence, applying the Markov renewal theorem to (4.14) (see Theorem 3 of Shurenkov [24]), we have

$$\lim_{n \rightarrow \infty} \tilde{\mathbf{x}}_n^{(\alpha)} = \frac{c(\alpha)}{\beta(\alpha)} \mathbf{e} \quad (4.17)$$

Premultiplying  $\Delta_{\boldsymbol{\mu}(\alpha)}$  to (4.17) and taking transposes, we get (4.5) since  $c(\alpha) = \boldsymbol{\omega}_0^{(\alpha)} \Delta_{\boldsymbol{\mu}(\alpha)}^{-1} \boldsymbol{\xi}(\alpha)^\top$ .  $\square$

## 5. Pilot examples

Theorem 4.1 is theoretically beautiful, but it may not be an easy task to find  $\alpha$  satisfying conditions (4.1), (4.2), (4.3) and (4.4). So, it is important to show its applicability. We can find such an example in [25], which considered a two-queue with shorter queue discipline. We consider other examples here. Our intention is to see how the conditions of Theorem 4.1 work and where we must be careful. Although the model is simple, we still include some interesting observations.

Suppose a discrete-time priority queue with a single server and two types of customers. At each time epoch, either one of the following three events occurs independently of everything else. (a) Type 1 customer singly arrives with probability  $p$ ; (b) a batch of type 2 customers of size  $n$  arrives with probability  $q_n$  for  $n = 1, 2, \dots$ ; or (c) service is completed with probability  $r$  (if no customer in the system, nothing occurs with this probability). We assume that

$$q = \sum_{n=1}^{\infty} q_n, \quad p + q + r = 1.$$

In the first two cases, we assume that type 1 customers have a priority over type 2 customers. That is, one type 1 customer is served with probability  $r$  if there are type 1 customers in the system. Otherwise, one type 2 customer can complete service with probability  $r$ . In the third case, this priority service is slightly modified in such a way that type 2 customers can be served with a given probability when the queue of type 2 customers is longer than the queue of type 1 customers. Obviously, these systems are stable if and only if  $p + q'_*(1) < r$ , where  $q_*(z) = \sum_{n=1}^{\infty} z^n q_n$ .

A natural description of this model is a discrete-time two dimensional Markov chain  $\{L_n^{(1)}, L_n^{(2)}\}$  with state space  $\mathbb{N}^2$ , where  $L^{(1)}$  and  $L^{(2)}$  stand for the numbers of type 1 and type 2 customers in the system, respectively. There are various ways to reformulate this Markov chain under the framework of the  $GI/G/1$  type of queues with a countable state space. Here we consider three of them. In the third case, we assume  $q_1 = q$ , i.e. single arrivals for type 2 customers as well, since this case is more complicated.

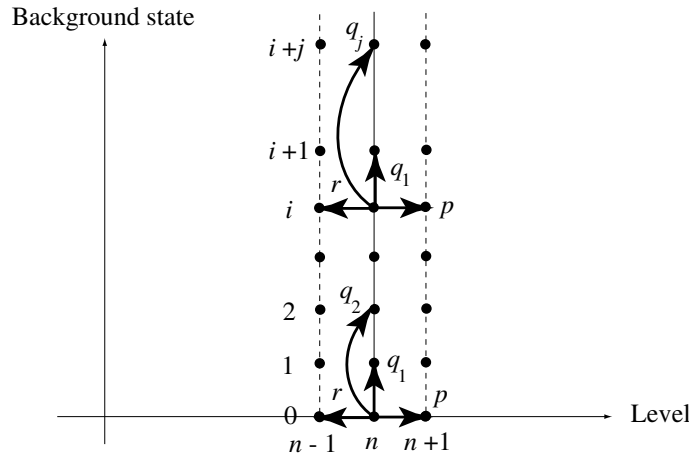


Figure 1: Background state transitions for Case 1

(Case 1) Take  $L^{(2)}$  as the background state (see Figure 1). So, we set

$$A_n(i, j) = \begin{cases} p1(j = i), & n = 1, \\ q_k 1(j = i + k) & n = 0, \\ r1(j = i), & n = -1, \\ 0, & \text{otherwise.} \end{cases}$$

Then, we have

$$A_*(\alpha) = \begin{pmatrix} p\alpha + r\alpha^{-1} & q_1 & q_2 & q_3 & q_4 & \cdots \\ 0 & p\alpha + r\alpha^{-1} & q_1 & q_2 & q_3 & \cdots \\ 0 & 0 & p\alpha + r\alpha^{-1} & q_1 & q_2 & \cdots \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \end{pmatrix}.$$

Note that the Markov chain  $A^{(\alpha)}$  does not satisfy (i). Obviously, there is no  $\alpha > 1$  satisfying conditions (4.1). Thus, we can not apply Theorem 4.1. On the other hand, we know that the marginal distribution of  $L^{(1)}$  geometrically decays with rate  $\alpha^{-1} = \frac{p}{r}$ . This is a rather trivial example, but it suggests that if the background state has no information on the decreasing of  $L^{(2)}$ , then Theorem 4.1 may not be applied.

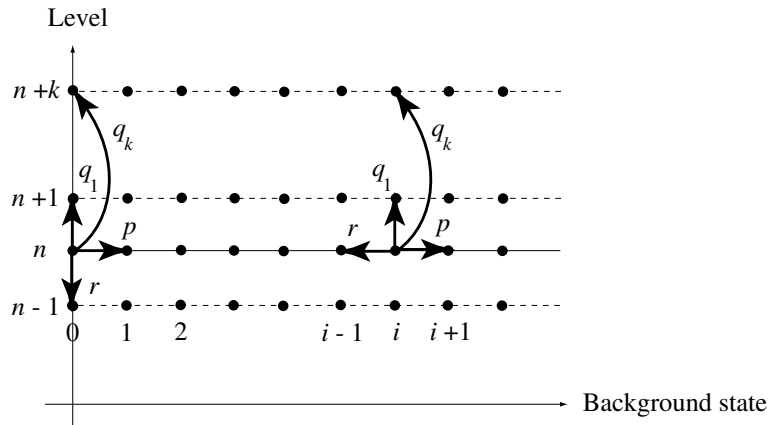


Figure 2: Background state transitions for Case 2

(Case 2) Take  $L^{(1)}$  as the background state (see Figure 2). In this case, we assume that there exists  $z_0 > 1$  (possibly,  $z_0 = \infty$ ) such that

$$q_*(z) < \infty \text{ for } 0 < z < z_0, \quad \sup\{q_*(z); z < z_0\} = \infty, \quad (5.1)$$

for the geometric decay of the stationary distribution of  $L_2$ . Clearly,

$$A_n(i, j) = \begin{cases} r1(j = i = 0), & n = -1, \\ p1(j = i + 1) + r1(j = i - 1 \geq 0) & n = 0, \\ q_n 1(j = i), & n \geq 1, \\ 0, & \text{otherwise.} \end{cases}$$

Then, we have

$$A_*(\alpha) = \begin{pmatrix} r\alpha^{-1} + q_*(\alpha) & p & 0 & 0 & 0 & \cdots \\ r & q_*(\alpha) & p & 0 & 0 & \cdots \\ 0 & r & q_*(\alpha) & p & 0 & \cdots \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \end{pmatrix}.$$

Let  $\boldsymbol{\mu}(\alpha) = (\mu_0, \mu_1, \dots)$  and  $\boldsymbol{h}(\alpha) = (h_0, h_1, \dots)^T$  be the right and left invariant vector of  $A_*(\alpha)$ . Then,

$$\mu_0 = r\mu_1 + (r\alpha^{-1} + q_*(\alpha))\mu_0, \quad (5.2)$$

$$\mu_n = r\mu_{n+1} + q_*(\alpha)\mu_n + p\mu_{n-1}, \quad n \geq 1. \quad (5.3)$$

Let  $\theta_1$  and  $\theta_2$  be the solutions of the following equation of  $\theta$ .

$$f(\theta) \equiv r\theta^2 - (1 - q_*(\alpha))\theta + p = 0. \quad (5.4)$$

Obviously, we have

$$\theta_1 + \theta_2 = \frac{1 - q_*(\alpha)}{r}, \quad \theta_1\theta_2 = \frac{p}{r} < 1. \quad (5.5)$$

Thus, if  $\theta_1 \neq \theta_2$ , then

$$\mu_n = \mu_0\theta_1^n + (\mu_1 - \theta_1\mu_0)\frac{\theta_2^n - \theta_1^n}{\theta_2 - \theta_1}, \quad n \geq 1. \quad (5.6)$$

Similarly, by changing the role of  $p$  and  $r$  and the role of  $\theta_1$  and  $\theta_2$  we get

$$h_n = h_0\eta_2^n + (h_1 - \eta_2h_0)\frac{\eta_1^n - \eta_2^n}{\eta_1 - \eta_2}, \quad n \geq 1, \quad (5.7)$$

where  $\eta_1$  and  $\eta_2$  are determined by

$$\eta_1 + \eta_2 = \frac{1 - q_*(\alpha)}{p}, \quad \eta_1\eta_2 = \frac{r}{p} > 1. \quad (5.8)$$

From (5.5) and (5.8), we can put  $\eta_1 = \frac{1}{\theta_1}$  and  $\eta_2 = \frac{1}{\theta_2}$ . If  $\theta_1$  and  $\theta_2$  are complex numbers, then their absolute values must be identical. Then, (4.3) is impossible because of (5.6) and (5.7). So, they must be real numbers. Furthermore,  $f(0) = p$  by (5.4) implies that both of them have the same sign. But, the negative sign is impossible again in the view of (5.3) because  $\mu_n > 0$ . Thus,  $\theta_1$  and  $\theta_2$  must be positive. Assume that  $0 < \theta_1 < \theta_2$ , which implies that  $\eta_2 < \eta_1$ . Then, we can see that (4.3) is satisfied if coefficients of  $\theta_2^n$  and  $\eta_1^n$  are zero in (5.6) and (5.7), respectively. This holds only if

$$\mu_1 - \theta_1\mu_0 = h_1 - \eta_2h_0 = 0.$$

Using (5.2) and the corresponding equation for  $h_0$ , this is equivalent to

$$1 - q_*(\alpha) = r(\theta_1 + \alpha^{-1}), \quad (5.9)$$

$$1 - q_*(\alpha) = \frac{p}{\theta_2} + r\alpha^{-1}. \quad (5.10)$$

Either one of these equations is obtained from the other due to (5.5). Hence, it is sufficient to find  $\alpha$  and  $\theta_1$  satisfying (5.9) and  $f(\theta_1) = 0$ . These equations and (5.5) imply that

$$\theta_1 = \frac{p\alpha}{r}, \quad \theta_2 = \frac{1}{\alpha}, \quad \eta_1 = \frac{r}{p\alpha}, \quad \eta_2 = \alpha. \quad (5.11)$$

Hence, we have

$$q_*(\alpha) = 1 - p\alpha - r\alpha^{-1}, \quad (5.12)$$

which determines  $\alpha$ . From (5.11) and (5.12), we can see that  $\alpha < \sqrt{\frac{r}{p}}$  is required for  $0 < \theta_1 < \theta_2$  to be satisfied. Since  $q'_*(1) < r - p$ , assumption (5.1) implies that (5.12) has a unique solution  $\alpha$  such that  $1 < \alpha < \sqrt{\frac{r}{p}}$  if and only if

$$q_*\left(\sqrt{\frac{r}{p}}\right) > 1 - 2\sqrt{pr}. \quad (5.13)$$

Obviously, (5.13) holds if  $p$  is sufficiently small. Let  $x = \sqrt{\frac{r}{p}}$ , then (5.13) is equivalent to

$$q_*(x) > 1 - 2px. \quad (5.14)$$

Since  $q_*(1) = q = 1 - p - r < 1 - 2p$ , it is easy to see from the convexity of  $q_*(x)$  that there exists a unique solution  $x_0 > 1$  of the equation:

$$q_*(x) = 1 - 2px,$$

and (5.14) holds for  $x > x_0$ . Thus, if the traffic intensity  $\rho_1 \equiv \frac{p}{r}$  of the higher priority customers satisfies

$$\rho_1 < x_0^{-2} \quad (\text{or equivalently, } \rho_2 \equiv \frac{q}{r} > \frac{1-r}{r} - x_0^{-2}), \quad (5.15)$$

then conditions (4.1), (4.2) and (4.3) are satisfied. To check condition (4.4), we note that  $\alpha < \frac{r}{p}$  and

$$B_*(\alpha) = \begin{pmatrix} q_*(\alpha) & p & 0 & 0 & 0 & \cdots \\ r & q_*(\alpha) & p & 0 & 0 & \cdots \\ 0 & r & q_*(\alpha) & p & 0 & \cdots \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \end{pmatrix} \leq A_*(\alpha).$$

These implies that

$$\begin{aligned} \boldsymbol{\nu}_0 B_*(\alpha) \mathbf{h}(\alpha) &\leq \alpha \boldsymbol{\nu}_0 \mathbf{h}(\alpha) \\ &= \alpha \sum_{n=0}^{\infty} P(L^{(1)} = n, L^{(2)} = 0) \eta_2^n \\ &\leq \alpha \sum_{n=0}^{\infty} P(L^{(1)} = n) \alpha^n \\ &= \alpha \sum_{n=0}^{\infty} \left(\frac{p}{r}\right)^n \alpha^n < \infty. \end{aligned}$$

The finiteness of the  $\beta(\alpha)$  is verified as

$$\begin{aligned} \boldsymbol{\mu}(\alpha) A'_*(\alpha) \mathbf{h}(\alpha) &= \boldsymbol{\mu}(\alpha) (A_*(\alpha) + (q'_*(\alpha) - q_*(\alpha))I) \mathbf{h}(\alpha) - r\alpha^{-2} \mu_0 h_0 \\ &= (1 + q'_*(\alpha) - q_*(\alpha)) \boldsymbol{\mu}(\alpha) \mathbf{h}(\alpha) - r\alpha^{-2} \mu_0 h_0 < \infty. \end{aligned}$$



Hence, if (5.15) is satisfied, then Theorem 4.1 concludes, for some constant  $c > 0$ , we have the geometric decay:

$$\lim_{n \rightarrow \infty} P(L^{(1)} = i, L^{(2)} = n) \alpha^n = c \left( \frac{p\alpha}{r} \right)^i, \quad i \geq 0. \quad (5.16)$$

From the above computation,  $c \leq \frac{r\alpha}{r-p\alpha}$ . For example, if  $q_1 = q$  and  $q_n = 0$  for  $n \geq 2$ , then we have

$$\alpha = \frac{r}{p+q}, \quad \theta_1 = \frac{p}{p+q}, \quad \theta_2 = \frac{p+q}{r}, \quad x_0^{-2} = \frac{1}{(1+p-r)^2}.$$

It may look strange that the geometric decay appears under the restriction (5.15). However, this is not surprising since it is known that, in the  $M/G/1$  priority queue with two types of customers, the tail of the marginal waiting time distribution of the lower priority customers has such a decay rate only if a similar traffic condition is satisfied (see Theorem 8.5 and Theorem 11.2 of [1]).

Another interesting fact of this example is to show that there exist many other  $\alpha$ ,  $\boldsymbol{\mu}(\alpha)$  and  $\boldsymbol{h}(\alpha)$ , satisfying (4.1) and (4.2) if the finiteness condition (4.3) is not required. To see this, we consider (5.6) and (5.7). From these equation,  $\mu_n$  and  $h_n$  are positive if

$$\mu_1 - \theta_1 \mu_0 > 0, \quad h_1 - \eta_2 h_0 > 0.$$

Similar to the derivations of (5.9) and (5.10), these are equivalent to

$$1 - q_*(\alpha) > r(\theta_1 + \alpha^{-1}), \quad 1 - q_*(\alpha) > \frac{p}{\theta_2} + r\alpha^{-1}.$$

Because of (5.5), these are equivalent to  $\theta_2 > \alpha^{-1}$ . Thus, we need to find  $\alpha > 1$  such that

$$\theta_2 = \frac{1}{2r} \left( 1 - q_*(\alpha) + \sqrt{(1 - q_*(\alpha))^2 - 4pr} \right) > \alpha^{-1}.$$

After some manipulations, this is satisfied if

$$q_*(\alpha) < \max \left( 1 - \left( p\alpha + \frac{r}{\alpha} \right), 1 - 2\sqrt{pr} \right).$$

Since  $p\alpha + r\alpha^{-1} \geq 2\sqrt{pr}$ , this is possible for all  $\alpha \in (1, \alpha_0)$ , where  $\alpha_0 > 1$  is the solution of (5.12). Thus, we found the  $\alpha$ 's which have positive invariant vectors. This fact is contrasted with that they are uniquely determined when the background space is finite. Of course, the latter case automatically satisfies (4.3), so there is no contradiction, but we must be careful to handle matrices with infinite dimensions.

(Case 3) In the previous cases, the tail behavior is considered when either one of the queue lengths goes to infinity. We here consider it when both queues go to infinity by choosing the minimum of them as an additive component. In this case, we can not apply Theorem 4.1 for the priority queue. To see this, we modify the priority queue in the following way. Type 2 customers singly arrive, they may be served with probability  $r\epsilon$  for  $0 \leq \epsilon \leq 1$  when  $0 < L_1 < L_2$ , but they are not served when  $L_2 \leq L_1$ . If  $\epsilon = 0$ , then this model is reduced to the ordinary priority queue, while, if  $\epsilon = 1$ , then the longer queue gets service. The state transitions of this queue is described in Figure 3. We consider the decay

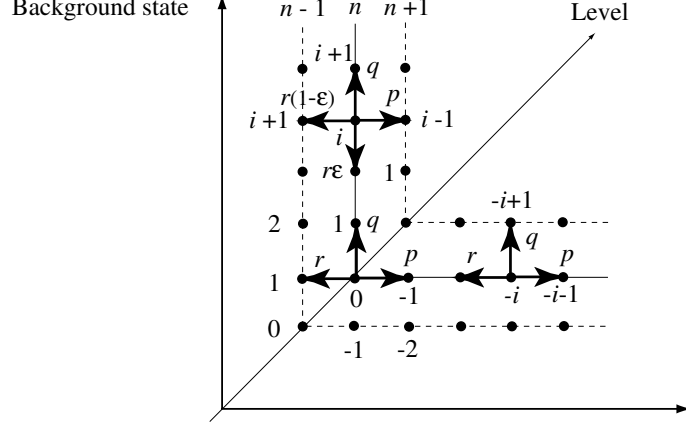


Figure 3: Background state transitions for Case 3

rate of the stationary distribution of the minimum of queue lengths. Let  $X_n = L^{(2)} - L^{(1)}$  and  $Y_n = \min(L_n^{(1)}, L_n^{(2)})$ . So,  $S_1 = \mathbb{Z}$  (see Figure 3). Then, we have

$$A_n(i, j) = \begin{cases} r(1 - \epsilon)1(j = i + 1 \geq 2) + r1(j = 1, i = 0), & n = -1, \\ p1(j = i - 1 \leq -1) + q1(j = i + 1 \geq 1) \\ \quad + r1(j = i + 1 \leq 0) + r\epsilon 1(j = i - 1 \geq 0) & n = 0, \\ p1(j = i - 1 \geq 0) + q1(j = i + 1 \leq 0), & n = 1, \\ 0, & \text{otherwise.} \end{cases}$$

Hence,  $A_*(\alpha)$  in the form of  $(\dots, -1, 0, 1, 2, \dots) \times (\dots, -1, 0, 1, 2, \dots)$  is

$$\begin{matrix} & \dots & -3 & -2 & -1 & 0 & 1 & 2 & 3 & \dots \\ \vdots & \begin{pmatrix} \ddots & \ddots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \dots & p & 0 & q\alpha + r & 0 & 0 & 0 & 0 & 0 & \dots \\ \dots & 0 & p & 0 & q\alpha + r & 0 & 0 & 0 & 0 & \dots \\ \dots & 0 & 0 & p & 0 & q + r\alpha^{-1} & 0 & 0 & 0 & \dots \\ \dots & 0 & 0 & 0 & p\alpha + r\epsilon & 0 & t(\alpha) & 0 & 0 & \dots \\ \dots & 0 & 0 & 0 & 0 & p\alpha + r\epsilon & 0 & t(\alpha) & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \ddots & \ddots & \ddots \end{pmatrix} & \end{matrix},$$

where  $t(\alpha) = q + r(1 - \epsilon)\alpha^{-1}$ .

Let  $\boldsymbol{\mu}(\alpha) = (\dots, \mu_{-1}, \mu_0, \mu_1, \dots)$  and  $\boldsymbol{h}(\alpha) = (\dots, h_{-1}, h_0, h_1, \dots)^T$  be the right and left invariant vector of  $A_*(\alpha)$ . Then,

$$\mu_n = p\mu_{n+1} + (q\alpha + r)\mu_{n-1}, \quad n \leq -1, \quad (5.17)$$

$$\mu_0 = (p\alpha + r\epsilon)\mu_1 + (q\alpha + r)\mu_{-1}, \quad (5.18)$$

$$\mu_1 = (p\alpha + r\epsilon)\mu_2 + (q + r\alpha^{-1})\mu_0, \quad (5.19)$$

$$\mu_n = (p\alpha + r\epsilon)\mu_{n+1} + (q + r(1 - \epsilon)\alpha^{-1})\mu_{n-1}, \quad n \geq 2. \quad (5.20)$$

Let

$$f_-(\theta) = p\theta^2 - \theta + q\alpha + r, \quad f_+(\theta) = (p\alpha + r\epsilon)\theta^2 - \theta + q + r(1 - \epsilon)\alpha^{-1},$$

and denote the solutions of  $f_-(\theta) = 0$  and  $f_+(\theta) = 0$  by  $\theta_{-1}, \theta_{-2}$  and  $\theta_1, \theta_2$ , respectively. Then, applying the same arguments in Case 2, we can see that, in order to satisfy condition (4.3), all the  $\theta_i$  must be real numbers, and  $\mu_n$  must be obtained as

$$\mu_n = \begin{cases} \mu_0 \theta_{-2}^n, & n \leq 0 \\ \mu_1 \theta_1^n, & n \geq 1, \end{cases}$$

where  $\theta_i$  are ordered as  $0 < \theta_1 < \theta_2, 0 < \theta_{-1} < \theta_{-2}$ . So we have

$$\mu_2 - \theta_1 \mu_1 = 0, \quad \mu_{-1} - \theta_{-2}^{-1} \mu_0 = 0.$$

Substituting the first expression into (5.19) and using the identify that  $\theta_1 - (p\alpha + r\epsilon)\theta_1^2 = q + r(1 - \epsilon)\alpha^{-1}$ , obtained from  $f_+(\theta_1) = 0$ , we have

$$(q + r(1 - \epsilon)\alpha^{-1})\mu_1 = (q + r\alpha^{-1})\theta_1 \mu_0.$$

Hence, substituting the expressions for  $\mu_0$  and  $\mu_1$  into (5.18) and using the identify that  $p\theta_{-2}^2 = \theta_{-2} - (q\alpha + r)$ , obtained from  $f_-(\theta_{-2}) = 0$ , we have

$$(q + r(1 - \epsilon)\alpha^{-1})p\theta_{-2} = (q + r\alpha^{-1})(p\alpha + r\epsilon)\theta_1. \quad (5.21)$$

By our definitions, we have the following expressions.

$$\theta_1 = \frac{1 - \sqrt{1 - 4(p\alpha + r\epsilon)(q + r(1 - \epsilon)\alpha^{-1})}}{p\alpha + r\epsilon}, \quad \theta_{-2} = \frac{1 + \sqrt{1 - 4p(q\alpha + r)}}{p},$$

Substituting these into (5.21) and after some manipulations, we have

$$r\epsilon = (q\alpha + r(1 - \epsilon))\sqrt{1 - 4p(q\alpha + r)} + (q\alpha + r)\sqrt{1 - 4(p\alpha + r\epsilon)(q + r(1 - \epsilon)\alpha^{-1})}. \quad (5.22)$$

This equation shows that, if  $\epsilon = 0$ , we can not find appropriate  $\theta_1$  and  $\theta_{-2}$ . Hence, we can not verify the conditions of Theorem 4.1 for the priority queue if the minimum of the queue lengths is taken as the level. For the modified priority queue with  $\epsilon > 0$ , we can find good  $\theta_1$  and  $\theta_{-2}$  for some  $\epsilon$ , but condition (4.4) is not obvious. In conclusion, we have less hope to apply Theorem 4.1 if the minimum of the queue lengths is the level.

## 6. Concluding remarks

The examples in the previous section show that Theorem 4.1 may be limited in applications, although it copes with interesting examples. In particular, condition (4.3) might be too strong. This is also closely related to effects of the boundary transitions. Namely, we need to incorporate the boundary behavior into the Markov renewal equation. To this end, it may be necessary to consider more specific models. For example, Miyazawa [19] conjectured decay rates in the generalized Jackson network and the batch movement network. It would be interesting to see whether the present result can be used to prove the conjectures. This remains in a future study.

Another issue is to find the  $\alpha$  and associated invariant vectors in applications of Theorem 4.1. This may not be so easy, and we may need to develop an algorithm to compute them.

We have only been concerned with the light tail case. However, it may be also interesting to consider the heavy tail case. Because of the convolution structure of the Markov renewal equation, this case can be considered under the same framework in principle. However, the infinite background states might cause difficulties, so this problem is also challenging.

**Acknowledgements** The authors are grateful to the anonymous referee for his comments and detailed suggestions. In particular, references [8, 22] on Markov additive processes are brought into our attention.

## References

- [1] Abate, J. and Whitt, W. (1997) Asymptotics for  $M/G/1$  low-priority waiting-time tail probabilities, *Queueing Systems* 25, 173-233.
- [2] Alsmeyer, G. (1994) On the Markov renewal theorem, *Stoch. Proc. their Appl.* 50, 37-56.
- [3] Alsmeyer, G. (2000) The ladder variable of a Markov random walk, *Probability and Mathematical Statistics* 20, 151-168.
- [4] Arjas, E. and Speed, T. P. (1973) Symmetric Wiener-Hopf factorizations in Markov additive processes, *Z. Wahrscheinlich. verw. Geb.* 26, 105-118.
- [5] Asmussen, S. (1987) *Applied Probability and Queues* 2nd Edition, Springer, New York.
- [6] Asmussen, S. (2000) *Ruin Probabilities*, World Scientific, Singapore.
- [7] Bertsimas, D., Paschalidis, I. C. and Tsitsiklis, J. N. (1999) Large deviations analysis of the generalized processor sharing policy, *Queueing Systems* 32, 319-349.
- [8] Chan, H. P. and Lai, T. L. (2003) Saddlepoint approximations and nonlinear boundary crossing probabilities of Markov random walks, *Annals of Applied Probability* 13, 395-429.
- [9] Çinlar, E. (1975) *Introduction to Stochastic Processes*, Prentice-Hall, Englewood Cliffs, New Jersey.
- [10] Collamore, J. F. (2002) Importance sampling techniques for the multidimensional ruin problem for general Markov additive sequences of random vectors, *Annals of Applied Probability* 12, 382-421.
- [11] Foley, R.D. and McDonald, D.R. (2001) Join the shortest queue: stability and exact asymptotics, *Annals of Applied Probability* 11(3), 569-607.

- [12] Grassmann, W. K. and Heyman, D. P. (1990) Equilibrium distribution of blocked-structured Markov chains with repeating rows, *J. Appl. Prob.* 27, 557–576.
- [13] Li, Q. and Zhao, Y. Q. (2002) Light-tailed asymptotics of stationary probability vector of Markov chains of  $GI/G/1$  type, submitted for publication.
- [14] Loynes, R. M. (1962) The stability of a queue with non-independent inter-arrival and service times, *Proceedings of Cambridge Philosophical Society* 58, 497–520.
- [15] McDonald, D.R. (1999) Asymptotic of first passage times for random walk in an orthant, *Annals of Applied Probability* 9, 110-145.
- [16] Miyazawa, M. (2002) A Markov renewal approach to the asymptotic decay of the tail probabilities in risk and queueing processes, *Probability in the Engineering and Informational Sciences* 16, 139–150.
- [17] Miyazawa, M. (2002) A paradigm of Markov additive processes for queues and their networks, in: *Matrix Analytic Methods Theory and Applications*, eds. Latouche, G. and Taylor, P. G., World Scientific, 265–289.
- [18] Miyazawa, M. (2002) The Markov renewal approach to  $M/G/1$  type queues with countably many background states, submitted for publication.
- [19] Miyazawa, M. (2003) Conjectures on decay rates of tail probabilities in generalized Jackson and batch movement networks, *Journal of the Operations Research Society of Japan* 46, 74–98.
- [20] Neuts, M. F. (1981) *Matrix-Geometric Solutions in Stochastic Models: An Algorithmic Approach*, The Johns University Press, Baltimore.
- [21] Neuts, M. F. (1989) *Structured Stochastic Matrices of  $M/G/1$  Type and Their Applications*, Marcel Dekker, New York.
- [22] Ney, P. and Nummelin, E. (1987) Markov additive processes I. Eigenvalue properties and limit theorems, *Annals of Probability* 15, 561-592.
- [23] Seneta, E. (1981) *Non-negative Matrices and Markov chains*, second edition, Springer-Verlag, New York.
- [24] Shurenkov, V. M. (1984) On the theory of Markov renewal, *Theory Probab. Appl.* 29, 247–265.
- [25] Takahashi, Y., Fujimoto, K. and Makimoto, N. (2001) Geometric decay of the steady-state probabilities in a quasi-birth-and-death process with a countable number of phases, *Stochastic Models* 17, No.1, 1–24.
- [26] Zhang, Z. L. (1998) Large deviations and the generalized processor sharing schedule for a multiple-queue systems, *Queueing Systems* 28, 349–376.
- [27] Zhao, Y. Q., Li, W. and Alfa, A. S. (1999) Duality results for block-structured transition matrices, *J. Appl. Prob.* 36, 1045–1057.

- [28] Zhao, Y. Q., Li, W. and Braun, W. J. (1998) Infinite block-structured transition matrices and their properties, *Adv. Appl. Prob.* 30, 365–384.
- [29] Zhao, Y. Q., Li, W. and Braun, W. J. (2003) Censoring, factorizations and spectral analysis for transition matrices with block-repeating entries, *Methodology and Computing in Applied Probability* 5, 35–58.