Light tail asymptotics in multidimensional reflecting processes for queueing networks

Masakiyo Miyazawa
Tokyo University of Science

Accepted March 21, 2010 (v 2.06)

Abstract

We are concerned with the stationary distributions of reflecting processes on multidimensional nonnegative orthants and other related processes, provided they exist. Such stationary distributions arise in performance evaluation for various queueing systems and their networks. However, it is very hard to obtain them analytically, so our interest is directed to analytically tractable characteristics. For this, we consider tail asymptotics of the stationary distributions.

The purpose of this paper is twofolds. We first overview the current approaches to attack the problem from a unified viewpoint. We then take up two approaches, Markov additive and analytic function approaches, which are recently developed by the author and his colleagues. We discuss their possible extensions. We mainly consider the tail asymptotics for two-dimensional reflecting processes, but also discuss how we can approach the case of more than two dimensions.

1 Introduction

Many queueing problems are related to networks, and in the present days have been studied using stochastic networks, which are stochastic models for describing stochastic flows on a graph with finitely many nodes. These models have been used to design and control network systems. We are often interested in performance measures observed over a long time period. For this, we first describe the model by a stochastic process, then consider its stationary distribution for computing performance measures of interest. Our primary interest is to see how those system performances depend on its modeling primitives, provided its stationary distribution exists.

However, except for special cases, computing the stationary distribution of a stochastic network is very difficult even for very simple models because their state spaces are multidimensional. For example, the Jackson network is an exceptional model, which has a product form stationary distribution. It is known that this nice analytical solution is destroyed by small structural changes such as server collaboration or batch arrivals. To overcome this difficulty, we consider the following two objects to be required.

(1a) A reasonably wide class of models which incorporate some structural changes;

(1b) Analytically tractable characteristics which are still useful to assess performance of models.

For (1a), we consider a discrete time reflecting process on the multidimensional nonnegative integer orthant, where the integer orthant means that all entries of its coordinate are nonnegative integers. This coordinate represents a state of a network. Here we allow flexible state transitions as long as possible while keeping analytical simplicity. To this end, we partition the orthant into two disjoint regions, called the interior and the boundary, where each state in the interior has a positive coordinate. We further partition the boundary into disjoint faces that are determined by the entries of the coordinate that vanish, and assume that state transitions within the interior and those within each face are homogeneous. That is, their increments at each transition instant do not depend on the current state as long as the process stays in the interior or in the same face. This process is referred to as a reflecting random walk on a nonnegative orthant. We will give its precise definition in Section 3. This model is used as a basic model, and we will discuss its extensions which allow multiple interiors and more complicated boundary.

As a related model of this reflecting random walk, we also discuss a semi-martingale reflecting Brownian motion, SRBM for short, which is a continuous time process with a continuous state space. This process is obtained as a limit of a sequence of reflecting random walks under suitable scaling in time and state space. In queueing terminology, it is obtained under the so-called heavy traffic condition (e.g., see Harrison and Williams [41]). The advantage of this model is its analytical simplicity. We need less primitive data to describe an SRBM, and its stationary equation is simpler than that of the reflecting random walk. Both classes of multidimensional processes have been studied for many years not only in the queueing theory but also in operations research and probability theory (see Harrison [39]).

Nevertheless, their theoretical studies are still at a primitive stage. For example, even the stability has not yet been fully answered for the more than three-dimensional reflecting random walks and SRBM. The tail asymptotics of the stationary distributions are only available for the two-dimensional processes except for special cases. Furthermore, Gamarnik [33, 34] argued for undecidability, that is, non-existence of a universal algorithm, for computing the stationary distribution of a multidimensional reflecting process, verifying its stability, as well as obtaining the tail decay rate (see Sipser [92] for details of undecidability). This does not exclude the possibility that the problems are solvable for some classes of models, but suggests that appropriate models should be chosen.

For (1b), we focus on tail asymptotic behaviors of the stationary distribution. As the undecidability discussed above suggests, this is still a hard problem, but greatly simplifies analysis compared with other characteristics. Furthermore, they are still important since they can be used to evaluate the probabilities of rare events which are generally preferred to be avoided. Thus, we consider the tail asymptotics of the stationary distribution of a multidimensional reflecting random walk. The aim of this paper is to give an overview of the methods to get these asymptotics and to discuss how they can be used in applications. Thus, this paper is basically a review paper, but includes some new suggestions as well. Namely, a new class of reflecting random walks is proposed in Section 3.3, while some new results are derived in Section 6. We also present a few conjectures.

In queueing theory, tail asymptotics have been studied for many years. The literature goes back at least to the early 1960s (e.g., see Feller [24]). Their main interest was in the exact or rough asymptotics by exponential (or geometric) functions for the stationary distributions of the workload (or queue length) in the $M/G/1$ and $GI/G/1$ queues, provided their stability was assumed (see Kingman [49]). Here, the tail distribution of a random variable X is said to have an exact exponential asymptotic if, for some constant $\alpha, b > 0$,

$$\lim_{x \to \infty} e^{\alpha x} \mathbb{P}(X > x) = b, \tag{1.1}$$

while it is said to have a rough exponential asymptotics if

$$\lim_{x \to \infty} \frac{1}{x} \log \mathbb{P}(X > x) = -\alpha. \tag{1.2}$$

Obviously, (1.1) implies (1.2). This α is called a decay rate. Those asymptotic results have been obtained using either the theory of analytic functions or the renewal theorem.

There were two streams for extending those results on a single queue with a single server. One direction is to cover more general arrival processes (see Glynn and Whitt [35]), and the other direction is to have many server queues (see, e.g., Takahashi [93], Neuts and Takahashi [77], Sadowsky [85] and Sadowsky and Szpankowski [86]). For them, there are three notable approaches. One is the large deviations technique (see Bertsimas, Paschalidis and Tsitsiklis [5], Chang [11], Dupuis and Ellis [18] and Shwartz and Weiss [91]). Another is the matrix-analytic method due to Neuts [75], which may be considered as an application of a Markov additive process and Wiener–Hopf factorization (e.g., see Arjas and Speed [3] and Miyazawa and Zwart [74]).

The studies in those lines generally allow neither simultaneous arrival at different queues nor dynamical changes in arrival and service mechanisms. However, there are some exceptions for them. They are two queues in parallel with Poisson arrivals and exponentially distributed service times. One variation is for arriving customers to join the shortest queue (see, e.g., Kingman [50]). Another is to allow simultaneous arrivals at two queues (see Flatto and Hahn [26]). These models can be also viewed as two-dimensional reflecting random walks, and analytic functions are used to get certain representations of their stationary distributions, which yield the exact geometric asymptotics. This may be considered as the third approach extending the classical analytic method. In this direction, we must acknowledge great contributions of the Russian school (Borovkov and Mogul'skii [7, 8, 9], Fayolle, Iasnogorodski and Malyshev [22] and Ignatyuk, Malyshev and Scherbako [45]).

All these three approaches, namely large deviations, Markov additive and analytic functions, have been further developed. We review them and discuss their possible extensions. For this, we start to reconsider the definitions (1.1) and (1.2) of the tail asymptotics in a more general context as well as for multidimensional distributions in Section 2. We then introduce the reflecting random walk and related models in Section 3. We discuss various approaches to get the tail asymptotics in Section 4. In those discussions, a particular interest is placed on what is difficult in studying the tail asymptotic problems.

Among those approaches, we detail the Markov additive approach in Section 5 and the analytic function approach using the convergence domain in Sections 6 and 7. We now have good answers for the two-dimensional reflecting processes. They are presented in Section 7. Those results are applied to some of modified Jackson networks and parallel queues with join the shortest queue in Section 8. We conclude this paper with various remarks for future study in Section 9. As you will see, there are so many problems to be open for further study, and some of them are going to be solved.

2 Tail asymptotics of distributions

In this section, we consider how we can define tail asymptotics. Our main interest is in multidimensional distributions, but we first consider one-dimensional distributions for simplicity.

Let X be a nonnegative random variable. We are interested in the tail probability $\mathbb{P}(X \geq x)$ for large x when the exact expression of $\mathbb{P}(X \geq x)$ is not available. We may think about approximating this tail distribution by a analytically tractable function. That is, a function h such that

$$\lim_{x \to \infty} \frac{\mathbb{P}(X \geq x)}{h(x)} = 1. \tag{2.1}$$

It may be questioned how this approximation by h is useful. For example, if the tail diminishes very quickly in the sense that it almost vanishes above a certain value of x or if it is very heavy, that is, it decreases very slowly, then it may not be meaningful to find such an approximating function. On the other hand, if the tail is between these two extremes, then the approximating function h of (2.1) may be useful in applications. To make these statements specific, we give the following definitions. Let $\varphi(\theta)$ be the moment generating function of X, that is,

$$\varphi(\theta) = \mathbb{E}(e^{\theta X}), \qquad \theta \in \mathbb{R},$$

as long as it exists.

Definition 2.1 The tail distribution $\mathbb{P}(X \geq x)$ is said to be

(2a) Small if $\varphi(\theta) < \infty$ for all $\theta > 0$,

(2b) Light if $\varphi(\theta) < \infty$ for some $\theta > 0$, but $\varphi(\theta) = \infty$ for some other $\theta > 0$,

(2c) Heavy if $\varphi(\theta) = \infty$ for all $\theta > 0$,

In this definition, X is real-valued, but the reflecting process is integer vector-valued as we discussed in Section 1 (see also Section 3). There is a good reason for this. If the state space of the reflecting random walk is one-dimensional, then we certainly do not need to consider a real-valued random variable. However, if it has the more than one dimension, the tail area to be considered may have various shapes. For example, a rectangle, a half-space separated by a hyperplane and a convex cone may be interesting. In these cases, the boundary of the tail area may not be well expressed by integers. The tail probability $\mathbb{P}(c_1X_1 + c_2X_2 > x)$ with positive numbers c_1 and c_2 is such an example.

As we have discussed, a light tail is ideal for studying the tail asymptotics. However, we do not know the tail type for the stationary distribution at the beginning. So far, the first step should be to consider which type of the tail distribution occurs under what conditions. We will consider this for $d = 2$, that is, the two-dimensional reflecting random walk in Section 6. In what follows, we heuristically consider how naturally a light tail arises in a queueing model with a single waiting line. This model is not a queueing network, but it may be also considered as one node in the queueing network.

Consider a single queue, and suppose that a smaller queue is more likely to increase than a larger queue. This may be intuitively expected. Let X be the size of such a queue in the steady state, then our supposition can be expressed as

$$\mathbb{P}(X \geq m+n | X \geq m) \leq \mathbb{P}(X \geq n), \qquad m, n \geq 1. \tag{2.2}$$

Let $f(n) = \mathbb{P}(X \geq n)$, then (2.2) is equivalent to

$$f(m+n) \leq f(m)f(n), \qquad m, n \geq 1. \tag{2.3}$$

Inequality (2.3) is termed submultiplicativity. Assume that $f(n) > 0$ for all $n \geq 1$. Then, taking logarithm of both sides, we have

$$\log f(m+n) \leq \log f(m) + \log f(n), \qquad m, n \geq 1.$$

This is called subadditivity. Then, it is well known that

$$\lim_{n\to\infty} \log f(n) = \inf_{n\geq 1} \frac{1}{n} \log f(n) < 0; \tag{2.4}$$

see, e.g., Lemma A.4 of Seneta [89] and Theorem 7.6.1 of Hille and Phillips [43]. Let

$$\alpha = -\inf_{n\geq 1} \frac{1}{n} \log f(n).$$

Obviously $\alpha > 0$. If $\alpha = \infty$, then the tail is small, while it is light if $\alpha < \infty$. This α is referred to as a decay rate of $\{f(n); n \geq 1\}$. Thus, under the assumption (2.2), the tail distribution is either light or small. This means that a heavy tail is impossible under this assumption.

In general, let X be a nonnegative real-valued random variable, and we define the decay rate α as

$$\lim_{x\to\infty} \frac{1}{x} \log \mathbb{P}(X > x) = -\alpha, \tag{2.5}$$

as long as it exists. In this case, $\mathbb{P}(X > x)$ is said to have rough asymptotic decay behavior with rate α. Note that this α is nonnegative, and may take the values 0 or ∞, which characterizes heavy and small tails.

As we have discussed, $0 < \alpha < \infty$ is preferable for making use of this asymptotics in applications. As we have already seen for the queue length distribution, we may expect this. It turns out that this is indeed the case in many stochastic network models. So far, we target the light tail asymptotics. It is also notable that the decay rate α can be characterized by the fact that, for any $\epsilon > 0$,

$$\mathbb{E}(e^{(\alpha-\epsilon)X}) < \infty, \qquad \mathbb{E}(e^{(\alpha+\epsilon)X}) = \infty.$$

This suggests that the moment generating function $\varphi(\theta) = \mathbb{E}(e^{\theta X})$ is useful for finding not only the tail type but also the decay rate.

We next consider a refinement of (2.5) in the form of exact asymptotics (2.1). However, we use a slightly weaker form to broaden its applicability.

Definition 2.2 If there exists a constant $b > 0$ and positive-valued function h on $\mathbb{R}_+$ such that

$$\lim_{x\to\infty} \frac{1}{h(x)} \mathbb{P}(X > x) = b, \tag{2.6}$$

then $\mathbb{P}(X > x)$ is said to have exact asymptotic function h. In this case, we also write $\mathbb{P}(X > x) \sim bh(x)$ or

$$\mathbb{P}(X > x) = bh(x) + o(h(x)), \qquad x \to \infty.$$

Note that we generally do not care about constant b in this definition. Of course, in application, this constant may be important, but theoretically it unduly restricts analytical study, so here we content ourselves with less fine asymptotics. In view of (2.5), $h(x)e^{\alpha x}$ would be a subexponential function, that is, the function which changes more slowly than an exponential function. It turns out that the following function

$$h(x) = x^{\kappa} e^{-\alpha x}, \qquad x > 0, \tag{2.7}$$

for finite $\alpha > 0$ and $\kappa \in \mathbb{R}$, occurs for the stationary distribution of the reflecting random walk. In particular, if $\kappa = 0$, then $\mathbb{P}(X > x)$ is said to have an exact exponential asymptotics with decay rate α. If X is integer-valued, then $\mathbb{P}(X \geq n)$ is said to have an exact geometric asymptotics with decay rate α. Note that this is equivalent to saying that $\mathbb{P}(X = n)$ has an exact geometric asymptotics with decay rate α.

One may wonder how to get the power κ and the decay rate α from the modeling primitives and how they change according to those primitives. These are not easy questions to answer when no analytic expression is available for the stationary distribution. Nevertheless, we can answer to them for the case of $d = 2$ to some extent, and we may expect to use the same idea for higher dimensions.

Until now, we have only considered one-dimensional distribution which captures a single queue. If such a queue belongs to a queueing network, we generally need to consider multiple queues at once. Thus, we may need to study the tail asymptotics of a multidimensional distribution. For such a distribution, we have to make the definition of a tail set clear. To consider this, let $\boldsymbol{X} \equiv (X_1, \ldots, X_d)$ be a d-dimensional random nonnegative vector for a positive integer d. Then, for a Borel measurable subset B of the d-dimensional Euclidean space $\mathbb{R}^d_+$ and a direction vector $\boldsymbol{c} \in \mathbb{R}^d$, that is, a vector $\boldsymbol{c}$ satisfying $\|\boldsymbol{c}\| = 1$ and $\boldsymbol{c} \geq \boldsymbol{0}$, we consider the tail asymptotics for

$$\mathbb{P}(\boldsymbol{X} \in x\boldsymbol{c} + B), \qquad x > 0, \tag{2.8}$$

where $\boldsymbol{u} + B = \{\boldsymbol{u} + \boldsymbol{y}; \boldsymbol{y} \in B\}$ for $\boldsymbol{u} \in \mathbb{R}^d$. In this case, $\boldsymbol{c}$ is called a direction vector, and $x\boldsymbol{c} + B$ is called a tail set.

In view of the tail distribution, we may require that

$$t(x\boldsymbol{c} + \boldsymbol{y}) \in x\boldsymbol{c} + B, \qquad \forall \boldsymbol{y} \in B, \forall t > 0,$$

that is, that $\boldsymbol{c} + B$ be a cone. Then, (2.8) is equivalent to

$$\mathbb{P}(\boldsymbol{X} \in x(\boldsymbol{c} + B)), \qquad x > 0. \tag{2.9}$$

This probability is generally used to consider the tail asymptotics. Since $\boldsymbol{c} + B \subset \mathbb{R}^d_+$, we may also use B itself instead of $\boldsymbol{c} + B$. This tail set is studied in the theory of large deviations.

Definition 2.3 If there exists a lower semi-continuous function $I(\boldsymbol{u})$ on $\mathbb{R}^d_+$ such that, for any measurable $B \subset \mathbb{R}^2_+$,

$$\limsup_{x\to\infty} \frac{1}{x} \log \mathbb{P}(\boldsymbol{X} \in xB) \le - \inf_{\boldsymbol{v}\in\overline{B}} I(\boldsymbol{v}), \tag{2.10}$$

$$\liminf_{x\to\infty} \frac{1}{x} \log \mathbb{P}(\boldsymbol{X} \in xB) \ge - \inf_{\boldsymbol{v}\in B^{\mathrm{o}}} I(\boldsymbol{v}) \tag{2.11}$$

where $\overline{B}$ and B^{o} are the closure and the interior of B, then $I(\boldsymbol{v})$ is called a rate function. This rate function is said to satisfy a large deviations principle for the distribution of $\boldsymbol{X}$.

We can again consider refinements of the rough asymptotics considered in this definition. That is, we may consider an asymptotic function for the tail probability $\mathbb{P}(\boldsymbol{X} \in xB)$ for each fixed B. For example, if, for $i \in J \equiv \{1, 2, \ldots, d\}$, we take $B = \{\boldsymbol{x} \in \mathbb{R}^d_+; x_i > 1\}$, then

$$\mathbb{P}(\boldsymbol{X} \in xB) = \mathbb{P}(X_i > x).$$

This is the marginal distribution of the ith component.

We may also consider the tail set:

$$B = \{\boldsymbol{u} \in \mathbb{R}^d_+; f(\boldsymbol{u}) > 1\} \tag{2.12}$$

by using a measurable function f from $\mathbb{R}^d_+$ to $\mathbb{R}_+$. This is particularly useful for considering tail types as in Definition 2.1. For this, we use

$$f(\boldsymbol{u}) = \sum_{i=1}^{n} c_i u_i \equiv \langle \boldsymbol{c}, \boldsymbol{u} \rangle,$$

for a directional vector $\boldsymbol{c} \ge \boldsymbol{0}$. In this case, (2.12) is the upper half-space over the hyper plane which is orthogonal to the vector $\boldsymbol{c}$. This set is analytically convenient because we can use the moment generating function as in Definition 2.1.

We define the joint moment generating function φ of a d-dimensional random vector $\boldsymbol{X}$ as

$$\varphi(\boldsymbol{\theta}) = \mathbb{E}\left(e^{\langle \boldsymbol{\theta}, \boldsymbol{X} \rangle}\right), \qquad \boldsymbol{\theta} \in \mathbb{R}^d.$$

We now classify the tail distribution of $\boldsymbol{X}$ as in Definition 2.1.

Definition 2.4 The distribution of a random vector $\boldsymbol{X}$ is said to have a small, light or heavy tail in direction $\boldsymbol{c} \geq \boldsymbol{0}$ if $\left\langle \boldsymbol{c}, \boldsymbol{X} \right\rangle$ has a small, light or heavy tail, respectively, in the sense of Definition 2.1. In particular, if the distribution of $\boldsymbol{X}$ has a light tail in all directions $\boldsymbol{c} \geq \boldsymbol{0}$, then $\boldsymbol{X}$ is said to have a light tail.

In view of this definition as well as the rough tail asymptotics, we may realize that it is important to consider the set of $\boldsymbol{\theta}$ for which $\varphi(\boldsymbol{\theta})$ is finite. For this, we define the convergence domain $\mathcal{D}$ as

$$\mathcal{D} = \text{the interior of } \{\boldsymbol{\theta} \in \mathbb{R}^d; \varphi(\boldsymbol{\theta}) < \infty\}. \tag{2.13}$$

Obviously, this domain plays a crucial role in finding the tail asymptotics. Furthermore, it can be used to characterize light tails. We note this as a lemma.

Lemma 2.1 The domain $\mathcal{D}$ is a convex subset of $\mathbb{R}^d$. (a) If there is some $\boldsymbol{\theta}^{(0)} \in \mathcal{D}$ such that $\boldsymbol{\theta}^{(0)} > \boldsymbol{0}$, then the tail distribution $\mathbb{P}(\left\langle \boldsymbol{c}, \boldsymbol{X} \right\rangle > u)$ $(u \geq 0)$ has a light tail or a small tail for each directional vector $\boldsymbol{c} \geq \boldsymbol{0}$. (b) If the assumption in (a) is satisfied and if $\mathcal{D}$ is bounded above by some hyperplane which is orthogonal to some vector $\boldsymbol{c} > \boldsymbol{0}$, then the tail distribution has a light tail.

Proof. Since the exponential function is convex, we have, for $\lambda \in (0,1)$ and $\boldsymbol{\theta}, \boldsymbol{\eta} \in \mathbb{R}^d$,

$$\begin{aligned}
\varphi(\lambda\boldsymbol{\theta} + (1-\lambda)\boldsymbol{\eta}) &= \mathbb{E}(e^{\lambda\left\langle \boldsymbol{\theta}, \boldsymbol{X} \right\rangle + (1-\lambda)\left\langle \boldsymbol{\eta}, \boldsymbol{X} \right\rangle}) \\
&\leq \mathbb{E}(\lambda e^{\left\langle \boldsymbol{\theta}, \boldsymbol{X} \right\rangle} + (1-\lambda) e^{\left\langle \boldsymbol{\eta}, \boldsymbol{X} \right\rangle}) \\
&= \lambda \mathbb{E}(e^{\left\langle \boldsymbol{\theta}, \boldsymbol{X} \right\rangle}) + (1-\lambda)\mathbb{E}(e^{\left\langle \boldsymbol{\eta}, \boldsymbol{X} \right\rangle}) \\
&= \lambda\varphi(\boldsymbol{\theta}) + (1-\lambda)\varphi(\boldsymbol{\eta}).
\end{aligned}$$

Thus, φ is a convex function, and therefore $\mathcal{D}$ is a convex set. Since $\boldsymbol{\theta}^{(0)} > \boldsymbol{0}$, we can find $u_0 > 0$ such that $\boldsymbol{\theta}^{(0)} > u_0\boldsymbol{c} > \boldsymbol{0}$. Then,

$$e^{uu_0}\mathbb{P}\left(\left\langle \boldsymbol{c}, \boldsymbol{X} \right\rangle > u\right) \leq \mathbb{E}\left(e^{u_0\left\langle \boldsymbol{c}, \boldsymbol{X} \right\rangle} 1(\left\langle u_0\boldsymbol{c}, \boldsymbol{X} \right\rangle > uu_0\right) \leq \mathbb{E}\left(e^{\left\langle \boldsymbol{\theta}^{(0)}, \boldsymbol{X} \right\rangle}\right) < \infty.$$

This implies that $\mathbb{P}\left(\left\langle \boldsymbol{c}, \boldsymbol{X} \right\rangle > u\right)$ decays at most exponentially fast. Thus, we have proved (a). If $\left\langle \boldsymbol{v}, \boldsymbol{X} \right\rangle$ has a small tail distribution for some directional vector $\boldsymbol{v} \geq \boldsymbol{0}$, then $\varphi(u\boldsymbol{v})$ must be finite for all $u \geq 0$. This contradicts the bounded assumption in (b). □

From this lemma, we can also see that, for any convex set $B \subset \mathbb{R}^+$ which does not contain a neighborhood of the origin, the tail distribution $\mathbb{P}(\boldsymbol{X} \in uB)$ $(u \geq 0)$ has a light or small tail under the condition of (a).

3 Reflecting processes on orthants

In this section, we introduce a unified model for a reflecting random walk. For this, we first discuss the Jackson network as a motivating example.

3.1 Motivating example: Jackson network

Consider a continuous time queueing network with d nodes, numbered as $1, 2, \ldots, d$. In Section 2, we have used the notation:

$$J = \{1, 2, \ldots, d\},$$

which is the set of nodes here. We assume that exogenous customers arrive at node i subject to a Poisson process with rate λ_i, and customers in node i have independent service times with an exponential distribution with mean $1/\mu_i$, and are served in first-in-first-out manner by a single server, which is independent of everything else. A customer who completes service at node i goes to node j with probability r_{ij} or leaves the network with probability r_{i0}, where

$$\sum_{j=0}^{d} r_{ij} = 1, \qquad i \in J.$$

We assume that all the movements are independent. Thus,

$$\lambda_i, \mu_i, r_{ij} \qquad i = 1, 2, \ldots, d, j = 0, 1, 2, \ldots, d$$

are modeling primitives. This model is referred to as a *Jackson network*.

This network model is usually described by a continuous time Markov chain. For this, let $L_i(t)$ be the number of customers in node i at time t. The d-dimensional vector-valued process $\boldsymbol{L}(t) \equiv (L_1(t), \ldots, L_d(t))$ is a continuous Markov chain, whose state space is the d-dimensional nonnegative integer orthant $S \equiv \mathbb{Z}_+^d$, where $\mathbb{Z}_+$ is the set of all nonnegative integers. It is not hard to see that its transition rate matrix $Q \equiv \{q(\boldsymbol{n}, \boldsymbol{n}'); \boldsymbol{n}, \boldsymbol{n}' \in S\}$ is given by, for $\boldsymbol{n} \neq \boldsymbol{n}'$,

$$q(\boldsymbol{n}, \boldsymbol{n}') = \begin{cases} \lambda_i & \boldsymbol{n}' = \boldsymbol{n} + \mathbf{e}_i, i \neq 0, \\ \mu_i r_{ij} & \boldsymbol{n}' = \boldsymbol{n} - \mathbf{e}_i + \mathbf{e}_j, n_i > 0, i, j \neq 0, \\ \mu_i r_{i0} & \boldsymbol{n}' = \boldsymbol{n} - \mathbf{e}_i, n_i > 0, i \neq 0, \\ 0 & \text{otherwise,} \end{cases} \tag{3.1}$$

where inequality of vectors stands for entry-wise inequalities, and

$$q(\boldsymbol{n}, \boldsymbol{n}) = -\sum_{\boldsymbol{n}' \neq \boldsymbol{n}} q(\boldsymbol{n}, \boldsymbol{n}'). \tag{3.2}$$

For notational convenience, we let $r_{00} = 0$ and

$$\mu_0 = \sum_{k=1}^{d} \lambda_k, \qquad r_{0i} = \lambda_i/\mu_0, \qquad i = 1, 2, \ldots, d.$$

Then, $(d+1) \times (d+1)$ matrix $R = \{r_{ij}\}$ is stochastic, and called a routing matrix. We can assume without loss of generality that R is irreducible.

If $\boldsymbol{L}(t)$ has the distribution which does not depend on t, it is called a stationary distribution. Denote this distribution by π if it exists. It is well known that this π is obtained as a nonnegative summable solution of the stationary equation:

$$\pi Q = 0,$$

and it is the product of the marginal distributions on nodes. To describe this distribution, let a_i be the solution of the following traffic equations:

$$a_i = \lambda_i + \sum_{j=1}^{d} a_j r_{ji}, \qquad i \in J.$$

Under the assumption that the routing matrix R is irreducible, the solution $a_1, a_2, \ldots, a_d$ exists uniquely. This a_i represents the total arrival rate at node i. Let $\rho_i = a_i/\mu_i$, and assume the stability condition:

$$\rho_i < 1, \qquad i \in J. \tag{3.3}$$

Then, the stationary distribution π is given by

$$\pi(\boldsymbol{n}) = \prod_{i=1}^{d} (1-\rho_i)\rho_i^{n_i}, \qquad \boldsymbol{n} \in S. \tag{3.4}$$

This distribution is said to have a product form. The details for this result can be found in standard textbooks (see, e.g., [12, 13, 90]).

Thus, for the Jackson network, we have a nice analytic expression for the stationary distribution, which accounts for its popularity in applications. However, this nice analytic result breaks down if there is even a small change of the modeling assumptions. For example, if we modify them in such a way that an idle server at node 1 helps a server at node 2 as long as node 1 is empty, which just increases μ_i when some other nodes are empty, the the product form solution is destroyed. Furthermore, there is no prospect of finding any analytic expression for the stationary distribution. We meet similar situations when customers simultaneously arrive at different nodes. We may want to see how the system performance is changed in those cases because such changes may naturally arise in applications.

This is exactly what we have discussed in Section 1. As a flexible model to handle these situations, we have proposed the reflecting random walk. We now formally introduce it.

3.2 Reflecting random walk on an orthant

We use some of standard notations for sets of numbers. Let $\mathbb{R}$ and $\mathbb{R}_+$ be the sets of all real nonnegative numbers, respectively. Similarly, let $\mathbb{Z}$ be the set of all integers. Let d be a positive integer. Then, $S \equiv \mathbb{Z}_+^d$ is referred to as a nonnegative orthant of $\mathbb{Z}^d$. The reflecting random walk is defined on this orthant. That is, it has state space S.

To describe a reflection mechanism, we partitioned this S into disjoint subsets. Let $J = \{1, 2, \ldots, d\}$. For each subset $A \subset J$, we define S_A as

$$S_A = \{\boldsymbol{x} \in S; x_i \geq 1, i \in A, x_j = 0, j \in J \setminus A\}.$$

If $A \neq J$, then S_A is called a boundary face. S_J represents the inside of S, and we also denote this inside by S_+. That is,

$$S_+ \equiv S_J = \{\boldsymbol{x} \equiv (x_1, \ldots, x_d) \in S; x_i > 0, i = 1, 2, \ldots, d\}.$$

The collection of all boundary faces is simply called the boundary, and denoted by ∂S. That is,

$$\partial S = \cup_{A \subset J, A \neq J} S_A.$$

We now define the reflecting random walk. For each $A \subset J$, let $\{\boldsymbol{X}_\ell^A; \ell = 1, 2, \ldots\}$ be a sequence of independent identically distributed random variables which are also independent of everything else. $\boldsymbol{X}_\ell^A$ represents a jump at time ℓ when the random walk is in S_A. We denote its distribution by $\{p_{\boldsymbol{x}}^A; \boldsymbol{x} \in \mathbb{R}^d\}$, that is,

$$p_{\boldsymbol{x}}^A = \mathbb{P}\left(\boldsymbol{X}_\ell^A = \boldsymbol{x}\right), \qquad \boldsymbol{x} \in \mathbb{Z}^d,$$

We omit the superscript A of $\boldsymbol{X}_\ell^A$ and $p_{\boldsymbol{x}}^A$ for $A = J$ when it is convenient. Thus, $\boldsymbol{X}_\ell$ and $p_{\boldsymbol{x}}$ may be used instead of them. We assume the following condition:

(3a) $p_{\boldsymbol{x}}^A = 0$ unless $x_i \geq -1$ for all $i \in A$ and $x_j \geq 0$ for all $j \in J \setminus A$.

This condition means that each jump in the side or face is skip-free downward.

Let $\boldsymbol{Z}_0$ be a random vector taking values in S, and inductively define a discrete time process $\{\boldsymbol{Z}_\ell; \ell = 0, 1, \ldots\}$ by

$$\boldsymbol{Z}_{\ell+1} = \boldsymbol{Z}_\ell + \sum_{A \subset J} \boldsymbol{X}_{\ell+1}^A 1(\boldsymbol{Z}_\ell \in S_A), \qquad \ell = 0, 1, \ldots. \tag{3.5}$$

By the assumption (3a), $\boldsymbol{Z}_\ell$ remains in S for all $\ell \geq 0$. We refer to this process as a *reflecting random walk on a nonnegative orthant with downward skip-free transitions*, or simply as a reflecting random walk. We may interpret $\boldsymbol{Z}_\ell$ as a state of a discrete time queueing network with d nodes, numbered as $1, 2, \ldots, d$, where the ith entry of the state is the number of customers in node i at time ℓ. Since, each entry $Z_{\ell,i}$ behaves like the $M/G/1$ queue at departure instants, this reflecting process is also referred to as a *multidimensional $M/G/1$-type queue*.

Clearly, $\{\boldsymbol{Z}_\ell\}$ is a discrete time Markov chain with state space S. Define its transition probability $p(\boldsymbol{n}, \boldsymbol{n}')$ as

$$p(\boldsymbol{n}, \boldsymbol{n}') = \mathbb{P}(\boldsymbol{Z}_{\ell+1} = \boldsymbol{n}' | \boldsymbol{Z}_\ell = \boldsymbol{n}), \qquad \boldsymbol{n}, \boldsymbol{n}' \in S,$$

where the right side of this equation does not depend on $\ell \geq 0$ by the modeling assumption. Let P be the infinite-dimensional matrix whose $(\boldsymbol{n}, \boldsymbol{n}')$th entry is $p(\boldsymbol{n}, \boldsymbol{n}')$. This P is a transition matrix, which is obviously stochastic.

We are interested in the stationary distribution of the reflecting random walk $\{\boldsymbol{Z}_\ell\}$. That is, we seek a distribution π on S such that

$$\mathbb{P}(\boldsymbol{Z}_\ell = \boldsymbol{n}) = \pi(\boldsymbol{n}), \qquad \boldsymbol{n} \in S, \ell = 0, 1, \ldots. \tag{3.6}$$

Let $\boldsymbol{Z}$ be a random vector subject to the distribution π, then it follows from (3.5) that

$$\boldsymbol{Z} \simeq \boldsymbol{Z} + \sum_{A \subset J} \boldsymbol{X}^A 1(\boldsymbol{Z} \in S_A), \qquad \ell = 0, 1, \ldots, \tag{3.7}$$

where "$\simeq$" stands for the equality in distribution. We can view the π as the row vector $\boldsymbol{\pi}$ whose $\boldsymbol{n}$th entry is $\pi(\boldsymbol{n})$. Then, (3.7) is equivalent to $\boldsymbol{\pi} = \boldsymbol{\pi} P$, and called a stationary equation. If P is irreducible, then it uniquely determines $\boldsymbol{\pi}$ as long as $\boldsymbol{\pi}$ exists. We assume this irreducibility throughout the paper. Algebraically our goal is to find the asymptotic behavior of the solution of this stationary equation. However, this will not be an easy task since $\boldsymbol{\pi}$ has an infinite-dimensional vector.

Even the existence of the stationary distribution is a big issue for the reflecting random walk. We will discuss it for $d = 2$. The problem is still open for $d \geq 4$. However, in its applications to queueing networks, we often easily find the stability condition by comparing the total arrival rate and the total service rate at each node.

We show how the reflecting random walk can accommodate the Jackson network and its modification for server collaboration by examples below.

Example 3.1 (Reflecting random walk for Jackson network) Let us show how the Jackson network is described by the reflecting random walk in discrete time. We first note that if we change time from t to bt for a constant $b > 0$, that is, time scale is changed by b, then λ_i and μ_i are also increased b times of them, so $\sum_{i=1}^{d} (\lambda_i + \mu_i)$ does so. However, this does not change the stationary distribution π. Hence, for studying the stationary distribution, we can assume without loss of generality that

$$\sum_{i=1}^{d} (\lambda_i + \mu_i) = 1.$$

By doing so, we closely look at the transition rate matrix Q of (3.2), and define $p^A_{\boldsymbol{n}}$ for each $A \subset J$ as

$$p^A_{\boldsymbol{n}} = \sum_{i \in A \cup \{0\}} \sum_{j=0}^{d} 1(\boldsymbol{n} = \mathbf{e}_j - \mathbf{e}_i) \mu_i r_{ij} + 1(\boldsymbol{n} = \mathbf{0}) \sum_{i \in J \setminus A} \mu_i,$$

where $\mathbf{e}_0 = \mathbf{0}$. Note that the second summation on the right hand side is a dummy transition for $\{p^A_{\boldsymbol{n}}\}$ to be a probability distribution. Thus, we have defined the reflecting random walk.

For $\boldsymbol{n} \in S_A$,

$$\begin{aligned} p(\boldsymbol{n}, \boldsymbol{n} + \mathbf{e}_j) &= p^A_{\mathbf{e}_j} = \mu_0 r_{0j} = \lambda_j, \qquad j = 1, \ldots, d; \\ p(\boldsymbol{n}, \boldsymbol{n}) &= p^A_{\mathbf{0}} = \sum_{i \in J \setminus A} \mu_i, \end{aligned}$$

and if $i \in A$ then

$$p(\boldsymbol{n}, \boldsymbol{n} + \mathbf{e}_j - \mathbf{e}_i) = p^A_{\mathbf{e}_j - \mathbf{e}_i} = \mu_i r_{ij}, \qquad j = 0, 1, \ldots, d.$$

Hence, for $\boldsymbol{n} \in S_A$,

$$p(\boldsymbol{n}, \boldsymbol{n}') = 1(\boldsymbol{n} \neq \boldsymbol{n}') q(\boldsymbol{n}, \boldsymbol{n}') + 1(\boldsymbol{n} = \boldsymbol{n}') \sum_{i \in J} \mu_i 1(n_i = 0).$$

Thus, $\pi Q = 0$ is equivalent $\pi P = \pi$, and this reflecting random walk indeed has the same distribution as the Jackson network.

This reflecting random walk can be also used to describe some modifications of the Jackson network. For example, let us change $p^A_{\boldsymbol{n}}$ for nonempty $A \neq J$ as

$$p^A_{\boldsymbol{n}} = \sum_{i \in A \cup \{0\}} \sum_{j=0}^{d} 1(\boldsymbol{n} = \mathbf{e}_j - \mathbf{e}_i)(\mu_i + \delta^A_i 1(i \neq 0)) r_{ij} + 1(\boldsymbol{n} = \mathbf{0}) \left(\sum_{i \in J \setminus A} \mu_i - \sum_{i \in A} \delta^A_i \right),$$

where δ^A_i for $i \in A$ is a nonnegative number, and it is assumed that

$$\sum_{i \in J \setminus A} \mu_i \geq \sum_{i \in A} \delta^A_i.$$

for $p^A_{\boldsymbol{n}}$ to be well defined. Then, this modification describes server collaboration when they are idle. Clearly, δ^A_i is an additional service rate for node i from idle servers. In this way, the reflecting random walk can be used to model the effect of server collaboration. Similarly, we can consider batch arrivals. □

Example 3.2 (Multiple QBD process) If the reflecting random walk $\{\boldsymbol{Z}_\ell\}$ is skip-free for all directions, that is, all entries of $\boldsymbol{X}^A_\ell$ take values 0, 1 or -1, then it is called a reflecting skip-free random walk. In queueing applications, it is also-called a multidimensional quasi-birth-and-death process, or multiple QBD process for short. This is because each entry of $\boldsymbol{Z}_\ell$ behaves like the birth-and-death process. This multiple QBD has simpler transitions, but still flexible for applications. For example, it can accommodate the Jackson network and some of its modifications. However, its tail asymptotics have not been well studied except for $d = 2$.

The multiple QBD process for $d = 2$ is called a double QBD process whose transition diagram is given below.

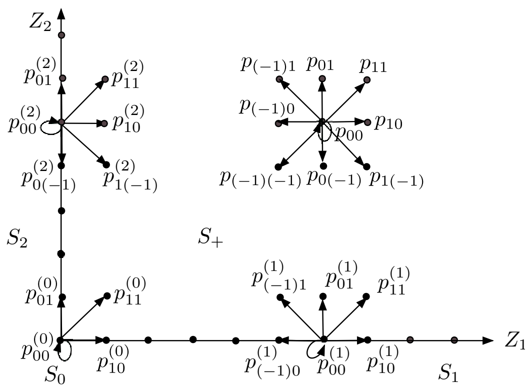

Figure 1: Transition diagram for the double QBD process

Even for this simple network model, some of tail asymptotics are still unknown. We will discuss them in Section 7. □

We will consider the tail asymptotic problem on the reflecting random walk in Section 6, and give some answers in Section 7.

3.3 Generalized reflecting random walk

We have considered the reflecting random walk on the orthant, but, for some applications, it may be convenient to have a more general state space and to allow some of boundary faces to be penetrable, that is, the boundary faces may be placed inside of the state space. We meet such a model when arriving customers join the shortest queue among parallel queues. For this model, the interior of the state space is partitioned, and each partitioned area has its own random walk.

It is straightforward to generalize the reflecting random walk on an orthant in this direction. We just replace the partitions $\{S^A\}$ according to all subsets of J by those according to an arbitrarily given index set. Denote this index set by $\mathcal{J}$, and let S be a subset of $\mathbb{Z}^d$. This S is used for a state space, which may not be an orthant. Similar to the reflecting random walk on an orthant, we partition S into disjoint subsets S_j for $j \in \mathcal{J}$. Thus, S_A is replaced by S_j. Similarly, we replace $\boldsymbol{X}_\ell^A$ by $\boldsymbol{X}_\ell^{(j)}$, and denote its distribution by $\{p_{\boldsymbol{x}}^{(j)}; \boldsymbol{x} \in S^{(j)}\}$. Thus, we define a discrete time process $\{\boldsymbol{Z}_\ell; \ell = 0, 1, \ldots\}$ by

$$\boldsymbol{Z}_{\ell+1} = \boldsymbol{Z}_\ell + \sum_{j \in \mathcal{J}} \boldsymbol{X}_{\ell+1}^{(j)} 1(\boldsymbol{Z}_\ell \in S_j), \qquad \ell = 0, 1, \ldots, \tag{3.8}$$

where we assume that distributions $\{p^{(j)}\}$ for $j \in \mathcal{J}$ are defined so that $\boldsymbol{Z}_{\ell+1} \in S$. We refer to this process as a generalized reflecting random walk, which is clearly a Markov chain. For the subset S_j, it is less meaningful to distinguish boundary and interior. Thus, we will not use them unless they are really needed. We give an example for this random walk below.

Example 3.3 (Two-sided double QBD) We consider a two-dimensional generalized reflecting random walk with two insides and four boundary faces, which is a special case of a two-sided QBD process introduced in Li, Miyazawa and Zhao [56]. Let $S \equiv \mathbb{Z} \times \mathbb{Z}_+$ and let $\mathcal{J} = \{-, +, 0, 1+, 1-, 2\}$. We define S_j as

$$\begin{aligned} &S_+ = \{(n_1, n_2) \in S; n_1, n_2 \geq 1\}, \quad S_- = \{(n_1, n_2) \in S; n_1 \leq -1, n_2 \geq 0\}, \quad S_0 = \{0\}, \\ &S_{1+} = \{(n, 0) \in S; n \geq 1\}, \quad S_{1-} = \{(n, 0) \in S; n \leq -1\}, \quad S_2 = \{0, n) \in S; n \geq 1\}. \end{aligned}$$

In this model, S_+ and S_- may be considered as interiors, and S_2 is a penetrable boundary face.

Assume that all the increments are skip-free. Then, the transition diagram of the Markov chain $\{\boldsymbol{Z}_\ell\}$ is given above. We refer to this model as a two-sided double QBD (quasi-birth-and-death) process. The tail asymptotics of this process is studied in Miyazawa [69]. We discuss them in Section 7.3. $\square$

3.4 Technical assumptions and stationary equations

We will consider the tail asymptotics of the stationary distribution of the d-dimensional reflecting random walk. We are interested in the case where it has a light tail. For this, we need some extra conditions on the modeling primitives. In this subsection, we first

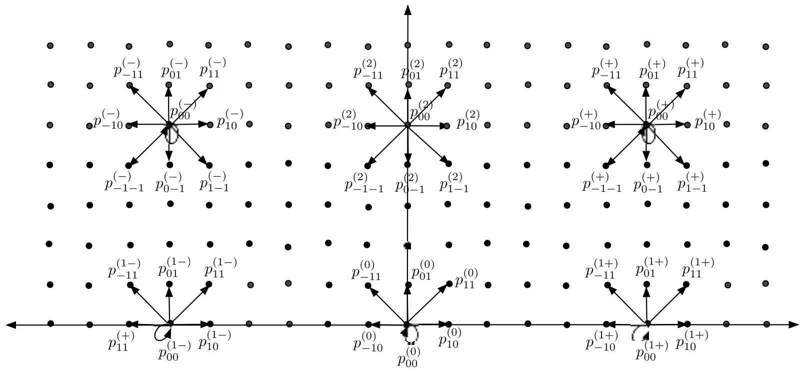

Figure 2: Transition diagram for the two-sided DQBD process.

give them in terms of the moment generating functions of distributions for the modeling primitives. We then derive the stationary equations in terms of generating functions.

For $A \subset J$ and $\boldsymbol{\theta} \in \mathbb{R}^d$, define the moment generating function of $\boldsymbol{X}^A$ as

$$\gamma_A(\boldsymbol{\theta}) = \mathbb{E}\left(e^{\left\langle \boldsymbol{\theta}, \boldsymbol{X}^A \right\rangle}\right), \qquad \boldsymbol{\theta} \in \mathbb{R}^d,$$

as long as it exists. Recall that $\boldsymbol{X}^A$ is an independent copy of $\boldsymbol{X}^A_\ell$. For the stationary distribution of the reflecting random walk to have a light tail, it is reasonable to assume that each increment at each transition instant has a light tailed distribution, that is,

(3b) For each $A \subset J$, $\gamma_A(\boldsymbol{\theta})$ is finite for some $\boldsymbol{\theta} > \boldsymbol{0}$.

In addition to this condition, we assume the following regularity condition:

(3b') For each $A \subset J$ and each $u > 0$, $\{\boldsymbol{\theta} \in \mathbb{R}^d; \gamma_A(\boldsymbol{\theta}) \le u\}$ is a closed set.

This condition is slightly stronger than what we really need, but proofs can be amended with minor technical arguments. So, we take an easy way.

We use one more assumption on the distributions of the increments also for making arguments simpler:

(3c) The random walk $\boldsymbol{Y}_\ell \equiv \sum_{i=1}^{\ell} \boldsymbol{X}^J_i$ is aperiodic and irreducible.

This irreducibility condition is equivalent to that the addition group generated by $\boldsymbol{n}$ such that $\mathbb{P}(\boldsymbol{X} = \boldsymbol{n}) > 0$ is identical with $\mathbb{Z}^d$. In certain applications, this is not satisfied, but we can again amend the arguments for such cases. By this assumption, the reflecting random walk $\{\boldsymbol{Z}_\ell\}$ is aperiodic and irreducible as a Markov chain. Hence, the stationary distribution is unique if

(3d) There exists a stationary distribution for the reflecting random walk $\{\boldsymbol{Z}_\ell\}$.

We recall that the stationary distribution is denoted by π. Assuming the condition (3d), let $\boldsymbol{Z}$ be a random vector subject to the stationary distribution π. We define a family of moment generating functions concerning $\boldsymbol{Z}$ as

$$\varphi_A(\boldsymbol{\theta}) = \mathbb{E}\left(e^{\langle \boldsymbol{\theta}, \boldsymbol{Z} \rangle} 1(\boldsymbol{Z} \in S_A)\right), \qquad \boldsymbol{\theta} \in \mathbb{R}^d, A \subset J.$$

Let $\varphi(\boldsymbol{\theta})$ be the moment generating function of π, then we have

$$\varphi(\boldsymbol{\theta}) = \sum_{A \subset J} \varphi_A(\boldsymbol{\theta}). \tag{3.9}$$

We take the moment generating functions of (3.7). Then, we get

$$\varphi(\boldsymbol{\theta}) = \sum_{A \subset J} \gamma_A(\boldsymbol{\theta}) \varphi_A(\boldsymbol{\theta}). \tag{3.10}$$

Hence, we have

$$\sum_{A \subset J} (1 - \gamma_A(\boldsymbol{\theta})) \varphi_A(\boldsymbol{\theta}) = 0. \tag{3.11}$$

as long as all $\varphi_A(\boldsymbol{\theta})$ are finite. We will use the pair of (3.9) and (3.11) so as to uniquely determine the stationary distribution π.

For the generalized reflecting random walk, we can similarly define φ_j and γ_j for $j \in \mathcal{J}$, and get the stationary equations. However, they may not be so useful because there is neither interior nor boundary. As we will see in Section 6, it is crucial to distinguish them for deriving the convergence domain of φ. Thus, we are probably better if we introduce interiors and boundary faces by adding extra random elements if necessary.

For example, let us consider the two-sided QBD process. In this case, we introduce

$$Z_{1-} = -Z_1 1(\boldsymbol{Z} \in S_- \cup S_{1-}), \qquad Z_{1+} = Z_1 1(\boldsymbol{Z} \in S_+ \cup S_{1+}),$$

and consider (Z_{1-}, Z_{1+}, Z_2) instead of $\boldsymbol{Z} \equiv (Z_1, Z_2)$. Similarly, the increments on S_0 and S_2 are partitioned as

$$\begin{aligned} X_{1-}^{(0)} &= -X_1^{(0)} 1(X_1^{(0)} = -1), \quad X_{1+}^{(0)} = X_1^{(0)} 1(X_1^{(0)} = 1), \\ X_{1-}^{(2)} &= -X_1^{(2)} 1(X_1^{(2)} = -1), \quad X_{1+}^{(2)} = X_1^{(2)} 1(X_1^{(2)} = 1). \end{aligned}$$

Thus, $(X_{1+}^{(j)}, X_{1-}^{(j)}, X_2^{(j)})$ replaces $\boldsymbol{X} \equiv (X_1^{(j)}, X_2^{(j)})$ for $j = 0, 1, 1-, 1+, 2$.

We appropriately define moment generating functions using three variables θ_{1-}, θ_{1+} and θ_2. For example,

$$\begin{aligned} &\varphi_-(\theta_1, \theta_2) = \mathbb{E}\left(e^{\theta_1 Z_{1-} + \theta_2 Z_2} 1(\boldsymbol{Z} \in S_-)\right), \\ &\varphi_{1-}(\theta_1) = \mathbb{E}\left(e^{\theta_1 Z_{1-}} 1(\boldsymbol{Z} \in S_{1-})\right), \quad \varphi_2(\theta_2) = \mathbb{E}\left(e^{\theta_2 Z_2} 1(\boldsymbol{Z} \in S_2)\right), \\ &\gamma_-(\theta_1, \theta_2) = \mathbb{E}\left(e^{\theta_1 X_{1-}^{(-)} + \theta_2 X_2^{(-)}}\right), \quad \gamma_{1-}(\theta_1, \theta_2) = \mathbb{E}\left(e^{\theta_1 X_{1-}^{(1-)} + \theta_2 X_2^{(1-)}}\right), \\ &\gamma_2(\theta_{1-}, \theta_{1+}, \theta_2) = \mathbb{E}\left(e^{\theta_{1-} X_{1-}^{(2)} + \theta_{1+} X_{1+}^{(2)} + \theta_2 X_2^{(2)}}\right). \end{aligned}$$

We then have the stationary equations similar to (3.9) and (3.11).

3.5 Semi-martingale reflecting Brownian motion (SRBM)

A reflecting Brownian motion on an orthant is a continuous time and space version of the multiple QBD. Let $\boldsymbol{X}(t)$ be a d-dimensional Brownian motion. We express it as $\boldsymbol{X}(t) = t\boldsymbol{\mu} + \boldsymbol{B}(t)$, where $\boldsymbol{\mu}$ is the mean drift vector, and $\boldsymbol{B}(t)$ is the null drift Brownian motion with $d \times d$ covariance matrix $\Sigma \equiv \{\sigma_{ij}\}$. We assume

(3-i) Σ is positive definite, that is, non-singular.

Let R be a $d \times d$ matrix. Then, we define the reflecting Brownian motion $\boldsymbol{Z}(t)$ as a solution of the following equation:

$$\boldsymbol{Z}(t) = \boldsymbol{X}(t) + R\boldsymbol{Y}(t), \qquad t \geq 0, \tag{3.12}$$

where $\boldsymbol{Y}(t)$ is a regulator, that is, a minimal continuous and nondecreasing process such that $\boldsymbol{Y}(0) = \boldsymbol{0}$ and its ith entry $Y_i(t)$ is increased only when $Z_i(t) = 0$ for each $i = 1, 2$. This $\boldsymbol{Z}(t)$ is referred to as a semi-martingale reflecting Brownian motion, SRBM for short (e.g., see Section 7.5 of Chen and Yao [13]).

Under the non-singularity assumption (3-i), it is known that this solution exists at least in distribution if and only if the following condition is satisfied (see Reiman and Williams [84] and Taylor and Williams [95]).

(3-ii) R is a complete-$\mathcal{S}$ matrix, that is, for each of its principal-submatrices, there is a nonnegative column vector which is transformed to a positive vector by this submatrix.

If R is an $\mathcal{M}$-matrix, that is, there is a nonnegative matrix G and a positive diagonal matrix D such that $R = (I - G)D$ and $(I - G)^{-1}$ exists, then the solution $\boldsymbol{Z}(t)$ of (3.12) can be expressed as a functional of $\boldsymbol{X}(t)$, that is, (3.12) has a strong solution. This functional is called a reflection mapping. Denote it by Ψ_t, then we have

$$\boldsymbol{Z}(t) = \Psi_t(\{\boldsymbol{X}(u); u \in [0, t]\}).$$

For the existence of the stationary distribution of the SRBM, it is known to be necessary that

(3-iii) R has an inverse R^{-1}, and $R^{-1}\boldsymbol{\mu} < \boldsymbol{0}$.

If R is an $\mathcal{M}$-matrix, then this condition is both necessary and sufficient.

For $d = 2$, the stationary distribution exists if and only if (3-iii) holds and R is a $\mathcal{P}$ matrix, that is, all principal submatrices of R have a positive determinant. Namely, the latter condition is written as

(3-iv) $r_{ii} > 0$ for $i = 1, 2$ and $r_{11}r_{22} - r_{12}r_{21} > 0$.

See Bramson, Dai and Harrison [10] and Harrison and Hasenbein [40] for recent developments.

We now assume that the stationary distribution exists, and denote it by π. We use Itô's integral formula for deriving the stationary equation. Denote, for a twice continuously differentiable function f of d variables, $\nabla f = (f'_1, \ldots, f'_d)^{\mathrm{T}}$ and

$$\mathcal{L}f(\boldsymbol{x}) = \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d}\sigma_{ij}\frac{\partial^2 f}{\partial x_i \partial x_j}(\boldsymbol{x}),$$

where σ_{ij} is (i,j)th entry of the covariance matrix Σ, and $\boldsymbol{x}^{\mathrm{T}}$ stands for the transpose of vector $\boldsymbol{x}$. Then, Itô's integral formula reads

$$\begin{aligned} f(\boldsymbol{Z}(u)) - f(\boldsymbol{Z}(0)) = &\int_0^1 \big\langle \nabla f(\boldsymbol{Z}(u)), \boldsymbol{\mu} du + d\boldsymbol{B}(u) \big\rangle \\ &+ \int_0^1 \big\langle \nabla f(\boldsymbol{Z}(u)), R d\boldsymbol{Y}(t) \big\rangle + \int_0^1 \mathcal{L}f(\boldsymbol{Z}(u))\, du, \end{aligned} \tag{3.13}$$

where $\big\langle \boldsymbol{x}, \boldsymbol{y} \big\rangle$ is the inner product of vectors $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$.

Assume that $\boldsymbol{Z}(t)$ is a stationary process with the initial distribution π. Define functions $\gamma(\boldsymbol{\theta})$ and $\gamma_{[j]}(\boldsymbol{\theta})$ as

$$\gamma(\boldsymbol{\theta}) = -\big\langle \boldsymbol{\theta}, \boldsymbol{\mu} \big\rangle - \frac{1}{2}\big\langle \boldsymbol{\theta}, \Sigma\boldsymbol{\theta} \big\rangle, \qquad \gamma_{[j]}(\boldsymbol{\theta}) = \sum_{i=1}^{d} \theta_i r_{ij}, \quad j = 1, 2, \ldots, d,$$

and denote the moment generating function of π by $\varphi(\boldsymbol{\theta})$. Here, $\gamma_{[j]}(\boldsymbol{\theta})$ corresponds to $\gamma_{J\setminus\{j\}}(\boldsymbol{\theta})$ of the reflecting random walk in Sect. 3.4. Let $\varphi_{[j]}(\boldsymbol{\theta}[j])$ denote the moment generating functions of $\boldsymbol{Z}(t)$ with respect to the Palm measure generated by the nondecreasing process $Y_j(t)$, where $\boldsymbol{\theta}[j]$ is the vector $\boldsymbol{\theta}$ whose jth entry is replaced by 0, that is,

$$\varphi_{[j]}(\boldsymbol{\theta}[j]) = \mathbb{E}_\pi\left(\int_0^1 e^{\big\langle \boldsymbol{\theta}[j], \boldsymbol{Z}(u) \big\rangle} dY_j(u)\right).$$

From (3.12), it is not hard to see that $\mathbb{E}_\pi(\boldsymbol{Y}(1)) = -R^{-1}\boldsymbol{\mu}$ is a finite and positive vector by (3-iii). Thus, $\varphi_{[j]}(\boldsymbol{\theta}[j])$ is well defined at least for $\boldsymbol{\theta}[j] \le 0$.

Let $f(\boldsymbol{x}) = \exp(\big\langle \boldsymbol{\theta}, \boldsymbol{x} \big\rangle)$ in (3.13), and taking the expectation with respect to the initial distribution π, we have the stationary equation:

$$\gamma(\boldsymbol{\theta})\varphi(\boldsymbol{\theta}) = \sum_{j=1}^{d} \gamma_{[j]}(\boldsymbol{\theta})\varphi_{[j]}(\boldsymbol{\theta}[j]), \tag{3.14}$$

as long as $\varphi(\boldsymbol{\theta})$, $\varphi_{[j]}(\boldsymbol{\theta}[j])$ are finite for all j, which holds at least for $\boldsymbol{\theta} \le \boldsymbol{0}$. This stationary equation corresponds to (3.11) for the reflecting random walk, which is obtained from the stationary equation (3.6).

We can see how (3.14) is simple compared with (3.11). This is another great advantage of an SRBM in applications. This suggests that we may also use an SRBM as a pilot model for the reflecting random walk in studying the stationary distribution and its tail asymptotics.

4 How to attack the problem

We now have all materials in our hands. The problem is how to derive the tail asymptotics. The difficulty of this problem comes from the fact that the reflecting boundary is not a bounded set. This means that the tail asymptotics can be influenced by the boundary even if the tail set is far away from the origin. Hence, we have to incorporate the influence into the tail asymptotics. For $d \geq 2$, there is more than one boundary face, so the influence from different faces also has to be simultaneously considered. Here we summarize four approaches which have been used to study those issues. Two of them, Markov additive and analytic function approaches in Sections 4.3 and 4.4, will be detailed in Sections 5, 6 and 7.

4.1 Brute force approach

If the stationary distribution is obtained in a closed form, then we may directly work with it to get its tail asymptotics. For example, a closed form expression is available for the stationary distribution of an SRBM for a two node tandem queue, and the tail asymptotics is obtained from it in Lieshout and Mandjes [59]. However, these cases are rather exceptional, and we cannot expect that this approach is generally applicable because analytical expressions of the stationary distributions are hardly ever obtained.

Because of this difficulty as well as its own interest, there have been numerous efforts to find analytically tractable solutions by modifying the modeling assumptions. Typically, the queueing networks are modified in such a way that they satisfy local balance, which produces product form solutions similarly to Jackson networks (e.g., see [12, 90]). Those modifications generally require unrealistic assumptions. However, in some cases, they can be used for stochastically bounding the stationary distributions (e.g., see Kella and Miyazawa [47] and Miyazawa and Taylor [72]). Thus, they may be useful to get rough asymptotics if lower bounds are available. However, even if they are found, it is hardly expected for them to be tight, that is, for the lower bounds to be identical with the upper bounds (see, e.g., [51]).

We refer to these two methods as a brute force approach. We should not exclude every approach for attacking the tail asymptotic problem, but we have to say that this approach is very limited in use.

4.2 Large deviations approach

A standard approach for the tail asymptotics is the theory of large deviations. This approach aims at finding a rate function that satisfies a large deviations principle for the stationary distribution (see Definition 2.3). In Majewski [61], it is obtained in two steps. First, we find tail asymptotics for a sequence of boundary free processes, which are usually input processes. Such a sequence is typically obtained through fluid scaling. This part is called a sample path large deviations. We consider this for the SRBM $\{\boldsymbol{Z}(t)\}$ discussed in Section 3.5. Let $\{\boldsymbol{B}(t); t \geq 0\}$ be the Brownian motion for this SRBM, which can be written as $\sqrt{\Sigma}\boldsymbol{W}(t)$ using the standard Brownian motion $\{\boldsymbol{W}(t); t \geq 0\}$ and covariance matrix Σ. For each $T > 0$, let $\mathcal{C}^d[0,T]$ be the set of all continuous functions from $[0,T]$ to $\mathbb{R}^d$. Then, the sample path large deviation principle for the fluid scaled process $\{\frac{1}{n}\boldsymbol{W}(nt); t \in [0,T]\}$ for the supremum is obtained for a closed set $A \subset \mathcal{C}^d[0,T]$ as

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\{\boldsymbol{W}(nt); t \in [0,T]\} \in nA\right) \leq - \inf_{\boldsymbol{\omega}\in A\cap\mathcal{H}^d} \frac{1}{2}\int_0^T \|\dot{\boldsymbol{\omega}}(t)\| dt, \tag{4.1}$$

where $\mathcal{H}^d$ is the set of all functions from $[0,\infty)$ to $\mathbb{R}^d$ which are absolutely continuous and have locally square integrable derivative. Taking the expression $\boldsymbol{X}(t) = t\boldsymbol{\mu} + \sqrt{\Sigma}\boldsymbol{W}(t)$ into account, we define the function I_T as

$$I_T(\boldsymbol{f}) = \frac{1}{2}\int_0^T \left\langle \dot{\boldsymbol{f}}(t) - \boldsymbol{\mu}, \Sigma^{-1}(\dot{\boldsymbol{f}}(t) - \boldsymbol{\mu})\right\rangle dt. \tag{4.2}$$

Then, (4.1) implies

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\{\boldsymbol{X}(nt); t \in [0,T]\} \in nA\right) \leq - \inf_{\boldsymbol{f}\in A\cap\mathcal{H}^d} I_T(\boldsymbol{f}). \tag{4.3}$$

Thus, we can see that I_T is the rate function for the fluid scaling $\{\frac{1}{n}\boldsymbol{X}(nt); t \in [0,T]\}$.

In the second step, we assume that R is an $\mathcal{M}$-matrix, which guarantees that the reflecting process $\boldsymbol{Z}(t)$ has a strong solution, that is, it is obtained from $\boldsymbol{X}(t)$ by the reflection mapping Ψ_t. We apply the contraction principle of large deviations for Ψ_t. This yields

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\{\boldsymbol{Z}(nt); t \in [0,T]\} \in nA\right) \leq - \inf_{\boldsymbol{f}\in\mathcal{H}^d, \{\Psi_t(\boldsymbol{f}); t\in[0,T]\}\in A} I_T(\boldsymbol{f}). \tag{4.4}$$

Then, letting $A = \{f \in \mathcal{C}[0,T]; \frac{1}{t}f(t) \in B, t \in (0,T]\}$ for a measurable closed set $B \subset \mathbb{R}^d_+$ and after some manipulations, we can prove that

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\boldsymbol{Z}(0) \in nB\right) \leq - \inf_{T>0, \boldsymbol{x}\in B} \inf_{\boldsymbol{f}\in\mathcal{H}^d, \Psi_T(\boldsymbol{f})=\boldsymbol{x}} I_T(\boldsymbol{f}). \tag{4.5}$$

Thus, the rate function for (2.10) is given by

$$I(\boldsymbol{x}) = \inf_{T>0} \inf_{\boldsymbol{f}\in\mathcal{H}^d, \Psi_T(\boldsymbol{f})=\boldsymbol{x}} \frac{1}{2}\int_0^T \left\langle \dot{\boldsymbol{f}}(t) - \boldsymbol{\mu}, \Sigma^{-1}(\dot{\boldsymbol{f}}(t) - \boldsymbol{\mu})\right\rangle dt. \tag{4.6}$$

It remains further work to get $I(\boldsymbol{x})$ in terms of the modeling primitives. This requires solving the variational problem in (4.6). For $d = 2$, this variational problem has been analytically solved in Avram, Dai and Hasenbein [4] under the assumption that R is an $\mathcal{M}$-matrix (see also [40]). Weaker conditions for this can be found in [19].

The basic idea of [4] is to reduce the function space for finding the optimal solution to a class of line graphs with two segments at most. Thus, the variational problem becomes an optimization problem with finite-dimensional variables. For $d = 3$, some studies in this line were made by El Kaharroubi, Yaacoubi, Tahar and Bichard [20], but the decay rate has not been obtained yet. An alternative expression of $I(\boldsymbol{x})$ was obtained by Dupuis and Ramanan [19], but it has not yet produced any explicit solution except for special cases.

This large deviations approach works also for the reflecting random walk as long as a reflecting mapping exists. However, these requirements are generally not satisfied because the reflections are not deterministic for the reflecting random walk, in general. Even if the reflecting mapping exists, it is very difficult to analytically solve the variational problem (e.g., see [63]).

4.3 Markov additive approach

Because of the limited availability of the sample path large deviations, another approaches have been explored. Among them, the Markov additive approach is the most popular for the reflecting random walk. A key ingredient of this approach is to extract a one-dimensional additive process removing one of the boundary faces. This enables us to apply limiting theorems, including large deviations.

This additive process itself is not Markov, so we add a background process for it to be Markov. This background process is generated by all components of the reflecting random walk except for the one corresponding to the additive process. Thus, it is a discrete time Markov chain, and the additive process with this background process is called a Markov additive process, which is formally defined in Section 5.2.

This Markov additive process is used to compute the mean sojourn time, that is, the mean visiting number, at each state before it returns to the level 0, which corresponds to the removed boundary face. The set of these conditional mean sojourn times is referred to as an occupation measure. The stationary distribution is obtained from:

(4a) The occupation measure of the Markov additive process,

(4b) The stationary measure on the removed boundary face.

Thus, we need to see tail asymptotics of these two quantities.

For $d = 1$, the Markov additive process is reduced to a renewal process. In this case, for (4a), we can apply the renewal theorem with help of the Wiener–Hopf factorization for a random walk while (4b) is trivial since the boundary is a single point. For $d \geq 2$, (4a) may be answered by either applying Markov renewal theorem or by computing the convergence parameter of the matrix moment generating function of the Markov additive transition kernel. The problem (4b) is much harder even for $d = 2$, and therefore strong conditions have often been used for suppressing this influence. There are many papers along this line (see, e.g., [29, 30, 32, 38, 42, 56]). The most general results in this line may be found in Miyazawa and Zhao [73], which are given in Theorem 5.2.

However, for $d = 2$, the problem for the tail asymptotics of (4b) has been solved by two different ways. The first is to combine two Markov additive processes along two different axes. This will be detailed in Section 5.6. The second is to find them through the convergence domain of the moment generating function of the stationary distribution, which will be discussed in Section 6 (see also Section 4.4.2).

The Markov additive approach appears under different formulations in the literature. We first summarize them, then discuss their features. The technical details of this approach will be discussed in Section 5.

4.3.1 Matrix analytic method

The matrix-analytic method originated in Neuts [75, 76] and has been applied to the tail asymptotic problems, particularly for single queues. A basic idea is to use vectors and matrices as if they are numbers. It was motivated by the possibility to avoid using transforms such as generating functions so as to numerically compute characteristics of interest directly. It was greatly succeeded for a quasi-birth-and-death process, QBD process for short, which is a Markov modulated birth-and-death process.

From its original motivation, this method has been mainly used when the background Markov chain has finitely many states. However, it is also well known that it can be used for countably many background states, although it generally looses the nice feature for numerical computations. Then, matrix manipulations as operators become more important (e.g., see Katou, Makimoto and Takahashi[46], Miyazawa [67] and Takahashi, Fijimoto and Makimoto [94]). In this respect, this approach can be considered as the Markov additive approach. Relations between those two approaches are also discussed in Miyazawa [65].

4.3.2 Borovkov–Mogul'skii approach

Borovkov and Mogul'skii have studied the tail asymptotic problem for many years [6, 7, 8, 9], and solved it for a two-dimensional reflecting random walk with a thick boundary and real vector-valued jumps, where a thick boundary means that the boundary has some bounded depth (Borovkov and Mogul'skii [9]). They combined various techniques for deriving exact asymptotics on a multidimensional renewal function, in which there are many excellent ideas for studying the tail asymptotics of the reflecting random walk. However, their results are not very explicit, and therefore they are not easy to use in applications. The essence of their approach is very close to the Markov additive approach. See the end of Section 5.6 for some remarks on this issue.

4.3.3 Foley–McDonald approach

Foley and McDonald have studied the tail asymptotic problem in a series of papers [28, 29, 30]. They mainly considered a skip-free reflecting random walk on the two-dimensional integer orthant, that is, a double QBD in our terminology. Their approach can also be considered as the Markov additive approach. However, there is one thing to be noted. As we will see, the Markov additive approach is generally useful to find exact geometric (or exponential) asymptotics, but not easy for finding other types of exact asymptotics. In [30], the authors challenged the latter problem using the ratio limit theorems for a Markov chain and the complex inversion of an analytic function around a branch point. This is rather connected to an analytic approach which will be discussed in Section 4.4.

4.3.4 Advantages and disadvantages of Markov additive approach

The Markov additive approach is very flexible for implementing extra information about the background states. For example, supplementary information on the arrival process and service times is easily incorporated. It also provides exact tail asymptotics for each fixed background state. There are also many studies on its own asymptotics (e.g., see Collamore [15], Ney and Nummelin [78, 79]). In this sense, the Markov additive approach has excellent features. However, it has two crucial limitations in application for the tail asymptotic problem.

One is the assumption that the additive component is one-dimensional. This enables us to apply the Wiener–Hopf factorization and Markov renewal theorem. For this, we can put necessary information into the background state space, but it may be complicated to compute eigenvalues and eigenvectors, particularly, for the reflecting random walk for $d \geq 3$. To prevent this difficulty, we may directly consider a multidimensional additive process. This formulation is studied in Miyazawa and Zwart [74]. One needs to generalize the Wiener–Hopf factorization. However, this approach has not yet been fully available to get the tail asymptotics.

The other is the strong conditions for the Markov renewal theorem to be applicable. For the double QBD process, there is a way to overcome this difficulty as shown in Section 5.6. However, it seems to be not applicable to higher dimensional reflecting processes. Thus, we may need another approach here. The analytic approach which will be discussed below seems to be a good candidate for this.

4.4 Analytic function approach

We may consider a multidimensional moment generating (or generating) function for the stationary distribution for the tail asymptotics problem. In Section 2, we have used them for categorizing the tail types. Here, we go one step further. The idea is to use complex variable functions and to apply the theory of analytic functions, where a complex-valued and complex variable function $f(z)$ is said to be analytic at $z = z_0$ if it is well defined on some neighborhood of z_0 on the complex plane $\mathbb{C}$ and it has a unique derivative at z_0 in all directions. The following fact is elementary, but it is the basis for this approach.

Lemma 4.1 Let $f(\theta)$ be the moment generating function of a measure on $\mathbb{R}_+$ with real variable θ, and let $\theta_0 = \sup\{\theta \in \mathbb{R}; f(\theta) < \infty\}$. Then, the complex variable function $f(z)$ is singular at $z = \theta_0$, and analytic on $\{z \in \mathbb{C}; \Re z < \theta_0\}$.

The corresponding theorem for a generating function is called Pringsheim's theorem, which is given bellow.

Lemma 4.2 (Pringsheim's theorem) Let $f(\theta)$ be the generating function of a measure on $\mathbb{Z}_+$, and let $\theta_0 = \sup\{\theta \geq 0; f(\theta) < \infty\}$. Then, the complex variable function $f(z)$ is singular at $z = \theta_0$, and analytic on $\{z \in \mathbb{C}; |z| < \theta_0\}$.

This lemma is less obvious and needs a proof (see, e.g., Theorem 17.13 in Volume 1 of Markushevich [64]). Because of these lemmas, we can expect that the leftmost singular point would be the decay rate. A significant feature of an analytic function is that it is uniquely determined by a set which has an accumulation point. For example, if two complex variable functions $f(z)$ and $g(z)$ agree on some open interval (a, b) of real numbers and if $f(z)$ is analytic on an open set G such that $(a, b) \subset G$, then g is uniquely extended on G in such a way that

$$g(z) = f(z), \qquad z \in G.$$

This is a classic result, but turns out to be very powerful for finding the domain of the moment generating function of the stationary distribution.

Another useful technique is the inversion formulas for a complex variable moment generating function at a leftmost singular point (see Doetsch [17]). They provide exact tail asymptotics. In the literature, two types of inversion formulas have been used according to the nature of the singularity, pole or branch point. The exact asymptotic function $h(x)$ has the form,

$$h(x) = x^{\kappa} e^{-\alpha x}.$$

If the singularity is caused by a pole, then κ is a nonnegative integer. On the other hand, if it is caused by a branch point, then κ is a rational number but not an integer. Their details can be found in Appendix C of Dai and Miyazawa [16]. For generating functions, similar results have been studied by researchers of combinatorics (see Flajolet and Sedqewick [25]), which are used in Li and Zhao [57, 58].

A problem with the analytic function approach is the difficulty in finding an analytic expression for the moment generating function of the stationary distribution, particularly for $d \geq 2$. This function is obtained as a solution of the stationary equations (3.9) and (3.10) (or (3.11)). Thus, we need to solve a functional equation for multivariable functions. This is generally a hard problem. Of course, there are some exceptional cases. For example, the moment generating functions can be analytically obtained for tandem and priority queues. In this case, the analytic function approach is well applied (see, e.g., [57, 60]). However, we cannot expect such nice solutions in general. Here, we need ideas to overcome this difficulty.

4.4.1 A method using Riemann surface

For $d = 2$, there have been some efforts to get a certain analytic expression for the generating function of the stationary distribution. Their essence is to reduce the problem to finding expressions for measures on the boundary faces, then getting the stationary distribution from those measures in terms of generating functions. In Fayolle, Iasnogorodski and Malyshev [22], either a Riemann surface generated by the null points of the generating function of the increments of the reflecting random walk in the interior or the solution for the boundary value problem is used for this derivation. The idea has already appeared in Kingman [50] and has been used in Flatto and McKean [27], Flatto and Hahn [26] and Fayolle and Iasnogorodski [21].

The current version of this approach is only applicable to the two-dimensional skip-free reflecting random walk, that is, the double QBD process. It has been used to derive rough asymptotics for this skip-free random walk in Ignatyuk, Malyshev and Scherbako [45]. This method can also be used to get exact asymptotics, but it has been limited to relatively simple models such as tandem or parallel queues (e.g., see [27]). Its recent studies can be found in Guillemin and Leeuwaarden [37] and Li and Zhao [58], where the approach is called a kernel method.

4.4.2 A method using the convergence domain

There is yet another analytic approach which is recently developed by Miyazawa and Rolski [71] and Dai and Miyazawa [16]. This approach uses not a generating function but a moment generating function. This is mainly because a moment generating function is convex and has nice analytic properties as a function of complex variable. Nevertheless, some basic ideas are very similar to the methods of Riemann surface and boundary value problem. Namely, this method is also based on the stationary equation, and the measures on the boundary faces play a key role.

A unique feature of this approach is to start with identifying the convergence domain of the moment generating function. The convexity and analytic properties of this function are particularly useful. For example, the domain has a nice geometric interpretation. Furthermore, the approach is potentially useful for $d \geq 3$. Once the domain is obtained, we can find singular points of the moment generating function on the boundary of the domain, then get tail asymptotics applying the analytic inversion formulas around the singular points. We will detail this approach in Sections 6 and 7.

5 Markov additive approach: technical details

We discuss technical ideas of the Markov additive approach discussed in some recent papers [53, 56, 68, 73, 74]. We start with a useful identity for a general Markov chain, which will be used not only for a Markov additive process but also for the reflecting random walk.

5.1 Pitman identity

As we have discussed in Section 4.3, the Markov additive approach uses two measures in (4a) and (4b). In this section, we derive a basic formula to produce the stationary distribution from them. Let S_b be a countable set, and let $S = \mathbb{Z} \times S_b$. Let $\{\boldsymbol{Z}_\ell\}$ be a S-valued Markov chain with transition kernel Q, and let $\{\mathcal{F}_\ell\}$ be its natural filtration. Let τ be a stopping time, that is, τ is a nonnegative integer-valued random variable such that $\{\tau \leq \ell\} \in \mathcal{F}_\ell$ for all $\ell \geq 0$. Define $S \times S$ matrices $G_\tau(s)$ and $H_\tau(s)$ as, for $\boldsymbol{m}, \boldsymbol{n} \in S$,

$$[\hat{G}_\tau(s)]_{\boldsymbol{m},\boldsymbol{n}} = \mathbb{E}_{\boldsymbol{m}}(s^\tau 1(\boldsymbol{Z}_\tau = \boldsymbol{n})1(\tau < \infty)),$$
$$[\hat{H}_\tau(s)]_{\boldsymbol{m},\boldsymbol{n}} = \mathbb{E}_{\boldsymbol{m}}\Big(\sum_{\ell=0}^{\tau-1} s^\ell 1(\boldsymbol{Z}_\ell = \boldsymbol{n})\Big),$$

where $\mathbb{E}_{\boldsymbol{m}}$ stands for the conditional expectation given $\boldsymbol{Z}_0 = \boldsymbol{m}$.

We consider the identity:

$$\sum_{\ell=0}^{\infty} s^\ell 1(\boldsymbol{Z}_\ell = \boldsymbol{n})1(\tau > \ell) + s^\tau 1(\boldsymbol{Z}_\tau = \boldsymbol{n})$$
$$= 1(\boldsymbol{Z}_0 = \boldsymbol{n}) + \sum_{\ell=0}^{\infty} s^{\ell+1} 1(\boldsymbol{Z}_{\ell+1} = \boldsymbol{n})1(\tau > \ell), \qquad \boldsymbol{n} \in S. \tag{5.1}$$

Since τ is a stopping time, we have

$$\begin{aligned}\mathbb{E}_{\boldsymbol{m}}(1(\boldsymbol{Z}_{\ell+1}=\boldsymbol{n})1(\tau>\ell)) &= \mathbb{E}_{\boldsymbol{m}}(1(\tau>\ell)\mathbb{E}(1(\boldsymbol{Z}_{\ell+1})=\boldsymbol{n}|\boldsymbol{Z}_{\ell}))\\ &= \mathbb{E}_{\boldsymbol{m}}\left(\sum_{\boldsymbol{n}'\in S}1(\tau>\ell,\boldsymbol{Z}_{\ell}=\boldsymbol{n}')Q_{\boldsymbol{n}',\boldsymbol{n}}\right).\end{aligned}$$

Hence, taking the conditional expectation of (5.1) yields

$$\hat{H}_\tau(s)+\hat{G}_\tau(s)=I+s\hat{H}_\tau(s)Q, \qquad 0\le s<1,$$

where I is the identity matrix. Rearranging terms in this equation, we have the so-called Pitman identity.

Lemma 5.1 (Pitman [82])

$$\hat{H}_\tau(s)(I-sQ)=I-\hat{G}_\tau(s), \qquad 0\le s<1. \tag{5.2}$$

We rewrite this identity as

$$\hat{G}_\tau(s)=I+\hat{H}_\tau(s)(I-sQ), \qquad 0\le s<1.$$

Then, it can be considered as a discrete time version of Dynkin's formula for a continuous time Markov process (e.g., see Ethier and Kurtz [31]).

5.2 Wiener–Hopf factorization

We now formally define a Markov additive process, and derive a useful identity on the ladder instants of the additive component applying Lemma 5.1. This identity is called RG decomposition or Wiener–Hopf factorization.

Let (X_ℓ,Y_ℓ) be an S-valued process satisfying the following condition:

$$\begin{aligned}&\mathbb{P}(X_{\ell+1}-X_\ell=n,Y_{\ell+1}=j|X_{k-1},Y_{k-1},k\le\ell,Y_\ell=i)\\ &\qquad\qquad=\mathbb{P}(X_{\ell+1}-X_\ell=n,Y_{\ell+1}=j|Y_\ell=i),\end{aligned}$$

for $n\in\mathbb{Z}$ and $i,j\in S_b$. Denote the right-hand side by $[A(n)]_{ij}$. Obviously, $\{Y_\ell\}$ is a Markov chain with transition probability matrix $\sum_{n=-\infty}^{+\infty}A(n)$. This $\{(X_\ell,Y_\ell)\}$ is said to be a discrete-time Markov additive process (MAP) with transition kernel $A(\cdot)$. $\{X_\ell\}$ is called an additive process while $\{Y_\ell\}$ is called a background process. The values of X_ℓ and Y_ℓ are referred to as level and background state, respectively. Define the stochastic kernel Q as

$$Q_{(m,i),(n,j)}=[A(n-m)]_{ij}.$$

Then, it is easy to see that $\{(X_\ell,Y_\ell)\}$ is a Markov chain with transition kernel Q. Thus, we can apply Lemma 5.1 to this Markov additive process, and (5.2) is available.

Let $\tau_y^{-0} = \inf\{n \geq 1; X_\ell - X_0 \leq y\}$. That is, τ_y^{-0} is the hitting time at or below level y from above. Define $\mathbb{Z}_+ \times S_b$ matrices $G_*^{-0}(s,\theta)$ and $H_*^+(s,\theta)$ by

$$[G_*^{-0}(s,\theta)]_{ij} = \mathbb{E}_i(s^{\tau_0^{-0}} e^{\theta(X_{\tau_0^{-0}} - X_0)} 1(Y_{\tau_0^{-0}} = j)),$$

$$[H_*(s,\theta)]_{ij} = \mathbb{E}_i\Big(\sum_{\ell=0}^{\tau_0^{-0}-1} s^\ell e^{\theta(X_\ell - X_0)} 1(Y_\ell = j)\Big).$$

Then, from (5.2) with $\tau = \tau_0^{-0}$, we have

$$H_*(s,\theta)(I - sA_*(\theta)) = I - G_*^{-0}(s,\theta), \tag{5.3}$$

where

$$[A_*(\theta)]_{ij} = \mathbb{E}(e^{\theta(X_1 - X_0)} 1(Y_1 = j)|Y_0 = i).$$

Define

$$[R_*^+(s,\theta)]_{ij} = \mathbb{E}_i\Big(\sum_{\ell=1}^{\infty} s^\ell e^{\theta(X_\ell - X_0)} 1(Y_\ell = j) 1(X_0 < X_\ell \leq \min(X_1, \ldots, X_{\ell-1}))\Big).$$

Then, it can be proved from a sample path decomposition that

$$H_*(s,\theta) = (I - R_*^+(s,\theta))^{-1}. \tag{5.4}$$

Hence, (5.3) implies

$$(I - sA_*(\theta)) = (I - R_*^+(s,\theta))(I - G_*^{-0}(s,\theta)). \tag{5.5}$$

This is called an RG decomposition (see, e.g., Grassmann and Heyman [36] and Zhao, Li and Braun [97]).

We convert (5.5) into another form, using a time reversed process of the Markov additive process under a suitable measure. For this, we need some further notions. Since $A \equiv A_*(1)$ is stochastic, it has a subinvariant vector $\boldsymbol{\pi}$, that is, $\boldsymbol{\pi} A \leq \boldsymbol{\pi}$. Then, we can define a substochastic matrix $\tilde{A}$ by

$$\tilde{A} = \Delta_{\boldsymbol{\pi}}^{-1} A^{\mathrm{T}} \Delta_{\boldsymbol{\pi}}^{-1},$$

where $\Delta_{\boldsymbol{\pi}}$ is the diagonal matrix whose diagonal entries are the entries of $\boldsymbol{\pi}$, and A^{T} denotes the transpose of A. Let $\{(\tilde{X}_\ell, \tilde{Y}_\ell)\}$ be the MAP generated by $\tilde{A}$. Define

$$[\tilde{G}_*^+(s,\theta)]_{ij} = \mathbb{E}_i\Big(s^{\tilde{\tau}_0^+} e^{\theta(\tilde{X}_{\tilde{\tau}_0^+} - \tilde{X}_0)} 1(\tilde{Y}_{\tilde{\tau}_0^+} = j)\Big),$$

where $\tilde{\tau}_y^+ = \inf\{n \geq 1; \tilde{X}_\ell - \tilde{X}_0 > 0\}$. Then, it can be shown that

$$(\tilde{G}_*^+(s,\theta))^{\mathrm{T}} = R_*^+(s,\theta). \tag{5.6}$$

Hence, we have

Theorem 5.1 (Wiener–Hopf factorization [3, 74])

$$I - sA_*(\theta) = (I - \tilde{G}_*^{+}(s,\theta)^{\mathrm{T}})(I - G_*^{-0}(s,\theta)), \qquad \theta \in \mathbb{R}, |s| \le 1, \tag{5.7}$$

as long as both sides exist and are finite.

This identity is known as the Wiener–Hopf factorization. In its applications, we need to carefully examine for which θ and s it is valid. This is thoroughly considered in Miyazawa and Zwart [74].

For $s = 1$, we simply denote $G_*^{-0}(s,\theta)$, $R_*^{+}(s,\theta)$ and $H_*^{+}(s,\theta)$ by $G_*^{-0}(\theta)$, $R_*(\theta)$ and $H_*(\theta)$. Then, (5.5) for $s = 1$ can be written as

$$(I - A_*(\theta)) = (I - R_*^{+}(\theta))(I - G_*^{-0}(\theta)). \tag{5.8}$$

This factorization formula plays a crucial role in the Markov additive approach because it relates the transformed occupation measure $H_*(\theta)$ to the transformed Markov additive kernel $A_*(\theta)$.

We have worked on the matrix moment generating functions, but it may be convenient to use probability or expectation matrices such as $A(n)$. For this, we introduce the following notation:

$$\begin{aligned}
&[G^{-0}(n)]_{ij} = \mathbb{P}_i(X_{\tau_0^{-0}} - X_0 = n, Y_{\tau_0^{-0}} = j),\\
&[H(n)]_{ij} = \mathbb{E}_i\Big(\sum_{\ell=0}^{\tau_0^{-0}-1} 1(X_\ell - X_0 = n, Y_\ell = j)\Big),\\
&[R^{+}(n)]_{ij} = \mathbb{E}_i\Big(\sum_{\ell=1}^{\infty} 1(X_\ell - X_0 = n, Y_\ell = j, X_0 < X_\ell \le \min(X_1, \ldots, X_{\ell-1}))\Big),\\
&[\tilde{G}^{+}(n)]_{ij} = \mathbb{P}_i(\tilde{X}_{\tau_0^{+}} - \tilde{X}_0 = n, \tilde{Y}_{\tau_0^{+}} = j).
\end{aligned}$$

Then, (5.6) and (5.8) can be written as

$$[\tilde{G}^{+}(n)]^{\mathrm{T}} = R^{+}(n), \qquad n \ge 1, \tag{5.9}$$

$$I - A(n) = (I - R^{+}) * (I - G^{-0})(n), \qquad n \in \mathbb{Z}, \tag{5.10}$$

where $A{*}B(n)$ represents the convolution of two sequences of matrices $\{A(n)\}$ and $\{B(n)\}$ of the same sizes, that is,

$$A * B(n) = \sum_{k=-\infty}^{+\infty} A(k)B(n-k),$$

where $B(k) = 0$ if it is not defined for k.

5.3 Reflecting Markov additive process

We next consider the description of the reflecting random walk using the Markov additive process. For this, we apply the Pitman identity (Lemma 5.1) and the Wiener–Hopf factorization (Theorem 5.1).

Let $\{\boldsymbol{Z}_\ell\}$ be the d-dimensional reflecting random walk that we introduced in Section 3.2. We choose $Z_{\ell 1}$ for the additive process. That is, let

$$\hat{X}_\ell = Z_{\ell 1}, \qquad \hat{Y}_\ell = (Z_{\ell 2}, Z_{\ell 3}, \ldots, Z_{\ell d}).$$

Here we do not use X_ℓ and Y_ℓ since $(\hat{X}_\ell, \hat{Y}_\ell)$ is not a Markov additive process. Since $\{(\hat{X}_\ell, \hat{Y}_\ell)\}$ is identical with $\boldsymbol{Z}_\ell$, it is a Markov chain. Let $\hat{Q} \equiv \{\hat{Q}_{(m,i),(n,j)}\}$ be its transition matrix. Then, we can see that $\hat{Q}_{(m,i),(n,j)}$ only depends on $n-m$ for $m, n \geq 1$. So, we define the $S_b \times S_b$ matrix $A(n)$ for $n = -1, 0, 1, \ldots$ as

$$[A(n-m)]_{i,j} = \hat{Q}_{(m,i),(n,j)}, \qquad m, n \geq 1, n-m \geq -1, i, j \in S_b.$$

Let $\{(X_\ell, Y_\ell)\}$ be the Markov additive process generated by the additive kernel $\{A(n)\}$. Similarly, we define $B(n)$ for $n \in \mathbb{Z}$ as

$$[B(n)]_{i,j} = \begin{cases} \hat{Q}_{(0,i),(n,j)}, & n \geq 0, \\ \hat{Q}_{(1,i),(0,j)}, & n = -1, \\ 0, & n \leq -2. \end{cases}$$

Thus, the reflecting random walk can be expressed by the Markov additive process and the boundary transitions $\{B(n)\}$.

We have constructed $\hat{Q}$ from the reflecting random walk. However, in this subsection, our arguments below do not depend on the random walk structure except for a few places. If we do not assume any special structure for the background state transitions in $\hat{Q}$, the Markov chain with transition kernel $\hat{Q}$ is referred to as a reflecting Markov additive process. In particular, it is called a quasi-birth-and-death process, QBD process for short, if the additive process is skip-free, that is, its increments are at most unit in absolute value.

We apply Pitman's identity (Lemma 5.1) to the reflecting MAP $\{(\hat{X}_\ell, \hat{Y}_\ell)\}$ with stopping time τ for each $m \geq 1$ defined by

$$\tau = \inf\{\ell \geq 1; \hat{X}_\ell \leq m-1\}.$$

Then the following corollary is immediate from (5.2).

Corollary 5.1 If $\hat{G}_\tau(1)$ has the stationary measure π_τ, that is,

$$\pi_\tau(m,i) = [\pi_\tau \hat{G}_\tau(1)](m,i), \qquad m \leq n-1, i \in S_b, \tag{5.11}$$

and if π on $\mathbb{Z}_+^d$ which is defined as

$$\pi(m,i) = [\pi_\tau \hat{H}_\tau(1)](m,i), \qquad m \geq n-1, i \in S_b, \tag{5.12}$$

is a measure, then π is the stationary measure of $\hat{Q}$.

We now assume that the reflecting random walk has the stationary distribution π. We decompose π as a sequence of vectors $\{\boldsymbol{\pi}_n\}$:

$$\boldsymbol{\pi}_n(i) = \pi(n,i), \qquad n \geq 0, i \in S_b.$$

To compute π_τ, we introduce the transition matrix $R^{0+}(n)$ defined as

$$[R^{0+}(n)]_{ij} = \sum_{\ell=1}^{\infty} \mathbb{P}(\hat{X}_\ell = n, \hat{Y}_\ell = j, 0 < \hat{X}_\ell \le \min(\hat{X}_1, \hat{X}_2, \ldots, \hat{X}_{\ell-1}) | \hat{X}_0 = 0, \hat{Y}_0 = i).$$

Using this definition and the fact that the transitions are homogeneous above or at level $n \ge 1$, we can write (5.12) with $m = n$ as

$$\boldsymbol{\pi}_n = \boldsymbol{\pi}_0 R^{0+}(n) + \sum_{k=1}^{n-1} \boldsymbol{\pi}_k R^{+}(n-k), \qquad n \ge 1. \tag{5.13}$$

Taking the transpose of this equation and using (5.9), we have

$$\boldsymbol{\pi}_n^{\mathrm{T}} = R^{0+}(n)^{\mathrm{T}} \boldsymbol{\pi}_0^{\mathrm{T}} + \sum_{k=1}^{n-1} \tilde{G}^{+}(n-k) \boldsymbol{\pi}_k^{\mathrm{T}}, \qquad n \ge 1. \tag{5.14}$$

This is a Markov renewal equation with transition kernel $\{\tilde{G}^{+}(n)\}$.

5.4 Exact geometric asymptotics

We are now ready to consider the tail asymptotics of the stationary distribution of the reflecting Markov additive process.

From (5.4) and (5.13), it follows that

$$\boldsymbol{\pi}_n = [\boldsymbol{\pi}_0 R^{0+} * H](n), \qquad n = 0, 1, \ldots. \tag{5.15}$$

This is a vector-and-matrix expression for the stationary distribution. In particular, if the additive process is skip-free, that is, the reflecting Markov additive process is a QBD process, then (5.15) can be written as

$$\boldsymbol{\pi}_n = \boldsymbol{\pi}_0 R^{0+} (R^{+})^{n-1}, \qquad n = 0, 1, \ldots. \tag{5.16}$$

This expression is well known as a matrix geometric form. If the background state space of the QBD process is finite, then matrix computations are feasible. This QBD process was firstly systematically studied by Neuts [75]. Here we do not assume that the background state space is finite, but many of the arguments are parallel to the finite case except for eigenvalues and eigenvectors. They are very hard to compute if the background state space is not finite, and we need further structure like a random walk.

In this and the next sections, we assume

(5a) The transition matrix Q of the MAP is irreducible and its additive process is 1-arithmetic (see Miyazawa and Zhao [73] for this definition).

In the view of (5.15), this is a natural condition for exact asymptotics to exist.

We further express (5.15) by a vector moment generating function. Let

$$\boldsymbol{\pi}_*(\theta) = \sum_{n=0}^{\infty} e^{\theta n} \boldsymbol{\pi}_n.$$

Then, from (5.4) and (5.15), we have

$$\boldsymbol{\pi}_*(\theta) = \boldsymbol{\pi}_0 R_*^{0+}(\theta) H_*(\theta) = \boldsymbol{\pi}_0 R_*^{0+}(\theta)(I - R_*^{+}(\theta))^{-1}. \tag{5.17}$$

From these formulas, we may see two scenarios for the asymptotics of the stationary distribution $\boldsymbol{\pi}_n$.

(5-i) The asymptotics are only determined by H. That is, the asymptotics of H dominates that of $\boldsymbol{\pi}_0 R_*^{0+}$.

(5-ii) The asymptotics are influenced by both H and $\boldsymbol{\pi}_0 R_*^{0+}$. In other words, $\boldsymbol{\pi}_0 R_*^{0+}$ controls H.

We first consider the case (5-i), and give sufficient conditions to have exact geometric asymptotics.

Assume $A_*(\theta)$ has the left and right positive invariant vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ such that $\boldsymbol{x}\boldsymbol{y} < \infty$. Define a Markov additive kernel by

$$\tilde{A}^{(\theta)}(n) = \Delta_{\boldsymbol{x}}^{-1}(e^{\theta n} A(n))^{\mathrm{T}} \Delta_{\boldsymbol{x}}.$$

Similarly, we define

$$\begin{aligned} \tilde{R}^{(\theta)-0}(n) &= \Delta_{\boldsymbol{x}}^{-1}(e^{\theta n} G^{-0}(n)^{\mathrm{T}}) \Delta_{\boldsymbol{x}}, \\ \tilde{G}^{(\theta)+}(n) &= \Delta_{\boldsymbol{x}}^{-1}(e^{\theta n} R^{+}(n))^{\mathrm{T}} \Delta_{\boldsymbol{x}}, \\ \tilde{G}^{(\theta)0+}(n) &= \Delta_{\boldsymbol{x}}^{-1}(e^{\theta n} R^{0+}(n))^{\mathrm{T}} \Delta_{\boldsymbol{x}}. \end{aligned}$$

These matrices are said to be twisted by θ.

We twist the RG decomposition (5.10), then we have

$$I - \tilde{A}^{(\theta)}(n) = [(I - \tilde{R}^{(\theta)-0}) * (I - \tilde{G}^{(\theta)+})](n). \tag{5.18}$$

It can be shown that $\tilde{G}_*^{(\theta)+}(1)$ is stochastic (positive recurrent) if and only if $\tilde{A}^{(\theta)} \equiv \tilde{A}_*^{(\theta)}(1)$ is stochastic (positive recurrent). For the stationary distribution $\boldsymbol{\pi} = (\boldsymbol{\pi}_0, \boldsymbol{\pi}_1, \ldots)$, let

$$\tilde{\boldsymbol{\pi}}_n^{(\theta)} = \Delta_{\boldsymbol{x}}^{-1}(e^{\theta n} \boldsymbol{\pi}_n^{\mathrm{T}}).$$

Then, from (5.13), we have

$$\tilde{\boldsymbol{\pi}}_n^{(\theta)} = \tilde{G}^{(\theta)0+}(n) \tilde{\boldsymbol{\pi}}_0 + \sum_{k=1}^{n-1} \tilde{G}^{(\theta)+}(k) \tilde{\boldsymbol{\pi}}_{n-k}^{(\theta)}. \tag{5.19}$$

Thus, we again have a Markov renewal equation. Define transition matrices for the background states by

$$\tilde{G}^{(\theta)+} = \sum_{n=1}^{\infty} \tilde{G}^{(\theta)+}(n), \qquad \tilde{G}^{(\theta)0+} = \sum_{n=1}^{\infty} \tilde{G}^{(\theta)0+}(n).$$

Let $\boldsymbol{\xi}^{(\theta)}$ be the left invariant positive vector of $\tilde{G}^{(\theta)+}$ if it exists. Assume that $\tilde{A}^{(\theta)}$ is positive recurrent and

$$\boldsymbol{\xi}^{(\theta)} \tilde{G}^{(\theta)0+} \tilde{\boldsymbol{\pi}}_0 < \infty,$$

and denote this θ by α. Let $\boldsymbol{r}^{(\alpha)} = \Delta_{\boldsymbol{x}}^{-1}(\boldsymbol{\xi}^{(\theta)})^{\mathrm{T}}$, then this condition is written as

$$\boldsymbol{\xi}^{(\alpha)}\tilde{G}^{(\alpha)0+}\tilde{\boldsymbol{\pi}}_0 = \boldsymbol{\pi}_0 R_*^{0+}(\alpha)\boldsymbol{r}^{(\alpha)}.$$

Then, applying the Markov renewal theorem (see, e.g., Alsmeyer [2] and Cinlar [14]) to (5.19) yields

$$\lim_{n\to\infty} e^{\alpha n}\boldsymbol{\pi}_n = \frac{1}{\beta}(\boldsymbol{\pi}_0 R_*^{0+}(\alpha)\boldsymbol{r}^{(\alpha)})\boldsymbol{x}, \tag{5.20}$$

where $\beta(\alpha) = \boldsymbol{\xi}^{(\alpha)}(\sum_{n=1}^{\infty} n\tilde{G}^{(\alpha)+}(n))\mathbf{1}$. Note that $\boldsymbol{r}^{(\alpha)} \equiv \Delta_{\boldsymbol{x}}^{-1}(\boldsymbol{\xi}^{(\alpha)})^{\mathrm{T}}$ is the right invariant vector of $R^{+}(\alpha)$. Thus, we get the following theorem.

Theorem 5.2 (Theorem 4.1 of Miyazawa and Zhao [73]) Assume that the reflecting Markov additive process has a stationary distribution and (5a) is satisfied. If there is an $\alpha > 0$ satisfying the following three conditions:

- (5b) $A_*(\alpha)$ has positive left and right invariant vectors $\boldsymbol{x}$ and $\boldsymbol{y}$,
- (5c) $A_*(\alpha)$ is positive, that is, $\langle \boldsymbol{x}, \boldsymbol{y}\rangle < \infty$ for the vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ of (5b),
- (5d) $\boldsymbol{\pi}_0 R_*^{0+}(\alpha)\boldsymbol{r}^{(\alpha)} < \infty$,

then $\boldsymbol{\pi}_n(i)$ has the exact geometric asymptotic (5.20) for each fixed $i \in S_b$ as $n \to \infty$.

This theorem does not need the background process to be a reflecting random walk. However, the three conditions are restrictive and may be hard to check. For the two-dimensional reflecting random walk, (5b) and (5c) can be checked, but there is some difficulty in verifying (5d) because it requires the stationary probabilities $\boldsymbol{\pi}_0$ on the boundary, which are unknown. Thus, its availability is limited, but there are many cases where it is still applicable (e.g., see [29, 32, 56, 94]). We remark that Theorem 5.2 does not cover even all the cases of (5-i) (see, e.g., Foley and McDonald [30]).

We next consider the case (5-ii). In this case, we have to know the tail asymptotics in different directions at once. This is generally a hard problem, and results are only known for the two-dimensional reflecting random walk. A key idea is to simultaneously consider two Markov additive processes in different directions, which are obtained from the reflecting random walk as discussed in Section 5.3, and derive certain fixed point equations. Their solutions give sufficient information on $\boldsymbol{\pi}_0$. This approach is used in Borovkov and Mogul'skii [9] and Miyazawa [68].

There is another way to verify (5d) for the two-dimensional reflecting random walk. The idea is to use the convergence domain of the moment generating function $\varphi_2(\theta)$ of $\{\boldsymbol{\pi}_0(n); n = 0, 1, \ldots\}$. Because $[\boldsymbol{y}]_i = e^{\eta i}$ for $i = 0, 1, \ldots$ is known, we have

$$\begin{aligned}
\boldsymbol{\pi}_0 R_*^{0+}(\alpha)\boldsymbol{r}^{(\alpha)} &= \boldsymbol{\pi}_0 R_*^{0+}(\alpha)(I - G_*^{0-}(\theta))\boldsymbol{y} \\
&\le \boldsymbol{\pi}_0 R_*^{0+}(\alpha)\boldsymbol{y} \\
&\le \sum_{i,j} \boldsymbol{\pi}_0(i)[B_*^{+}(\alpha)]_{ij} e^{\eta_2 j} \\
&= \sum_{i=0}^{\infty} \boldsymbol{\pi}_0(i) e^{\eta_1 j} \sum_{j=0}^{\infty} [B_*^{+}(\alpha)]_{0,j-i} e^{\eta_2 (j-i)} \\
&= \varphi_2(\eta_1) \times \text{ constant.}
\end{aligned}$$

Hence, if $\varphi_2(\eta_1)$ is finite, then (5d) is satisfied. The finiteness of $\varphi_2(\eta_1)$ would be obtained from the convergence domain of the moment generating function of the stationary distribution. Thus, the problem is reduced to finding the convergence domain, which will be discussed in Section 6.

5.5 Lower bound for the decay rate

It may be questioned whether any further tail asymptotics can be obtained in the framework of the Markov additive process. There are some ways to get different type of exact asymptotics using ratio limit theorems of Markov chain (see, e.g., Foley and McDonald [30]). However, they are still limited in use. Here we consider the problem from a different viewpoint.

We reconsider the expression (5.15) through the occupation measure $H(\ell)$. Since (5.15) implies that

$$\log \boldsymbol{\pi}_n(i) \geq \log[\boldsymbol{\pi}_0 R^{0+}(1)]_j + \log H_{ji}(n-1), \qquad i, j \in S_b,$$

we can get a lower bound for the tail decay rate of the stationary distribution if the decay rate of $H(n)$ is available. Indeed, it is shown in Theorem 4.1 of Kobayashi, Miyazawa and Zhao [53] that

$$\lim_{n\to\infty} \frac{1}{n} \log[H(n)]_{ij} = -d_H, \qquad i, j \in S_b,$$

where

$$d_H = \sup\{\theta \geq 0; H_*(\theta) < \infty\}.$$

Thus, we have

$$\liminf_{n\to\infty} \frac{1}{n} \log \boldsymbol{\pi}_n(i) \geq -d_H, \qquad i \in S_b. \tag{5.21}$$

It remains to get d_H from the modeling primitives. The following idea is standard for this (see, e.g., Ignatiouk-Robert [44]). For a square matrix A, define its convergence parameter $c_p(A)$ as

$$c_p(A) = \sup\left\{ s \geq 0; \sum_{\ell=0}^{\infty} s^\ell A^\ell < \infty \right\}.$$

Since $H_*(\theta) = (I - R_*^+(\theta))^{-1}$, d_H should be obtained from θ such that $c_p(R_*^+(\theta)) = 1$. This is equivalent to $c_p(A_*(\theta)) = 1$ from the RG factorization (5.8). Since $A_*(\theta)$ is an infinite-dimensional matrix, there are multiple θ satisfying $c_p(A_*(\theta)) = 1$. Let

$$\theta^{(c)} = \sup\{\theta \geq 0; c_p(A_*(\theta)) = 1\}.$$

Then, Theorem 4.1 of [53] shows that $d_H = \theta^{(c)}$. In computing this $\theta^{(c)}$, we usually investigate a subinvariant vector $\boldsymbol{x}$, which is a positive vector such that

$$\boldsymbol{x} A_*(\theta) \leq \boldsymbol{x}$$

It is known that the existence of this subinvariant vector is equivalent to $c_p(A_*(\theta)) = 1$ (see Nummelin [80] and Seneta [89]). This fact is compatible with the condition (5b) in Theorem 5.2.

Hence, (5.21) is written as

$$\liminf_{n\to\infty} \frac{1}{n} \log \boldsymbol{\pi}_n(i) \ge -\theta^{(c)}, \qquad i \in S_b. \tag{5.22}$$

Thus, we have a lower bound $\theta^{(c)}$ for the decay rate. In the view of the two scenarios (5-i) and (5-ii), we can expect that the exact decay rate $\theta^{(c)}$ is tight for (5-i), but not for (5-ii). However, the lower bound is still useful to find the decay rate in the other direction.

5.6 The decay rate for a QBD process

As we have discussed, it is hard to get the tail asymptotics for the reflecting Markov additive process without extra conditions (5c) and (5d). Even for the decay rate in coordinate directions, we cannot get a complete answer. However, if the additive process is skip-free, that is, the precess is QBD, then there is a way to overcome this difficulty. For this, we use the matrix geometric expression (5.16), which is rewritten as

$$\boldsymbol{\pi}_n = \boldsymbol{\pi}_1 (R^+)^{n-1}, \qquad n = 1, 2, \ldots. \tag{5.23}$$

The idea is that if $\boldsymbol{\pi}_1$ is asymptotically identical to the left eigenvector of R^+ then the eigenvalue for this eigenvector can determine the decay rate because of the matrix geometric form (5.23). The following fact is a key for this, which is obtained in Li, Miyazawa and Zhao [56].

Lemma 5.2 (Theorem 2.1 of [56]) For the discrete-time QBD process with background state space $S_b = \mathbb{Z}_+$, if there exist a positive left invariant vector $\boldsymbol{x} = (x_k)$ of $A_*(\alpha)$ for some $\alpha > 0$ and some finite $c \ge 0$ such that

$$\lim_{k\to\infty} \frac{1}{x_k} \boldsymbol{\pi}_1(k) = c, \tag{5.24}$$

then, for any nonnegative column vector $\boldsymbol{h}$ satisfying $\langle \boldsymbol{x}, \boldsymbol{h} \rangle < \infty$,

$$\lim_{n\to\infty} e^{\alpha n} \langle \boldsymbol{\pi}_n, \boldsymbol{h} \rangle = c\, e^{\alpha} \langle \boldsymbol{x}, \boldsymbol{h} \rangle. \tag{5.25}$$

In particular, if $0 < c < \infty$, then $\langle \boldsymbol{\pi}_n, \boldsymbol{h} \rangle$ decays geometrically with rate α as n goes to infinity.

Note that this result does not require the positivity assumption (5c) on $A_*(\alpha)$. Instead of this assumption, we need to find an appropriate α so that (5.24) holds, and the background state space must be totally ordered. In some special models, this works well as reported in [1, 48, 56, 96]. However, the conditions may not be easily verified since we generally do not know the tail asymptotic of $\{\boldsymbol{\pi}_1(k)\}$. However, if the background process is also a QBD, this difficulty can be overcome. This is exactly what has been done for the double QBD process in Miyazawa [69].

In what follows, we briefly introduce ideas presented in [69]. Consider the double QBD process, and generate the Markov additive process as we have done in Section 5.3. Then, the decay rates are derived in the following steps:

1. (Theorem 3.1 of [69]) Find a region for $A_*(\theta)$ to have the left and right positive invariant vectors. Compute these invariant vectors for each θ in the obtained region. Perform the same procedure for the other direction.
2. (Proposition 3.1, Corollary 3.1 of [69]) Find upper bounds for the decay rates in both directions using an extended version of Lemma 5.2.
3. (Theorem 4.1 of [69]) Derive an optimization problem to determine the decay rates using the upper bounds in Step 2.
4. (Corollary 4.1 of [69]) Solve this optimization problem and get the decay rates in both directions at once.

Here we cite results for Steps 3 and 4 from [69], which will be compared with another derivations in Sections 6 and 7. Let

$$\Gamma = \{\boldsymbol{\theta} \in \mathbb{R}^2; \gamma_{\{1,2\}}(\boldsymbol{\theta}) < 1\}, \qquad \Gamma_k = \{\boldsymbol{\theta} \in \mathbb{R}^2; \gamma_{\{k\}}(\boldsymbol{\theta}) < 1\}, \quad k = 1, 2.$$

Theorem 5.3 (Theorem 4.1 of Miyazawa [69]) For the double QBD process satisfying conditions (3c) and (3d), define α_i for $i = 1, 2$ as

$$\alpha_1 = \sup\{\theta_1; \eta_1 \le \theta_1, \theta_2 \le \eta_2, (\theta_1, \theta_2) \in \Gamma \cap \Gamma_1, (\eta_1, \eta_2) \in \Gamma \cap \Gamma_2\}, \tag{5.26}$$
$$\alpha_2 = \sup\{\eta_2; \eta_1 \le \theta_1, \theta_2 \le \eta_2, (\theta_1, \theta_2) \in \Gamma \cap \Gamma_1, (\eta_1, \eta_2) \in \Gamma \cap \Gamma_2\}. \tag{5.27}$$

Then, α_1 and α_2 are the decay rates of $\boldsymbol{\pi}_n(i)$ as $n \to \infty$ for each fixed $i \ge 0$.

Corollary 5.2 (Corollary 4.1 of Miyazawa [69]) For $k = 1, 2$, let

$$\boldsymbol{\theta}^{(k,c)} = \arg_{\boldsymbol{\theta} \in \mathbb{R}^2} \sup\{\theta_1 \ge 0; \boldsymbol{\theta} \in \Gamma \cap \Gamma_k\},$$

then the solution of (5.26) and (5.27) is obtained as

$$\alpha_1 = \begin{cases} \theta_1^{(1,c)}, & \theta_2^{(1,c)} < \theta_2^{(2,c)}, \\ \overline{\xi}_1(\theta_2^{(2,c)}), & \theta_2^{(1,c)} \ge \theta_2^{(2,c)}, \end{cases} \qquad \alpha_2 = \begin{cases} \theta_2^{(2,c)}, & \theta_1^{(2,c)} < \theta_1^{(2,c)}, \\ \overline{\xi}_2(\theta_1^{(1,c)}), & \theta_1^{(2,c)} \ge \theta_1^{(2,c)}, \end{cases} \tag{5.28}$$

where $\overline{\xi}_k(\theta_{3-k}) = \sup\{\theta_k; \gamma(\theta_1, \theta_2) = 1\}$ for $k = 1, 2$.

We will see that these α_1 and α_2 exactly correspond to τ_1 and τ_2 of (6.13) in Lemma 6.8 for a more general reflecting random walk. This means that the optimization problems (5.26) and (5.27) in Step 3 can be reduced to the fixed point problem with equation (6.12). This fixed point equation has also been used to combine the asymptotics of two Markov additive processes in the coordinate directions in Borovkov and Mougl'skii [9]. However, they neither explicitly obtained the fixed point equation nor solved it. This is the drawback of their general modeling assumptions (see [69] for more discussions on this issue).

6 Domain for the analytic function approach

In this and next sections, we discuss the analytic function approach using the convergence domain for the reflecting random walk $\{\boldsymbol{Z}_\ell\}$ defined in Section 3.2. We will obtain the tail asymptotics only for $d = 2$, but we start with a general $d \geq 2$ to see how our framework works. For this, we assume (3b), (3b'), (3c), and (3d). That is, the stationary distribution of the reflecting random walk uniquely exists, and the distributions of all increments at each time have light tails.

In this section, we consider the domain $\mathcal{D}$ of the moment generating function of the stationary distribution, then the tail asymptotics will be considered in Section 7.

6.1 Stationary inequalities

We first observe two key facts. We like to use the stationary equation (3.11) for finding the convergence domain of the moment generating function $\varphi(\boldsymbol{\theta})$, but it is valid only when $\varphi(\boldsymbol{\theta})$ is finite. This is something like a circular argument. Here we need a clue to expand the region of those $\boldsymbol{\theta} \in \mathbb{R}^d$ for which $\varphi(\boldsymbol{\theta})$ is known to be finite. The next lemma gives us this clue. We recall that $J = \{1, 2, \ldots, d\}$, and denote the set of all subsets of J by 2^J.

Lemma 6.1 For the reflecting random walk $\{\boldsymbol{Z}_\ell\}$ defined in Section 3.2, assume the conditions (3b), (3b'), (3c), and (3d). For $\mathcal{C} \subset 2^J$, if the following two conditions hold:

$$\gamma_A(\boldsymbol{\theta}) < 1 \quad \text{for all } A \in \mathcal{C}, \tag{6.1}$$

$$\varphi_A(\boldsymbol{\theta}) < \infty \quad \text{for all } A \in 2^J \setminus \mathcal{C}, \tag{6.2}$$

then

$$0 \leq \sum_{A \in \mathcal{C}} (1 - \gamma_A(\boldsymbol{\theta}))\varphi_A(\boldsymbol{\theta}) \leq \sum_{A \in 2^J \setminus \mathcal{C}} (\gamma_A(\boldsymbol{\theta}) - 1)\varphi_A(\boldsymbol{\theta}) < \infty, \tag{6.3}$$

and therefore $\varphi_A(\boldsymbol{\theta}) < \infty$ for all $A \in \mathcal{C}$, and $\varphi(\boldsymbol{\theta}) < \infty$.

PROOF. We apply truncation arguments for the stationary equation (3.7). For each $n = 1, 2, \ldots$, let

$$f_n(x) = \min(x, n), \qquad x \in \mathbb{R}.$$

It is not hard to see that, for any $x \geq 0$ and $y \in \mathbb{R}$,

$$f_n(x + y) \leq f_n(x) + \begin{cases} y, & x \leq n, \\ 0, & x > n. \end{cases} \tag{6.4}$$

Hence, we have, using the independence of $\boldsymbol{Z}$ and $\boldsymbol{X}^A$,

$$\begin{aligned} E(e^{f_n(\langle \boldsymbol{\theta}, (\boldsymbol{Z}+\boldsymbol{X}^A) \rangle)} 1(\boldsymbol{Z} \in S_A)) \;\leq\;& E(e^{f_n(\langle \boldsymbol{\theta}, \boldsymbol{Z} \rangle)} 1(\boldsymbol{Z} \in S_A, \langle \boldsymbol{\theta}, \boldsymbol{Z} \rangle \leq n)) E(e^{f_n(\langle \boldsymbol{\theta}, \boldsymbol{X}^A \rangle)}) \\ &+ E(e^{f_n(\langle \boldsymbol{\theta}, \boldsymbol{Z} \rangle)} 1(\boldsymbol{Z} \in S_A, \langle \boldsymbol{\theta}, \boldsymbol{Z} \rangle > n)). \end{aligned}$$

On the other hand, it follows from (3.7) that

$$e^{f_n(\langle \boldsymbol{\theta}, \boldsymbol{Z} \rangle)} \simeq \sum_{A \subset J} e^{f_n(\langle \boldsymbol{\theta}, \boldsymbol{Z} \rangle + \langle \boldsymbol{\theta}, \boldsymbol{X}^A \rangle)} 1(\boldsymbol{Z} \in S_A).$$

Taking the expectation of this distributional equation and applying the above inequality, we have

$$\begin{aligned} E(e^{f_n(\langle \boldsymbol{\theta}, \boldsymbol{Z} \rangle)}) &= \sum_{A \subset J} E(e^{f_n(\langle \boldsymbol{\theta}, \boldsymbol{Z} \rangle + \langle \boldsymbol{\theta}, \boldsymbol{X}^A \rangle)} 1(\boldsymbol{Z} \in S_A)) \\ &\le \sum_{A \subset J} E(e^{f_n(\langle \boldsymbol{\theta}, \boldsymbol{Z} \rangle)} 1(\boldsymbol{Z} \in S_A, \langle \boldsymbol{\theta}, \boldsymbol{Z} \rangle \le n)) E(e^{f_n(\langle \boldsymbol{\theta}, \boldsymbol{X}^A \rangle)}) \\ &\quad + \sum_{A \subset J} E(e^{f_n(\langle \boldsymbol{\theta}, \boldsymbol{Z} \rangle)} 1(\boldsymbol{Z} \in S_A, \langle \boldsymbol{\theta}, \boldsymbol{Z} \rangle > n)). \end{aligned} \tag{6.5}$$

Substituting

$$E(e^{f_n(\langle \boldsymbol{\theta}, \boldsymbol{Z} \rangle)}) = \sum_{A \subset J} E(e^{f_n(\langle \boldsymbol{\theta}, \boldsymbol{Z} \rangle)} 1(\boldsymbol{Z} \in S_A))$$

into (6.5) and rearranging terms, we have

$$\begin{aligned} &\sum_{A \in \mathcal{C}} (1 - E(e^{f_n(\langle \boldsymbol{\theta}, \boldsymbol{X}^A \rangle)})) E(e^{f_n(\langle \boldsymbol{\theta}, \boldsymbol{Z} \rangle)} 1(\boldsymbol{Z} \in S_A, \langle \boldsymbol{\theta}, \boldsymbol{Z} \rangle \le n)) \\ &\qquad \le \sum_{A \in 2^J - \mathcal{C}} (E(e^{f_n(\langle \boldsymbol{\theta}, \boldsymbol{X}^A \rangle)}) - 1) E(e^{f_n(\langle \boldsymbol{\theta}, \boldsymbol{Z} \rangle)} 1(\boldsymbol{Z} \in S_A, \langle \boldsymbol{\theta}, \boldsymbol{Z} \rangle \le n)). \end{aligned}$$

Let n go to infinity in this inequality, then conditions (6.1)–(6.2) and the monotone convergence theorem yield (6.3) since $f_n(x)$ is nondecreasing in x. □

We next consider bounding the domain $\mathcal{D}$. For this, let

$$\Gamma_{\max} = \{\boldsymbol{\theta} \in \mathbb{R}^d; \boldsymbol{\theta} \le \exists \boldsymbol{\theta}', \gamma_J(\boldsymbol{\theta}') < 1\}.$$

The following lemma shows that $\mathcal{D}$ is upper bounded by $\Gamma_{\max}$. Its proof for $d = 2$ can be found in Kobayashi and Miyazawa [51] (see Lemma 3.3 there), and a less complete proof for a general d is found in Borovkov [6]. Since the proof in [51] can be easily adapted for a general d, we omit a proof of the following lemma.

Lemma 6.2 For the reflecting random walk $\{\boldsymbol{Z}_\ell\}$ defined in Section 3.2, assume the conditions (3b), (3b'), (3c), and (3d). For any direction vector $\boldsymbol{c} > \boldsymbol{0}$ and a nonempty open set $B \subset \mathbb{R}^d_+$,

$$\liminf_{n \to \infty} \frac{1}{n} \log \mathbb{P}(\boldsymbol{Z} \in n\boldsymbol{c} + B) \ge -\sup\{\langle \boldsymbol{\theta}, \boldsymbol{c} \rangle; \gamma_J(\boldsymbol{\theta}) \le 1\}, \tag{6.6}$$

and therefore $\mathcal{D} \subset \Gamma_{\max}$.

Note that $\Gamma_{\max}$ is bounded from above, that is, there is a $\boldsymbol{\theta}' \in \mathbb{R}^d$ such that $\boldsymbol{\theta} \le \boldsymbol{\theta}'$ for all $\boldsymbol{\theta} \in \Gamma_{\max}$ because of (3b'). Also note that, if $E(X_i^J) = 0$ for all $i \in J$, then

$$\{\boldsymbol{\theta} \in \mathbb{R}^d; \gamma_J(\boldsymbol{\theta}) \le 1\} = \{\mathbf{0}\},$$

by the convexity of function $\gamma_J(\boldsymbol{\theta})$ and $\gamma_J(\mathbf{0}) = 1$. Hence, the right hand side of (6.6) equals 0, and we immediately have the following facts from Lemma 6.2.

Lemma 6.3 For the reflecting random walk satisfying the same assumptions as Lemma 6.2, (a) the stationary distribution cannot have a small tail in any direction (see Definition 2.4 for the small tail; (b) If $E(X_i^J) = 0$ for all $i \in J$, then the stationary distribution has a heavy tails in all directions.

6.2 A program for identifying the convergence domain

Recall that the domain $\mathcal{D}$ is defined as

$$\mathcal{D} = \text{the interior of } \{\boldsymbol{\theta} \in \mathbb{R}^d; \varphi(\boldsymbol{\theta}) < \infty\}.$$

We consider an iterative algorithm to find this domain. For this, we use Lemma 6.1, and need some notations. Let 2^J be the set of all subsets of $J \equiv \{1, 2, \dots, d\}$ including the empty set. For each $A \in 2^J$, we let

$$\Gamma_A = \{\boldsymbol{\theta} \in \mathbb{R}^d; \gamma_A(\boldsymbol{\theta}) < 1\}.$$

For convenience, we often write $\gamma_J(\boldsymbol{\theta})$ and Γ_J as $\gamma(\boldsymbol{\theta})$ and Γ, respectively. Note that Γ is a bounded convex set by (3c). However, we cannot use $\varphi(\boldsymbol{\theta})$ for $\varphi_J(\boldsymbol{\theta})$ since they are different. We define $\boldsymbol{\theta}_A$ as the d-dimensional vector whose ith entry is θ_i for $i \in A$ and vanishes for $i \in J \setminus A$ for $\boldsymbol{\theta} \equiv (\theta_1, \theta_2, \dots, \theta_d)$. Note that $\varphi_A(\boldsymbol{\theta}) = \varphi_A(\boldsymbol{\theta}_A)$.

Let $\mathcal{C}$ be an arbitrary collection of subsets of J. That is, $\mathcal{C}$ is a subset of 2^J. We allow $\mathcal{C}$ to be the empty set. Let $\mathcal{G}$ be an arbitrary subset of $\mathbb{R}^d$. For these $\mathcal{C}$ and $\mathcal{G}$, we define

$$\mathcal{D}_{\mathcal{C}}(\mathcal{G}) = \{\boldsymbol{\theta} \in \cap_{A \in \mathcal{C}} \Gamma_A; \forall B \in 2^J \setminus \mathcal{C}, \exists \boldsymbol{\eta} \in \mathcal{G}, \boldsymbol{\theta}_B < \boldsymbol{\eta}_B\},$$

which is an open set, where $\cap_{B \in \emptyset} \Gamma_B = \mathbb{R}^d$. By this definition and Lemma 6.1, if $\mathcal{G} \subset \mathcal{D}$, then $\mathcal{D}_{\mathcal{C}}(\mathcal{G}) \subset \mathcal{D}$. We also note that

$$\begin{aligned} \mathcal{D}_{\emptyset}(\mathcal{G}) &= \{\boldsymbol{\theta} \in \mathbb{R}^d; \forall B \in 2^J, \exists \boldsymbol{\eta} \in \mathcal{G}, \boldsymbol{\theta}_B < \boldsymbol{\eta}_B\} \\ &= \cup_{\boldsymbol{\eta} \in \mathcal{G}} \{\boldsymbol{\theta} \in \mathbb{R}^d; \boldsymbol{\theta} < \boldsymbol{\eta}\}. \end{aligned}$$

Denote the convex hull of a subset $\mathcal{A}$ of $\mathbb{R}^d$ by $\mathrm{conv}(\mathcal{A})$. Then, $\mathcal{G} \subset \mathcal{D}$ implies that

$$\mathcal{G} \subset \mathrm{conv}(\cup_{\mathcal{C} \subset 2^J} \mathcal{D}_{\mathcal{C}}(\mathcal{G})) \subset \mathcal{D}.$$

This suggests iteratively using a mapping from $\mathcal{G}$ to $\mathrm{conv}(\cup_{\mathcal{C} \subset 2^J} \mathcal{D}_{\mathcal{C}}(\mathcal{G}))$.

We let $\mathcal{G}_0 = \{\boldsymbol{\theta} \in \mathbb{R}^d; \boldsymbol{\theta} < \mathbf{0}\}$, and inductively define, for $n = 0, 1, \dots$,

$$\mathcal{G}_{n+1} = \mathrm{conv}(\cup_{\mathcal{C} \subset 2^J} \mathcal{D}_{\mathcal{C}}(\mathcal{G}_n)). \tag{6.7}$$

Clearly, $\mathcal{G}_0 \subset \mathcal{D}$ and $\mathcal{G}_n$ increases in n. Hence,

$$\mathcal{G}_\infty \equiv \lim_{n\to\infty} \mathcal{G}_n$$

exists, and $\mathcal{G}_\infty \subset \mathcal{D}$. It is easy to see that $\mathcal{G}_\infty$ is an open set and a solution of the following fixed set equation:

$$\mathcal{G} = \mathrm{conv}(\cup_{\mathcal{C}\subset 2^J} \mathcal{D}_{\mathcal{C}}(\mathcal{G})). \tag{6.8}$$

We summarize the above arguments in as a theorem.

Theorem 6.1 For the reflecting random walk $\{\boldsymbol{Z}_\ell\}$ defined in Section 3.2, assume the conditions (3b), (3b'), (3c) and (3d), then we have

(6-i) $\mathcal{D}$ is the solution of (6.8).

(6-ii) $\mathcal{G}_\infty$ is the minimal solution of (6.8) such that $\boldsymbol{\theta} \in \mathcal{G}_\infty$ for all $\boldsymbol{\theta} < \boldsymbol{0}$.

(6-iii) $\mathcal{G}_\infty$ and $\mathcal{D}$ are convex open sets.

(6-iv) $\mathcal{G}_\infty \subset \mathcal{D} \subset \Gamma_{\max}$.

Hence, the stationary distribution has light tails in all directions if $\boldsymbol{0} \in \mathcal{G}_\infty$.

Remark 6.1 (6-ii) and (6-iii) are valid without conditions (3c) and (3d) since nothing are involved with the stationary distribution. We only need (3b) and (3b') for them.

Based on this lemma, we propose either computing $\mathcal{G}_\infty$ or to find the minimal solution of the fixed set equation (6.8). In particular, we conjecture the following claim.

Conjecture 6.1 Under the assumptions of Theorem 6.1, $\mathcal{G}_\infty = \mathcal{D}$.

In view of Remark 6.1, we also have another conjecture.

Conjecture 6.2 Under the assumptions of (3b) and (3b'), the stationary distribution with a light tail exists if and only if $\boldsymbol{0} \in \mathcal{G}_\infty$.

Both conjectures hold true for $d = 2$ under certain skip-free conditions as we will see in Sections 6.3 and 6.4. Furthermore, $\mathcal{D}$ is explicitly obtained using extreme points of Γ_A' in this case. This suggests a similar characterization for $d \geq 3$. These ideas are also considered for the multidimensional SRBM in Miyazawa and Kobayashi [70]. However, they remain as conjectures for $d \geq 3$.

Once the domain $\mathcal{D}$ is identified, we can get the following upper bound for a similar but slightly different tail set from that of (2.10). Let $\boldsymbol{c} \geq 0$ be a direction vector, and let B be a bounded measurable subset of $\mathbb{R}^d_+$. Then, we have

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\boldsymbol{Z} \in n\boldsymbol{c} + B) \leq -\sup\{\langle \boldsymbol{\theta}, \boldsymbol{c}\rangle; \boldsymbol{\theta} \in \mathcal{D}\}. \tag{6.9}$$

The proof of this upper bound for $d = 2$ is given for the two-dimensional SRBM in [16] and the two-dimensional reflecting random walk in [51]. We note that their proofs can be used for a general $d \geq 3$.

6.3 Light tail conditions for $d = 2$

In the previous section, we proposed a program to find the domain $\mathcal{D}$ for the d-dimensional reflecting random walk. We will show that it indeed works for $d = 2$. For this, we assume (3b), (3b') and (3c), but do not assume (3d). Instead of it, we use stability conditions given below, which turns out to be necessary and sufficient for the stationary distribution to have a light tail.

For $d = 2$, the stability is completely characterized using the expectations of the increments in Fayolle, Malyshev and Menshikov [23]. To describe these stability conditions, we introduce some notation. Let, for $i = 1, 2$,

$$m_i = \mathbb{E}(X_i^{\{1,2\}}), \qquad m_i^{(1)} = \mathbb{E}(X_i^{\{1\}}), \qquad m_i^{(2)} = \mathbb{E}(X_i^{\{2\}}).$$

Define the vectors

$$\begin{aligned} &\boldsymbol{m} = (m_1, m_2), \qquad \boldsymbol{m}^{(1)} = (m_1^{(1)}, m_2^{(1)}), \qquad \boldsymbol{m}^{(2)} = (m_1^{(2)}, m_2^{(2)}), \\ &\boldsymbol{m}_\perp^{(1)} = (m_2^{(1)}, -m_1^{(1)}), \qquad \boldsymbol{m}_\perp^{(2)} = (-m_2^{(2)}, m_1^{(2)}). \end{aligned}$$

Obviously, $\boldsymbol{m}_\perp^{(k)}$ is orthogonal to $\boldsymbol{m}^{(k)}$ for each $k = 1, 2$. Note that $\boldsymbol{m}$ is orthogonal to the tangent of the convex curve $\gamma(\boldsymbol{\theta}) \equiv \gamma_{\{1,2\}}(\boldsymbol{\theta}) = 1$ at the origin. Similarly, $\boldsymbol{m}^{(k)}$ is orthogonal to the tangent of the convex curve $\gamma_{\{k\}}(\boldsymbol{\theta}) = 1$ at the origin. Note that $m_2^{(1)} \geq 0$ and $m_1^{(2)} \geq 0$ because $X_2^{\{1\}} \geq 0$ and $X_1^{\{2\}} \geq 0$; see Figure 3.

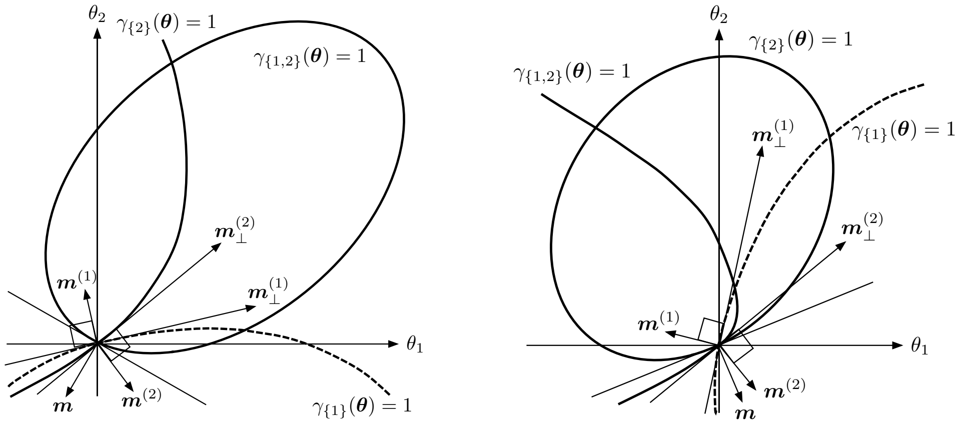

Figure 3: Tangent hyperplanes and orthogonal vectors for (6a) and (6b)

In Fayolle, Malyshev and Menshikov [23], the stability conditions are separately studied for $\boldsymbol{m} = \mathbf{0}$ and $\boldsymbol{m} \neq \mathbf{0}$. When $\boldsymbol{m} = \mathbf{0}$, the stationary distribution has a heavy tail in all directions by Lemma 6.3. Thus, we assume $\boldsymbol{m} \neq \mathbf{0}$ for light tail. Unfortunately, Theorem 3.3.1 of [23] for this case is incorrect. That is, it misses one case. We correct it as follows.

Lemma 6.4 (Corrected Theorem 3.3.1 of Fayolle, Malyshev and Menshikov [23]) If $\boldsymbol{m} \neq \mathbf{0}$, then the two-dimensional reflecting random walk $\{\boldsymbol{Z}_\ell\}$ has the stationary distribution if and only if one of the following three conditions holds:

(6a) $m_1 < 0, m_2 < 0$, $\langle \boldsymbol{m}, \boldsymbol{m}_\perp^{(1)} \rangle < 0$, and $\langle \boldsymbol{m}, \boldsymbol{m}_\perp^{(2)} \rangle < 0$;

(6b) $m_1 \geq 0, m_2 < 0$, $\langle \boldsymbol{m}, \boldsymbol{m}_\perp^{(1)} \rangle < 0$, and $m_2^{(2)} < 0$ for $m_1^{(2)} = 0$;

(6c) $m_1 < 0, m_2 \geq 0$, $\langle \boldsymbol{m}, \boldsymbol{m}_\perp^{(2)} \rangle < 0$ and $m_1^{(1)} < 0$ for $m_2^{(1)} = 0$.

Remark 6.2 The last conditions in (6b) and (6c) are obviously required for the stability since $m_1^{(2)} = 0$ ($m_2^{(1)} = 0$) implies that $P(X_1^{(2)} = 0) = 1$ ($P(X_2^{(1)} = 0) = 1$, respectively). However, they are missing in Theorem 3.3.1 of [23].

The conditions in Lemma 6.4 have geometric interpretations. First, $\langle \boldsymbol{m}, \boldsymbol{m}_\perp^{(1)} \rangle < 0$ means that $\boldsymbol{m}_\perp^{(1)}$ is above the hyperplane which is orthogonal to $\boldsymbol{m}$. This implies that $\Gamma \cap \Gamma_{\{1\}} (\equiv \Gamma_J \cap \Gamma_{\{1\}})$ contains a vector $\boldsymbol{\theta}$ such that $\theta_1 > 0$. It also contains a vector $\boldsymbol{\theta}$ such that $\theta_1 > 0$ and $\theta_2 \leq 0$ if $m_1 < 0$. $\langle \boldsymbol{m}, \boldsymbol{m}_\perp^{(2)} \rangle < 0$ has a similar interpretation (see the left picture of Figure 3). On the other hand, if $m_1 \geq 0$ and $m_2 < 0$, then $\Gamma \cap \Gamma_{\{2\}}$ always contains a vector $\boldsymbol{\theta}$ such that $\langle \boldsymbol{\theta}, \boldsymbol{m}^{(2)} \rangle < 0$ and $\theta_2 > 0$ (see the right picture of Figure 3). Note that $\boldsymbol{\theta} \in \Gamma \cap \Gamma_{\{k\}}$ implies $\langle \boldsymbol{\theta}, \boldsymbol{m} \rangle < 0$ and $\langle \boldsymbol{\theta}, \boldsymbol{m}^{(k)} \rangle < 0$ for $k = 1, 2$. These arguments conclude the following lemma, which is formally proved in [51].

Lemma 6.5 (Lemma 2.2 of [51]) Either one of the stability conditions of Lemma 6.4 holds if and only if $\gamma(\boldsymbol{\theta}) = 1$ and $\gamma_k(\boldsymbol{\theta}) = 1$ has a solution $\boldsymbol{\theta}$ such that $\theta_k > 0$ for each $k = 1, 2$. Furthermore, at least for either one of $k = 1, 2$, there exists a $\boldsymbol{\theta} \in \Gamma \cap \Gamma_k$ such that $\theta_k > 0$ and $\theta_{3-k} < 0$.

We are now ready to determine the tail type of the stationary distribution.

Theorem 6.2 For the two-dimensional reflecting random walk $\{\boldsymbol{Z}_\ell\}$ satisfying the conditions (3b), (3b') and (3c), the following conditions are equivalent:

(6d) The stationary distribution of this process exists and has a light tail.

(6e) Either one of the three conditions (6a), (6b) and (6c) holds.

Proof. We first prove that (6e) implies (6d). The existence of the stationary distribution is immediate from Lemma 6.4. We need to show that $\varphi(\boldsymbol{\theta}) < \infty$ for some $\boldsymbol{\theta} > \boldsymbol{0}$. For this, we apply operation (6.7). By Lemma 6.5, $\mathcal{G}_0$ is not empty, and we can find $\boldsymbol{\theta} \equiv (\theta_1, \theta_2) \in \mathcal{G}_1$ such that either $\theta_1 > 0, \theta_2 \leq 0$ or $\theta_1 \leq 0, \theta_2 > 0$. We then repeat the operation, and find a point $\boldsymbol{\theta} > \boldsymbol{0}$ in $\mathcal{G}_2$. This point must be in $\mathcal{D}$ by Theorem 6.1. This proves the claim. We next prove the reverse direction. Since the stationary distribution exists, by Lemma 6.4, we have (6e) if $\boldsymbol{m} \neq \boldsymbol{0}$. If $\boldsymbol{m} = \boldsymbol{0}$, then the stationary distribution has a heavy tail by Lemma 6.2. This contradicts (6d). Thus, the converse is proved. □

6.4 The convergence domain for $d = 2$

We now consider the convergence domain $\mathcal{D}$ of the moment generating function $\varphi(\boldsymbol{\theta})$. For this, we introduce the following shorthand notation:

$$\begin{aligned}
&\varphi_+ = \varphi_J, \qquad && \varphi_k = \varphi_{\{k\}}, \qquad && k = 0, 1, 2,\\
&\gamma = \gamma_J, && \gamma_k = \gamma_{\{k\}}, && k = 0, 1, 2,\\
&\Gamma = \Gamma_J, && \Gamma_k = \Gamma_{\{k\}}, && k = 1, 2.
\end{aligned}$$

We also use the following notation for the boundaries of Γ and Γ_k.

$$\partial\Gamma = \{\boldsymbol{\theta} \in \mathbb{R}^2; \gamma(\boldsymbol{\theta}) = 1\}, \qquad \partial\Gamma_k = \{\boldsymbol{\theta} \in \mathbb{R}^2; \gamma_k(\boldsymbol{\theta}) = 1\}, \quad k = 1, 2.$$

By Theorem 6.2 and Lemma 6.5, Γ contains a $\boldsymbol{\theta} > \mathbf{0}$ such that $\varphi(\boldsymbol{\theta}) < \infty$, and both of Γ_1 and Γ_2 are not empty. From the conditions (3b), (3b') and (3c), Γ is a bounded convex set, while Γ_1 is convex and bounded from above but unbounded from below in the second component since the second component of $\boldsymbol{X}^{\{1\}}$ is nonnegative. Similarly, Γ_2 is convex and bounded from above but unbounded from below in the first component.

The decay rates will be determined through extreme points of the bounded convex sets $\Gamma \cap \Gamma_k$. For this, it will be convenient to use the following notations for $k = 1, 2$.

$$\boldsymbol{\theta}^{(k,\max)} = \arg \max_{(\theta_1,\theta_2)} \{\theta_k; \gamma(\theta_1, \theta_2) = 1\}, \quad \boldsymbol{\theta}^{(k,\min)} = \arg \min_{(\theta_1,\theta_2)} \{\theta_k; \gamma(\theta_1, \theta_2) = 1\}, \quad (6.10)$$

$$\boldsymbol{\theta}^{(k,\mathrm{c})} = \arg \sup_{(\theta_1,\theta_2)} \{\theta_k; \boldsymbol{\theta} \in \Gamma \cap \Gamma_k\}, \qquad \boldsymbol{\theta}^{(k,\mathrm{e})} = \arg \sup_{(\theta_1,\theta_2)} \{\theta_k; \boldsymbol{\theta} \in \partial\Gamma \cap \partial\Gamma_k\}. \qquad (6.11)$$

From the definitions, it is easy to see that, for $k = 1, 2$,

$$\boldsymbol{\theta}^{(k,\mathrm{c})} = \begin{cases} \boldsymbol{\theta}^{(k,\mathrm{e})}, & \gamma_k(\boldsymbol{\theta}^{(k,\max)}) > 1, \\ \boldsymbol{\theta}^{(k,\max)}, & \gamma_k(\boldsymbol{\theta}^{(k,\max)}) \le 1. \end{cases}$$

We will also use the following functions:

$$\begin{aligned} \overline{\xi}_1(\theta_2) &= \max\{\theta; \gamma(\theta, \theta_2) = 1\}, & \overline{\xi}_2(\theta_1) &= \max\{\theta; \gamma(\theta_1, \theta) = 1\}, \\ \underline{\xi}_1(\theta_2) &= \min\{\theta; \gamma(\theta, \theta_2) = 1\}, & \underline{\xi}_2(\theta_1) &= \min\{\theta; \gamma(\theta_1, \theta) = 1\}. \end{aligned}$$

The following lemmas are keys for our arguments, which are immediate from Lemma 6.1.

Lemma 6.6 If $\boldsymbol{\theta} \in \mathbb{R}^2$ satisfies the condition that $\boldsymbol{\theta} \in \Gamma$ and $\varphi_k(\theta_k) < \infty$ for $k = 1, 2$, then $\varphi(\boldsymbol{\theta})$ is finite.

Lemma 6.7 For each $k = 1, 2$, choose $\boldsymbol{\theta} \in \Gamma \cap \Gamma_k$ such that $\varphi_{3-k}(\theta_{3-k})$ is finite, then $\varphi(\boldsymbol{\theta})$ and $\varphi_k(\theta_k)$ are finite.

We can execute the program proposed in Section 6.2 by producing $\mathcal{G}_n$ inductively. Instead of doing so, we here inductively produce a sequence of points which converges to extreme points of the closure of $\mathcal{G}_\infty$ following Kobayashi and Miyazawa [51]. Our aim is to show how these points are related to $\mathcal{G}_n$ with the help of Lemmas 6.6 and 6.7.

Let $\Gamma_k^{(0)} = \{\mathbf{0}\}$ for $k = 1, 2$. Obviously, $\mathbf{0}$ is the extreme point of $\mathcal{G}_0$. Let $\Gamma_k^{(1)} = \{\boldsymbol{\theta} \in \Gamma \cap \Gamma_k; \theta_{3-k} \le 0\}$ for $k = 1, 2$. Obviously, $\varphi_{3-k}(\theta_{3-k})$ is finite. Hence, by Lemma 6.7, $\varphi(\boldsymbol{\theta})$ and $\varphi_k(\theta_k)$ are finite for $\boldsymbol{\theta} \in \Gamma_k^{(1)}$. Thus, $\varphi(\boldsymbol{\theta})$, $\varphi_1(\theta_1)$ and $\varphi_2(\theta_2)$ are finite for $\boldsymbol{\theta} \in \Gamma_1^{(1)} \bigcup \Gamma_2^{(1)}$. We define $\boldsymbol{\theta}_\triangle^{(1)} \equiv (\theta_1^{(1,1)}, \theta_2^{(2,1)})$ by

$$\theta_k^{(k,1)} = \sup\{\theta_k; (\theta_1, \theta_2) \in \Gamma_k^{(1)}\}, \qquad k = 1, 2.$$

Here we use subscript k in addition to superscript k because it will be used as the kth entry of the two-dimensional vector $\boldsymbol{\theta}^{(k,1)}$, where $\theta_{3-k}^{(k,1)} = \underline{\xi}_{3-k}(\theta_k^{(k,1)})$. It is easy to see that at least one of $\theta_1^{(1,1)}$ and $\theta_2^{(2,1)}$ is positive by Lemma 6.5, and

$$\boldsymbol{\theta}^{(1,1)}, \boldsymbol{\theta}^{(2,1)} \text{ are extreme points of } \overline{\mathcal{G}}_1,$$

where $\overline{\mathcal{G}}_1$ is the closure of $\mathcal{G}_1$. We inductively define $\boldsymbol{\theta}_{\triangle}^{(n)} = (\theta_1^{(1,n)}, \theta_2^{(2,n)})$ for $n \geq 1$ by

$$\theta_k^{(k,n)} = \sup\{\theta_k; \boldsymbol{\theta} \in \Gamma \cap \Gamma_k, \theta_{3-k} \leq \theta_{3-k}^{(3-k,n-1)}\},$$

and let

$$\boldsymbol{\theta}^{(1,n)} = (\theta_1^{(1,n)}, \underline{\xi}_2(\theta_1^{(1,n)})), \qquad \boldsymbol{\theta}^{(2,n)} = (\underline{\xi}_1(\theta_2^{(2,n)}), \theta_2^{(1,n)}).$$

It is easy to see that $\boldsymbol{\theta}^{(1,n)}, \boldsymbol{\theta}^{(2,n)}$ are extreme points of $\overline{\mathcal{G}}_n$.

Then, $\boldsymbol{\theta}_{\triangle}^{(n)}$ is nondecreasing in n, and $\boldsymbol{\theta}_{\triangle}^{(n)} \leq \boldsymbol{\theta}^{\max}$ from our definition. Thus, the sequence $\boldsymbol{\theta}_{\triangle}^{(n)}$ converges to a finite vector. Denote this limit by $\boldsymbol{\tau} \equiv (\tau_1, \tau_2)$. We can see that

$$\tau_k = \sup\{\theta_k; \boldsymbol{\theta} \in \Gamma \cap \Gamma_k, \theta_{3-k} \leq \tau_{3-k}\}, \qquad k = 1, 2. \tag{6.12}$$

This can be considered as a fixed point equation. We illustrate these iterations in Figure 4.

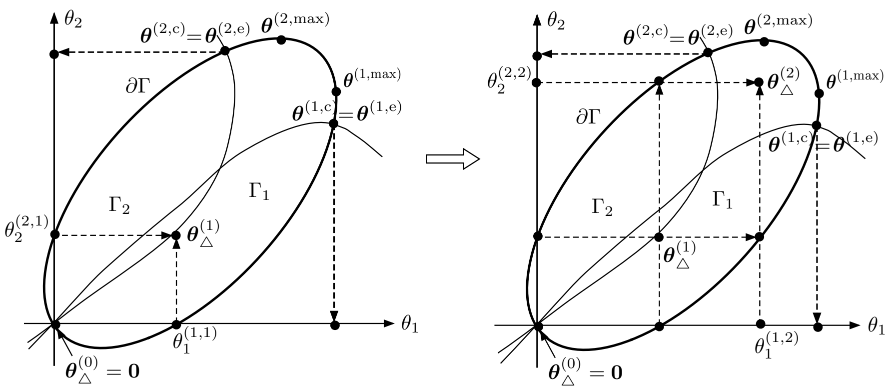

Figure 4: The first two steps of the iterations

To solve (6.12), the following classifications will be convenient:

$$\text{(D1) } \theta_1^{(2,c)} < \theta_1^{(1,c)} \text{ and } \theta_2^{(1,c)} < \theta_2^{(2,c)}, \qquad \text{(D2) } \boldsymbol{\theta}^{(2,c)} \leq \boldsymbol{\theta}^{(1,c)}, \qquad \text{(D3) } \boldsymbol{\theta}^{(1,c)} \leq \boldsymbol{\theta}^{(2,c)}.$$

Note that it is impossible to have $\theta_1^{(2,c)} > \theta_1^{(1,c)}$ and $\theta_2^{(1,c)} > \theta_2^{(2,c)}$. The following solution is obtained for (6.12) in [51].

Lemma 6.8 (Lemma 3.1 of Kobayashi and Miyazawa [51]) The limit $\boldsymbol{\tau}$ of the sequence $\boldsymbol{\theta}_{\triangle}^{(n)}$ is given by

$$(\tau_1, \tau_2) = \begin{cases} (\theta_1^{(1,c)}, \theta_2^{(2,c)}) & \text{if (D1) holds,} \\ (\overline{\xi}_1(\theta_2^{(2,c)}), \theta_2^{(2,c)}) & \text{if (D2) holds,} \\ (\theta_1^{(1,c)}, \overline{\xi}_2(\theta_1^{(1,c)})) & \text{if (D3) holds.} \end{cases} \tag{6.13}$$

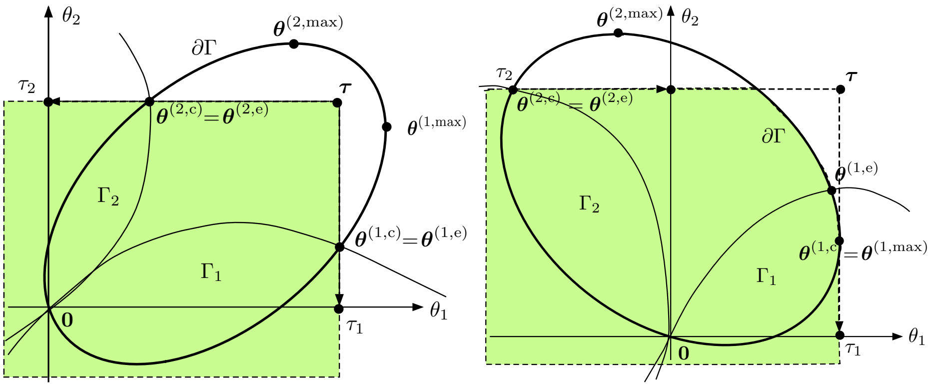

Figure 5: The domains $\mathcal{D}$ (green color on line) for (D1)

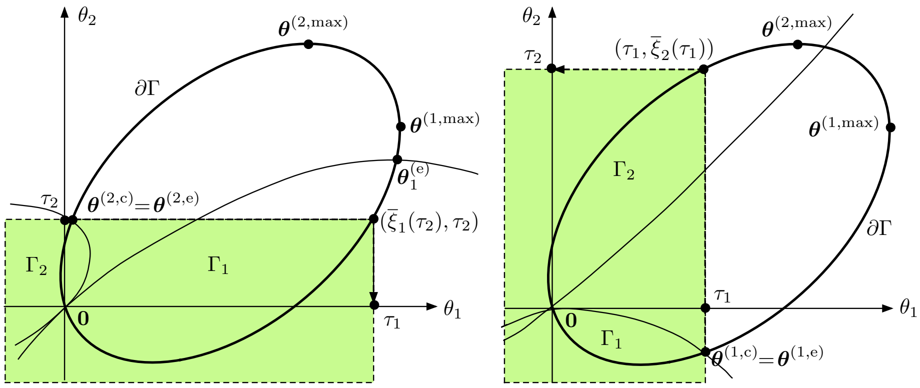

Figure 6: The domains $\mathcal{D}$ (green color on line) for (D2) and (D3)

We have an answer to the domain $\mathcal{D}$ in the following theorem.

Theorem 6.3 (Theorem 3.1 of Kobayashi and Miyazawa [51]) Under the conditions (3b), (3b'), (3c), and the stability condition given in Lemma 6.4, we have

$$\mathcal{D} = \{\boldsymbol{\theta} \in \Gamma_{\max}; \boldsymbol{\theta} < \boldsymbol{\tau}\}, \tag{6.14}$$

where we recall that

$$\Gamma_{\max} = \{\boldsymbol{\theta} \in \mathbb{R}^2; \boldsymbol{\theta} < \exists \boldsymbol{\theta}', \gamma(\boldsymbol{\theta}') < 1\}.$$

This theorem is proved in [51]. Here we outline this proof. From the observation that $\boldsymbol{\theta}^{(1,n)}, \boldsymbol{\theta}^{(2,n)} \in \overline{\mathcal{G}}_n$ and Theorem 6.1 and Lemma 6.8, it is not hard to see that

$$\mathcal{D}_0 \equiv \{\boldsymbol{\theta} \in \Gamma_{\max}; \boldsymbol{\theta} < \boldsymbol{\tau}\} \subset \mathcal{D}. \tag{6.15}$$

Thus, we need to show the opposite inclusion, that is, $\varphi(\boldsymbol{\theta}) = \infty$ for $\boldsymbol{\theta} \not\in \mathcal{D}_0$. This is verified by the following lemmas, all of which assume the conditions of Theorem 6.3.

Lemma 6.9 $\boldsymbol{\theta} \not\in \Gamma_{\max}$ implies that $\varphi(\boldsymbol{\theta}) = \infty$.

Lemma 6.10 For $k = 1, 2$, $\theta_k > \theta_k^{(\mathrm{k,c})}$ implies that $\varphi_k(\theta_k) = \infty$, and therefore $\varphi(\boldsymbol{\theta}) = \infty$.

Lemma 6.11 For $k = 1, 2$, $\theta_k > \tau_k$ implies that $\varphi_k(\theta_k) = \infty$, and therefore $\varphi(\boldsymbol{\theta}) = \infty$.

All of these lemmas have been proved in Kobayashi and Miyazawa [51]. Here we explain how they can be obtained. Lemma 6.9 is immediate from Lemma 6.2 (see also Lemma 3.6 of [51]). Lemma 6.10 is Lemma 3.7 of [51], which requires the Markov additive approach. Lemma 6.10 is also a special case of Lemma 6.11 because $\theta_k^{(\mathrm{k,c})} \le \tau_k$. However, we require Lemma 6.10 for proving Lemma 6.11. Lemma 6.11 requires the analytic extension of a complex variable function. We outline its idea below.

Since $\tau_k \le \theta_k^{(k,\mathrm{c})}$, we only need to consider the case that $\tau_k < \theta_k^{(k,\mathrm{c})}$ in view of Lemmas 6.10 and 6.11. Suppose that $\tau_1 < \theta_1^{(1,\mathrm{c})}$. This occurs only when (D2) holds. We claim that $\varphi_1(\theta_1) = \infty$ for $\theta_1 > \tau_1$. This proves Theorem 6.3.

From (3.11) and (6.15), we have, $(z_1, z_2) \in \mathbb{C}^2$ such that $(\Re z_1, \Re z_2) \in \mathcal{D}_0$,

$$\begin{aligned}(1 - \gamma(z_1, z_2))\varphi_+(z_1, z_2)\\ = (\gamma_1(z_1, z_2) - 1)\varphi_1(z_1) + (\gamma_2(z_1, z_2) - 1)\varphi_2(z_2) + (\gamma_0(z_1, z_2) - 1)\mathbb{P}_{\mathbf{0}}.\end{aligned} \tag{6.16}$$

To make the left side of this equation to vanish, we introduce an analytic extension $\underline{\xi}_2(z)$ of $\underline{\xi}_2(\theta)$ for real $\theta \in (\theta_2^{(2,\min)}, \theta_2^{(2,\max)})$ such that

$$\gamma(z, \underline{\xi}_2(z)) = 1.$$

By the implicit function theorem, $\underline{\xi}_2(z)$ is analytic at least on some open set G including the real interval $(\theta_2^{(2,\min)}, \theta_2^{(2,\max)})$. Plugging $(z_1, z_2) = (z, \underline{\xi}_2(z))$ into (6.16), we have, for $z \in G$ satisfying $\Re z \in (0, \tau_1)$,

$$(1 - \gamma_1(z, \underline{\xi}_2(z)))\varphi_1(z) = (\gamma_2(z, \underline{\xi}_2(z)) - 1)\varphi_2(\underline{\xi}_2(z)) + (\gamma_0(z, \underline{\xi}_2(z)) - 1)\mathbb{P}_{\mathbf{0}}. \tag{6.17}$$

Note that $\varphi_2(\underline{\xi}_2(z))$ is analytic for $\Re z < \tau_1$ but singular at $z = \tau_1$ since $\underline{\xi}_2(\tau_1) = \tau_2$. Since $1 - \gamma_1(z, \underline{\xi}_2(z)) \ne 0$ and $\gamma_2(z, \underline{\xi}_2(z)) - 1 \ne 0$ for $\Re z \in (0, \theta_1^{(2,\mathrm{c}))})$, this singularity implies that of $\varphi_1(z)$ at $z = \tau_1$. This proves the claim.

7 Deriving the tail asymptotics for $d = 2$

Once the domain $\mathcal{D}$ is obtained, we can use the stationary equations (3.11) and (3.14) of moment generating functions on $\mathcal{D}$. This enables us to find tail decay rates at the boundary of $\mathcal{D}$. We demonstrate this for the reflecting random walk for $d = 2$ and a two-dimensional SRBM. Here we employ two methods.

First, we refine the Markov additive approach in Section 5 using the information of the domain. This will be discussed in Section 7.1 for the two-dimensional reflecting random walks with unbounded jumps. Second, we directly work on the stationary equation (3.14) for the two-dimensional SRBM. This will be discussed in Section 7.2.

7.1 Two dimensional reflecting random walk

We continue to use the random vector $\boldsymbol{Z}$ in Section 6, which is subject to the stationary distribution of the two-dimensional reflecting random walk. We consider two types of the tail sets, $\{Z_k \ge n, Z_{3-k} = 0\}$ and $\{\langle \boldsymbol{c}, \boldsymbol{Z} \rangle \ge n\}$ for a directional vector $\boldsymbol{c} \ge \boldsymbol{0}$. For convenience, we will use unit vectors $\mathbf{e}_1 \equiv (1,0)$ and $\mathbf{e}_2 \equiv (0,1)$. All results in this subsection is cited from Kobayashi and Miyazawa [51]. So, we omit their proofs except for the following lemma, which suggests what decay rate we can expect.

Lemma 7.1 Under the assumptions of Theorem 6.3,

$$\limsup_{n\to\infty} \frac{1}{n} \log P(Z_k \ge n, Z_{3-k} = 0) \le -\tau_k, \qquad k = 1, 2, \tag{7.1}$$

and, for any directional vector $\boldsymbol{c} \ge \boldsymbol{0}$,

$$\limsup_{n\to\infty} \frac{1}{n} \log P(\langle \boldsymbol{c}, \boldsymbol{Z} \rangle \ge n) \le -\sup\{x \ge 0; x\boldsymbol{c} \in \mathcal{D}\}. \tag{7.2}$$

Proof. (7.1) is immediate from Theorem 6.3. To see (7.2), we use Markov inequality:

$$e^{xn} P(\langle \boldsymbol{c}, \boldsymbol{Z} \rangle \ge n) \le E(e^{\langle x\boldsymbol{c}, \boldsymbol{Z} \rangle}), \qquad x \ge 0, n = 0, 1, \ldots.$$

Taking logarithms of both sides, dividing by $n \ge 1$ and letting $n \to \infty$, we have (7.2). □

By this lemma, the decay rates of $P(Z_k \ge n, Z_{3-k} = 0)$ and $P(\langle \boldsymbol{c}, \boldsymbol{Z} \rangle \ge n)$ are expected to be respectively τ_k and

$$\alpha_{\boldsymbol{c}} = \sup\{x \ge 0; x\boldsymbol{c} \in \mathcal{D}\}, \qquad \text{for directional vector } \boldsymbol{c} \ge \boldsymbol{0}.$$

For the double QBD process, we have used the procedure given in Section 5.6. Its key ingredient is Lemma 5.2, but we cannot use this lemma because of unbounded jumps. On the other hand, we have the domain $\mathcal{D}$, which enables us to directly use the results such as Theorem 7.2 in Section 5. This has been performed with help of Lemma 7.1 and generalizing Lemma 5.2 by Kobayashi and Miyazawa [51].

Theorem 7.1 (Theorem 4.1 of [51]) Under the conditions of Theorem 6.3, we have,

$$\lim_{n\to\infty} \frac{1}{n} \log P(Z_k \ge n, Z_{3-k} = 0) = -\tau_k, \qquad k = 1, 2. \tag{7.3}$$

Theorem 7.2 (Theorem 4.2 of [51]) Under the same conditions of Theorem 7.1, we have, for any directional vector $\boldsymbol{c} \ge \boldsymbol{0}$,

$$\lim_{n\to\infty} \frac{1}{n} \log P(\langle \boldsymbol{c}, \boldsymbol{Z} \rangle \ge n) = -\alpha_{\boldsymbol{c}}, \tag{7.4}$$

where we recall that $\alpha_{\boldsymbol{c}} = \sup\{x \ge 0; x\boldsymbol{c} \in \mathcal{D}\}$. Furthermore, if $\gamma(\alpha_{\boldsymbol{c}}\boldsymbol{c}) = 1$ and if $\gamma_k(\alpha_{\boldsymbol{c}}\boldsymbol{c}) \ne 1$ and $\alpha_{\boldsymbol{c}} c_k \ne \tau_k$ for $k = 1, 2$, then we have the following exact asymptotics:

$$\lim_{n\to\infty} e^{\alpha_{\boldsymbol{c}} n} P(\langle \boldsymbol{c}, \boldsymbol{Z} \rangle \ge n) = b_{\boldsymbol{c}}. \tag{7.5}$$

Remark 7.1 Since $\alpha_{\mathbf{e}_k}$ may be less than τ_k, the decay rate of $P(Z_k \ge n)$ may be different from that of $P(Z_k \ge n, Z_{3-k} = 0)$.

We note that the tail asymptotic of $\mathbb{P}(\big\langle \boldsymbol{c}, \boldsymbol{Z} \big\rangle > x)$ is generally different from that of $\mathbb{P}(\boldsymbol{Z} > x\boldsymbol{c})$. The latter may be different from $\mathbb{P}(\boldsymbol{Z} \in x(\boldsymbol{c} + B)$ for a bounded set $B \subset \mathbb{R}_+^2$. Nevertheless, the tail asymptotics obtained in Theorem 7.2 have some similarity to the asymptotics of $\mathbb{P}(\boldsymbol{Z} \in x\boldsymbol{c} + U_1)$ for the two-dimensional reflecting skip-free random walk obtained in Borovkov and Mogul'skii [9], where U_1 is a unit square.

7.2 Two dimensional SRBM

We consider the two-dimensional SRBM under the conditions (3-ii) and (3-iii) in Section 3.5. These are necessary and sufficient conditions for the existence of the stationary distribution. Then, we have the stationary equation (3.14) in terms of moment generating functions. For $d = 2$, it is written as

$$\gamma(\boldsymbol{\theta})\varphi(\boldsymbol{\theta}) = \gamma_{[1]}(\boldsymbol{\theta})\varphi_{[1]}(\theta_2) + \gamma_{[2]}(\boldsymbol{\theta})\varphi_{[2]}(\theta_1), \qquad \boldsymbol{\theta} \in \mathcal{D}. \tag{7.6}$$

We only use this equation for deriving the tail asymptotics. A key idea is to directly connect $\varphi_{[1]}$ and $\varphi_{[2]}$. For this, we take a path obtained from $\gamma(\theta_1, \theta_2) = 0$ on $\mathbb{R}^2$. On this path, $\gamma_{[1]}(\theta_1)$ and $\gamma_{[2]}(\theta_2)$ are directly related by (7.6), where θ_2 is a function of θ_1 and vice versa.

Denote this function $f_2(\theta_1)$. Then, it can be shown that $f_2(z)$ with complex variable z is analytic for $\Re z \in \mathcal{D}$. Analytically extending this complex variable function, we can find analytic behaviors of $\varphi_{[1]}(z)$ and $\varphi_{[2]}(z)$, respectively, at the singular points whose real parts are smallest. Then, the complex variable version of (7.6):

$$\gamma(z_1, z_2)\varphi(z_1, z_2) = \gamma_{[1]}(z_1, z_2)\varphi_{[1]}(z_2) + \gamma_{[2]}(z_1, z_2)\varphi_{[2]}(z_1), \qquad (\Re z_1, \Re z) \in \mathbb{C}^2, \tag{7.7}$$

and analytic inversions yield the exact tail asymptotics of $\mathbb{P}(\langle \boldsymbol{c}, \boldsymbol{Z}\rangle > x)$ as $x \to \infty$ for each directional vector $\boldsymbol{c} \geq \boldsymbol{0}$.

This scenario has been recently completed by Dai and Miyazawa [16]. Here we briefly introduce their results. The domain $\mathcal{D}$ has essentially the same form given by Theorem 6.3. The regions Γ, Γ_1 and Γ_2 are defined as

$$\Gamma = \{\boldsymbol{\theta} \in \mathbb{R}^2; \gamma(\boldsymbol{\theta}) > 0\}, \qquad \Gamma_k = \{\boldsymbol{\theta} \in \mathbb{R}^2; \gamma_{[k]}(\boldsymbol{\theta}) < 0\}, \quad k = 1, 2.$$

Vectors $\boldsymbol{\theta}^{(k,\max)}$, $\boldsymbol{\theta}^{(k,\min)}$, $\boldsymbol{\theta}^{(k,\mathrm{c})}$ and $\boldsymbol{\theta}^{(k,\mathrm{e})}$ are defined by (6.10) and (6.11), respectively. The classifications (D1), (D2) and (D3) are also the same as those in Section 6.4. See Theorem 2.1 of [16] for details, in which slightly different notation is used, but they exactly correspond to those introduced in this paper.

Because of simplicity of the stationary equation (7.6), we can successfully apply complex analysis and get sharper results. We denote a random vector subject to the stationary distribution by $\boldsymbol{Z}$. We also use the following notation. If $\boldsymbol{\tau} \not\in \Gamma$, then

$$\boldsymbol{\eta}^{(1)} = (\tau_1, \overline{\xi}_2(\tau_1)), \qquad \boldsymbol{\eta}^{(2)} = (\overline{\xi}_1(\tau_2), \tau_2).$$

Otherwise, we let $\boldsymbol{\eta}^{(1)} = \boldsymbol{\eta}^{(2)} = \boldsymbol{\tau}$. Typical figures of the domain $\mathcal{D}$ are drawn in Figure 7.

The exact tail asymptotics are derived for the one-dimensional stationary distribution in each direction in the following two theorems. In what follows, for a non-zero vector $\boldsymbol{u} \in \mathbb{R}^2_+$, the line $t\boldsymbol{u}$ with $t \geq 0$ is referred to as the ray $\boldsymbol{u}$, and ray $\boldsymbol{u} \equiv (u_1, u_2)$ is said to be below (on, above) ray $\boldsymbol{v} \equiv (v_1, v_2)$ if $u_1 = sv_1$ for some $s > 0$ implies $u_2 < sv_2$ ($u_2 = sv_2$, $u_2 > sv_2$, respectively).

Theorem 7.3 (Theorem 2.2 of Dai and Miyazawa [16]) Assume that conditions (3-ii) and (3-iii) hold and that the SRBM data is in case (D1). Let $\boldsymbol{c} \in \mathbb{R}^2_+$ be a direction.

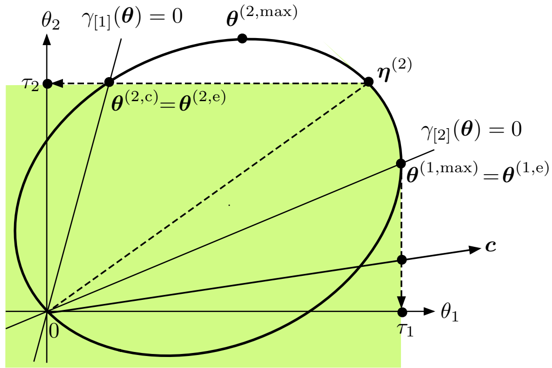

Figure 7: The domains $\mathcal{D}$ (green color on line) for (D1) and (D2)

Then, $\mathbb{P}(\langle \boldsymbol{c}, \boldsymbol{Z}\rangle > x)$ has the exact asymptotic $bh_{\boldsymbol{c}}(x)$ with some constant $b>0$ and $h_{\boldsymbol{c}}(x)$ being given below. (a) When the ray $\boldsymbol{c}$ is below ray $\boldsymbol{\eta}^{(1)}$,

$$h_{\boldsymbol{c}}(x) = \begin{cases} e^{-\alpha_c x} & \text{if } \boldsymbol{\eta}^{(1)} \neq \boldsymbol{\theta}^{(1,\max)}, \\ x^{-1/2}e^{-\alpha_c x} & \text{if } \boldsymbol{\eta}^{(1)} = \boldsymbol{\theta}^{(1,\max)} = \boldsymbol{\theta}^{(1,e)}, \\ x^{-3/2}e^{-\alpha_c x} & \text{if } \boldsymbol{\eta}^{(1)} = \boldsymbol{\theta}^{(1,\max)} \neq \boldsymbol{\theta}^{(1,e)}. \end{cases} \tag{7.8}$$

(b) When the ray c is above or on the ray $\boldsymbol{\eta}^{(1)}$ and below the ray $\boldsymbol{\eta}^{(2)}$,

$$h_{\boldsymbol{c}}(x) = \begin{cases} x^{-1/2}e^{-\alpha_c x}, & \text{if } c \text{ is on line } \boldsymbol{\eta}^{(1)} = \boldsymbol{\theta}^{(1,\max)} \neq \boldsymbol{\theta}^{(1,e)}, \\ xe^{-\alpha_c x}, & \text{if } c \text{ is on line } \boldsymbol{\eta}^{(1)} \neq \boldsymbol{\theta}^{(1,\max)}, \\ e^{-\alpha_c x}, & \text{otherwise.} \end{cases} \tag{7.9}$$

(c) When the ray c is above the ray $\boldsymbol{\eta}^{(2)}$, the case is symmetric to (a) and a part of (b).

For cases (D2) and (D3), we only consider (D2) II because of their symmetry. In (D2), $\tau_2 = \theta_2^{(2,e)}$, $\tau_1 = \overline{\xi}_1(\tau_2)$, $\boldsymbol{\eta}^{(1)} = \boldsymbol{\eta}^{(2)} = \boldsymbol{\tau} = (\tau_1, \tau_2)$, and the condition $\boldsymbol{\eta}^{(1)} \neq \boldsymbol{\theta}^{(1,\max)}$ is equivalent to the condition $\tau_1 < \theta_1^{(1,\max)}$.

Theorem 7.4 (Theorem 2.3 of Dai and Miyazawa [16]) Assume that conditions (3-ii) and (3-iii) hold and that the SRBM data is in case (D2). Let $\boldsymbol{c} \in \mathbb{R}^2_+$ be a direction. Then, $\mathbb{P}(\langle \boldsymbol{c}, \boldsymbol{Z}\rangle > x)$ has the exact asymptotic $bh_{\boldsymbol{c}}(x)$ with some constant $b>0$ and $h_{\boldsymbol{c}}(x)$ being given below.
(a) When the ray $\boldsymbol{c}$ is below the ray $\boldsymbol{\tau}$,

$$h_{\boldsymbol{c}}(x) = \begin{cases} e^{-\alpha_c x}, & \text{if } \boldsymbol{\tau} \neq \boldsymbol{\theta}^{(1,e)} \text{ and } \boldsymbol{\tau} \neq \boldsymbol{\theta}^{(1,\max)} \text{ or if } \boldsymbol{\tau} = \boldsymbol{\theta}^{(1,\max)} = \boldsymbol{\theta}^{(1,e)}, \\ xe^{-\alpha_c x}, & \text{if } \boldsymbol{\tau} = \boldsymbol{\theta}^{(1,e)} \neq \boldsymbol{\theta}^{(1,\max)}, \\ x^{-1/2}e^{-\alpha_c x}, & \text{if } \boldsymbol{\tau} = \boldsymbol{\theta}^{(1,\max)} \neq \boldsymbol{\theta}^{(1,e)}. \end{cases} \tag{7.10}$$

(b) When the ray c is on the ray $\boldsymbol{\tau}$,

$$h_{\boldsymbol{c}}(x) = xe^{-\alpha_c x}. \tag{7.11}$$

(c) When the ray c is above the ray $\boldsymbol{\tau}$,

$$h_{\boldsymbol{c}}(x) = e^{-\alpha_c x}. \tag{7.12}$$

The results in Theorems 7.3 and 7.4 exactly correspond to those in Theorem 7.2. As we noted there, the tail asymptotic of $\mathbb{P}(\langle \boldsymbol{c}, \boldsymbol{Z} \rangle > x)$ is generally different from that of $\mathbb{P}(\boldsymbol{Z} > x\boldsymbol{c})$ and $\mathbb{P}(\boldsymbol{Z} \in x(\boldsymbol{c} + B))$ for a bounded open set $B \subset \mathbb{R}^2_+$. We may discuss their difference using the large deviations rate function obtained by Avram, Dai and Hasenbein [4]. This will be done elsewhere.

7.3 Two-sided double QBD

This process is introduced in Example 3.3. It is determined by six distributions of the increments, $\{p^{(j)}_{\boldsymbol{n}}\}$ for $j = 0, +, -, 1+, 1-, 2$. As we have discussed in Example 3.3, we have to work with the three-dimensional variable $(\theta_{1-}, \theta_{1+}, \theta_2)$ for the generating functions.

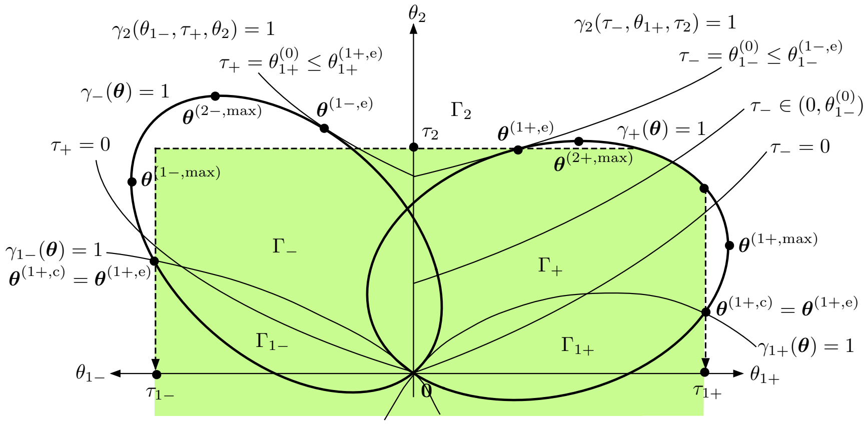

Figure 8: The domain $\mathcal{D}$ (green colored area) projected to θ_{1-}–θ_2 and θ_{1+}–θ_2 quadrants for two-sided DQBD

Similar to the double QBD process, let

$$\Gamma_{1-} = \{\boldsymbol{\theta} \in \mathbb{R}^2; \gamma_-(\boldsymbol{\theta}) < 1, \gamma_{1-}(\boldsymbol{\theta}) < 1\}, \quad \Gamma_{1+} = \{\boldsymbol{\theta} \in \mathbb{R}^2; \gamma_+(\boldsymbol{\theta}) < 1, \gamma_{1+}(\boldsymbol{\theta}) < 1\},$$
$$\Gamma_2 = \{(\theta_{1-}, \theta_{1+}, \theta_2) \in \mathbb{R}^3; \gamma_-(\theta_{1-}, \theta_2) < 1, \gamma_{1+}(\theta_{1+}, \theta_2) < 1, \gamma_2(\theta_{1-}, \theta_{1+}, \theta_2) < 1\}.$$

We then get the extreme points τ_{1-}, τ_{1+} and τ_2 of the domain in the coordinate directions as the solution of the following fixed point equations:

$$\tau_{1-} = \sup\{\theta_1 \geq 0; \boldsymbol{\theta} \in \Gamma_{1-}, \theta_2 \leq \tau_2\}, \tag{7.13}$$
$$\tau_{1+} = \sup\{\theta_1 \geq 0; \boldsymbol{\theta} \in \Gamma_{1+}, \theta_2 \leq \tau_2\}, \tag{7.14}$$
$$\tau_2 = \sup\{\theta_2 \geq 0; (\theta_{1-}, \theta_{1+}, \theta_2) \in \Gamma_2, \theta_{1-} \leq \tau_{1-}, \theta_{1+} \leq \tau_{1+}\}. \tag{7.15}$$

To find this solution, we compute the following extreme points (see Figure 8):

$$\boldsymbol{\theta}^{(1-,c)} = \arg_{\boldsymbol{\theta}} \sup\{\theta_1 \geq 0; (\theta_1, \theta_2) \in \Gamma_- \cap \Gamma_{1-}\},$$
$$\boldsymbol{\theta}^{(1+,c)} = \arg_{\boldsymbol{\theta}} \sup\{\theta_1 \geq 0; (\theta_1, \theta_2) \in \Gamma_+ \cap \Gamma_{1+}\},$$
$$\boldsymbol{\theta}^{(2,c)} = \arg_{(\theta_{1-}, \theta_{1+}, \theta_2)} \sup\{\theta_2 \geq 0; (\theta_{1-}, \theta_{1+}, \theta_2) \in \Gamma_2\}.$$

Then, we have

$$\tau_{1-} = \begin{cases} \theta^{(1-,c)}_1, & \theta^{(1-,c)}_2 < \theta^{(2,c)}_2, \\ \theta^{(2,c)}_1, & \theta^{(1-,c)}_2 \geq \theta^{(2,c)}_2, \end{cases} \qquad \tau_{1+} = \begin{cases} \theta^{(1+,c)}_1, & \theta^{(1+,c)}_2 < \theta^{(2,c)}_2, \\ \theta^{(2,c)}_1, & \theta^{(1+,c)}_2 \geq \theta^{(2,c)}_2, \end{cases} \tag{7.16}$$

$$\tau_2 = \begin{cases} \theta_2^{(2,c)}, & \theta_{1-}^{(2,c)} < \theta_1^{(1-,c)}, \theta_{1+}^{(2,c)} < \theta_1^{(1+,c)}, \\ \min(\theta_2^{(1-,c)}, \theta_2^{(1+,c)}) & \theta_{1-}^{(2,c)} \geq \theta_1^{(1-,c)} \text{ or } \theta_{1+}^{(2,c)} \geq \theta_1^{(1+,c)}. \end{cases} \tag{7.17}$$

These τ_{1-}, τ_{1+} and τ_2 are obtained as the decay rates of the three coordinate directions in Miyazawa [68] (see Theorem 1.3 there), which uses the Markov additive approach. Here we derive them by the analytic function approach using the program proposed in Section 6.2. Similar to the SRBM case in Section 7.2, it should not be very difficult to find the exact tail asymptotics in an arbitrary given direction. However, the proposed program in Section 6.2 has not yet been verified except for the two-dimensional reflecting process on the orthant. Hence, we need some more work for using the analytic function approach.

8 Applications

In this section, we collect some examples to see how the decay rates obtained through the domain $\mathcal{D}$ are useful in applications.

8.1 Modified Jackson networks

8.1.1 Jackson network

We have considered the Jackson network in Section 3.1. Here we consider the case for $d = 2$. For this model, the moment generating functions γ and γ_k are given by

$$\gamma(\boldsymbol{\theta}) = \lambda_1 e^{\theta_1} + \lambda_2 e^{\theta_2} + \mu_1 e^{-\theta_1}(r_{12} e^{\theta_2} + r_{10}) + \mu_2 e^{-\theta_2}(r_{21} e^{\theta_1} + r_{20}), \tag{8.1}$$
$$\gamma_1(\boldsymbol{\theta}) = \lambda_1 e^{\theta_1} + \lambda_2 e^{\theta_2} + \mu_1 e^{-\theta_1}(r_{12} e^{\theta_2} + r_{10}) + \mu_2. \tag{8.2}$$

Assume the stability condition (3.3). Then, the stationary distribution for $d = 2$ is given by

$$\pi(n, m) = (1 - \rho_1)(1 - \rho_2)\rho_1^m \rho_2^n, \qquad m, n \in \mathbb{Z}_+.$$

Let $\tau_i = -\log \rho_i$ for $i = 1, 2$. Then, the moment generating functions of the π, $\{\pi(0, n); n \in \mathbb{Z}_+\}$ and $\{\pi(m, 0); m \in \mathbb{Z}_+\}$ are computed as

$$\varphi_+(\boldsymbol{\theta}) = \varphi(\boldsymbol{\theta}) = \frac{(e^{\tau_1} - 1)(e^{\tau_2} - 1)}{(e^{\tau_1} - e^{\theta_1})(e^{\tau_2} - e^{\theta_2})},$$
$$\varphi_i(\theta_i) = \frac{e^{\tau_i} - 1}{e^{\tau_i} - e^{\theta_i}}(1 - e^{-\tau_{3-i}}), \quad i = 1, 2.$$

Plugging these into the stationary equation, we derive

$$\begin{aligned}(1 - \gamma(\boldsymbol{\theta})) + (1 - \gamma_1(\boldsymbol{\theta}))(e^{\tau_2} - e^{\theta_2})e^{-\tau_2} + (1 - \gamma_2(\boldsymbol{\theta}))(e^{\tau_1} - e^{\theta_1})e^{-\tau_1} \\ = (\gamma_0(\boldsymbol{\theta}) - 1)\pi(0, 0)(e^{\tau_1} - e^{\theta_1})(e^{\tau_2} - e^{\theta_2})e^{-\tau_1 - \tau_2}.\end{aligned}$$

Let $\boldsymbol{\theta} = \boldsymbol{\tau}$ in this equation, then we immediately see that $\gamma(\boldsymbol{\tau}) = 1$. Thus, $\boldsymbol{\tau} \equiv (\tau_1, \tau_2)$ must be on the boundary of Γ, that is, $\boldsymbol{\tau} \in \partial\Gamma$. This is illustrated in Figure 9.

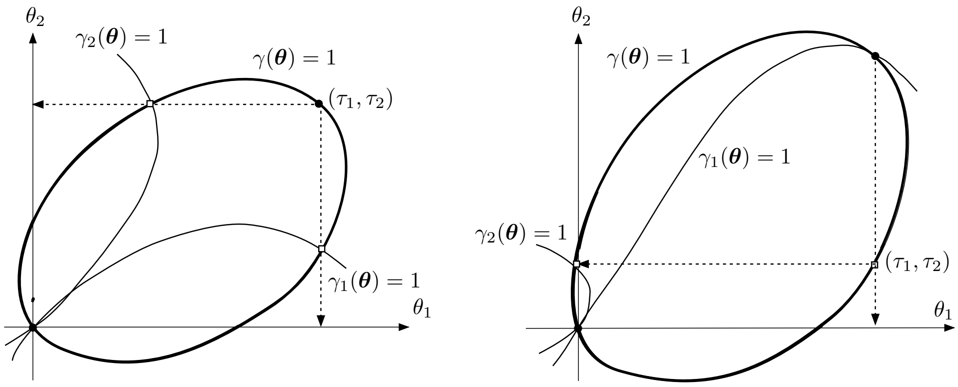

Figure 9: The location of $\boldsymbol{\tau}$ for (D1) and (D2)

We next compute $\boldsymbol{\theta}^{(1,\mathrm{e})}$, which is the solution of the following equations:

$$\gamma(\boldsymbol{\theta}) = 1, \qquad \gamma_1(\boldsymbol{\theta}) = 1, \qquad \boldsymbol{\theta} \neq \mathbf{0}.$$

From (8.1), (8.2) and $e^{\theta_2} = r_{21}e^{\theta_1} + r_{20}$, we can find

$$\boldsymbol{\theta}^{(1,\mathrm{e})} = \left(-\log \rho_1, -\log \frac{\rho_1}{r_{21} + (1 - r_{21})\rho_1} \right).$$

Hence, $\gamma(\tau_1, \theta_2) = 1$ has two solutions:

$$\theta_2 = \tau_2 (= -\log \rho_2) \text{ and } -\log \frac{\rho_1}{r_{21} + (1 - r_{21})\rho_1}.$$

Thus, if $\rho_2 \geq \frac{\rho_1}{r_{21}+(1-r_{21})\rho_1}$, which is equivalent to

$$\frac{\rho_1}{1 - \rho_1} \leq \frac{r_{21}\rho_2}{1 - \rho_2}, \tag{8.3}$$

then $\tau_2 \leq \theta_2^{(1,\mathrm{e})}$, and therefore we have the case (D2). Otherwise, the case (D1) or (D3) occurs, and we have $\boldsymbol{\theta}^{(1,\mathrm{c})} = \boldsymbol{\theta}^{(1,\mathrm{e})}$ (see Figure 9). Similar results are obtained for $\boldsymbol{\theta}^{(2,\mathrm{e})}$. That is, (D3) occurs if

$$\frac{\rho_2}{1 - \rho_2} \leq \frac{r_{12}\rho_1}{1 - \rho_1}. \tag{8.4}$$

Otherwise, (D1) or (D2) occurs.

We cannot have (8.3) and (8.4) simultaneously because they imply $r_{12}r_{21} \geq 1$, which contradicts the stability condition. Thus, we have the following three classifications:

(J1) If neither (8.3) nor (8.4) holds, then we have the case (D1).

(J2) If (8.3) holds but (8.4) does not hold, then we have the case (D2).

(J3) If (8.3) does not hold but (8.4) holds, then we have the case (D3).

8.1.2 Server collaboration

Similarly to Miyazawa [69], we next modify the Jackson network in such a way that servers at nodes 1 and 2 help service at the other nodes when their own nodes are empty. We describe this by increasing μ_k up to μ_k^* in $\gamma_k(\boldsymbol{\theta})$ for $k = 1, 2$. To make arguments simple, we assume that

$$\mu_1 + \mu_1^* \le \mu_1 + \mu_2, \qquad \mu_2 + \mu_2^* \le \mu_1 + \mu_2.$$

This only changes $\gamma_k(\boldsymbol{\theta})$ for $k = 1, 2$. We denote this modified functions by $\gamma_k^*(\boldsymbol{\theta})$. From (8.2) and the corresponding formulas for $\gamma_2(\boldsymbol{\theta})$, we can see that the curve of $\gamma_1^*(\boldsymbol{\theta}) = 1$ is located above that of $\gamma_1(\boldsymbol{\theta}) = 1$. Similarly, the curve of $\gamma_2^*(\boldsymbol{\theta}) = 1$ is located at the right side of that of $\gamma_2(\boldsymbol{\theta}) = 1$. See Figure 10.

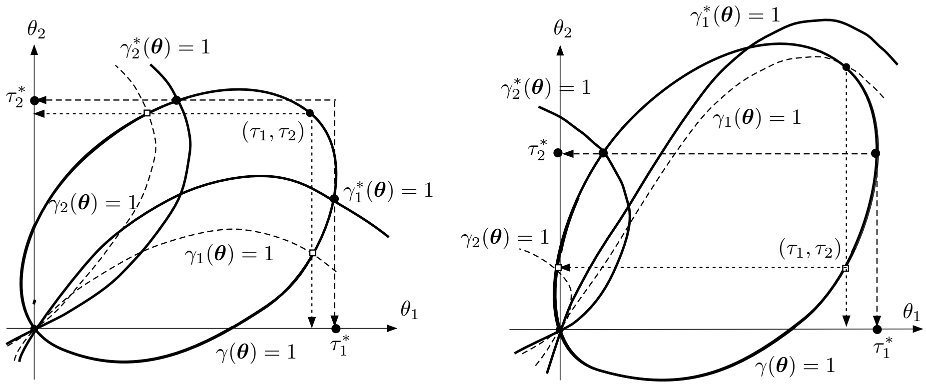

Figure 10: The effect of server collaboration for (D1) and (D2)

From this figure, we can see how the decay rate τ_k of the queue length distribution at node k is increased to τ_k^*, which is the decay rate of the modified network. It is noted that increasing the service rate at node 1 does not gain any improvement for the case (D2) while that at node 2 improves service at both nodes. These improvements stop if increased service rates are larger than the thresholds which correspond to the maximum points $\boldsymbol{\theta}^{(1,\max)}$ and $\boldsymbol{\theta}^{(2,\max)}$. This is detailed in Miyazawa [69]. This problem is also discussed by Foley and McDonald [30] and Khanchi [48].

8.1.3 Batch arrival Jackson network

We can also modify the two node Jackson network by exogenous batch arrivals. This is considered in Kobayashi and Miyazawa [51]. Contrary to the server collaboration, its effect is negative as one may expect. In this case, the closed curve of $\gamma(\boldsymbol{\theta}) = 1$ becomes smaller and the curves of $\gamma_1(\boldsymbol{\theta}) = 1$ and $\gamma_2(\boldsymbol{\theta}) = 1$ are shifted downward and to the left, respectively.

For this batch arrival model, Miyazawa and Taylor [72] derived a stochastic upper bound of product form for the stationary distribution. Kobayashi and Miyazawa [51] show that this bound is generally not tight, that is, the decay rates of the upper bound in coordinate directions are generally smaller than the decay rates for the stationary distribution.

In our formulation, we can also allow simultaneous batch arrivals at two nodes whose sizes can have an arbitrary joint distribution. For example, the bath room problem of Flatto and McKean [27] can be answered.

8.2 Join the shortest queue

It is very natural to join the shortest queue if there are parallel queues for identical service. This service system is called a join the shortest queue. A typical assumption is that customers arrive subject to the Poisson process and join the shortest queue with tie breaking, and service times are *i.i.d.* with the exponential distribution. This queueing model may look simple, but the stationary distribution of its joint queue length is very hard to get even for two parallel queues. Thus, the tail asymptotics have been studied. It has a long history from starting from Kingman [50] in 1961, but satisfactory answers are only available for two parallel queues (e.g., see Foley and McDonald [28], Kurkova and Suhov [55], Li, Miyazawa and Zhao [56], Sakuma and Miyazawa [88] and Takahashi, Fujimoto and Makimoto [94]). The large deviations principle is derived for the stationary distributions of joint queue lengths under very general assumption for general $d \geq 2$ in Puhalskii and Vladimirov [83]. However, the result is not easy to use for applications because we have to solve the variational problem (see Section 4.2).

We consider such a general model but for $d = 2$. This model has two parallel queues, numbered as queues 1 and 2. For each $i = 1, 2$, queue i serves customers in the first-come first-served manner with *i.i.d.* service times subject to the exponential distribution with rate μ_i. There are three exogenous Poisson arrival streams. The first and second streams go to queues 1 and 2 with rates λ_1 and λ_2, respectively, while arriving customers in the third stream with rate δ choose the shorter queue with tie breaking. The probability that a customer with tie breaking chooses queue 1 does not change the tail decay rate, so we simply assume it to be 1/2. This model is referred to as a generalized join shortest queue (see Figure 11).

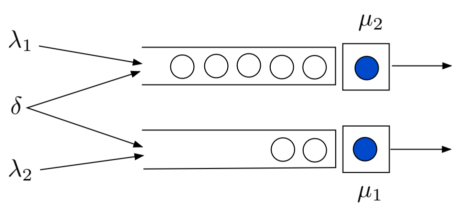

Figure 11: Generalized join shortest queue

The tail asymptotic problem for this generalized join shortest queue was studied by Foley and McDonald [28]. However, they mainly considered the case where the moment generating function of the Markov additive kernel is positive (see condition (5c) in Theorem 5.2). For some other cases, the exact geometric asymptotics were obtained in Li, Miyazawa and Zhao [56]. However, those two papers have not yet completely solved the tail asymptotic problem even for the rough asymptotics, that is, the decay rate problem. For the latter problem, a complete solution was recently obtained in Miyazawa [68] by using the Markov additive approach and the optimization technique developed in Miyazawa

[69]. In this section, we revisit this result using the analytic function approach based on the convergence domain.

Similar to the Jackson network, we can formulate this continuous time model as a discrete time Markov chain. Following Miyazawa [68], we formulate it as a two-sided double QBD. For this, we assume without loss of generality that

$$\lambda_1 + \lambda_2 + \mu_1 + \mu_2 + \delta = 1.$$

Let $L_{1\ell}$ and $L_{2\ell}$ be the queue lengths including customers being served at time $\ell = 0, 1, \ldots$, and let $L_{1\ell} = L_{2\ell} - L_{1\ell}$ and $Z_{2\ell} = \min(L_{1\ell}, L_{2\ell})$. It is not hard to see that $\boldsymbol{Z}_\ell \equiv (Z_{1\ell}, Z_{2\ell})$ is the two-sided double QBD process introduced in Example 3.3. For example, it is a skip-free random walk on each region $S_+ \equiv (\mathbb{Z}_+ \cup \backslash\{0\}) \times (\mathbb{Z}_+ \setminus \{0\})$ reflected at the boundary $S_{1+} \equiv (\mathbb{Z}_+ \setminus \{0\}) \times \{0\}$ (see Fig. 12).

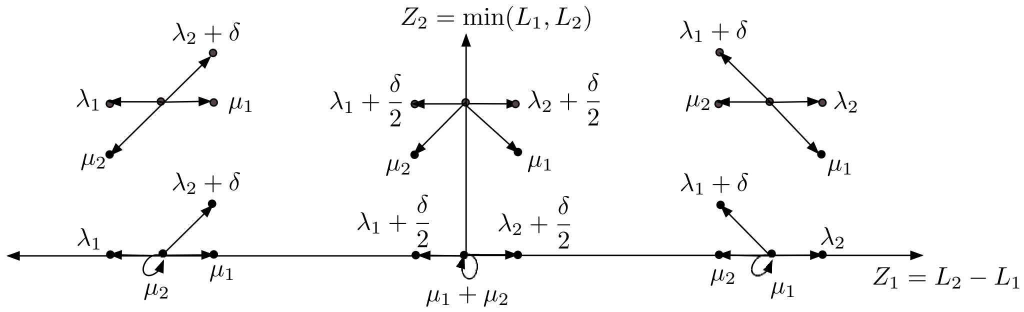

Figure 12: State transitions for the generalized shortest queue.

Then, the transition probabilities are give by

$$\begin{aligned}
&p^{(-)}_{(-1)0} = \lambda_1, \quad p^{(-)}_{(-1)(-1)} = \mu_2, \quad p^{(-)}_{10} = \mu_1, \quad p^{(-)}_{11} = \lambda_2 + \delta,\\
&p^{(+)}_{10} = \lambda_2, \quad p^{(+)}_{1(-1)} = \mu_1, \quad p^{(+)}_{(-1)0} = \mu_2, \quad p^{(+)}_{(-1)1} = \lambda_1 + \delta,\\
&p^{(2)}_{10} = \lambda_2 + \frac{\delta}{2}, \quad p^{(2)}_{1(-1)} = \mu_1, \quad p^{(2)}_{(-1)(-1)} = \mu_2, \quad p^{(2)}_{(-1)0} = \lambda_1 + \frac{\delta}{2},
\end{aligned}$$

$$\begin{aligned}
&p^{(1-)}_{(-1)0} = \lambda_1, \quad p^{(1-)}_{00} = \mu_2, \quad p^{(1+)}_{10} = \mu_1, \quad p^{(1-)}_{11} = \lambda_2 + \delta,\\
&p^{(1+)}_{10} = \lambda_2, \quad p^{(1+)}_{00} = \mu_1, \quad p^{(1+)}_{(-1)0} = \mu_2, \quad p^{(1+)}_{(-1)1} = \lambda_1 + \delta,\\
&p^{(0)}_{10} = \lambda_2 + \frac{\delta}{2}, \quad p^{(0)}_{00} = \mu_1 + \mu_2, \quad p^{(0)}_{(-1)0} = \lambda_1 + \frac{\delta}{2}
\end{aligned}$$

where all other transitions are null. To exclude obvious cases, we assume that δ, μ_1, μ_2 are all positive.

Denote traffic intensities by

$$\rho_1 = \frac{\lambda_1}{\mu_1}, \qquad \rho_2 = \frac{\lambda_2}{\mu_2}, \qquad \rho = \frac{\lambda_1 + \lambda_2 + \delta}{\mu_1 + \mu_2}.$$

Then, it is known that this generalized join shortest queue is stable if and only if $\rho_1 < 1, \rho_2 < 1$ and $\rho < 1$ (e.g., see Foley and McDonald [28]). This stability condition is assumed throughout this section. We will also use the following notation, which was introduced and shown to be very useful in computations in Li, Miyazawa and Zhao [56]:

$$\beta_1 = \mu_1 \rho^2 + \lambda_2, \qquad \beta_2 = \mu_2 \rho^2 + \lambda_1.$$

We need to compute $\theta_1^{(1-,\mathrm{c})}$, $\theta_1^{(1+,\mathrm{c})}$ and $\theta_2^{(2,\mathrm{c})}$, which are defined as

$$\begin{aligned}
\theta_1^{(1-,\mathrm{c})} &= \sup\{\theta_1; \gamma_-(\boldsymbol{\theta}) \le 1, \gamma_{1-}(\boldsymbol{\theta}) \le 1\},\\
\theta_1^{(1+,\mathrm{c})} &= \sup\{\theta_1; \gamma_+(\boldsymbol{\theta}) \le 1, \gamma_{1+}(\boldsymbol{\theta}) \le 1\},\\
\theta_2^{(2,\mathrm{c})} &= \sup\{\theta_2; \gamma_-(\theta_{1-}, \theta_2) \le 1, \gamma_+(\theta_{1+}, \theta_2) \le 1, \gamma_2(\theta_{1-}, \theta_{1+}, \theta_2) \le 1\}.
\end{aligned}$$

They are obtained if we can solve the following three sets of equations:

$$\gamma_-(\boldsymbol{\theta}) = 1, \qquad \gamma_{1-}(\boldsymbol{\theta}) = 1, \tag{8.5}$$
$$\gamma_+(\boldsymbol{\theta}) = 1, \qquad \gamma_{1+}(\boldsymbol{\theta}) = 1, \tag{8.6}$$
$$\gamma_-(\theta_{1-}, \theta_2) = 1, \qquad \gamma_+(\theta_{1+}, \theta_2) = 1, \qquad \gamma_2(\theta_{1-}, \theta_{1+}, \theta_2) = 1. \tag{8.7}$$

For convenience, let $z = e^{\theta_1}$ and $\xi = e^{\theta_2}$ in (8.5). Then, we have

$$\lambda_1 z + \mu_2 z \xi^{-1} + \mu_1 z^{-1} + (\lambda_2 + \delta) z^{-1} \xi = 1, \tag{8.8}$$
$$\lambda_1 z + \mu_2 + \mu_1 z^{-1} + (\lambda_2 + \delta) z^{-1} \xi = 1. \tag{8.9}$$

Solving these equations for $z \neq 1$, we have $z = \xi = \rho_1^{-1}$. For $z = \rho_1^{-1}$, (8.8) yields $\xi = \rho_1^{-1}, \frac{\mu_2}{\lambda_2+\delta}\rho_1^{-1}$. Note that $\rho_1^{-1} < \frac{\mu_2}{\lambda_2+\delta}\rho_1^{-1}$ if and only if $\mu_2 > \lambda_2 + \delta$. Hence, using notation $\theta_i^{(1-,\max)}$:

$$\theta_1^{(1-,\max)} = \max\{\log z;\ (8.8) \text{ holds}\}, \qquad \theta_2^{(1-,\max)} = \max\{\log \xi;\ (8.8) \text{ holds}\},$$

we have

$$\left(\theta_1^{(1-,\mathrm{c})}, \theta_2^{(1-,\mathrm{c})}\right) = \begin{cases} (\log \rho_1^{-1}, \log \rho_1^{-1}), & \mu_2 > \lambda_2 + \delta, \\ (\theta_1^{(1-,\max)}, \theta_2^{(1-,\max)}), & \mu_2 \le \lambda_2 + \delta. \end{cases} \tag{8.10}$$

It is also noted that $\theta_1^{(1-,\max)} \ge \log \rho_1^{-1}$, so we always have that $\theta_1^{(1-,\mathrm{c})} \ge \log \rho_1^{-1}$.

Remark 8.1 The $\theta_i^{(1-,\max)}$ for $i = 1, 2$ are computed from their definitions as

$$\begin{aligned}
\theta_1^{(1-,\max)} &= \log \frac{1}{2\lambda_1}\left(1 - 2\sqrt{\mu_2(\lambda_2+\delta)} + \zeta_1^{(-)}\right),\\
\theta_2^{(1-,\max)} &= \log \frac{1 - 4(\lambda_1\mu_1 + (\lambda_2+\delta)\mu_2) + \zeta_2^{(-)}}{8\lambda_1(\lambda_2+\delta)}
\end{aligned}$$

where

$$\begin{aligned}
\zeta_1^{(-)} &= \sqrt{1 + 4(\mu_2(\lambda_2+\delta) - \sqrt{\mu_2(\lambda_2+\delta)} - \lambda_1\mu_1)}\,,\\
\zeta_2^{(-)} &= \sqrt{(1 - 4(\lambda_1\mu_1 + (\lambda_2+\delta)\mu_2))^2 - 64(\lambda_2+\delta)\lambda_1\mu_1\mu_2}\,.
\end{aligned}$$

Similarly, letting $z = e^{\theta_1}$ and $\xi = e^{\theta_2}$ in (8.6),

$$\lambda_2 z + \mu_1 z\xi^{-1} + \mu_2 z^{-1} + (\lambda_1 + \delta)z^{-1}\xi = 1, \tag{8.11}$$
$$\lambda_2 z + \mu_1 + \mu_2 z^{-1} + (\lambda_1 + \delta)z^{-1}\xi = 1. \tag{8.12}$$

Solving these equations for $z \neq 1$, we have $z = \xi = \rho_2^{-1}$. For $z = \rho_2^{-1}$, (8.11) yields $\xi = \rho_2^{-1}, \frac{\mu_1}{\lambda_1+\delta}\rho_2^{-1}$. Recalling that

$$\theta_1^{(1+,\max)} = \max\{\log z;\ (8.11) \text{ holds}\}, \qquad \theta_2^{(1+,\max)} = \max\{\log \xi;\ (8.11) \text{ holds}\},$$

we have that $\theta_1^{(1+,c)} \geq \log \rho_2^{-1}$ and

$$\left(\theta_1^{(1+,c)}, \theta_2^{(1+,c)}\right) = \begin{cases} (\log \rho_2^{-1}, \log \rho_2^{-1}), & \mu_1 > \lambda_1 + \delta \\ (\theta_1^{(1+,\max)}, \theta_2^{(1+,\max)}), & \mu_1 \leq \lambda_1 + \delta. \end{cases} \tag{8.13}$$

We also consider the solution of (8.7). In this case, let $\xi = e^{\theta_2}$, $z_1 = e^{\theta_{1-}}$ and $z_2 = e^{\theta_{1+}}$. Then, (8.7) becomes

$$\lambda_1 z_1 + \mu_2 z_1 \xi^{-1} + \mu_1 z_1^{-1} + (\lambda_2 + \delta) z_1^{-1}\xi = 1, \tag{8.14}$$
$$\lambda_2 z_2 + \mu_1 z_2 \xi^{-1} + \mu_2 z_2^{-1} + (\lambda_1 + \delta) z_2^{-1}\xi = 1, \tag{8.15}$$
$$\left(\lambda_1 + \frac{\delta}{2}\right) z_1 + \mu_2 z_1 \xi^{-1} + \mu_1 z_2 \xi^{-1} + \left(\lambda_2 + \frac{\delta}{2}\right) z_2 = 1. \tag{8.16}$$

These equations have been solved in Li, Miyazawa and Zhao [56]. That is, if $z \neq 1$, then $\xi = \rho^{-2}$ and $z_1 = z_2 = \rho^{-1}$. For $\xi = \rho^{-2}$, the first equation has solutions $z_1 = \rho^{-1}, \frac{\beta_1+\delta}{\beta_2}\rho^{-1}$, and the second equation yields $z_2 = \rho^{-1}, \frac{\beta_2+\delta}{\beta_1}\rho^{-1}$. In this case, $\theta_2^{(2,c)}$ is obtained as the maximum ξ that satisfies (8.14), (8.15) and

$$\left(\lambda_1 + \frac{\delta}{2}\right) z_1 + \mu_2 z_1 \xi^{-1} + \mu_1 z_2 \xi^{-1} + \left(\lambda_2 + \frac{\delta}{2}\right) z_2 \leq 1. \tag{8.17}$$

Thus, we need to solve a convex optimization problem. We already know that $(z_1, z_2, \xi) = (1, 1, 1), (\rho^{-1}, \rho^{-1}, \rho^{-2})$ are the extreme points of the set of all the points (z_1, z_2, ξ) that satisfy the constraints (8.17). To identify the latter point on the convex curves (8.14) and (8.15), it is convenient to introduce the following classifications:

$$\beta_2 + \delta > \beta_1, \qquad \beta_1 + \delta > \beta_2, \tag{8.18}$$
$$\beta_2 + \delta \leq \beta_1, \qquad \beta_1 + \delta > \beta_2, \tag{8.19}$$
$$\beta_2 + \delta > \beta_1, \qquad \beta_1 + \delta \leq \beta_2, \tag{8.20}$$

where we exclude the case that $\beta_2 + \delta \leq \beta_1$ and $\beta_1 + \delta \leq \beta_2$, which is impossible since $\delta > 0$. Note that (8.18) is equivalent to

$$|\beta_1 - \beta_2| < \delta,$$

which is introduced under the nomenclature "a strongly pooled condition" in Foley and McDonald [28].

We now find $\theta_2^{(2,c)}$ by solving the convex optimization problem.

Lemma 8.1 (Lemma 1.8 of Miyazawa [68]) If the strongly pooled condition (8.18) holds, then

$$\theta_2^{(2,c)} = \log \rho^{-2}, \qquad \theta_{1-}^{(2,c)} = \theta_{1+}^{(2,c)} = \log \rho^{-1}.$$

Otherwise, if (8.19) holds, then

$$(\theta_2^{(2,c)}, \theta_{1-}^{(2,c)}, \theta_{1+}^{(2,c)}) = \left(\theta_2^{(1-,\max)}, \log \frac{e^{\theta_2^{(2,c)}}}{2(\lambda_1 e^{\theta_2^{(2,c)}} + \mu_2)}, \arg \max_{(\theta_1, \theta_2^{(2,c)}) \in \mathcal{D}_0^{(+)}} \theta_1 \right),$$

and, if (8.20) holds, then

$$(\theta_2^{(2,c)}, \theta_{1-}^{(2,c)}, \theta_{1+}^{(2,c)}) = \left(\theta_2^{(1+,\max)}, \arg \max_{(\theta_1, \theta_2^{(2,c)}) \in \mathcal{D}_0^{(-)}} \theta_1, \log \frac{e^{\theta_2^{(2,c)}}}{2(\lambda_2 e^{\theta_2^{(2,c)}} + \mu_1)} \right).$$

We need another classification:

$$\rho_1 < \rho, \qquad \rho_2 < \rho, \tag{8.21}$$
$$\rho_1 \geq \rho, \qquad \rho_2 < \rho, \tag{8.22}$$
$$\rho_1 < \rho, \qquad \rho_2 \geq \rho, \tag{8.23}$$

where we do not consider the case that $\rho_1 \geq \rho$ and $\rho_2 \geq \rho$, which is impossible since $\delta > 0$. The condition (8.21) is referred to as a weakly pooled condition in Foley and McDonald [28].

Under the conditions (8.18) and (8.21), the asymptotic decay of

$$\mathbb{P}(\min(L_1, L_2) = n, L_1 - L_2 = \ell), \qquad n \to \infty,$$

is shown to be exactly geometric with decay rate $-\log(\rho^2)$ for each fixed ℓ in Foley and McDonald [28] while some other cases are obtained in Li, Miyazawa and Zhao [56]. We are ready to present a complete answer.

Theorem 8.1 (Theorem 1.5 of Miyazawa [68]) For the generalized join the shortest queue with two queues, suppose that the stability conditions $\rho < 1$, $\rho_1 < 1$ and $\rho_2 < 1$ are satisfied. Then, the geometric decay rate $r_2 \equiv e^{-\theta_2^{(2,c)}}$ exists for the minimum of the two queues in the sense of marginal distribution as well as jointly with each fixed difference of the two queues, and one of the following three cases occurs.

(S1) If (8.18) holds, then either one of the following cases occurs:

- (S1a) (8.21) implies $r_2 = \rho^2$.
- (S1b) (8.22) implies $r_2 = \dfrac{\lambda_2 + \delta}{\mu_2} \rho_1$.
- (S1c) (8.23) implies $r_2 = \dfrac{\lambda_1 + \delta}{\mu_1} \rho_2$.

(S2) If (8.19) holds, then either one of the following cases occurs:

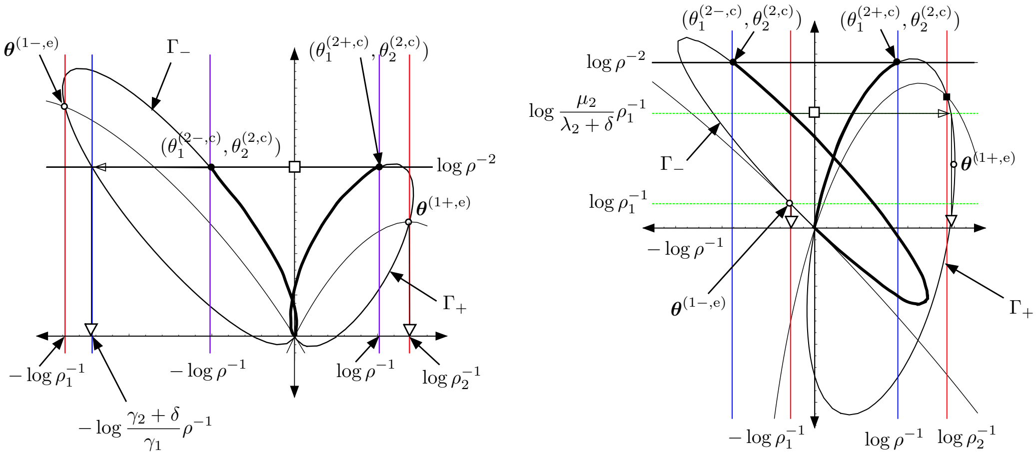

Figure 13: The decay rates for strongly pooled (8.18): case (S1a) for (8.21) and case (S1b) for (8.22).

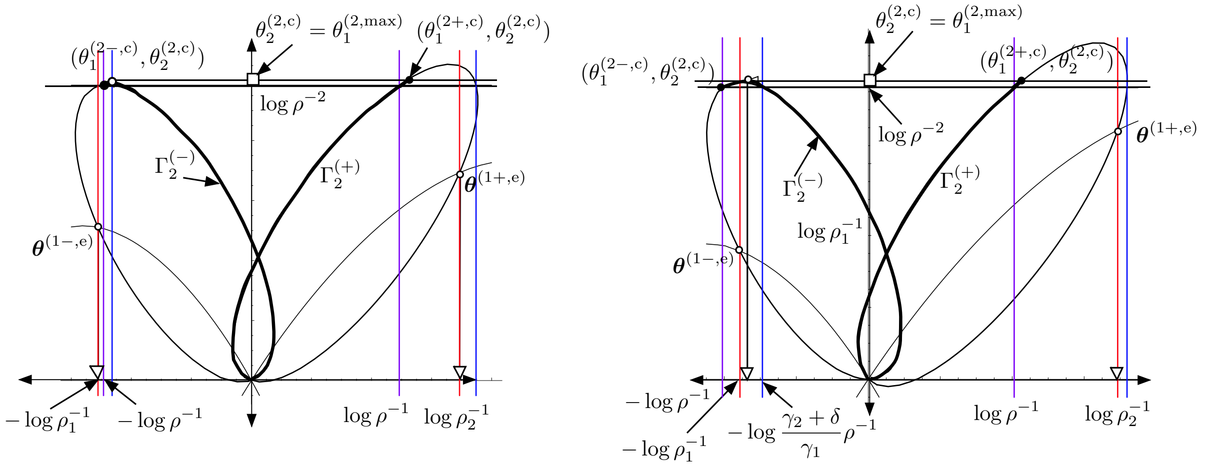

Figure 14: The decay rates for not strongly pooled (8.19): case (S2a) for (8.21) and case (S2b) for (8.22).

(S2a) (8.21) implies $r_2 = \begin{cases} e^{-\theta_2^{(1-,\max)}}, & \theta_{1+}^{(2,\mathrm{c})} \leq \theta_1^{(1+,\mathrm{c})}, \\ \dfrac{\lambda_1+\delta}{\mu_1}\rho_2, & \theta_{1+}^{(2,\mathrm{c})} > \theta_1^{(1+,\mathrm{c})}. \end{cases}$

(S2b) (8.22) implies

$$r_2 = \begin{cases} e^{-\theta_2^{(1-,\max)}}, & \theta_{1-}^{(2,\mathrm{c})} < \log \rho_1^{-1}, \theta_{1+}^{(2,\mathrm{c})} < \theta_1^{(1+,\mathrm{c})}, \\ \dfrac{\lambda_2+\delta}{\mu_2}\rho_1, & \theta_{1-}^{(2,\mathrm{c})} \geq \log \rho_1^{-1}, \theta_1^{(1+,\mathrm{c})} < \theta_1^{(1+,\mathrm{c})}, \\ \dfrac{\lambda_1+\delta}{\mu_1}\rho_2, & \theta_{1-}^{(2,\mathrm{c})} < \log \rho_1^{-1}, \theta_{1+}^{(2,\mathrm{c})} \geq \theta_1^{(1+,\mathrm{c})}, \\ \min\left(\dfrac{\lambda_2+\delta}{\mu_2}\rho_1, \dfrac{\lambda_1+\delta}{\mu_1}\rho_2\right), & \theta_{1-}^{(2,\mathrm{c})} \geq \log \rho_1^{-1}, \theta_{1+}^{(2,\mathrm{c})} \geq \theta_1^{(1+,\mathrm{c})}. \end{cases}$$

(S2c) (8.23) implies $r_2 = \dfrac{\lambda_1+\delta}{\mu_1}\rho_2.$

(S3) If (8.20) holds, then either one of the following cases occurs.

(S3a) (8.21) implies $r_2 = \begin{cases} e^{-\theta_2^{(1+,\max)}}, & \theta_{1-}^{(2,\mathrm{c})} \le \theta_1^{(1-,\mathrm{c})} \\ \dfrac{\lambda_2+\delta}{\mu_2}\rho_1, & \theta_{1-}^{(2,\mathrm{c})} > \theta_1^{(1-,\mathrm{c})}. \end{cases}$

(S3b) (8.22) implies $r_2 = \dfrac{\lambda_2+\delta}{\mu_2}\rho_1$.

(S3c) (8.23) implies

$$r_2 = \begin{cases} e^{-\theta_2^{(1+,\max)}}, & \theta_{1-}^{(2,\mathrm{c})} < \theta_1^{(1-,\mathrm{c})}, \theta_{1+}^{(2,\mathrm{c})} < \log\rho_2^{-1} \\ \dfrac{\lambda_2+\delta}{\mu_2}\rho_1, & \theta_{1-}^{(2,\mathrm{c})} \ge \theta_1^{(1-,\mathrm{c})}, \theta_{1+}^{(2,\mathrm{c})} < \log\rho_2^{-1} \\ \dfrac{\lambda_1+\delta}{\mu_1}\rho_2, & \theta_{1-}^{(2,\mathrm{c})} < \theta_1^{(1-,\mathrm{c})}, \theta_{1+}^{(2,\mathrm{c})} \ge \log\rho_2^{-1} \\ \min\left(\dfrac{\lambda_2+\delta}{\mu_2}\rho_1, \dfrac{\lambda_1+\delta}{\mu_1}\rho_2\right), & \theta_{1-}^{(2,\mathrm{c})} \ge \theta_1^{(1-,\mathrm{c})}, \theta_{1+}^{(2,\mathrm{c})} \ge \log\rho_2^{-1}. \end{cases}$$

Furthermore, the decay rates are exactly geometric for the cases (S1), (S2) unless $\theta_{1-}^{(2,\mathrm{c})} = \theta_1^{(1-,\max)}$ and (S3) unless $\theta_{1+}^{(2,\mathrm{c})} = \theta_1^{(1+,\max)}$.

Note that the cases (S2) and (S3) are symmetric. See Figure 13 for (S1) and Figure 14 for (S2).

9 Concluding remarks

This paper focuses on the tail asymptotic problem for the stationary distribution of reflecting processes. We have mainly considered two-dimensional reflecting random walks and two-dimensional SRBM, and had some discussions for higher-dimensional reflecting processes. However, we have not discussed much about relaxing the modeling assumptions and said nothing about limiting behaviors of a sequence of reflecting processes constrained in bounded (or partially bounded) regions as the regions are expanded. The latter is important for applications because buffers cannot have infinite capacity in actual queueing networks. In this section, we address these two issues.

We first list possible changes about the modeling assumptions.

(i) Relaxing technical assumptions such as the aperiodicity and irreducibility of the random walk and the non-singularity of the covariance matrix of the Brownian motion (see conditions (3c), (3d) and (3-i)).

(ii) Removing the skip-free assumption in all directions.

(iii) Allowing the reflecting random walk to be a real vector-valued.

(iv) Modulating the reflecting process by a background process.

As for (i), some of them were discussed in Miyazawa [69]. Li and Zhao [57] studied the priority queue with two types of customers, which is the case that the random walk is not irreducible. We may need more systematic studies on this issue.

To implement (ii) and (iii), we need to appropriately define a reflection mechanism. One such attempt was made by Borovkov and Mogul'skii [9]. They use a thick boundary, but it considerably complicates analysis. Another way is to introduce a function to return to the boundary when the boundary is overshot. This is something similar to the reflection mapping of an SRBM. The third one is to change the transition probabilities from the interior to the boundary. This has been considered for a reflecting Markov additive process in Miyazawa and Zwart [74].

As for (iv), a reflecting Markov additive process is proposed in [74]. This class of models is important to more accurately describe queueing networks. Even a finite background space is useful. There are some studied in this direction for $d = 2$ (see, e.g., [32, 46, 88]) for a so-called generalized Jackson network, which replaces the Poisson arrivals by the renewal arrivals and allows service times to be generally distributed. No satisfactory answer had been obtained, but Ozawa [81] very recently solved this problem in a certain way using the framework presented in Section 5.6.

Of course, it is much more interesting to extend the tail asymptotic results of the two-dimensional processes to higher-dimensional cases. We have already proposed one program in Section 6.2. There are some related conjectures. Miyazawa [66] conjectured the decay rates for the generalized Jackson network. Miyazawa and Kobayashi [70] make a similar conjecture for an SRBM, which is in the same line as that conjectured in Section 6.2. In a very recent work, Kobayashi, Sakuma and Miyazawa [52] solved the join the shortest queue problem for an arbitrary number of parallel queues using a similar technique presented in Section 6. This is a sign for the multidimensional problem for $d \geq 3$ to be solvable.

We finally consider a sequence of the reflecting random walks with bounded state spaces whose limit is unbounded. In general, those processes with bounded state spaces may be interesting for numerical computations, but are less interesting for theoretical study because analytically tractable results cannot be expected. This leads us to consider their limiting behavior.

There are two papers studying this issue. Kroese, Scheinhardt and Taylor [54] considered the effect of buffer truncation for the two-node Jackson tandem queue, and found the condition that the buffer full probabilities converge to the decay rate of that of no buffer truncation. They also found the limiting buffer full probability when this condition is not satisfied. Those results are extended for the two-node Jackson network by Sakuma and Miyazawa [87]. It would be nice to investigate this limiting behavior for the general reflecting random walks on $\mathbb{Z}_+^d$.

Acknowledgement

I got into the area of the tail asymptotic problem on the stationary distribution around 2000. Since then, I have benefited from colleagues in this and related areas. In particular, I am grateful to Jim Dai, Robert Foley, David McDonald, Tomasz Rolski and Yiqiang

Zhao for numerous discussions and great inputs. During this period, I was invited to give special or plenary talks at four conferences in Tokyo in 2000, Adelaide in 2002, Toledo in 2010, and Ottawa in 2010. I am grateful to their organizers, Masakazu Kojima, Peter Taylor, Jesus Artalejo, and Yiqiang Zhao. This paper would not be completed without encouragements from Jesus Artalejo, and it much benefited by thorough corrections of Peter Taylor. So a special thanks is dedicated to both of them.

A part of this paper was prepared during my stay at Cambridge in 2010 for attending the research program "Stochastic Processes in Communication Sciences". I am grateful to its organizers including Sergey Foss for inviting me to this stimulating program, and I thank for financial support from the Isaac Newton Institute for Mathematical Sciences. This research is supported in part by by the Japan Society for the Promotion of Science under grant No. 21510165.

References

[1] I. Adan. R.D. Foley and D.R. McDonald (2009) Exact asymptotics for the stationary distribution of a Markov chain: a production model, *Queueing Systems* 62, 311–344.

[2] G. Alsmeyer (1994) On the Markov renewal theorem, *Stoch. Proc. their Appl.* 50, 37–56.

[3] E. Arjas and T.P. Speed (1973) Symmetric Wiener–Hopf factorizations in Markov additive processes, *Z. Wahrcheinlich. verw. Geb.* 26, 105–118.

[4] F. Avram, J.G. Dai and J.J. Hasenbein (2001) Explicit solutions for variational problems in the quadrant, *Queueing Systems* 37, 259–289.

[5] D. Bertsimas, I. C. Paschalidis and J. N. Tsitsiklis (1998) On the large deviations behavior of acyclic networks of $G/G/1$ queues. *Annals of Applied Probability* 8, 1027–1069.

[6] A.A. Borovkov (1998) *Ergodicity and Stability of Stochastic Processes*, translated from Russian, Wiley, New York.

[7] A.A. Borovkov and A.A. Mogul'skii (1996) The second rate function and the asymptotic problems of renewal and hitting the boundary for multidimensional random walks, *Siberian Mathematical Journals* 38, 647–682.

[8] A.A. Borovkov and A.A. Mogul'skii (2000) Integro-local limit theorems including large deviations for sums of random vectors. II, *Theory of Probability and its applications* 45, 3–22.

[9] A.A. Borovkov and A.A. Mogul'skii (2001) Large deviations for Markov chains in the positive orthant, *Russian Math. Surveys* 56, 803–916.

[10] M. Bramson, J.G. Dai and J.M. Harrison (2010) Positive recurrence of reflecting Brownian motion in three dimensions, *The Annals of Applied Probability* 20, 753–783.

[11] C.S. Chang (1995) Sample path large deviations and intree networks. *Queueing Systems* 20. 7–36.

[12] X. Chao, M. Miyazawa and M. Pinedo (1999) *Queueing Networks; customers, signals and product form solutions*, Wiley, Chichester.

[13] H. Chen and D.D. Yao (2001) *Fundamentals of Queueing Networks, Performance, Asymptotics, and Optimization*, Springer-Verlag, New York.

[14] E. Çinlar (1975) *Introduction to Stochastic Processes*, Prentice-Hall, Englewood Cliffs, New Jersey.

[15] J. Collamore (1996) Hitting probabilities and large deviations, *Annals of Probability* 24, 2065–2078.

[16] J.G. Dai and M. Miyazawa (2010) Reflecting Brownian motion in two dimensions: Exact asymptotics for the stationary distribution. Working paper.

[17] G. Doetsch (1974) *Introduction to the Theory and Application of the Laplace Transformation*, Springer, Berlin.

[18] P. Dupuis and R.S. Ellis (1997), *A Weak Convergence Approach to the Theory of Large Deviations*, John Wiley & Sons, INC., New York.

[19] P. Dupuis and K. Ramanan (2002), A time-reversed representation for the tail probabilities of stationary reflected Brownian motion, *Stochastic Processes and Their Applications* 98, 253–287.

[20] A. El Kaharroubi, A. Yaacoubi, A.B. Tahar and K. Bichard (2010) Variational problem in the non-negative orthant of $\mathbb{R}^3$: Reflective faces and boundary influence, submitted for publication.

[21] G. Fayolle and R. Iasnogorodski (1979) Two coupled processors: The reduction to a Riemann–Hilbert problem, *Z. Wahrscheinlichkeitstheorie* 47, 325–351.

[22] G. Fayolle, R. Iasnogorodski and V.A. Malyshev (1999) *Random Walks in the Quarter-Plane: Algebraic Methods, Boundary Value Problems and Applications*, Springer, New York.

[23] G. Fayolle, V.A. Malyshev and M.V. Menshikov (1995) *Topics in the Constructive Theory of Countable Markov Chains*, Cambridge University Press, Cambridge.

[24] W.L. Feller (1971) *An Introduction to Probability Theory and Its Applications*, 2nd edition, John Wiley & Sons, New York.

[25] P. Flajolet and R. Sedqewick (2009) *Analytic Combinatorics*, Cambridge University Press, Cambridge, UK.

[26] L. Flatto and S Hahn (1984) Two parallel queues by arrivals with two demands I. *SIAM Journal on Applied Mathematics* 44, 1041–1053.

[27] L. Flatto and H.P. McKean (1977) Two queues in parallel. *Communications on Pure and Applied Mathematics* 30, 255–263.

[28] R.D. Foley and D.R. McDonald (2001) Join the shortest queue: stability and exact asymptotics, *Annals of Applied Probability* 11(3), 569–607.

[29] R.D. Foley and D.R. McDonald (2005) Large deviations of a modified Jackson network: Stability and rough asymptotics, *The Annals of Applied Probability* 15, 519–541.

[30] R.D. Foley and D.R. McDonald (2005) Bridges and networks: exact asymptotics, *The Annals of Applied Probability* 15, 542–586.

[31] S.N. Ethier and T.G. Kurtz (1986) *Markov Processes: Characterization and Convergence*, Wiley, New York.

[32] K. Fujimoto, Y. Takahashi and N. Makimoto (1998) Asymptotic properties of stationary distributions in two stage tandem queueing systems. *J. Opns. Res. Soc. Japan* 41, 118–141.

[33] D. Gamarnik (2002) On deciding stability of constrained homogeneous random walks and queueing systems, *Mathematics of Operations Research* 27, 272–293.

[34] D. Gamarnik (2007) Computing stationary probability distribution and large deviations rates for constrained homogeneous random walks. The undecidability results, *Mathematics of Operations Research* 32, 257–265

[35] P. Glynn and W. Whitt (1994) Logarithmic asymptotics for steady state tail probabilities in a single-server queue. *J. Appl. Prob.* 31A, 131–156.

[36] W.K. Grassmann and D.P. Heyman (1990) Equilibrium distribution of blocked-structured Markov chains with repeating rows, *J. Appl. Prob.* 27, 557–576.

[37] F. Guillemin and J.S.H. van Leeuwaarden (2011) Rare event asymptotics for a random walk in the quarter plane, *Queueing Systems* 67, 1–32.

[38] L. Haque, L. Liu and Y.Q. Zhao (2005) Sufficient conditions for a geometric tail in a QBD process with countably many levels and phases, *Stochastic Models* 21, 77–99.

[39] J.M. Harrison (2003) A broader view of Brownian networks, *Annals of Applied Probability* 13, 1119–1150.

[40] J.M. Harrison and J.J. Hasenbein (2009) Reflected Brownian motion in the quadrant: tail behavior of the stationary distribution, *Queueing Systems* 61, 113–138

[41] J.M. Harrison and R.J. Williams (1987) Brownian models of open queueing networks with homogeneous customer populations, *Stochastics* 22, 77–115.

[42] Q. He, H. Li and Y.Q. Zhao (2009) Light-tailed behavior in QBD process with countably many phases, *Stochastic Models* 25, 50–75.

[43] E. Hille and R. S. Phillips (1957) *Functional Analysis and semi-groups*, American Mathematical Society, Rhode Island.

[44] I. Ignatiouk-Robert (2001) Sample path large deviations and convergence parameters, *The Annals of Applied Probability* 11, 1292–1329.

[45] I.A. Ignatyuk, V.A. Malyshev and V.V. Scherbakov (1994) Boundary effects in a large deviation problems, *Russian Mathematical Survey*, 49, 41–99.

[46] K. Katou, N. Makimoto and Y. Takahashi (2008) Upper bound for the decay rate of the joint queue-length distribution in a two-node Markovian queueing system, *Queueing Systems* 58, 161–189.

[47] O. Kella and M. Miyazawa: Parallel fluid queues with constant inflows and simultaneous random reductions. *J. Appl Prob.*, 38 (2001) 609–620.

[48] A. Khanchi (2010) Asymptotic hitting distribution for a reflected random walk in the positive quadrant. To appear in *Stochastic Models*.

[49] J.F.C. Kingman (1970) Inequalities in the theory of queues. *J. Roy. Stat. Soc. B* 32, 883–909.

[50] J.F.C. Kingman (1961) Two similar queues in parallel, *Annals of Mathematical Statistics* 32, 1314–1323.

[51] M. Kobayashi and M. Miyazawa (2010) The tail asymptotic behavior of the stationary distribution of a double $M/G/1$ process and their applications to a batch arrival Jackson network, submitted for publication (http://queue3.is.noda.sut.ac.jp/miyazawa/mm-paper).

[52] M. Kobayashi, Y. Sakuma and M. Miyazawa (2011) Join the shortest queue among k parallel queues: tail asymptotics of its stationary distribution, the proceedings of the Queueing Symposium; Stochastic models and their applications, 143–152, Kyoto, Japan.

[53] M. Kobayashi, M. Miyazawa and Y.Q. Zhao (2010) Tail asymptotics of the occupation measure for a Markov additive process with an $M/G/1$-type background process, *Stochastic Models* 26. 463–486.

[54] Kroese D.P., Scheinhardt W.R.W. and Taylor P.G. (2003) Spectral properties of the tandem Jackson network seen as a quasi-birth-and-death process, *Annals of Applied Probability* 14, 2057–2089.

[55] L.A. Kurkova and Y.M. Suhov (2003) Malyshev's theory and JS-queues. Asymptotics of stationary probabilities. *Annals of Applied Probability* 13, 1313–1354.

[56] H. Li, M. Miyazawa and Y.Q. Zhao (2007) Geometric decay in a QBD process with countable background states with applications to shortest queues, *Stochastic Models* 23, 413–438.

[57] H. Li and Y.Q. Zhao (2009) Exact tail asymptotics in a priority queue–characterizations of the preemptive model, *Queueing Systems* 63, 355–381.

[58] H. Li and Y.Q. Zhao (2010) Tail asymptotics for a generalized two-demand queueing models – A kernel method. Submitted.

[59] P. Lieshout and M. Mandjes (2007) Brownian tandem queues, *Mathematical Methods in Operations Research* 66, 275–298.

[60] P. Lieshout and M. Mandjes (2008) Asymptotic analysis of Lévy-driven tandem queues, *Queueing Systems* 60, 203–226.

[61] K. Majewski (1996) Large Deviations of Stationary Reflected Brownian Motions, in *Stochastic Networks: Theory and Applications*, edited by F. P. Kelly and S. Zachary and I. Ziedins, Oxford University Press.

[62] K. Majewski (1998) Large deviations of the steady state distribution of reflected processes with applications to queueing systems, *Queueing Systems* 29, 351–381.

[63] K. Majewski (2004) Large deviation bounds for single class queueing networks and their calculation, *Queueing Systems* 48, 103–134.

[64] A.I. Markushevich (1977) *Theory of functions*, 2nd edition, translated by R.A. Silverman, reprinted by American Mathematical Society.

[65] M. Miyazawa (2002) A paradigm of Markov additive processes for queues and their networks. *The Proceedings of the Fourth International Conference on Matrix-Analytic Methods in Stochastic Models: Matrix-Analytic Methods, Theory and Applications* 265–289,World Scientific, Singapore.

[66] M. Miyazawa (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.

[67] M. Miyazawa (2004) The Markov renewal approach for the stationary distributions in M/G/1 type queues with countably many background states, *Queueing Systems* 46, 177–196.

[68] M. Miyazawa (2009) Two Sided DQBD Process and Solutions to the Tail Decay Rate Problem and Their Applications to the Generalized Join Shortest Queue, in *Advances in Queueing Theory and Network Applications*, 3–33.

[69] M. Miyazawa (2009) Tail Decay Rates in Double QBD Processes and Related Reflected Random Walks, *Mathematics of Operations Research* 34, 547–575.

[70] M. Miyazawa and M. Kobayashi (2010). Conjectures on tail asymptotics of the stationary distribution for a multidimensional SRBM, submitted for publication (http://queue3.is.noda.sut.ac.jp/miyazawa/mm-paper).

[71] M. Miyazawa and T. Rolski (2009) Exact asymptotics for a Levy-driven tandem queue with an intermediate input, *Queueing Systems* 63, 323–353.

[72] M. Miyazawa and P.G. Taylor (1997) A geometric product-form distribution for a queueing network with non-standard batch arrivals and batch transfers. *Adv.Appl.Prob.* 29, 523–544

[73] M. Miyazawa and Y.Q. Zhao (2004) The stationary tail asymptotics in the $GI/G/1$ type queue with countably many background states, *Advances in Applied Probability* 36, 1231–1251.

[74] M. Miyazawa and B. Zwart (2009) Wiener–Hopf factorizations for a multidimensional Markov additive process and their applications to reflected processes, submitted for publication (http://queue3.is.noda.sut.ac.jp/miyazawa/mm-paper).

[75] M.F. Neuts (1981) *Matrix-Geometric Solutions in Stochastic Models: An Algorithmic Approach,* The Johns University Press, Baltimore.

[76] M.F. Neuts (1989) *Structured Stochastic Matrices of $M/G/1$ Type and Their Applications*, Mercel Dekker, New York.

[77] M.F. Neuts and Y. Takahashi (1997) Asymptotic behavior of the stationary distributions in the $GI/PH/c$ queue with heterogeneous servers. *Zeitschrift Wahrscheinlichkeitstheorie* 57 441–452.

[78] P. Ney and E. Nummelin (1987) Markov additive processes I. Eigenvalue properties and limit theorems, *The Annals of Probability* 15, 561–592.

[79] P. Ney and E. Nummelin (1987) Markov additive processes II. Large deviations, *The Annals of Probability* 15, 593–609.

[80] E. Nummelin (1984) *General irreducible Markov chains and non-negative operators*, Cambridge University Press, Cambridge.

[81] T. Ozawa (2011) Asymptotics for the stationary tail distributions in discrete-time two-dimensional QBD processes, the proceedings of the Queueing Symposium; Stochastic models and their applications, 254–263, Kyoto, Japan.

[82] J.W. Pitman (1974) An identity for stopping times of a Markov process. In: *Studies in Probability and Statistics*, ed. E.J. Williams. Jerusalem Academic Press 41–57.

[83] A.A. Puhalskii and A.A. Vladimirov (2007) A Large Deviation Principle for Join the Shortest Queue, *Mathematics of Operations Research* 32, 700–710.

[84] M.I Reiman and R.J. Williams (1988). A boundary property of semimartingale reflecting Brownian motions. *Probability Theory and Related Fields* 77, 87–97. Correction: 80, 633 (1989).

[85] J. S. Sadowsky (1995) The probability of large queue lengths and waiting times in a heterogeneous multiserver queue I: Positive recurrence and logarithmic limits. *Adv. Appl. Prob.* 27, 567–583.

[86] J. S. Sadowsky and W. S. Szpankowski (1995) The probability of large queue lengths and waiting times in a heterogeneous multiserver queue I: Tight limits. *Adv. Appl. Prob.* 27, 532-566.

[87] Y. Sakuma and M. Miyazawa (2005) On the effect of a finite buffer in a two node Jackson network, *Journal of Applied Probability* 42, 199–222.

[88] Y. Sakuma, M. Miyazawa and Y. Q. Zhao (2006) Decay rate for a $PH/M/2$ queue with shortest queue discipline. *Queueing Systems* 53, 189–201.

[89] E. Seneta(1981) *Non-negative Matrices and Markov chains*, second edition, Springer-Verlag: New York.

[90] R. F. Serfozo (1999) *Introduction to Stochastic Networks*, Springer-Verlag, New York.

[91] A. Shwartz and A. Weiss (1995) *Large Deviations For Performance Analysis, Queues, communications, and computing*, Chapman & hall, London.

[92] M. Sipser (1997) *Introduction to the Theory of Computability*, PWS Publishing Company, Boston, MA.

[93] Y. Takahashi (1981) Asymptotic exponentiality of the tail of the waiting-time distribution in a $PH/PH/c$ queue. *Adv. Appl. Prob.* 13, 619–630.

[94] Y. Takahashi, K. Fujimoto and N. Makimoto, (2001) Geometric decay of the steady-state probabilities in a quasi-birth-and-death process with a countable number of phases, *Stochastic Models* 17, 1–24.

[95] L.M. Taylor and R.J. Williams (1993) Existence and uniqueness of semimartingale reflecting Brownian motions in an orthant. *Probability Theory and Related Fields* 96, 283–317.

[96] J. Tang and Y.Q. Zhao (2008) Stationary tail asymptotics of a tandem queue with feedback, *Annals of Operations Research* 160, 173–189.

[97] Y.Q. Zhao, W. Li and W.J. Braun (2003) Censoring, factorizations and spectral analysis for transition matrices with block-repeating entries, *Methodology and Computing in Applied Probability* 5, 35–58.