MARTINGALE APPROACH FOR TAIL ASYMPTOTIC PROBLEMS IN THE GENERALIZED JACKSON NETWORK

BY

MASAKIYO MIYAZAWA[*] (NODA)

Dedicated to Professor Tomasz Rolski for his 70th birthday

Abstract. We study the tail asymptotic of the stationary joint queue length distribution for a generalized Jackson network (GJN for short), assuming its stability. For the two-station case, this problem has recently been solved in the logarithmic sense for the marginal stationary distributions under the setting that arrival processes and service times are of phase-type. In this paper, we study similar tail asymptotic problems on the stationary distribution, but problems and assumptions are different. First, the asymptotics are studied not only for the marginal distribution but also the stationary probabilities of state sets of small volumes. Second, the interarrival and service times are generally distributed and light tailed, but of phase-type in some cases. Third, we also study the case that there are more than two stations, although the asymptotic results are less complete. For them, we develop a martingale method, which has been recently applied to a single queue with many servers by the author.

2010 AMS Mathematics Subject Classification: Primary: 60K25, 90B15; Secondary: 60G44, 60F10.

Key words and phrases: Generalized Jackson network, stationary distribution, tail asymptotic, piecewise deterministic Markov process, martingale, change of measure, instability.

1. INTRODUCTION

Asymptotic analyses have been actively studied in the recent queueing theory. This is because queueing models, particularly, queueing networks, become very complicated and their exact analyses are getting harder. We are interested in asymptotic analyses for large queues in a generalized Jackson network and aim to understand their asymptotic behaviors through its modeling primitives.

There are two different types of asymptotic analyses for large queues. One is for a given fixed model. Large deviations are typically studied for this. Another is to study them through an approximating model. For example, such a model is obtained as the limit of a sequence of models under heavy traffic by scaling time, space and/or modeling primitives. It is called *heavy traffic approximation* (e.g., see [20], [22]). Here, large queues are caused by heavy traffic. In this paper, we focus on the large deviations for a fixed model. Among them, we are particularly interested in the logarithmic tail asymptotics of the stationary distribution for a generalized Jackson network, GJN for short.

This problem has been studied by the standard approach of large deviations, but the decay rates are hard to obtain analytically by using modeling primitives (e.g., see [14]). The author [17] recently studied it by a matrix analytic method, and derived the decay rates for the marginal stationary distributions in an arbitrary direction for a two-station GJN, assuming phase-type distributions for service times and arrival processes, called a *phase-type setting*. We aim to generalize this result under a more general setting by a different approach.

Let $d \geqslant 2$ be the number of stations in the GJN. For $d = 2$, we relax the phase-type assumption, and consider the decay rates of the stationary probabilities for state sets of small volumes in addition to those of the marginal stationary distribution. For $d \geqslant 3$, we derive upper and lower bounds for those decay rates.

Our basic idea is to simplify the derivation of those asymptotic results in such a way that they are obtained in a similar manner to a reflecting random walk on a multidimensional orthant. This simplification greatly benefits for analysis although the decay rate problems for the reflecting random walk have not been fully solved even for $d = 2$. To this end, we take an approach studied for a single queue with heterogeneous servers in [18], and modify it for a queueing network. In this approach, we first describe the GJN by a piecewise deterministic Markov process, PDMP for short. We then derive martingales for change of measures, and formulate the asymptotic problems under a new measure.

PDMP is a continuous time Markov process whose sample path is composed of two parts, a continuous part, which is deterministic, and a discontinuous part, called *jumps*, by which randomness is created. Thus, PDMP is particularly suitable for queueing models. However, jump instants are random, and state changes at them are complicated. Because of this, PDMP is hard for analysis. So, other methods have been employed in queueing theory. For example, the state space is discretized by using phase-type distributions, and a Markov chain is obtained. Then, matrix analysis is applicable. This phase-type approach is numerically powerful but analytically less explicit because of state space description. Furthermore, it is getting harder to apply as a queueing model becomes complicated like a queueing network. We will not use such a matrix analysis. Nevertheless, it turns out that the phase-type assumption is helpful in our asymptotic analysis in some cases.

Contrary to the analytical difficulty, the PDMP has a simple sample path. Its time evolution is easily presented by a stochastic integral equation using a test

function, which maps the states of the PDMP to real values (see (2.7) below). In this stochastic equation, state changes at the jump instants cause difficulty for analysis as we mentioned above. Davis [9], who introduced PDMP, replaces those state changes at jump instants by a martingale and the so-called boundary condition on the test function.

However, it is not easy to find a good class of the test functions which characterize a distribution on the state space of the PDMP. The idea of [18] is to choose a smaller class of test functions to overcome those difficulties. We then have a semi-martingale which cannot characterize a distribution on the state space, but still retains full information to study large queues. Once the semi-martingale is obtained, we use the standard technique for change of measure through constructing an exponential martingale, called a *multiplicative functional*.

In applying this martingale approach to the GJN, we need to know how the network model is changed under the new measure. Intuitively, some of its stations must be unstable for the tail asymptotic analysis to work. To study this instability problem, we will use the fact that the network structure is unchanged under the change of measure, and therefore the stability of each station is characterized by the traffic intensity at that station. These traffic intensities are obtained from the traffic equations, but they are non-linear because of unstable stations. Thus, the instability problem is not obvious. We challenge it, and find some sufficient conditions for the GJN to be partly unstable under the new measure, which depends on the choice of a martingale for change of measure.

This paper is made up of four sections. In Section 2, the GJN (generalized Jackson network) is described by a PDMP, and a martingale for change of measure is derived. This section also considers geometric interpretations of the stability condition of the GJN, and present main results for the asymptotic problems. Section 3 discusses the method of change of measure, and considers how the GJN is changed under a new measure. In Section 4, the main results are proved. For this, we first list major steps for deriving upper and lower bounds, then prepare several lemmas to complete the proofs.

In this paper, we will use real vectors in the following way. Column and row vectors and their dimensions are not specified as long as they can be identified in the context where they are used. Their inequality holds in entrywise; $\mathbf{e}_k$ is the unit vector whose k-th entry is one while all the other entries vanish; $\mathbf{1}$ is the vector all of whose entries are one. The inner product $\sum_i x_i y_i$ of vectors $\boldsymbol{x}, \boldsymbol{y}$ of the same dimension is denoted by $\langle \boldsymbol{x}, \boldsymbol{y} \rangle$, and $\|\boldsymbol{x}\| = \sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle}$; $\boldsymbol{x}$ is said to be a unit direction vector if $\boldsymbol{x} \geqslant 0$ and $\|\boldsymbol{x}\| = 1$. We denote the set of all unit direction vectors in $\mathbb{R}_+^d$ by $\overrightarrow{U}_d$. For $\boldsymbol{x}$ in a finite-dimensional vector space $\mathcal{S}$ and its subset B, we will use the convention that $\boldsymbol{x} + B = \{\boldsymbol{x} + \boldsymbol{y} \in \mathcal{S}; \boldsymbol{y} \in B\}$. For a finite set A, its cardinality is denoted by $|A|$.

year. I have benefited from those conferences about not only academic collaborations but also personal interactions. This paper is one of such outcomes, and I am grateful to Tomasz's great academic contribution as well as his warm hospitality.

2. GENERALIZED JACKSON NETWORK

We are concerned with a queueing network which has a finite number of stations with single servers and single class of customers. At each station, there is an infinite buffer, exogenous customers arrive subject to a renewal process if any, and customers are served in first-come-first-served manner by independent and identically distributed service times. Furthermore, the renewal process and service times are independent of everything else. Customers who complete service at a station are independently routed to the next stations or leave the network according to a given probability. We refer to this queueing network as a GJN (*generalized Jackson network*).

2.1. Notation and assumptions. Let us introduce the notation for a GJN. Let d be the total number of stations. We index stations by elements in $\mathcal{J} \equiv \{1, 2, \ldots, d\}$, and let $\mathcal{E}$ be the set of the stations which have exogenous arrivals. For each station i, let $F_{e,i}$ for $i \in \mathcal{E}$ be the interarrival time distribution of exogenous customers, and let $F_{s,i}$ for $i \in \mathcal{J}$ be the service time distribution. Let p_{ij} be the probability that a customer completing service at station i is routed to station j for $i, j \in \mathcal{J}$, where those customers leave the outside of the network with probability

$$p_{i0} \equiv 1 - \sum_{i \in \mathcal{J}} p_{ij}.$$

To exclude trivial cases, we assume that $d \times d$ matrix $P \equiv \{p_{ij}; i, j \in \mathcal{J}\}$ is strictly substochastic, and $(d+1) \times (d+1)$ matrix $\overline{P} \equiv \{p_{ij}; i, j \in \{0\} \cup \mathcal{J}\}$ is irreducible, where $p_{00} = 0$, and $p_{0i} > 0$ only if $i \in \mathcal{E}$, where the value of p_{0i} is specified later. We call P a *routing matrix*, while $\overline{P}$ is called an *over all routing matrix*.

At time t, let $L_i(t)$ be the number of customers in station $i \in \mathcal{J}$, and let $R_{s,i}(t)$ be the residual service time of a customer being served there if any, where we set up a new service time just after service completion, and this service time is unchanged as long as station i is empty. Thus, all $R_{s,i}(t)$ are always positive because of the right continuity, and $R_{s,i}(t-)$ vanishes only at service completion instants. For $i \in \mathcal{E}$, let $R_{e,i}(t)$ be the residual time to the next exogenous arrival at station i.

Denote the vectors whose i-th entries are $L_i(t)$, $R_{s,i}(t)$ for $i \in \mathcal{J}$ and $R_{e,i}(t)$ for $i \in \mathcal{E}$ by $\boldsymbol{L}(t)$, $\boldsymbol{R}_s(t)$, $\boldsymbol{R}_e(t)$, respectively, and define $X(t)$ and $\boldsymbol{R}(t)$ as

$$X(t) = \big(\boldsymbol{L}(t), \boldsymbol{R}_e(t), \boldsymbol{R}_s(t)\big), \quad \boldsymbol{R}(t) = \big(\boldsymbol{R}_e(t), \boldsymbol{R}_s(t)\big), \quad t \geqslant 0.$$

Then, $\{p_{ij}; i, j \in \mathcal{J}\}$, $\{F_{e,i}; i \in \mathcal{E}\}$ and $\{F_{s,i}; i \in \mathcal{J}\}$ are the modeling primitives, and the state space S for $X(t)$ is given by

$$S = \{(\boldsymbol{z}, \boldsymbol{y}_e, \boldsymbol{y}_s); \boldsymbol{z} \in \mathbb{Z}_+^d, \boldsymbol{y}_e \in \mathbb{R}_+^{\mathcal{E}}, \boldsymbol{y}_s \in \mathbb{R}_+^d\},$$

where $\mathbb{Z}_+$ and $\mathbb{R}_+$ are the sets of all nonnegative integers and all nonnegative real numbers, respectively. As usual, we assume that $X(t)$ is right-continuous and has left-hand limits. Let $\{\mathcal{F}_t; t \geqslant 0\}$ be a filtration generated by histories of all the sample paths of $X(\cdot)$; then $\mathcal{F}_t$ is right-continuous, and $\{X(t); t \geqslant 0\}$ is an $\mathcal{F}_t$-Markov process.

Let $\widehat{F}_{e,i}$ and $\widehat{F}_{s,i}$ be the moment generating functions, MGF for short, of the distributions $F_{e,i}$ and $F_{s,i}$, respectively. We define $\beta_{w,i}$ and $\theta_{w,i}$ for $w = e, s$ as

$$\beta_{w,i} = \sup\{\theta \in \mathbb{R}; \widehat{F}_{w,i}(\theta) < \infty\}, \quad \theta_{w,i} = \inf\{\theta \in \mathbb{R}; e^{-\theta} < \widehat{F}(\beta_{w,i})\}.$$

We will assume that $\beta_{w,i} > 0$ and $\theta_{w,i} = \infty$ for all $w = e, s$ and i. That is, all the distributions, $F_{w,i}$, have light tail and their moment generating functions diverge at their singular points. These conditions are assumed for technical simplicity.

For some important cases, we have to restrict these distributions in the following class. A positive random variable T or its distribution is said to have a *conditional MGF with a uniform bound* if there is a function h of $\theta > 0$ such that $\mathbb{E}(e^{\theta T}) < \infty$ implies that

$$(2.1) \qquad\qquad \mathbb{E}(e^{\theta(T-t)}|T > t) \leqslant h(\theta), \quad \theta > 0,$$

as long as $\mathbb{P}(T > t) > 0$. Obviously, if T is bounded, it satisfies the condition (2.1). Another obvious example is an NBU distribution, which is characterized by $\mathbb{P}(T > s + t | T > s) \leqslant \mathbb{P}(T > t)$ for $s, t > 0$. An important class for our application is of phase-type, which is defined as

$$(2.2) \qquad\qquad F(t) \equiv \mathbb{P}(T \leqslant t) = 1 - \boldsymbol{a}e^{tU}\boldsymbol{1}, \quad t \geqslant 0,$$

where $\boldsymbol{a}$ is a finite-dimensional probability vector, and U is a defective transition rate matrix with the same dimension as $\boldsymbol{a}$ such that $(-U)^{-1}$ is finite.

LEMMA 2.1. *A phase-type distribution has a conditional MGF with a uniform bound.*

P r o o f. Assume that F is given by (2.2). Let T be a random variable subject to F, and let $\boldsymbol{b}(s) = \boldsymbol{a}e^{sU}/(\boldsymbol{a}e^{sU}\boldsymbol{1})$. Then $\boldsymbol{b}(s)$ is a probability vector, and

$$\mathbb{P}(T > s + t | T > t) = \boldsymbol{b}(t)e^{sU}\boldsymbol{1}, \quad s, t > 0,$$

and therefore

$$\mathbb{E}(e^{\theta(T-t)}|T > t) = \boldsymbol{b}(t)(-U)(\theta I + U)^{-1}\boldsymbol{1}, \quad t \geqslant 0, \theta > 0,$$

which is finite as long as $\mathbb{E}(e^{\theta T}) = \boldsymbol{a}(-U)(\theta I + U)^{-1}\boldsymbol{1}$ is finite. Hence, we have (2.1) by letting $h(\theta)$ be the maximum of all the entries of the vector $(-U)(\theta I + U)^{-1}\boldsymbol{1}$. $\blacksquare$

Thus, we consider the tail asymptotic problem for the GJN assuming the distributions of $T_{e,i}, T_{s,j}$ to have light tails, and, in some cases, we assume the following:

(A1) All the $T_{e,i}$ for $i \in \mathcal{E}$ and $T_{s,i}$ for $j \in \mathcal{J}$ have conditional MGF with uniform bounds, that is, satisfy (2.1).

Let $\lambda_i = 1/\mathbb{E}(T_{e,i})$ for $i \in \mathcal{E}$ and $\mu_i = 1/\mathbb{E}(T_{s,i})$ for $i \in \mathcal{J}$. For convenience, we put $\lambda_i = 0$ for $i \in \mathcal{J} \setminus \mathcal{E}$. We now put $p_{0i} = \lambda_i / \sum_{j \in \mathcal{J}} \lambda_j$ for $i \in \mathcal{J}$ in $\overline{P}$. Let α_i for $i \in \mathcal{J}$ be the solutions of the following traffic equation:

$$(2.3) \qquad \alpha_i = \lambda_i + \sum_{j \in \mathcal{J}} \alpha_j p_{ji}, \quad i \in \mathcal{J}.$$

It is easy to see that the solutions uniquely exist because the routing matrix P is strictly substochastic and $\overline{P}$ is irreducible. Let $\rho_i = \alpha_i \mathbb{E}(T_{s,i})$, and assume the stability condition, that is,

$$(2.4) \qquad \rho_i < 1, \quad i \in \mathcal{J}.$$

In Section 2.5, we will consider the case where some of stations are unstable. This case occurs under change of measure, and ρ_i is no longer a right traffic intensity.

2.2. Piecewise deterministic Markov process (PDMP). In this paper, we consider $\{X(t); t \geqslant 0\}$ as a piecewise deterministic Markov process, PDMP for short, introduced by Davis [9]. PDMP is a Markov process with piecewise deterministic and continuously differentiable sample path and finitely many discontinuous epochs in each finite time interval. Its randomness arises at discontinuous epochs, which are uniquely determined by hitting times when the deterministic sample path gets into a specified state set. The set of those discontinuous epochs constitutes a counting process, and the piecewise deterministic sample path is randomly changed at those times. We here assume that there is no other discontinuous state change. This slightly changes the standard description of PDMP due to [9], but it is a matter of formulation since Davis' PDMP can be described by the present formulation as well.

We now introduce the notation to describe $X(t)$ as a PDMP. Let N be a counting process for the expiring times of all the remaining times. That is,

$$N(t) = \sum_{u \in (0,t]} \Big(\sum_{i \in \mathcal{E}} 1\big(\Delta R_{e,i}(u) > 0\big) + \sum_{i \in \mathcal{J}} 1\big(\Delta R_{s,i}(u) > 0\big) \Big), \quad t \geqslant 0,$$

where Δ is the difference operator such that $\Delta f(t) = f(t) - f(t-)$ for a function f which is right-continuous and has left-hand limits. Clearly, N counts all the discontinuous points of $X(t)$. However, it may multiply counts at the same instant, and therefore $\Delta N(t)$ may be greater than one. To avoid this, we define a simplification of N as

$$N^*(t) = \sum_{u \in (0,t]} 1\big(\Delta N(u) > 0\big), \quad t \geqslant 0.$$

We then let $t_0 = 0$, and inductively define $t_n = \inf\{u > t_{n-1}; \Delta N^*(u) > 0\}$ for $n = 1, 2, \ldots$ Thus, t_n is the n-th discontinuous epoch of $X(t)$, and a stopping time with respect to $\mathcal{F}_t$.

Between times t_{n-1} and t_n, $X(t)$ is linearly changing, so continuously differentiable in such a way that

$$\frac{d}{dt}L_i(t) = 0, \quad \frac{d}{dt}R_{e,i}(t) = -1(i \in \mathcal{E}), \quad \frac{d}{dt}R_{s,i}(t) = 1\big(R_{s,i}(t) > 0\big).$$

This differentiation can be described by an operator $\mathcal{A}$ on $C^1(S)$, which is the set of all continuously differentiable functions from S to $\mathbb{R}$. Namely, $\mathcal{A}$ is defined as

$$(2.5) \quad \mathcal{A}f(\boldsymbol{x}) = -\sum_{i \in \mathcal{E}} \frac{\partial}{\partial y_{e,i}} f(\boldsymbol{z}, \boldsymbol{y}_e, \boldsymbol{y}_s) - \sum_{i \in \mathcal{J}} \frac{\partial}{\partial y_{s,i}} f(\boldsymbol{z}, \boldsymbol{y}_e, \boldsymbol{y}_s) 1(z_i \geqslant 1).$$

Since the PDMP is a strong Markov process, the conditional distribution of $X(t_n)$ given $\mathcal{F}_{t_n-}$ is a function of $X(t_n-)$ for each $n \geqslant 1$, which is characterized by the transition kernel $\mathcal{K}$ given by

$$(2.6) \qquad \mathcal{K}f\big(X(t-)\big) = \mathbb{E}\Big(f\big(X(t)\big)|X(t-)\Big), \quad X(t-) \in \Gamma,$$

for $f \in \mathcal{M}(S)$, where Γ is the set of $\boldsymbol{x} \equiv (\boldsymbol{z}, \boldsymbol{y}_e, \boldsymbol{y}_s) \in S$ such that

$$\exists i \in \mathcal{E}, y_{e,i} = 0 \quad \text{or} \quad \exists i \in \mathcal{J}, z_i \geqslant 1, y_{s,i} = 0.$$

This Γ is referred to as a terminal set, while $\mathcal{K}$ is referred to as a jump kernel.

2.3. Martingale decomposition of the PDMP. From (2.5) and (2.6) and the counting process N^*, we have a time evolution equation for $f \in C^1(S)$, that is,

$$(2.7) \quad f\big(X(t)\big) = f\big(X(0)\big) + \int_0^t \mathcal{A}f\big(X(u)\big)du + \int_0^t \Delta f\big(X(u)\big)dN^*(u).$$

We refer to f as a test function, as is typically called.

We apply the same martingale method as discussed in [18]. We repeat it here briefly for this paper to be self-contained. We first note that

$$M_0(t) \equiv \int_0^t \Big(f\big(X(u)\big) - \mathcal{K}f\big(X(u-)\big)\Big)dN^*(u), \quad t \geqslant 0,$$

is an $\mathcal{F}_t$-martingale if $\mathbb{E}\big(|M(t)|\big) < \infty$. Since

$$\Delta f\big(X(u)\big) = f\big(X(u)\big) - \mathcal{K}f(u-) + \mathcal{K}f\big(X(u-)\big) - f\big(X(u-)\big),$$

it follows from (2.7) that

$$(2.8) \quad M_0(t) = f\big(X(t)\big) - f\big(X(0)\big)$$
$$- \Big[\int\limits_0^t \mathcal{A}f\big(X(s)\big)ds + \int\limits_0^t \Big(\mathcal{K}f\big(X(s-)\big) - f\big(X(s-)\big)\Big)dN^*(s) \Big].$$

We define $M(\cdot)$ and $D(\cdot)$ as

$$(2.9) \qquad M(t) = f\big(X(t)\big) - f\big(X(0)\big) - \int\limits_0^t \mathcal{A}f\big(X(s)\big)ds, \quad t \geqslant 0,$$

$$(2.10) \qquad D(t) = \int\limits_0^t \Big(\mathcal{K}f\big(X(s-)\big) - f\big(X(s-)\big)\Big)dN^*(s).$$

Since

$$(2.11) \qquad\qquad M(t) = M_0(t) + D(t), \quad t \geqslant 0,$$

we have the following fact.

LEMMA 2.2. *For the PDMP $X(\cdot)$, if the condition*

$$(2.12) \qquad\qquad \mathcal{K}f(\boldsymbol{x}) = f(\boldsymbol{x}), \quad \forall \boldsymbol{x} \in \Gamma,$$

is satisfied and if $\mathbb{E}\big(|M(t)|\big) < \infty$ for all $t \geqslant 0$, then $M(\cdot)$ is an $\mathcal{F}_t$-martingale.

We refer to (2.12) as a terminal condition following the terminology of [18].

2.4. Terminal condition for the GJN. A key of our arguments is to find a set of test functions satisfying the terminal condition (2.12). For this, we mainly apply the following test function, parameterized by $\boldsymbol{\theta} \in \mathbb{R}^d$,

$$(2.13) \qquad f_{\boldsymbol{\theta}}(\boldsymbol{x}) = e^{\langle\boldsymbol{\theta},\boldsymbol{z}\rangle - \langle\boldsymbol{\gamma}_e(\boldsymbol{\theta}),\boldsymbol{y}_e\rangle - \langle\boldsymbol{\gamma}_s(\boldsymbol{\theta}),\boldsymbol{y}_s\rangle}, \quad \boldsymbol{x} \equiv (\boldsymbol{z}, \boldsymbol{y}_e, \boldsymbol{y}_s) \in S,$$

using some vector-valued functions $\boldsymbol{\gamma}_e(\boldsymbol{\theta}) \in \mathbb{R}^{\mathcal{E}}$ and $\boldsymbol{\gamma}_s(\boldsymbol{\theta}) \in \mathbb{R}^{\mathcal{J}}$, where we recall that $\langle\boldsymbol{a}, \boldsymbol{b}\rangle$ is the inner product of vectors $\boldsymbol{a}, \boldsymbol{b}$ of the same dimensions. In some cases, it needs to truncate some of $y_{e,i}$ and $y_{s,j}$ as $y_{e,i} \wedge v$ and $y_{s,j} \wedge v$ for $v > 0$, which causes the change of $\gamma_{e,i}(\theta_i)$ to $\gamma_{e,i}(v, \theta_i)$, as we will see, where $a \wedge b = \min(a, b)$ for $a, b \in \mathbb{R}$. By $J_e(v)$ we denote the set of $i \in \mathcal{E}$ such that $y_{e,i}$ is truncated by v. Similarly, $J_s(v)$ denotes the set of $i \in \mathcal{J}$ such that $y_{s,i}$ is truncated by v for $i \in J_s(v)$. Let $\boldsymbol{J}(v) = \big(J_e(v), J_s(v)\big) \subset \mathcal{E} \times \mathcal{J}$. Then, the test function $f_{\boldsymbol{\theta}}$ is changed to

$$(2.14) \qquad\qquad f_{\boldsymbol{J}(v),\boldsymbol{\theta}}(\boldsymbol{x}) = e^{\langle\boldsymbol{\theta},\boldsymbol{z}\rangle - w_{\boldsymbol{J}(v)}(\boldsymbol{\theta},\boldsymbol{y})}, \quad \boldsymbol{x} \equiv (\boldsymbol{z}, \boldsymbol{y}) \in S,$$

where $\boldsymbol{y} = (\boldsymbol{y}_e, \boldsymbol{y}_s)$ and

$$(2.15) \qquad w_{\boldsymbol{J}(v)}(\boldsymbol{\theta}, \boldsymbol{y}) = \sum_{i \in J_e(v)} \gamma_{e,i}(v, \theta_i)(y_{e,i} \wedge v) + \sum_{i \in \mathcal{E} \setminus J_e(v)} \gamma_{e,i}(\theta_i) y_{e,i}$$
$$+ \sum_{i \in J_s(v)} \gamma_{s,i}(v, \theta_i)(y_{s,i} \wedge v) + \sum_{i \in \mathcal{J} \setminus J_e(v)} \gamma_{s,i}(\theta_i) y_{s,i}.$$

Obviously, $f_{\emptyset, \boldsymbol{\theta}}(\boldsymbol{x}) = f_{\boldsymbol{\theta}}(\boldsymbol{x})$.

Our first task is to determine functions $\gamma_{e,i}(v, \cdot)$ and $\gamma_{s,j}(v, \cdot)$ for $v \in (0, \infty]$ so that the terminal condition (2.12) is satisfied, where $\gamma_{u,i}(\cdot) = \gamma_{u,i}(\infty, \cdot)$ for $u = e, s$. For this, we first consider a prototype for them as we have done in Section 2.3 of [18]. Let T be a positive-valued random variable, and denote its distribution by F. We truncate T by a positive number v as $T^{(v)} \equiv T \wedge v$, and denote the distribution of $T^{(v)}$ by $F^{(v)}$. We denote the moment generating functions of F and $F^{(v)}$ by $\widehat{F}$ and $\widehat{F}^{(v)}$, respectively.

Note that $\widehat{F}^{(v)}(\theta)$ exists and is finite for all $\theta \in \mathbb{R}$, but this may not be true for $\widehat{F}(\theta) \equiv \widehat{F}^{(\infty)}(\theta)$. For $v \in (0, \infty]$, let

$$(2.16) \qquad \beta_F^{(v)} = \sup\{\theta \in \mathbb{R}; \widehat{F}^{(v)}(\theta) < \infty\},$$

$$(2.17) \qquad \theta_F^{(v)} = \inf\{\theta \in \mathbb{R}; e^{-\theta} < \widehat{F}^{(v)}(\beta_F^{(v)})\};$$

then $\beta_F^{(v)} = \infty$ and $\theta_F^{(v)} = -\infty$ for $v < \infty$, while they may be finite for $v = \infty$, where $\theta_F^{(v)} \leqslant 0$, since $\beta_F^{(v)} \geqslant 0$. Note that $\widehat{F}(\theta)$ is finite for $\theta < \beta_F^{(\infty)}$. Define ξ to be a solution of the equation

$$(2.18) \qquad e^{\theta} \widehat{F}^{(v)}(\xi) = 1, \quad \theta \in \mathbb{R}, v > 0.$$

Obviously, ξ uniquely exists for each $\theta \in \mathbb{R}$ and $v > 0$. We denote it by $\xi_F(v, \theta)$. It has the following properties, which are proved in Lemma 2.4 of [18].

LEMMA 2.3. *For each fixed $v > 0$:*
(a) *$\xi_F(v, 0) = 0$, and $\xi_F(v, \theta)$ is strictly decreasing and concave in $\theta \in \mathbb{R}$.*
(b) *$\xi_F(v, \theta)$ is positive and decreasing in v for each fixed $\theta < 0$.*
(c) *$\xi_F(v, \theta)$ is negative and increasing in v for each fixed $\theta > 0$.*
(d) *$\xi_F(v, \theta)$ is differentiable in θ, and*

$$\frac{\partial}{\partial \theta} \xi_F(v, \theta) = -\frac{e^{-\theta}}{(\widehat{F}^{(v)})'(\xi_F(v, \theta))}.$$

We define $\xi_F^{(\triangle)}(\theta)$ and $\xi_F(\theta)$ as

$$\xi_F^{(\triangle)}(\theta) = \lim_{v \uparrow \infty} \xi_F(v, \theta), \ \theta \in \mathbb{R}, \quad \xi_F(\theta) = \xi_F(+\infty, \theta), \ \theta > \theta_F^{(\infty)},$$

which exist and are finite. These functions have some nice properties. For them, we cite Lemma 2.5 of [18] in which $\overline{\theta} = \infty$ in the present case.

LEMMA 2.4. (a) $\xi_F^{(\Delta)}(\theta)$ is nonincreasing and concave for all $\theta \in \mathbb{R}$.

(b) $\xi_F^{(\Delta)}(\theta)$ has the following properties:

$$\xi_F^{(\Delta)}(\theta) = \begin{cases} \beta_F^{(\infty)}, & \theta \leqslant \theta_F^{(\infty)}, \\ \xi_F(\theta), & \theta > \theta_F^{(\infty)}, \end{cases} \qquad \frac{d}{d\theta}\xi_F^{(\Delta)}(\theta) = \begin{cases} 0, & \theta < \theta_F^{(\infty)}, \\ (\xi_F^{(\infty)})'(\theta), & \theta > \theta_F^{(\infty)}, \end{cases}$$

where $(\xi_F^{(\infty)})'(\theta_F^{(\infty)})$ is the derivative from the right if $\theta_F^{(\infty)}$ is finite, and

$$(2.19) \qquad e^\theta \widehat{F}\big(\xi_F^{(\Delta)}(\theta)\big) \begin{cases} \leqslant 1, & \theta < \theta_F^{(\infty)}, \\ = 1, & \theta \geqslant \theta_F^{(\infty)}. \end{cases}$$

Throughout the paper, we assume that

$$(2.20) \qquad \theta_{F_{e,i}}^{(\infty)} = -\infty, \ i \in \mathcal{E}, \qquad \theta_{F_{s,i}}^{(\infty)} = -\infty, \ i \in \mathcal{J},$$

which means that $F_{e,i}$ and $F_{s,j}$ have light tails, and their moment generating functions diverge at the upper boundaries of their convergence domains. This assumption can be removed by using $\xi_F^{(\Delta)}$ as shown in [18] for a single queue. However, it will be complicated for a queueing network. This is the reason why we assume (2.20).

Let

$$q_i(\boldsymbol{\theta}) = e^{-\theta_i}\Big(\sum_{j \in \mathcal{J}} p_{ij}e^{\theta_j} + p_{i0}\Big), \qquad \boldsymbol{\theta} \in \mathbb{R}^d, \ i \in \mathcal{J}.$$

Then $q_i(\boldsymbol{\theta}) > 0$ and is convex in $\boldsymbol{\theta}$, and it is easy to see that

$$\log q_i(\boldsymbol{\theta}) = -\theta_i + \log\Big(\sum_{j \in \mathcal{J}} p_{ij}e^{\theta_j} + p_{i0}\Big)$$

is a convex function of $\boldsymbol{\theta} \in \mathbb{R}^d$ because $\sum_{j \in \mathcal{J}} p_{ij}e^{\theta_j} + p_{i0}$ is a sum of convex functions (see the lemma of [12]).

We now define $\gamma_{e,i}(v, \theta_i)$ and $\gamma_{s,i}(v, \boldsymbol{\theta})$ for $v \in (0, \infty]$ as

$$(2.21) \qquad \begin{aligned} \gamma_{e,i}(v, \theta_i) &= -\xi_{F_{e,i}^{(v)}}(\theta_i), \quad i \in \mathcal{E}, \\ \gamma_{s,i}(v, \boldsymbol{\theta}) &= -\xi_{F_{s,i}^{(v)}}\big(\log q_i(\boldsymbol{\theta})\big), \quad i \in \mathcal{J}. \end{aligned}$$

As informally mentioned, we let $\gamma_{e,i}(\theta_i) = \gamma_{e,i}(\infty, \theta_i)$ and $\gamma_{s,i}(\boldsymbol{\theta}) = \gamma_{s,i}(\infty, \boldsymbol{\theta})$. Due to the assumption (2.20), these functions are well defined for all $\boldsymbol{\theta} \in \mathbb{R}^d$. Clearly, their definitions are equivalent to

$$(2.22) \qquad \begin{aligned} e^{\theta_i} \widehat{F}_{e,i}^{(v)}\big(-\gamma_{e,i}(v, \theta_i)\big) &= 1, \quad i \in \mathcal{E}, \\ q_i(\boldsymbol{\theta})\widehat{F}_{s,i}^{(v)}\big(-\gamma_{s,i}(v, \boldsymbol{\theta})\big) &= 1, \quad i \in \mathcal{J}. \end{aligned}$$

These equations mean that $\Delta R_{e,i}(t)$ and $\Delta R_{s,j}(t)$ at the jump instants are compensated by the change of the queue lengths so that the terminal condition (2.12) is satisfied. This is an intuitive background for the definitions of $\gamma_{e,i}, \gamma_{s,j}$.

REMARK 2.1. *The reader may wonder why the minus signs are needed in* (2.21) *because* $\gamma_{e,i}, \gamma_{s,j}$ *in the test functions* $f_{\boldsymbol{\theta}}$, *and* $f_{\boldsymbol{\theta}}$ *also have the minus signs, and they can be cancelled. The reason for this is that they have nice interpretations for large deviations. For example, let* $N_{e,i}(t)$ *be the number of arrivals at station* $i \in \mathcal{E}$ *by time t. Then* $N_{e,i}(\cdot)$ *is a renewal process, and Glynn and Whitt* [10] *show that*

$$(2.23) \qquad \lim_{t\to\infty} \frac{1}{t} \log \mathbb{E}(e^{\theta_i N_{e,i}(t)}) = \gamma_{e,i}(\theta_i), \qquad \theta_i > \theta_{F_{e,i}}^{(\infty)},$$

for any initial distribution for $N_{e,i}(\cdot)$. *This suggests that* $\gamma_{e,i}(\theta_i)$ *must be one of the key information for the tail asymptotic of our problem. However, we will not use this property of* $\gamma_{e,i}$ *because the definition* (2.21) *is sufficiently informative for our analysis.*

Note that $\gamma_{e,i}(v, \theta_i)$ and $\gamma_{s,i}(v, \boldsymbol{\theta})$ are convex in θ_i and $\boldsymbol{\theta}$, respectively, because $\xi_{F_{e,i}^{(v)}}(\theta)$ and $\xi_{F_{s,i}}(\boldsymbol{\theta})$ are decreasing and concave in $\theta \in \mathbb{R}$ and $\log q_i(\boldsymbol{\theta})$ is convex. For $v \in (0, \infty]$, and $\boldsymbol{J}(v) \equiv \big(J_e(v), J_s(v)\big) \subset \mathcal{E} \times \mathcal{J}$, let, for $\boldsymbol{\theta} \in \mathbb{R}^d$,

$$(2.24) \qquad \gamma_{\boldsymbol{J}(v)}(\boldsymbol{\theta}) = \sum_{i \in J_e(v)} \gamma_{e,i}(v, \theta_i) + \sum_{i \in \mathcal{E} \setminus J_e(v)} \gamma_{e,i}(\theta_i)$$
$$+ \sum_{i \in J_s(v)} \gamma_{s,i}(v, \theta_i) + \sum_{i \in \mathcal{J} \setminus J_s(v)} \gamma_{s,i}(\theta_i),$$

and $\gamma(\boldsymbol{\theta}) = \gamma_{\emptyset}(\boldsymbol{\theta})$, that is,

$$\gamma(\boldsymbol{\theta}) = \sum_{i \in \mathcal{E}} \gamma_{e,i}(\theta_i) + \sum_{i \in \mathcal{J}} \gamma_{s,i}(\boldsymbol{\theta}).$$

Furthermore, $\gamma_{\boldsymbol{J}(v)}(\boldsymbol{\theta})$ converges to $\gamma(\boldsymbol{\theta})$ for each $\boldsymbol{\theta} \in \mathbb{R}^d$ as $v \to \infty$, which is uniform on a compact set of $\boldsymbol{\theta}$. The next lemma is a key for our arguments, and easily follows from Lemma 3.2 in [4]. We also remarked its intuitive meaning below (2.22). So far, its proof is omitted.

LEMMA 2.5. *For* $v \in (0, \infty]$, *the test function* $f_{\boldsymbol{J}(v), \boldsymbol{\theta}}$ *of* (2.14) *satisfies the terminal condition* (2.12) *with equality for all* $\boldsymbol{\theta} \in \mathbb{R}^d$, *respectively.*

We next consider a martingale for the test functions $f_{\boldsymbol{J}(v), \boldsymbol{\theta}}$. Denote the probability measure for $X(\cdot)$ with the initial state $\boldsymbol{x} \in S$ by $\mathbb{P}_{\boldsymbol{x}}$, and let $\mathbb{E}_{\boldsymbol{x}}$ stand for the expectation under $\mathbb{P}_{\boldsymbol{x}}$. We first note that

$$(2.25) \qquad \mathbb{E}_{\boldsymbol{x}}\Big(f_{\boldsymbol{J}(v), \boldsymbol{\theta}}\big(X(t)\big)\Big) < \infty, \qquad t \geqslant 0,$$

always holds for each $\boldsymbol{x} \in S$ and $\boldsymbol{\theta} \in \mathbb{R}^d$ because the total number of exogenous arrivals and service completions in each finite time interval has a super-light tail (lighter than any exponential decay); see, e.g., Lemma 4.1 of [18] for the single queue case. Hence, Lemmas 2.2 and 2.5 immediately imply the following fact.

LEMMA 2.6. *Fix* $\boldsymbol{\theta} \in \mathbb{R}^d$ *and* $\boldsymbol{x} \in S$. *For the PDMP* $X(\cdot)$ *and the test function* $f_{\boldsymbol{\theta}}$ *of* (2.14), *let, for* $t \geqslant 0$,

$$(2.26) \quad M_{\boldsymbol{J}(v),\boldsymbol{\theta}}(t) = f_{\boldsymbol{J}(v),\boldsymbol{\theta}}\big(X(t)\big) - f_{\boldsymbol{J}(v),\boldsymbol{\theta}}\big(X(0)\big)$$

$$+ \int\limits_0^t \gamma_{J(v)}(\boldsymbol{\theta}) f_{\boldsymbol{J}(v),\boldsymbol{\theta}}\big(X(u)\big) du$$

$$- \sum_{i \in J_e(v)} \gamma_{e,i}(v,\theta_i) \int\limits_0^t 1\big(R_{e,i}(u) > v\big) f_{\boldsymbol{J}(v),\boldsymbol{\theta}}\big(X(u)\big) du$$

$$- \sum_{i \in J_s(v)} \gamma_{s,i}(v,\boldsymbol{\theta}) \int\limits_0^t 1\big(R_{s,i}(u) > v\big) f_{\boldsymbol{J}(v),\boldsymbol{\theta}}\big(X(u)\big) du$$

$$- \int\limits_0^t \sum_{i \in J_e(v)} \gamma_{s,i}(v,\boldsymbol{\theta}) 1\big(L_i(u) = 0\big) f_{\boldsymbol{J}(v),\boldsymbol{\theta}}\big(X(u)\big) du$$

$$- \int\limits_0^t \sum_{i \in \mathcal{J} \setminus J_e(v)} \gamma_{s,i}(\boldsymbol{\theta}) 1\big(L_i(u) = 0\big) f_{\boldsymbol{J}(v),\boldsymbol{\theta}}\big(X(u)\big) du.$$

Then $M_{\boldsymbol{J}(v),\boldsymbol{\theta}}(\cdot)$ *is an* $\mathcal{F}_t$-*martingale under* $\mathbb{P}_{\boldsymbol{x}}$.

As always, $M_{\emptyset,\boldsymbol{\theta}}(\cdot)$ is simply denoted by $M_{\boldsymbol{\theta}}(\cdot)$, which also is an $\mathcal{F}_t$-martingale under $\mathbb{P}_{\boldsymbol{x}}$. Note that (2.26) may read as a semi-martingale representation of $f_{\boldsymbol{J}(v),\boldsymbol{\theta}}\big(X(t)\big)$.

2.5. Stability condition and geometric interpretation. As we mentioned in Section 2.1, the GJN (generalized Jackson network) is *stable* if the stability condition (2.4) holds. Except for trivial cases, it is also necessary. We will consider this network under change of measure, which is generally unstable, and it is important to see under what condition which station is unstable. To make these arguments clear, we formally define stability and instability for each station. Station i is said to be *weakly stable* (*stable*) if $L_i(t)$ is recurrent (positive recurrent, respectively), and to be *weakly unstable* (*unstable*) if $L_i(t)$ is null recurrent or transient (transient, respectively).

In this subsection, we so far do not assume the stability condition (2.4), and consider conditions for a station to be unstable (or stable). For this, we first need to compute an arrival rate at each station, which is obtained as the maximal solution $\{\widetilde{\alpha}_i; i \in \mathcal{F}\}$ of the following traffic equation (e.g., see [5], [6]):

$$(2.27) \qquad \widetilde{\alpha}_i = \lambda_i + \sum_{j \in \mathcal{J}} (\widetilde{\alpha}_j \wedge \mu_j) p_{ji}, \quad i \in \mathcal{J},$$

where we recall that $\mu_i = 1/\mathbb{E}(T_{s,i})$. Let $\widetilde{\rho}_i = \widetilde{\alpha}_i/\mu_i$, which may be different from ρ (see at the end of Section 2.1). Under appropriate conditions such as $T_{s,i}$ has a spread out distribution (see [1] for its definition), station i is weakly stable (stable) if and only if $\widetilde{\rho}_i \leqslant 1$ ($\widetilde{\rho}_i < 1$), and weakly unstable (unstable) if and only if $\widetilde{\rho}_i \geqslant 1$ ($\widetilde{\rho}_i > 1$, respectively).

It is easy to see that $\widetilde{\alpha}_i \leqslant \alpha_i$ for all $i \in \mathcal{J}$, where recall that α_i is the solution of the standard traffic equation (2.3). The $\widetilde{\alpha}_i$ can be numerically obtained from (2.27) in finite steps, but it is hard to get its analytic expression. For us, it is particularly important to give sufficient conditions in terms of $\gamma_{e,i}, \gamma_{s,i}$ for a station to be unstable or weakly unstable because these functions are well handled under change of measure. We first give sufficient conditions for instability in terms of λ_i, μ_i, p_{ij} and α_i.

LEMMA 2.7. (a) *For each* $j \in \mathcal{J}$, *if either* $\alpha_j \leqslant \mu_j$ *or*

$$(2.28) \qquad \lambda_j + \sum_{k \in \mathcal{J}} \mu_k p_{kj} \leqslant \mu_j$$

holds, then $\widetilde{\rho}_j \leqslant 1$. *That is, station* j *is weakly stable.*

(b) *If* $\mu_j < \alpha_j$ *and if* $\widetilde{\rho}_i \leqslant 1$ *for all* $i \in \mathcal{J} \setminus \{j\}$, *then* $\widetilde{\rho}_j > 1$, *that is, station* j *is unstable.*

(c) *If, for all* $j \in \mathcal{J}$,

$$(2.29) \qquad \lambda_j + \sum_{k \in \mathcal{J}} \mu_k p_{kj} \geqslant \mu_j,$$

then $\widetilde{\rho}_j \geqslant 1$ *for all* $j \in \mathcal{J}$. *That is, all stations are weakly unstable. If* (2.29) *holds with strict inequality for* $j = i$, *then* $\widetilde{\rho}_i > 1$, *that is, station* i *is unstable.*

REMARK 2.2. *For our application, it would be nice if* (b) *can be generalized in such a way that, for* $A \subset \mathcal{J}$, *if* $\mu_j < \alpha_j$ *for all* $j \in A$ *and if* $\widetilde{\rho}_i \leqslant 1$ *for all* $i \in \mathcal{J} \setminus A$, *then* $\widetilde{\rho}_j > 1$ *for all* $j \in A$. *Unfortunately, this is generally not true. A counterexample is easily obtained, for example, for a three-station tandem queue* (see Section 4 of [5] for some related discussions). *We need to update* α_j *using the information on the unstable station to be available to get such a generalization, but it would be less analytically tractable. Thus, we will not pursue it in this paper.*

P r o o f. (a) Since $\widetilde{\alpha}_j \leqslant \alpha_j$, it follows from $\alpha_j \leqslant \mu_j$ that $\widetilde{\rho}_j \leqslant 1$. If (2.28) holds, (2.27) implies that $\widetilde{\alpha}_j \leqslant \mu_j$, which is equivalent to $\widetilde{\rho}_j \leqslant 1$.

(b) Suppose that $\widetilde{\rho}_j \leqslant 1$, contrary to the claim; then $\widetilde{\rho}_k \leqslant 1$ for all $k \in \mathcal{J}$ by the second assumption. Hence, $\widetilde{\alpha}_k \leqslant \mu_k$ for all $k \in \mathcal{J}$, and therefore the non-linear traffic equation (2.27) is identical with the linear traffic equation (2.3). Thus, $\alpha_k = \widetilde{\alpha}_k \leqslant \mu_k$ for all $k \in \mathcal{J}$. This contradicts the assumption that $\mu_i < \alpha_i$, and therefore (b) is proved.

(c) Let $A = \{i \in \mathcal{J}; \widetilde{\alpha}_i < \mu_i\}$; then (2.27) can be written as

$$\widetilde{\alpha}_j = \lambda_j + \sum_{i \in A} \widetilde{\alpha}_i p_{ij} + \sum_{i \in \mathcal{J} \setminus A} \mu_i p_{ij}.$$

Hence, (2.29) implies that

$$\widetilde{\alpha}_j - \mu_j = \lambda_j - \mu_j + \sum_{i \in A} (\widetilde{\alpha}_i - \mu_i) p_{ij} + \sum_{i \in \mathcal{J}} \mu_i p_{ij} \geqslant \sum_{i \in A} (\widetilde{\alpha}_i - \mu_i) p_{ij}.$$

We then sum up both sides of this inequality for all $j \in A$, which yields

$$\sum_{i \in A} (\widetilde{\alpha}_i - \mu_i)\Big(1 - \sum_{j \in A} p_{ij}\Big) \geqslant 0.$$

Since $\widetilde{\alpha}_i - \mu_i < 0$ for $i \in A$, we must have

$$\sum_{j \in A} p_{ij} = 1, \quad i \in A,$$

which contradicts the irreducibility of the over all routing matrix $\overline{P}$, and therefore $A = \emptyset$. This proves the first half of (c). It also implies that $\widetilde{\alpha}_j = \mu_j$ for all $j \in \mathcal{J}$. Hence, if (2.29) holds with strict inequality, then (2.27) implies that

$$\widetilde{\alpha}_j = \lambda_j + \sum_{k \in \mathcal{J}} \mu_k p_{kj} > \mu_j.$$

This proves the remaining part of (c). ∎

We next characterize the conditions in Lemma 2.7 by the gradient vector of $\gamma(\boldsymbol{\theta})$ and $\gamma_{s,i}(\boldsymbol{\theta})$ at $\boldsymbol{\theta} = \mathbf{0}$. Define the gradient operator ∇ as

$$(2.30) \qquad \nabla\gamma(\boldsymbol{\theta}) = \left(\frac{\partial}{\partial \theta_1}\gamma(\boldsymbol{\theta}), \frac{\partial}{\partial \theta_2}\gamma(\boldsymbol{\theta}), \dots, \frac{\partial}{\partial \theta_d}\gamma(\boldsymbol{\theta}) \right).$$

Since

$$(2.31) \qquad \frac{\partial}{\partial \theta_i}\gamma_{e,i}(\theta_i)\Big|_{\theta_i=0} = \lambda_i, \qquad \frac{\partial}{\partial \theta_i}\gamma_{s,i}(\boldsymbol{\theta})\Big|_{\boldsymbol{\theta}=\mathbf{0}} = -\mu_i,$$

$$(2.32) \qquad \frac{\partial}{\partial \theta_j}\gamma_{s,i}(\boldsymbol{\theta})\Big|_{\boldsymbol{\theta}=\mathbf{0}} = \mu_i p_{ij}, \quad i, j \in \mathcal{J},$$

and $\boldsymbol{\alpha} = \boldsymbol{\lambda}(I - P)^{-1}$, we have

$$(2.33) \qquad \nabla\gamma(\mathbf{0}) = \boldsymbol{\lambda} - \boldsymbol{\mu}(I - P), \quad \nabla\gamma(\mathbf{0})(I - P)^{-1} = \boldsymbol{\alpha} - \boldsymbol{\mu}.$$

Using these facts, we have geometric interpretations for the conditions in Lemma 2.7 by the curves of $\gamma(\boldsymbol{\theta}) = 0$ and $\gamma_{s,i}(\boldsymbol{\theta}) = 0$ for $i \in \mathcal{J}$. For this, we introduce vectors $\boldsymbol{t}_i \in \mathbb{R}^d$ for $i \in \mathcal{J}$ such that

$$(2.34) \qquad \langle \nabla\gamma_{s,j}(\mathbf{0}), \boldsymbol{t}_i \rangle = 0, \ j \neq i, \quad \langle \nabla\gamma_{s,i}(\mathbf{0}), \boldsymbol{t}_i \rangle > 0.$$

Note that this $\boldsymbol{t}_i$ is uniquely determined except for its length $\|\boldsymbol{t}_i\|$.

LEMMA 2.8. *Let $T = (\boldsymbol{t}_1, \boldsymbol{t}_2, \dots, \boldsymbol{t}_d)$; then, for some positive vector $\boldsymbol{a}$,*

$$(2.35) \qquad T = -(I - P)^{-1}\Delta_{\boldsymbol{a}},$$

and therefore T is non-singular, and $\boldsymbol{t}_i \leqslant \mathbf{0}$ with $t_{ii} < 0$ for all $i \in \mathcal{J}$.

P r o o f. Since (2.31) yields

(2.36) $$\nabla\gamma_{s,j}(\mathbf{0}) = \mu_j\big((p_{j1}, p_{j2}, \ldots, p_{jd}) - \mathbf{e}_j\big), \quad j \in \mathcal{J},$$

(2.35) is immediate from (2.34). ∎

LEMMA 2.9. (a) *For $j \in \mathcal{J}$, the condition* (2.28) *holds if and only if the j-th entry of the gradient vector $\nabla\gamma(\mathbf{0})$ is not positive.*

(b) *For each $k \in \mathcal{J}$, $\mu_k < (=, >) \alpha_k$ if and only if $\langle\nabla\gamma(\mathbf{0}), \mathbf{t}_k\rangle < (=, >,$ respectively$)$ 0.*

REMARK 2.3. $\gamma_{s,i}(\boldsymbol{\theta}) = 0\ (> 0)$ *if and only if $q_i(\boldsymbol{\theta}) = 1\ (> 1, respectively)$, by the definition* (2.21), *so they present the same geometric curve. However, the gradients $\nabla\gamma_{s,i}(\boldsymbol{\theta})$ and $\nabla q_i(\boldsymbol{\theta})$ may not be identical. In particular, $\nabla\gamma_{s,i}(\mathbf{0}) = \mu_i\nabla q_i(\mathbf{0})$.*

P r o o f. (a) is immediate from the first equation of (2.33).
(b) It follows from (2.35) that

$$\boldsymbol{\alpha} - \boldsymbol{\mu} = \nabla\gamma(\mathbf{0})(I - P)^{-1} = -\nabla\gamma(\mathbf{0})(\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_d)\Delta_{\boldsymbol{a}}^{-1}.$$

Thus, $\mu_k - \alpha_k$ and $\langle\nabla\gamma(\mathbf{0}), \mathbf{t}_k\rangle$ have the same sign, which proves the claim. ∎
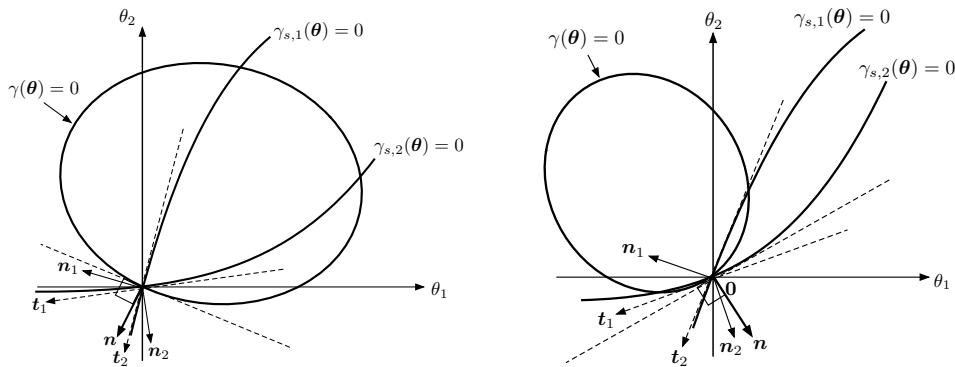

FIGURE 1. Geometric objects for $d = 2$, where $\mathbf{n} = \nabla\gamma(\mathbf{0})$, $\mathbf{n}_i = \nabla\gamma_i(\mathbf{0})$ for $i = 1, 2$.

Figure 1 illustrates the two cases for $d = 2$. The left panel shows that $\mathbf{n} < \mathbf{0}$ and $\langle\mathbf{n}, \mathbf{t}_i\rangle > 0$ for $i = 1, 2$, and both stations are stable by Lemma 2.9. The right panel shows that $n_1 > 0, n_2 < 0$, $\langle\mathbf{n}, \mathbf{t}_1\rangle < 0$ and $\langle\mathbf{n}, \mathbf{t}_2\rangle > 0$, and station 1 is unstable while station 2 is stable by Lemmas 2.7 and 2.9.

Lemma 2.9 together with Lemmas 2.7 and 2.8 provides us geometric interpretations of the stability and instability conditions for stations through curves $\gamma(\boldsymbol{\theta}) = 0$ and $\gamma_{s,i}(\boldsymbol{\theta}) = 0$ for $i \in \mathcal{J}$. We will use them for the GJN before and after change of measure.

2.6. Tail asymptotics. We now return to the assumption that the GJN is stable, that is, (2.4) holds. Under this assumption, we will use the following sets for considering the tail asymptotics of the stationary distribution. Let

$$\Gamma^{\text{IN}} = \{\boldsymbol{\theta} \in \mathbb{R}^d; \gamma(\boldsymbol{\theta}) < 0\}, \quad \Gamma^{\text{OUT}} = \{\boldsymbol{\theta} \in \mathbb{R}^d; \gamma(\boldsymbol{\theta}) > 0\},$$

$$\overleftarrow{\Gamma}^{\text{IN}} = \{\boldsymbol{\theta} \in \mathbb{R}^d; \boldsymbol{\theta} < \boldsymbol{\theta}', \exists \boldsymbol{\theta}' \in \Gamma^{\text{IN}}\}, \quad \overrightarrow{\Gamma}^{\text{OUT}} = \{\boldsymbol{\theta} \in \mathbb{R}^d; \boldsymbol{\theta}' < \boldsymbol{\theta}, \exists \boldsymbol{\theta}' \notin \overleftarrow{\Gamma}^{\text{IN}}\}.$$

For $A \subset \mathcal{J}$, let

$$\Gamma_A = \{\boldsymbol{\theta} \in \mathbb{R}^d; \gamma_{s,i}(\boldsymbol{\theta}) > 0, \forall i \in \mathcal{J} \setminus A\},$$

and put $\Gamma_A^{\text{IN}} = \Gamma^{\text{IN}} \cap \Gamma_A$ and

$$\overleftarrow{\Gamma}_A^{\text{IN}} = \{\boldsymbol{\theta} \in \overleftarrow{\Gamma}^{\text{IN}}; \boldsymbol{\theta} < \boldsymbol{\theta}', \exists \boldsymbol{\theta}' \in \Gamma_A^{\text{IN}}\}.$$

In particular, for $A = \{k\}$ with $k \in \mathcal{J}$ and $u = \text{\small IN}$ or $\text{\small OUT}$, Γ_A^u is simply denoted by Γ_k^u. Those sets are open and connected. We denote their boundaries by putting the operator ∂ like $\partial \Gamma^{\text{IN}}$, which is $\{\boldsymbol{\theta} \in \mathbb{R}^d; \gamma(\boldsymbol{\theta}) = 0\}$. Obviously, $\partial \Gamma^{\text{IN}} = \partial \Gamma^{\text{OUT}}$, and $\partial \overleftarrow{\Gamma}^{\text{IN}} = \partial \overrightarrow{\Gamma}^{\text{OUT}}$.

Note that Γ^{IN} is a non-empty bounded and convex set because $\gamma(\boldsymbol{\theta})$ is convex and diverges as $\|\boldsymbol{\theta}\|$ goes to infinity in any direction, and therefore Γ^{OUT} is also not empty. We check below that Γ_A^{IN} is not empty for $A \neq \emptyset$.

LEMMA 2.10. *Assume that the GJN is stable, and let $A \subset \mathcal{J}$.*

(a) *If $A \neq \emptyset$, then Γ_A^{IN} is not empty, and contains some $\boldsymbol{\theta} \geqslant \mathbf{0}$ with $\theta_i > 0$ for all $i \in A$.*

(b) *Define*

$$\Gamma_A^{\text{CX}} = \Big\{\boldsymbol{\theta} \in \Gamma^{\text{IN}}; \sum_{i \in \mathcal{E}} \gamma_{e,i}(\boldsymbol{\theta}) + \sum_{j \in A} \gamma_{s,j}(\boldsymbol{\theta}) < 0\Big\};$$

then Γ_A^{CX} is convex, $\Gamma_A^{\text{IN}} \subset \Gamma_A^{\text{CX}}$, and $\partial \Gamma^{\text{IN}} \cap \Gamma_A \subset \partial \Gamma_A^{\text{CX}}$.

P r o o f. (a) We note two facts. Firstly, $\langle \nabla \gamma(\mathbf{0}), \boldsymbol{t}_i \rangle > 0$ for all $i \in \mathcal{J}$ by Lemma 2.9 and the stability condition (2.4). Secondly, $\boldsymbol{t}_i \leqslant \mathbf{0}$ with $t_{ii} < 0$ by Lemma 2.8. These facts imply $b_i(-\boldsymbol{t}_i) \in \Gamma$ for some $b_i > 0$. Let $H_i^{\text{IN}} = \{\boldsymbol{x} \in \Gamma; \langle \nabla \gamma_{s,i}(\mathbf{0}), \boldsymbol{x} \rangle \geqslant 0\}$. Since H_i^{IN} is a convex set (half space), $H_{\mathcal{J} \setminus A}^{\text{IN}} \equiv \bigcap_{i \in \mathcal{J} \setminus A} H_i^{\text{IN}}$ is also convex, and contains $b_j(-\boldsymbol{t}_j)$ for $j \in A$ because $\boldsymbol{t}_i \in H_i^{\text{IN}}$ and $\boldsymbol{t}_j \notin H_i^{\text{IN}}$ for $j \neq i$. Hence, their convex combination is also in $H_{\mathcal{J} \setminus A}^{\text{IN}}$, and nonnegative with positive entries for $j \in A$ because $b_j(-\boldsymbol{t}_j) \geqslant \mathbf{0}$ and $b_j(-t_{jj}) > 0$ for all $j \in A$. Furthermore, $H_{\mathcal{J} \setminus A}^{\text{IN}} \subset \Gamma_A^{\text{IN}}$ because $\boldsymbol{x} \in H_i^{\text{IN}}$ implies that $\gamma_{s,i}(\boldsymbol{x}) > 0$ for $\boldsymbol{x} \neq \mathbf{0}$. Thus, (a) is proved.

(b) Since $\gamma_{e,i}$ and $\gamma_{s,j}$ are convex functions, Γ_A^{CX} is a convex set. Since $\gamma_{s,i}(\boldsymbol{\theta}) > 0$ for all $i \in \mathcal{J} \setminus A$ for $\boldsymbol{\theta} \in \Gamma_A^{\text{IN}}$, we have, for $\boldsymbol{\theta} \in \Gamma^{\text{IN}}$,

$$\sum_{i \in \mathcal{E}} \gamma_{e,i}(\boldsymbol{\theta}) + \sum_{j \in A} \gamma_{s,j}(\boldsymbol{\theta}) = \gamma(\boldsymbol{\theta}) - \sum_{i \in \mathcal{J} \setminus A} \gamma_{s,i}(\boldsymbol{\theta}) < 0,$$

which proves that $\Gamma_A^{\text{IN}} \subset \Gamma_A^{\text{CX}}$. If $\boldsymbol{\theta} \in \partial\Gamma^{\text{IN}} \cap \Gamma_A$, then $\gamma(\boldsymbol{\theta}) = 0$ and $\gamma_{s,i}(\boldsymbol{\theta}) > 0$ for all $i \in \mathcal{J} \setminus A$, and therefore $\boldsymbol{\theta} \in \partial\Gamma_A^{\text{CX}}$. ∎

REMARK 2.4. *Since $\gamma_{s,i}(\boldsymbol{\theta}) = 0$ is equivalent to $q_i(\boldsymbol{\theta}) = 1$ by (2.21), $\boldsymbol{\theta} \in \partial\Gamma_A$, equivalently, $\gamma_{s,i}(\boldsymbol{\theta}) = 0$ for $i \in \mathcal{J} \setminus A$, if and only if*

$$(I - P^{(\mathcal{J}\setminus A)})e^{\boldsymbol{\theta}_{\mathcal{J}\setminus A}} = P^{(\mathcal{J}\setminus A, A)}e^{\boldsymbol{\theta}_A} + \boldsymbol{p}_{\mathcal{J}\setminus A, 0},$$

where $P^{(\mathcal{J}\setminus A)} \equiv \{p_{ij}; i, j \in \mathcal{J} \setminus A\}$ and $P^{(\mathcal{J}\setminus A, A)} \equiv \{p_{ij}; i \in \mathcal{J} \setminus A, j \in A\}$, $e^{\boldsymbol{\theta}_A}$ is the $|A|$-dimensional vector whose i-th entry is e^{θ_i} for $i \in A$, and $\boldsymbol{p}_{\mathcal{J}\setminus A, 0}$ is the $|\mathcal{J} \setminus A|$-dimensional vector whose i-th entry is p_{i0} for $i \in \mathcal{J} \setminus A$. Since $P^{(\mathcal{J}\setminus A)}$ is strictly substochastic, if $\theta_i \geqslant 0$ for $i \in A$ and $\boldsymbol{\theta} \in \partial\Gamma_A$, then

$$e^{\boldsymbol{\theta}_{\mathcal{J}\setminus A}} = (I - P^{(\mathcal{J}\setminus A)})^{-1}(P^{(\mathcal{J}\setminus A, A)}e^{\boldsymbol{\theta}_A} + \boldsymbol{p}_{\mathcal{J}\setminus A, 0}) \geqslant \boldsymbol{1}_{\mathcal{J}\setminus A},$$

which implies that $\theta_j \geqslant 0$ for $j \in \mathcal{J} \setminus A$.

We now present main results, which are proved in Section 4.3. For this, we use the following notation. For $\boldsymbol{x} \in \mathbb{R}^d$ and $A \subset \mathcal{J}$, let $\boldsymbol{x}_A$ be the $|A|$-dimensional vector which is obtained from $\boldsymbol{x}$ by dropping its i-entry of $\boldsymbol{x}$ for all $i \in \mathcal{J} \setminus A$. Let

$$\varphi_k(\boldsymbol{\theta}) = \mathbb{E}\big(e^{\langle \boldsymbol{\theta}, \boldsymbol{L}\rangle}1(L_k = 0)\big), \ k \in \mathcal{J}, \quad \varphi(\boldsymbol{\theta}) = \mathbb{E}(e^{\langle \boldsymbol{\theta}, \boldsymbol{L}\rangle}),$$
$$r_*(\mathbf{e}_k) = \sup\big\{\theta_k; \boldsymbol{\theta} \in \Gamma_k^{\text{IN}} \cap \Theta_k, \varphi_k(\boldsymbol{\theta}) < \infty\big\}, \quad k \in \mathcal{J},$$

where $\Theta_k = \big\{\boldsymbol{\theta} \in \mathbb{R}^d; \theta_i \geqslant 0, \forall i \in \mathcal{J} \setminus \{k\}\big\} \cup \big\{\boldsymbol{\theta} \in \mathbb{R}^d; \theta_i < 0, \forall i \in \mathcal{J} \setminus \{k\}\big\}$. Note that $\Theta_k = \mathbb{R}^2$ for $d = 2$, and therefore $r_*(\mathbf{e}_k) = r_{\{k\}}(\mathbf{e}_k)$. For $A \subset \mathcal{J}$ and $\boldsymbol{c} \in \overrightarrow{U}_d$, that is, the unit direction vector $\boldsymbol{c}$, let

$$r_A(\boldsymbol{c}) = \sup\big\{\langle \boldsymbol{\theta}, \boldsymbol{c}\rangle; \boldsymbol{\theta} \in \Gamma_A^{\text{IN}}, \varphi_i(\boldsymbol{\theta}) < \infty, \forall i \in A\big\},$$
$$m_A(\boldsymbol{c}) = \sup\big\{u; u\boldsymbol{c} \in \overleftarrow{\Gamma}_A^{\text{IN}}, \varphi_i(\boldsymbol{\theta}) < \infty, \forall i \in A\big\}.$$

Note that $r_A(\boldsymbol{c}) \leqslant m_A(\boldsymbol{c})$ because $\|\boldsymbol{c}\| = 1$.

THEOREM 2.1. *Assume that the GJN is stable, and let B_0 be a compact subset of $\mathbb{R}_+^d$.*
(a) *For $k \in \mathcal{J}$,*

$$(2.37) \qquad \limsup_{x \to \infty} \frac{1}{x}\log\mathbb{P}(\boldsymbol{L} \in x\mathbf{e}_k + B_0) \leqslant -r_*(\mathbf{e}_k).$$

(b) *If the uniformly bounded assumption* (A1) *is satisfied, then, for $\boldsymbol{c} \in \overrightarrow{U}_d$,*

$$(2.38) \quad \limsup_{x \to \infty} \frac{1}{x}\log\mathbb{P}(\boldsymbol{L} \in x\boldsymbol{c} + B_0) \leqslant -\max\{r_A(\boldsymbol{c}); \boldsymbol{c}_A > \boldsymbol{0}_A, A \in 2^{\mathcal{J}} \setminus \emptyset\},$$

$$(2.39) \qquad \limsup_{x \to \infty} \frac{1}{x}\log\mathbb{P}(\langle \boldsymbol{c}, \boldsymbol{L}\rangle > x) \leqslant -\max\{m_A(\boldsymbol{c}); A \in 2^{\mathcal{J}} \setminus \emptyset\}.$$

For $B \subset \mathbb{R}_+^d$, define a convex corn as

$$\text{Corn}(B) = \{\boldsymbol{x} \in \mathbb{R}_+^d; u\boldsymbol{x} \in B, \exists u > 0\}.$$
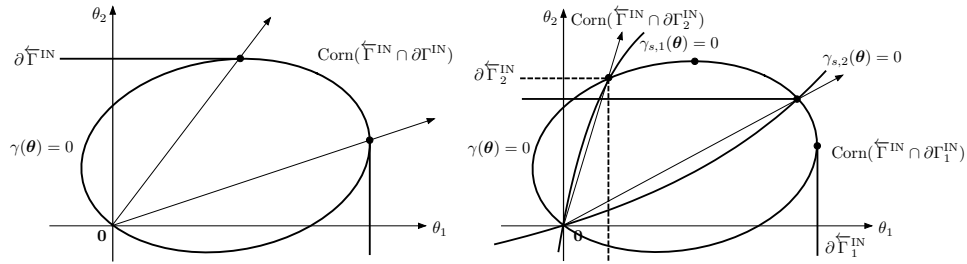

FIGURE 2. Corns used in Theorem 2.2 below.

THEOREM 2.2. *Assume that the GJN is stable.*
(a) *For $d = 2$, let B_0 be a compact set of $\mathbb{R}_+^2$; then, for $k = 1, 2$,*

$$(2.40) \qquad \liminf_{x \to \infty} \frac{1}{x} \log \mathbb{P}(\boldsymbol{L} \in x\mathbf{e}_k + B_0) \geqslant -r_{\{k\}}(\mathbf{e}_k).$$

(b) *For general $d \geqslant 2$ and $\boldsymbol{c} \in \overrightarrow{U}_d$, if $\boldsymbol{c} \in \text{Corn}(\overleftarrow{\Gamma}^{\text{IN}} \cap \partial\Gamma^{\text{IN}})$, then*

$$(2.41) \qquad \liminf_{x \to \infty} \frac{1}{x} \log \mathbb{P}(\langle \boldsymbol{c}, \boldsymbol{L} \rangle > x) \geqslant -\sup\{u \geqslant 0; u\boldsymbol{c} \in \Gamma^{\text{IN}}\}.$$

(c) *For $d = 2$, $k = 1, 2$ and $\boldsymbol{c} \in \overrightarrow{U}_2$, if $\boldsymbol{c} \in \text{Corn}(\overrightarrow{\Gamma}^{\text{IN}} \cap \partial\Gamma_k^{\text{IN}})$, then*

$$(2.42) \qquad \liminf_{x \to \infty} \frac{1}{x} \log \mathbb{P}(\langle \boldsymbol{c}, \boldsymbol{L} \rangle > x) \geqslant -\sup\{u \geqslant 0; u\boldsymbol{c} \in \overleftarrow{\Gamma}_k^{\text{IN}}\}.$$

For $d = 2$, we can get bounds explicitly. For this, let

$$(2.43) \qquad \delta_1 = \sup\{\theta_1 \geqslant 0; \boldsymbol{\theta} \in \Gamma_1^{\text{IN}}, \theta_2 \leqslant \delta_2\},$$

$$(2.44) \qquad \delta_2 = \sup\{\theta_2 \geqslant 0; \boldsymbol{\theta} \in \Gamma_2^{\text{IN}}, \theta_1 \leqslant \delta_1\},$$

which are known to have a unique solution $\boldsymbol{\delta} = (\delta_1, \delta_2)$ (see the proof of Corollary 2.1 in Section 4.3), and define

$$\mathcal{D}_2 = \{\boldsymbol{\theta} \in \overleftarrow{\Gamma}^{\text{IN}}; \theta_i < \delta_i, i = 1, 2\}.$$

Then we have the following corollary.

COROLLARY 2.1. *Assume the stable GJN has two stations* ($d = 2$).
(a) *For* $k = 1, 2,$

$$(2.45) \qquad \lim_{x \to \infty} \frac{1}{x} \log \mathbb{P}(\boldsymbol{L} \in x\mathbf{e}_k + B_0) = -\sup\{\theta_k; \boldsymbol{\theta} \in \mathcal{D}_2\}.$$

(b) *If* (A1) *is satisfied, then, for* $\boldsymbol{c} \in \overrightarrow{U}_2,$

$$(2.46) \qquad \limsup_{x \to \infty} \frac{1}{x} \log \mathbb{P}(\boldsymbol{L} \in x\boldsymbol{c} + B_0) \leqslant -\sup\{\langle \boldsymbol{\theta}, \boldsymbol{c} \rangle; \boldsymbol{\theta} \in \mathcal{D}_2\},$$

$$(2.47) \qquad \lim_{x \to \infty} \frac{1}{x} \log \mathbb{P}(\langle \boldsymbol{c}, \boldsymbol{L} \rangle > x) = -\sup\{u; u\boldsymbol{c} \in \mathcal{D}_2\}.$$

It is notable that $\mathcal{D}_2$ have been obtained as the convergence domain of $\varphi(\boldsymbol{\theta})$ $\equiv \mathbb{E}(e^{\langle \boldsymbol{\theta}, \boldsymbol{L} \rangle})$, and used to derive (2.47) for the two-station JGN with phase-type $F_{e,i}, F_{s,j}$ in Theorem 4.2 of [17]. The asymptotic (2.45) in the coordinate directions is not derived in [17], but can be obtained from Theorem 3.2 there. We here have the asymptotic (2.45) without the phase-type assumption. We conjecture that the assumption (A1) can be removed from all the results, but it seems a quite hard problem.

Similar results to (2.45) and (2.47) are known for a reflecting random walks on the quarter plane (e.g., see [15], [16]) and semi-martingale reflecting two-dimensional Brownian motions, SRBM for short (see [7]). On the other hand, the asymptotic (2.46) is new for the GJN, but known for the two-dimensional SRBM (see [2], [8]), where (2.46) is sharpened.

For $d \geqslant 3$, there is very little known about the tail asymptotics of the stationary distribution not only for the GJN but also a reflecting random walk and SRBM. There are some studies in the framework of sample path large deviations, but those results need to solve certain optimization problems, which are hard to solve even numerically (e.g., see [14]). Contrary to them, (2.38) and (2.39) may be applied to get explicit bounds, by using ideas for a reflecting random walk (see Theorem 6.1 of [16]).

3. CHANGE OF MEASURE FOR GJN

In this section, we present some preliminary results for proving Theorems 2.1 and 2.2 and Corollary 2.1. A change of measure is typically used in the theory of large deviations. We also use it, and construct a new measure using a multiplicative functional, which is obtained from the martingale $M_{\boldsymbol{J}(v),\boldsymbol{\theta}}(\cdot)$ in Section 2.4. However, we assume $\boldsymbol{J}(v) = \emptyset$ in this section for making arguments simpler. It also suffices for major applications in the later sections.

Thus, the new measure is constructed from $M_{\boldsymbol{\theta}}(\cdot) \equiv M_{\emptyset,\boldsymbol{\theta}}(\cdot)$. For this, we first drive a multiplicative functional. Its derivation is rather standard, but will be detailed because it is crucial for our arguments. Our major interest in this section

is to see how the GJN is modified under the new measure. It is important for us to specifically identify its modeling parameters, which has not been studied in the literature except for the single queue case (see [18]), and may have an independent interest.

3.1. Multiplicative functional. Let $Y(t)$ be a left-continuous process, which is called *predictable* because $Y(t-)$ is $\mathcal{F}_{t-}$-measurable. Assume that $Y(\cdot)$ is bounded in each finite interval. Recall that $M_0(\cdot)$, $M(\cdot)$ and $D(\cdot)$ are defined by (2.8), (2.9) and (2.10), respectively. Assume that the terminal condition (2.12) is satisfied. Assume that $M(\cdot)$ is an $\mathcal{F}_t$-martingale under $\mathbb{P}_{\boldsymbol{x}}$ for each $\boldsymbol{x} \in S$.

We define the integral of $Y(\cdot)$ with respect to martingale $M(\cdot)$ by

$$Y \cdot M(t) \equiv 1 + \int_0^t Y(u) dM(u),$$

where integration is a natural extension of a Riemann–Stieltjes integral (see Section 4d of Chapter I of [11]). For a positive-valued test function f, choose $Y(t)$ as

$$Y(t) = \frac{1}{f\big(X(0)\big)} \exp\bigg(-\int_0^t \frac{\mathcal{A}f\big(X(u)\big)}{f\big(X(u)\big)} du \bigg),$$

which is obviously positive and continuous in t and adapted to $\mathcal{F}_t$. Hence, $Y \cdot M(\cdot)$ is a martingale. We denote it by $E^f(\cdot)$. Thus, it follows from (2.9) that

$$(3.1) \quad E^f(t) = 1 + \int_0^t Y(u)\Big(df\big(X(u)\big) - \mathcal{A}f\big(X(u)\big) du \Big)$$

$$= 1 + \int_0^t Y(u) df\big(X(u)\big) + \int_0^t f\big(X(u)\big) Y'(u) du$$

$$= 1 + \big[Y(u) f\big(X(u)\big) \big]_0^t = \frac{f\big(X(t)\big)}{f\big(X(0)\big)} \exp\bigg(-\int_0^t \frac{\mathcal{A}f\big(X(u)\big)}{f\big(X(u)\big)} du \bigg),$$

which is an $\mathcal{F}_t$-martingale under $\mathbb{P}_{\boldsymbol{x}}$.

On the other hand, $E^f(\cdot)$ is a multiplicative functional because it is right-continuous, $E^f(0) = 1$, $\mathbb{E}_{\boldsymbol{x}}\big(E^f(t) \big) = 1$ and

$$E^f(s+t) = E^f(s) \Theta_s \circ E^f(t), \quad s, t \geqslant 0,$$

where

$$\Theta_s \circ E^f(t) = \frac{f\big(X(s+t)\big)}{f\big(X(s)\big)} \exp\bigg(-\int_s^{s+t} \frac{\mathcal{A}f\big(X(u)\big)}{f\big(X(u)\big)} du \bigg).$$

Thus, we can define a probability measure $\widetilde{\mathbb{P}}_{\boldsymbol{x}}^f$ for an initial state $\boldsymbol{x} \in S$ by

$$(3.2) \qquad \frac{d\widetilde{\mathbb{P}}_{\boldsymbol{x}}^f}{d\mathbb{P}_{\boldsymbol{x}}}\bigg|_{\mathcal{F}_t} = E^f(t), \quad t \geqslant 0,$$

because $E^f(\cdot)$ is a martingale (see [13] for details). We refer to (3.2) as exponential change of measure. Let $\mathbb{P}_\nu$ and $\widetilde{\mathbb{P}}_\nu^f$ be probability measures such that $\mathbb{P}_\nu(C) = \int_S \mathbb{P}_{\boldsymbol{x}}(C)\nu(d\boldsymbol{x})$ and $\widetilde{\mathbb{P}}_\nu^f(C) = \int_S \widetilde{\mathbb{P}}_{\boldsymbol{x}}^f(C)\nu(d\boldsymbol{x})$ for $X(0)$ to have a probability distribution ν on S, (3.2) implies that, for a nonnegative $\mathcal{F}_t$-measurable random variable U with finite expectation, we have

$$(3.3) \qquad \widetilde{\mathbb{E}}_\nu^f(U) = \mathbb{E}_\nu\big(U E^f(t)\big), \quad \mathbb{E}_\nu(U) = \widetilde{\mathbb{E}}_\nu^f\big(E^f(t)^{-1}U\big),$$

where $\mathbb{E}_\nu$ and $\widetilde{\mathbb{E}}_\nu^f$ represent the expectations concerning $\mathbb{P}_\nu$ and $\widetilde{\mathbb{P}}_\nu^f$, respectively. Similarly, for conditional expectations, we have, for $0 \leqslant s < t$,

$$(3.4) \quad \widetilde{\mathbb{E}}(U|\mathcal{F}_s) = \mathbb{E}\left(U\frac{E^f(t)}{E^f(s)}\bigg|\mathcal{F}_s\right), \quad \mathbb{E}(U|\mathcal{F}_s) = \widetilde{\mathbb{E}}\left(U\frac{E^f(s)}{E^f(t)}\bigg|\mathcal{F}_s\right).$$

One can easily check this equation from the definition of a conditional expectation (see, e.g., Section III.3 of [11]).

When $f = f_{\boldsymbol{\theta}}$ of (2.13) and $M = M_{\boldsymbol{\theta}}$ of (2.26) with $\boldsymbol{J}(v) = \emptyset$, we denote $\widetilde{\mathbb{P}}_{\boldsymbol{x}}^{f_{\boldsymbol{\theta}}}$ by $\widetilde{\mathbb{P}}_{\boldsymbol{x}}^{(\boldsymbol{\theta})}$. If $\boldsymbol{J}(v) \neq \emptyset$, then the new measure is denoted by $\widetilde{\mathbb{P}}_{\boldsymbol{x}}^{(\boldsymbol{J}(v),\boldsymbol{\theta})}$.

3.2. GJN under the new measure. Let us consider how the GJN is modified under the new measure $\widetilde{\mathbb{P}}_{\boldsymbol{x}}^{(\boldsymbol{\theta})}$. A general principle for change of measure is considered for a PDMP in [19], but we need to compute specific modeling parameters. For this, we follow the method of [18] studied for a single queue with many heterogeneous servers. We modify it here for the GJN. Since the differential operator $\mathcal{A}$ is unchanged because it works on a deterministic part of the sample path of $X(\cdot)$, we only need to consider the jump kernel $\mathcal{K}$. Denote it under $\widetilde{\mathbb{P}}_{\boldsymbol{x}}^{(\boldsymbol{\theta})}$ by $\widetilde{\mathcal{K}}^{(\boldsymbol{\theta})}$.

We first write $E^{f_{\boldsymbol{\theta}}}(t)$ explicitly as

$$(3.5) \quad E^{f_{\boldsymbol{\theta}}}(t) = \exp\Big(\langle\boldsymbol{\theta}, \boldsymbol{L}(t) - \boldsymbol{L}(0)\rangle - w\big(\boldsymbol{\theta}, \boldsymbol{R}(t) - \boldsymbol{R}(0)\big) - \gamma(\boldsymbol{\theta})t$$
$$+ \int_0^t \sum_{i\in\mathcal{J}} \gamma_{s,i}(\boldsymbol{\theta})1\big(L_i(u) = 0\big)du\Big),$$

where $w(\boldsymbol{\theta}, \boldsymbol{y}) = w_\emptyset(\boldsymbol{\theta}, \boldsymbol{y})$ (see (2.15)), that is,

$$w(\boldsymbol{\theta}, \boldsymbol{y}) = \sum_{i\in\mathcal{E}} \gamma_{e,i}(\theta_i)y_{e,i} + \sum_{i\in\mathcal{J}} \gamma_{s,i}(\theta_i)y_{s,i}.$$

Our first task is to compute the distributions of $T_{e,i}, T_{s,j}$ under $\widetilde{\mathbb{P}}_{\boldsymbol{x}}^{(\boldsymbol{\theta})}$. These distributions (moment generating functions) are denoted, respectively, by $F_{e,i}^{(\boldsymbol{\theta})}$ ($\widehat{F}_{e,i}^{(\boldsymbol{\theta})}$) and $F_{s,j}^{(\boldsymbol{\theta})}$ ($\widehat{F}_{s,j}^{(\boldsymbol{\theta})}$). Recall $\beta_F^{(v)}$ of (2.16), and denote $\beta_{F_{e,i}}^{(\infty)}$ and $\beta_{F_{s,j}}^{(\infty)}$ simply by $\beta_{e,i}^{(\infty)}$ and $\beta_{s,j}^{(\infty)}$, respectively. The following lemma is similar to Lemma 4.4 of [18].

LEMMA 3.1. *For each $\boldsymbol{\theta} \in \mathbb{R}^d$, $v \in (0, \infty]$ and $\eta \in \mathbb{R}$,*

$$(3.6) \quad \widehat{F}_{e,i}^{(\boldsymbol{\theta})}(\eta) = e^{\theta_i} \widehat{F}_{e,i}\big(\eta - \gamma_{e,i}(\theta_i)\big), \quad \eta \leqslant \beta_{e,i}^{(\infty)} + \gamma_{e,i}(\theta_i), \ i \in \mathcal{E} \setminus J_e(v),$$

$$(3.7) \quad \widehat{F}_{s,i}^{(\boldsymbol{\theta})}(\eta) = q_i(\boldsymbol{\theta}) \widehat{F}_{s,i}\big(\eta - \gamma_{s,i}(\boldsymbol{\theta})\big), \quad \eta \leqslant \beta_{s,i}^{(\infty)} + \gamma_{s,i}(\boldsymbol{\theta}), \ i \in \mathcal{J} \setminus J_s(v).$$

Since $\widehat{F}_{e,i}^{(\boldsymbol{\theta})}(0) = \widehat{F}_{s,i}^{(\boldsymbol{\theta})}(0) = 1$ by (2.19)–(2.21), $F_{e,i}^{(\boldsymbol{\theta})}$ and $F_{s,j}^{(\boldsymbol{\theta})}$ are proper distribution functions. Let

$$\lambda_i^{(\boldsymbol{\theta})} = \big(\widetilde{\mathbb{E}}_{e,i}^{(\boldsymbol{\theta})}(T_{e,i})\big)^{-1}, \quad \mu_i^{(\boldsymbol{\theta})} = \big(\widetilde{\mathbb{E}}_{s,i}^{(\boldsymbol{\theta})}(T_{s,i})\big)^{-1},$$

where $\widetilde{\mathbb{E}}_{e,i}^{(\boldsymbol{\theta})}$ and $\widetilde{\mathbb{E}}_{s,i}^{(\boldsymbol{\theta})}$ represent the conditional expectations under $\widetilde{\mathbb{E}}_{\boldsymbol{x}}^{(\boldsymbol{\theta})}$ just before time when external arrivals and service completion, respectively, at station i occur. Then, by Lemma 3.1, we have

$$(3.8) \quad \lambda_i^{(\boldsymbol{\theta})} = \Big(e^{\theta_i} \widehat{F}_{e,i}'\big(-\gamma_{e,i}(\theta_i)\big)\Big)^{-1}, \quad i \in \mathcal{E},$$

$$(3.9) \quad \mu_i^{(\boldsymbol{\theta})} = \Big(q_i(\boldsymbol{\theta}) \widehat{F}_{s,i}'\big(-\gamma_{s,i}(\theta_i)\big)\Big)^{-1}, \quad i \in \mathcal{J}.$$

The jump kernel $\mathcal{K}$ is changed to $\widetilde{\mathcal{K}}^{(\boldsymbol{\theta})}$ as

$$(3.10) \quad \widetilde{\mathcal{K}}^{(\boldsymbol{\theta})} 1_{B_\ell \times B_e \times B_s}(\boldsymbol{x})$$
$$= \begin{cases} \widetilde{\mathbb{P}}_{e,i}^{(\boldsymbol{\theta})}(\boldsymbol{z} + \boldsymbol{e}_i \in B_\ell, \boldsymbol{y}_e + T_{e,i}\boldsymbol{e}_i \in B_e, \boldsymbol{y}_s \in B_s), & y_{e,i} = 0, \\ \widetilde{\mathbb{P}}_{s,i}^{(\boldsymbol{\theta})}(\boldsymbol{z} - \boldsymbol{e}_i + \boldsymbol{e}_j \in B_\ell, \boldsymbol{y}_e \in B_e, \boldsymbol{y}_s + T_{s,i}\boldsymbol{e}_i \in B_s), & y_{s,i} = 0, \end{cases}$$

where $\theta_0 = 0$. Hence, the routing probability from station i to j under $\widetilde{\mathbb{P}}_{\boldsymbol{x}}^{(\boldsymbol{\theta})}$ is

$$(3.11) \quad p_{ij}^{(\boldsymbol{\theta})} \equiv e^{-\theta_i + \theta_j} p_{ij} / q_i(\boldsymbol{\theta}).$$

Thus, the GJN keeps the same network structure under the new probability measure $\widetilde{\mathbb{P}}_{\boldsymbol{x}}^{(\boldsymbol{\theta})}$, but its modeling primitives, $F_{e,i}$, $F_{s,j}$ and p_{ij}, are changed to $F_{e,i}^{(\boldsymbol{\theta})}$, $F_{s,j}^{(\boldsymbol{\theta})}$ and $p_{ij}^{(\boldsymbol{\theta})}$, respectively, which do not depend on the initial state $\boldsymbol{x}$. Let

$$q_i^{(\boldsymbol{\theta})}(\boldsymbol{\eta}) = e^{-\eta_i} \Big(\sum_{j \in \mathcal{J}} p_{ij}^{(\boldsymbol{\theta})} e^{\eta_j} + p_{i0}^{(\boldsymbol{\theta})}\Big), \quad i \in \mathcal{J},$$

which corresponds to $q_i(\boldsymbol{\eta})$ in the original model, where $\boldsymbol{\eta} \in \mathbb{R}^d$ is a variable. From this definition and (3.11), we have

$$q_i^{(\boldsymbol{\theta})}(\boldsymbol{\eta}) = \frac{q_i(\boldsymbol{\eta} + \boldsymbol{\theta})}{q_i(\boldsymbol{\theta})}.$$

As in the original network model, we define $\gamma_{e,i}^{(\boldsymbol{\theta})}(\eta_i)$, $\gamma_{s,j}^{(\boldsymbol{\theta})}(\boldsymbol{\eta})$ as the unique solutions of the following equations:

$$e^{\eta_i} \widehat{F}_{e,i}^{(\boldsymbol{\theta})}\big(-\gamma_{e,i}^{(\boldsymbol{\theta})}(\eta_i)\big) = 1, \ i \in \mathcal{E}, \quad q_i^{(\boldsymbol{\theta})}(\boldsymbol{\eta}) \widehat{F}_{s,i}^{(\boldsymbol{\theta})}\big(-\gamma_{s,i}^{(\boldsymbol{\theta})}(\boldsymbol{\eta})\big) = 1, \ i \in \mathcal{J},$$

for $v \in (0, \infty]$. These definitions yield

$$\gamma_{e,i}^{(\boldsymbol{\theta})}(\eta_i) = \gamma_{e,i}(\eta_i + \theta_i) - \gamma_{e,i}(\theta_i), \quad \gamma_{s,j}^{(\boldsymbol{\theta})}(\boldsymbol{\eta}) = \gamma_{s,j}(\boldsymbol{\eta} + \boldsymbol{\theta}) - \gamma_{s,j}(\boldsymbol{\theta}),$$

and define $\gamma^{(\boldsymbol{\theta})}(\boldsymbol{\eta})$ as

$$\gamma^{(\boldsymbol{\theta})}(\boldsymbol{\eta}) = \sum_{i \in \mathcal{E}} \gamma_{e,i}^{(\boldsymbol{\theta})}(\eta_i) + \sum_{i \in \mathcal{J}} \gamma_{s,i}^{(\boldsymbol{\theta})}(\boldsymbol{\eta}) = \gamma(\boldsymbol{\eta} + \boldsymbol{\theta}) - \gamma(\boldsymbol{\theta}).$$

Let $\boldsymbol{\alpha}^{(\boldsymbol{\theta})} = \boldsymbol{\lambda}^{(\boldsymbol{\theta})}(I - P^{(\boldsymbol{\theta})})^{-1}$, and $\widetilde{\boldsymbol{\alpha}}^{(\boldsymbol{\theta})}$ is defined similarly to $\widetilde{\boldsymbol{\alpha}}$.

We immediately see from these formulas that

$$(3.12) \qquad \nabla \gamma_{s,i}^{(\boldsymbol{\theta})}(0) = \nabla \gamma_{s,i}(\boldsymbol{\theta}), \quad \nabla \gamma^{(\boldsymbol{\theta})}(0) = \nabla \gamma(\boldsymbol{\theta}).$$

Similarly to (2.33) and (2.24), we have

$$(3.13) \qquad \nabla \gamma^{(\boldsymbol{\theta})}(\mathbf{0}) = \boldsymbol{\lambda}^{(\boldsymbol{\theta})} - \boldsymbol{\mu}^{(\boldsymbol{\theta})}(I - P^{(\boldsymbol{\theta})}),$$

$$(3.14) \qquad \nabla \gamma_{s,j}^{(\boldsymbol{\theta})}(\mathbf{0}) = \mu_j^{(\boldsymbol{\theta})}\big((p_{j1}^{(\boldsymbol{\theta})}, p_{j2}^{(\boldsymbol{\theta})}, \ldots, p_{jd}^{(\boldsymbol{\theta})}) - \mathbf{e}_j\big), \quad j \in \mathcal{J}.$$

Hence, we can update Lemmas 2.7 and 2.9 in exactly the same way as for the network model under $\widetilde{\mathbb{P}}_{\boldsymbol{x}}^{(\boldsymbol{\theta})}$, where $\alpha_j, \widetilde{\alpha}_j$ are updated to $\alpha_j^{(\boldsymbol{\theta})}, \widetilde{\alpha}_j^{(\boldsymbol{\theta})}$, respectively.

The following lemma is almost immediate from (3.11) and (3.14), but will be useful to check the conditions in Lemma 2.9 in which μ_k, α_k are updated to $\mu_k^{(\boldsymbol{\theta})}, \alpha_k^{(\boldsymbol{\theta})}$, respectively. As $\boldsymbol{t}_i$ of (2.34), we define $\boldsymbol{t}_i^{(\boldsymbol{\theta})} \in \mathbb{R}^d$ by

$$\langle \nabla \gamma_{s,j}(\boldsymbol{\theta}), \boldsymbol{t}_i^{(\boldsymbol{\theta})} \rangle = 0, \ j \neq i, \quad \langle \nabla \gamma_{s,i}(\boldsymbol{\theta}), \boldsymbol{t}_i^{(\boldsymbol{\theta})} \rangle > 0.$$

Hence, the following lemma can be obtained similarly to Lemma 2.8.

LEMMA 3.2. *Let $T^{(\boldsymbol{\theta})}$ be the matrix whose i-th column is $\boldsymbol{t}_i^{(\boldsymbol{\theta})}$; then $T^{(\boldsymbol{\theta})}$ is non-singular and not positive, that is, $\boldsymbol{t}_i^{(\boldsymbol{\theta})} \leqslant \mathbf{0}$ with $t_{ii}^{(\boldsymbol{\theta})} < 0$ for all $i \in \mathcal{J}$.*

4. PROOFS

The goal of this section is to prove the theorems and their corollary. A main idea is to use the new measure introduced in Section 3.2 by appropriate choosing the parameter $\boldsymbol{\theta}$. Some of its arguments are parallel to those in Section 4 of [18], but we require more lemmas because the state space for the queue lengths is multidimensional. We start to represent the stationary tail probability under the new measure.

4.1. A procedure for deriving tail asymptotics. Recall the notation $w(\boldsymbol{\theta}, \boldsymbol{y})$, and, for $\boldsymbol{R}(t) = \big(\boldsymbol{R}_e(t), \boldsymbol{R}_s(t)\big)$,

$$w\big(\boldsymbol{\theta}, \boldsymbol{R}(t)\big) = \langle \boldsymbol{\gamma}_e(\boldsymbol{\theta}), \boldsymbol{R}_e(t) \rangle + \langle \boldsymbol{\gamma}_s(\boldsymbol{\theta}), \boldsymbol{R}_s(t) \rangle.$$

Then, it follows from (3.3) and (3.5) that, for a given initial distribution ν,

$$(4.1) \quad d\mathbb{P}_\nu = \big(E^{f_{\boldsymbol{\theta}}}(t)\big)^{-1} d\widetilde{\mathbb{P}}_\nu^{(\boldsymbol{\theta})} = f_{\boldsymbol{\theta}}\big(X(0)\big) \exp\Big(- \langle \boldsymbol{\theta}, \boldsymbol{L}(t) \rangle + w\big(\boldsymbol{\theta}, \boldsymbol{R}(t)\big)$$
$$+ \gamma(\boldsymbol{\theta})t - \sum_{i \in \mathcal{J}} \gamma_{s,i}(\boldsymbol{\theta}) \int\limits_0^t 1\big(L_i(u) = 0\big) du \Big) d\widetilde{\mathbb{P}}_\nu^{(\boldsymbol{\theta})}$$

on $\mathcal{F}_t$, where we recall that $f_{\boldsymbol{\theta}}\big(X(0)\big) = e^{\langle \boldsymbol{\theta}, \boldsymbol{L}(0) \rangle - w(\boldsymbol{\theta}, \boldsymbol{R}(0))}$.

We take the initial distribution ν in the following way. Let $S_{\mathrm{L}} = \mathbb{Z}_+^d$, $S_{\mathrm{R}} = \mathbb{R}_+^{\mathcal{E}} \times \mathbb{R}^d$. For each $A \subset \mathcal{J}$, let $\tau_A^{\mathrm{EX}}, \tau_A^{\mathrm{RE}}$ be the first exit from and return times to $\partial_A S_{\mathrm{L}}$ concerning $\boldsymbol{L}(t)$ such that $\tau_A^{\mathrm{EX}} < \tau_A^{\mathrm{RE}}$, where

$$\partial_A S_{\mathrm{L}} = \bigcup_{i \in A} \{\boldsymbol{z} \in S_{\mathrm{L}}; z_i = 0\}.$$

Let ν_A^- be the distribution of $X(\tau_A^{\mathrm{EX}}-)$ given that $X(0)$ is subject to the normalized stationary distribution limited on $\partial_A^{\mathrm{L}} S \equiv \partial_A S_{\mathrm{L}} \times S_{\mathrm{R}}$. This ν_A^- is taken for ν in (4.1). Denote a random vector subject to the stationary distribution of $X(t)$ by $X \equiv (\boldsymbol{L}, \boldsymbol{R}_e, \boldsymbol{R}_s)$. Then, the cycle formula yields, for $x > 0$ and $B(x) \subset S_{\mathrm{L}} \setminus \partial_A S_{\mathrm{L}}$,

$$(4.2) \qquad \mathbb{P}\big(\boldsymbol{L} \in B(x)\big) = b(A) \mathbb{E}_{\nu_A^-}\Big(\int\limits_0^{\tau_A^{\mathrm{RE}}} 1\big(\boldsymbol{L}(u) \in B(x)\big) du \Big),$$

where $b(A) = \mathbb{P}\big(\boldsymbol{L} \in S_{\mathrm{L}} \setminus \partial_A S_{\mathrm{L}}\big) / \mathbb{E}_{\nu_A^-}(\tau_A^{\mathrm{RE}} - \tau_A^{\mathrm{EX}})$. We are interested here in the asymptotic of $\mathbb{P}\big(\boldsymbol{L} \in B(x)\big)$ as $x \to \infty$.

We apply change of measure to (4.2). For this, let τ_x be a stopping time which satisfies

$$(4.3) \qquad \tau_x \leqslant \tau_{B(x)}^{\mathrm{IN}} \equiv \inf\{t \geqslant 0; \boldsymbol{L}(t) \in B(x)\}, \quad x > 0.$$

We do not make any extra assumption for this τ_x in this section, but it will be either $\tau_{B(x)}^{\mathrm{IN}}$ or $\tau_{C(x)}^{\mathrm{IN}}$ for some $C(x) \subset S_{\mathrm{L}} \setminus \partial_A S_{\mathrm{L}}$ in Sections 4.2 and 4.3. In what follows, except for Steps 5 and 5', we fix $A \subset \mathcal{J}$. Let

$$Y(t) = \mathbb{E}_{\nu_A^-}\Big(\int\limits_t^{\tau_A^{\mathrm{RE}}} 1\big(\boldsymbol{L}(u) \in B(x)\big) du \big| \mathcal{F}_t \Big).$$

Then it follows from (4.1) with $\nu = \nu_A^-$ and (4.2) that

$$(4.4) \quad \mathbb{P}\big(\boldsymbol{L} \in B(x)\big) = b(A)\mathbb{E}_{\nu_A^-}\big(Y(\tau_x)\big)$$

$$= b(A)\widetilde{\mathbb{E}}_{\nu_A^-}^{(\boldsymbol{\theta})}\Big[f_{\boldsymbol{\theta}}\big(X(0)\big)Y(\tau_x)1(\tau_x < \infty)e^{-\langle\boldsymbol{\theta},\boldsymbol{L}(\tau_x)\rangle + w(\boldsymbol{\theta},\boldsymbol{R}(\tau_x))}$$

$$\times \exp\Big(\gamma(\boldsymbol{\theta})\tau_x - \sum_{i\in\mathcal{J}\backslash A}\gamma_{s,i}(\boldsymbol{\theta})\int_0^{\tau_x}1\big(L_i(u)=0\big)du\Big)\Big].$$

We are now ready to consider the asymptotic of $\mathbb{P}\big(\boldsymbol{L} \in B(x)\big)$ as $x \to \infty$. For its upper bound, we take the following steps.

S t e p 1. Choose $\boldsymbol{\theta} \in \Gamma_A^{\text{IN}}$; then $\gamma(\boldsymbol{\theta}) < 0$ and $\gamma_{s,i}(\boldsymbol{\theta}) > 0$ for all $i \in \mathcal{J} \setminus A$, and therefore

$$\exp\Big(\gamma(\boldsymbol{\theta})\tau_x - \sum_{i\in\mathcal{J}\backslash A}\gamma_{s,i}(\boldsymbol{\theta})\int_0^{\tau_x}1\big(L_i(u)=0\big)du\Big) < 1.$$

In Steps 2–4 below, we keep the same $\boldsymbol{\theta} \in \Gamma_A^{\text{IN}}$.

S t e p 2. Verify that there is a constant C_1 such that, if $\tau_x < \infty$, then

$$(4.5) \qquad\qquad Y(\tau_x)e^{w(\boldsymbol{\theta},\boldsymbol{R}(\tau_x))} < C_1.$$

S t e p 3. Verify that $\widetilde{\mathbb{E}}_{\nu_A^-}^{(\boldsymbol{\theta})}\big(f_{\boldsymbol{\theta}}\big(X(0)\big)\big)$ is finite if $\varphi_i(\boldsymbol{\theta}) < \infty$ for all $i \in A$.

S t e p 4. Find finite real-valued functions $\overline{a}_0(\boldsymbol{\theta})$ and $\overline{a}_1(\boldsymbol{\theta}) > 0$ such that

$$(4.6) \qquad\qquad \overline{a}_1(\boldsymbol{\theta})x - \langle\boldsymbol{\theta},\boldsymbol{L}(\tau_x)\rangle \leqslant \overline{a}_0(\boldsymbol{\theta});$$

then $e^{-\langle\boldsymbol{\theta},\boldsymbol{L}(\tau_x)\rangle}$ is bounded above by $e^{\overline{a}_0(\boldsymbol{\theta})-\overline{a}_1(\boldsymbol{\theta})x}$.

S t e p 5. Derive an inequality from (4.4) using Steps 1–4. Take the logarithms of both sides of this inequality, and divide by x; then letting $x \to \infty$ yields the upper bound $-\overline{a}_1(\boldsymbol{\theta})$ for each $\boldsymbol{\theta}$ and $A \subset \mathcal{J}$. We then take the infimum of $-\overline{a}_1(\boldsymbol{\theta})$ for all $\boldsymbol{\theta}$ and $A \subset \mathcal{J}$ for which Steps 1–4 work well.

To derive the lower bounds, we will use the martingale $M_{\boldsymbol{J}(v),\boldsymbol{\theta}}$ of (2.26) in Lemma 2.6 for change of measure, choosing the index set for truncation: $\boldsymbol{J}(v) \equiv \big(J_e(v), J_s(v)\big)$ for each fixed $\boldsymbol{\theta} \in \mathbb{R}^d$ as

$$(4.7) \qquad J_e(v) = \{i \in \mathcal{E}; \gamma_{e,i}(\theta_i) < 0\}, \qquad J_s(v) = \{i \in \mathcal{E}; \gamma_{s,i}(\boldsymbol{\theta}) < 0\},$$

and we choose a sufficiently large v such that $\gamma_{e,i}(v,\theta_i) < 0$ for all $i \in J_e(v)$ and $\gamma_{s,j}(v,\boldsymbol{\theta}) < 0$ for all $j \in J_s(v)$, which is possible by Lemma 2.4 and the assumption (2.20). Then, $w_{\boldsymbol{J}(v)}(\boldsymbol{\theta},\boldsymbol{y})$ of (2.15) is bounded below for all $\boldsymbol{y} = (\boldsymbol{y}_e, \boldsymbol{y}_s) \in \mathbb{R}^{\mathcal{E}} \times \mathbb{R}^d$. Namely,

$$(4.8) \qquad w_{\boldsymbol{J}(v)}(\boldsymbol{\theta},\boldsymbol{y}) \geqslant v\Big(\sum_{i\in J_e(v)}\gamma_{e,i}(v,\theta_i) + \sum_{i\in J_s(v)}\gamma_{s,i}(v,\boldsymbol{\theta})\Big) > -\infty.$$

Then, using the martingale (2.26), we see that (4.4) is changed as

$$(4.9) \quad \mathbb{P}\big(\boldsymbol{L} \in B(x)\big) = b(A)\widetilde{\mathbb{E}}_{\nu_A^-}^{(\boldsymbol{J}(v),\boldsymbol{\theta})}\Big[f_{\boldsymbol{J}(v),\boldsymbol{\theta}}\big(X(0)\big)Y(\tau_x)1(\tau_x < \infty)$$

$$\times \exp\Big(-\langle\boldsymbol{\theta}, \boldsymbol{L}(\tau_x)\rangle + w_{\boldsymbol{J}(v)}\big(\boldsymbol{\theta}, \boldsymbol{R}(\tau_x)\big) + \gamma(\boldsymbol{\theta})\tau_x$$

$$-\sum_{i \in J_e(v)} \int_0^{\tau_x} \gamma_{e,i}(v, \theta_i)1\big(R_{e,i}(u) > v\big)du$$

$$-\sum_{i \in J_s(v)} \int_0^{\tau_x} \gamma_{s,i}(v, \boldsymbol{\theta})1\big(R_{s,i}(u) > v\big)du$$

$$-\sum_{i \in \mathcal{J}\setminus A} \int_0^{\tau_x} \gamma_{J_s(v),i}(\boldsymbol{\theta})1\big(L_i(u) = 0\big)du\Big)\Big],$$

where $\gamma_{J_s(v),i}(\boldsymbol{\theta}) = \gamma_{s,i}(v, \boldsymbol{\theta})1\big(i \in J_s(v)\big) + \gamma_{s,i}(\boldsymbol{\theta})1\big(i \neq J_s(v)\big)$. Note that the first integration term with minus sign in the exponent of (4.9) is bounded below by zero by the choice of $\boldsymbol{J}(v)$. We now take the following steps for the lower bounds.

S t e p 1′. Choose $\boldsymbol{\theta} \in \Gamma^{\text{OUT}}$ such that $\gamma_{s,i}(\boldsymbol{\theta}) < 0$ for all $i \in \mathcal{J} \setminus A$, which implies $i \in J_s(v)$ and, for sufficiently large $v > 0$,

$$\exp\Big(\gamma(\boldsymbol{\theta})\tau_x - \sum_{i \in \mathcal{J}\setminus A} \int_0^{\tau_x} \gamma_{s,i}(v, \boldsymbol{\theta})1\big(L_i(u) = 0\big)du\Big) \geqslant 1.$$

The lower bounds are only used for Theorem 2.2. Thus, $A = \mathcal{J}$ for general d and $A = \{k\}$ for $d = 2$. We also keep the same $\boldsymbol{\theta}$ in Steps 2′–4′ below.

S t e p 2′. Verify that there is a constant C_2 such that, if $\tau_x < \infty$, then

$$(4.10) \quad Y(\tau_x) > C_2.$$

This inequality can be considered opposite to (4.5) in Step 2 because of (4.8).

S t e p 3′. Find finite-valued functions $\underline{a}_0(\boldsymbol{\theta}), \underline{a}_1(\boldsymbol{\theta})$ such that

$$(4.11) \quad \underline{a}_1(\boldsymbol{\theta})x - \langle\boldsymbol{\theta}, \boldsymbol{L}(\tau_x)\rangle \geqslant \underline{a}_0(\boldsymbol{\theta});$$

then $e^{-\langle\boldsymbol{\theta}, \boldsymbol{L}(\tau_x)\rangle}$ is bounded below by $e^{\underline{a}_0(\boldsymbol{\theta})-\underline{a}_1(\boldsymbol{\theta})x}$.

S t e p 4′. Find a subset U of $\partial_A S_{\mathrm{L}}$ such that

$$(4.12) \quad \liminf_{x\to\infty} \widetilde{\mathbb{P}}_{\nu_A^-}^{(\boldsymbol{\theta})}\big(\boldsymbol{L}(0) \in U, \tau_x < \infty\big) > 0,$$

$$(4.13) \quad \widetilde{\mathbb{E}}_{\nu_A^-}^{(\boldsymbol{\theta})}\Big(1\big(\boldsymbol{L}(0) \in U\big)f_{\boldsymbol{\theta}}\big(X(0)\big)\Big) < \infty.$$

S t e p 5′. The final step is similar to Step 5 of the upper bound.

In this procedure, we first need to find appropriate $B(x)$ and τ_x so that (4.11) and (4.12) hold, then go through steps. Among them, (4.12) is technically most demanding.

4.2. Lemmas for tail asymptotics. We start with the following lemma.

LEMMA 4.1. *For each $A \subset \mathcal{J}$, let us choose an open or closed set $B(x) \subset S_{\mathrm{L}} \setminus \partial_A S_{\mathrm{L}}$ for $x > 0$, and let $\boldsymbol{\theta} \in \Gamma_A^{\mathrm{IN}}$. If there is a positive constant c_0 to be independent of x such that*

$$(4.14) \qquad \sup\{|\langle \boldsymbol{\theta}, (\boldsymbol{z} - \boldsymbol{z}') \rangle|; \boldsymbol{z}, \boldsymbol{z}' \in B(x)\} < c_0 \|\boldsymbol{\theta}\|,$$

then (4.5) with $\tau_x = \tau_{B(x)}^{\mathrm{IN}}$ holds, where C_1 is independent of x.

P r o o f. We follow the proving method of Lemma 4.6 of [18]. We replace $\boldsymbol{L}(\cdot)$ by $\boldsymbol{H}(\cdot)$ such that $\boldsymbol{H}(\cdot)$ is obtained from $\boldsymbol{L}(\cdot)$ by removing the reflecting boundary $\partial_A S_{\mathrm{L}}$. Hence, the state space of $\boldsymbol{H}(\cdot)$ has no limitation concerning entries with indices in A. For $t > 0$, let

$$\sigma_x(t) = \inf\{u \geqslant t + \tau_{B(x)}^{\mathrm{IN}}; \boldsymbol{H}(u) \in B(x)\};$$

then, on $\{\tau_{B(x)}^{\mathrm{IN}} < \infty\}$,

$$t \leqslant \int_{\tau_{B(x)}^{\mathrm{IN}}}^{\tau_A^{\mathrm{RE}}} 1\big(\boldsymbol{L}(u) \in B(x)\big) du \leqslant \int_{\tau_{B(x)}^{\mathrm{IN}}}^{\infty} 1\big(\boldsymbol{H}(u) \in B(x)\big) du$$

implies that $t + \tau_{B(x)}^{\mathrm{IN}} \leqslant \sigma_x(t) < \infty$. Hence, we have, on $\{\tau_{B(x)}^{\mathrm{IN}} < \infty\}$,

$$(4.15) \quad Y(\tau_{B(x)}^{\mathrm{IN}}) e^{w(\boldsymbol{\theta}, \boldsymbol{R}(\tau_{B(x)}^{\mathrm{IN}}))}$$

$$= \mathbb{E}_{\nu_A^-}\Big(\int_{\tau_{B(x)}^{\mathrm{IN}}}^{\tau_A^{\mathrm{RE}}} 1\big(\boldsymbol{H}(u) \in B(x)\big) du \, e^{w(\boldsymbol{\theta}, \boldsymbol{R}(\tau_{B(x)}^{\mathrm{IN}}))} \big| \mathcal{F}_{\tau_{B(x)}^{\mathrm{IN}}} \Big)$$

$$= \mathbb{E}_{\nu_A^-}\Big(\int_0^{\infty} 1\Big(\int_{\tau_{B(x)}^{\mathrm{IN}}}^{\tau_A^{\mathrm{RE}}} 1\big(\boldsymbol{H}(u) \in B(x)\big) du > t \Big) dt \, e^{w(\boldsymbol{\theta}, \boldsymbol{R}(\tau_{B(x)}^{\mathrm{IN}}))} \big| \mathcal{F}_{\tau_{B(x)}^{\mathrm{IN}}} \Big)$$

$$\leqslant \int_0^{\infty} \mathbb{E}_{\nu_A^-}\Big(e^{w(\boldsymbol{\theta}, \boldsymbol{R}(\tau_{B(x)}^{\mathrm{IN}}))} 1\big(\sigma_x(t) < \infty\big) \big| \mathcal{F}_{\tau_{B(x)}^{\mathrm{IN}}} \Big) dt.$$

We evaluate the integrand in the last integral using change of measure by $\boldsymbol{H}(\cdot)$ similar to $\boldsymbol{L}(\cdot)$. Let

$$J_e(v) = \{i \in \mathcal{E}; \gamma_{e,i}(\theta_i) > 0\}, \qquad J_s(v) = \{i \in \mathcal{J}; \gamma_{s,i}(\boldsymbol{\theta}) > 0\},$$

and we choose a sufficiently large v such that $\gamma_{e,i}(v, \theta_i) > 0$ for all $i \in J_e(v)$ and $\gamma_{s,j}(v, \boldsymbol{\theta}) > 0$ for all $j \in J_s(v)$, which is possible by the same reason as that used for (4.7).

For change of measure, we use the test function $f_{J(v),\theta}$ of (2.14) and the martingale $M_{J(v),\theta}$ of (2.26), where $L(t)$ is replaced by $H(t)$. Then, the exponential martingale $E^f(\cdot)$ is obtained as

$$
(4.16) \quad E^{f_{J(v),\theta}}(t) = \frac{e^{\langle \theta, H(t) \rangle - w_{J(v)}(\theta, R(t))}}{e^{\langle \theta, H(0) \rangle - w_{J(v)}(\theta, R(0))}}
$$
$$
\times \exp \Big(-\gamma_{J(v)}(\theta)t + \sum_{i \in \mathcal{J} \backslash A} \gamma_{s,i}(\theta) \int_0^t 1\big(H_i(u) = 0\big)du
$$
$$
+ \sum_{i \in J_e(v)} \gamma_{e,i}(v, \theta_i) \int_0^t 1\big(R_{e,i}(u) > v\big)du
$$
$$
+ \sum_{i \in J_s(v)} \gamma_{s,i}(v, \theta_i) \int_0^t 1\big(R_{s,i}(u) > v\big)du \Big),
$$

and define the new measure $\widetilde{\mathbb{P}}_\nu^{(J(v), \theta)}$. Since $\theta \in \Gamma_A^{\text{IN}}$ implies that $\gamma(\theta) < 0$, there is a sufficiently large v such that $\gamma_{J(v)}(\theta) < 0$. We choose this v; then it follows from its conditional expectation version (3.4) that

$$
(4.17) \quad \mathbb{E}_{\nu_A^-} \Big(e^{w_{J(v)}(\theta, R(\tau_{B(x)}^{\text{IN}}))} 1\big(\sigma_x(t) < \infty\big) \big| \mathcal{F}_{\tau_{B(x)}^{\text{IN}}} \Big)
$$
$$
= \widetilde{\mathbb{E}}_{\nu_A^-}^{(J(v), \theta)} \Big[\frac{e^{\langle \theta, H(\tau_{B(x)}^{\text{IN}}) \rangle - w(\theta, R(\tau_{B(x)}^{\text{IN}}))}}{e^{\langle \theta, H(\sigma_x) \rangle - w(\theta, R(\sigma_x))}} 1\big(\sigma_x(t) < \infty\big)
$$
$$
\times \exp \Big(w_{J(v)}\big(\theta, R(\tau_{B(x)}^{\text{IN}})\big) + \gamma_{J(v)}(\theta)\big(\sigma_x(t) - \tau_{B(x)}^{\text{IN}}\big)
$$
$$
- \sum_{i \in J_e(v)} \gamma_{e,i}(v, \theta_i) \int_{\tau_{B(x)}^{\text{IN}}}^{\sigma_x(t)} 1\big(R_{e,i}(u) > v\big)du
$$
$$
- \sum_{i \in J_s(v)} \gamma_{s,i}(v, \theta_i) \int_{\tau_{B(x)}^{\text{IN}}}^{\sigma_x(t)} 1\big(R_{s,i}(u) > v\big)du
$$
$$
- \sum_{i \in \mathcal{J} \backslash A} \gamma_{s,i}(\theta) \int_{\tau_{B(x)}^{\text{IN}}}^{\sigma_x(t)} 1\big(H_i(u) = 0\big)du \Big) \Big| \mathcal{F}_{\tau_{B(x)}^{\text{IN}}} \Big]
$$
$$
\leqslant \widetilde{\mathbb{E}}_{\nu_A^-}^{(J(v), \theta)} \Big(e^{-\langle \theta, H(\sigma_x(t)) \rangle + \langle \theta, H(\tau_{B(x)}^{\text{IN}}) \rangle}
$$
$$
\times e^{w_{J(v)}(\theta, R(\sigma_x(t))) + \gamma_{J(v)}(\theta)t} 1\big(\sigma_x(t) < \infty\big) \big| \mathcal{F}_{\tau_{B(x)}^{\text{IN}}} \Big)
$$
$$
\leqslant \widetilde{\mathbb{E}}_{\nu_A^-}^{(J(v), \theta)} \Big(e^{-\langle \theta, H(\sigma_x(t)) \rangle + \langle \theta, H(\tau_{B(x)}^{\text{IN}}) \rangle} \Big)
$$
$$
\times \exp \Big(\sum_{i \in J_e(v)} \gamma_{e,i}(v, \theta_i)v + \sum_{i \in J_s(v)} \gamma_{s,i}(v, \theta_i)v + \gamma_{J(v)}(\theta)t \Big),
$$

since $\sigma_x(t) - \tau_{B(x)}^{\text{IN}}) \geqslant t$ on $\{\sigma_x(t) < \infty\}$; $\gamma_{e,i}(\boldsymbol{\theta}) > 0$ if and only if $i \in J_e(v)$; $\gamma_{s,j}(\boldsymbol{\theta}) > 0$ if and only if $j \in J_s(v)$; and $\gamma_{s,i}(\boldsymbol{\theta}) > 0$ for $i \in \mathcal{J} \setminus A$ by $\boldsymbol{\theta} \in \Gamma_A^{\text{IN}}$. We note here that the condition (4.14) implies

$$(4.18) \qquad -\langle \boldsymbol{\theta}, \boldsymbol{H}(\sigma_x(t)) \rangle + \langle \boldsymbol{\theta}, \boldsymbol{H}(\tau_{B(x)}^{\text{IN}}) \rangle \leqslant \|\boldsymbol{\theta}\| c_0,$$

from which the last term in (4.17) is bounded by

$$c(v) e^{\gamma_{\boldsymbol{J}(v)}(\boldsymbol{\theta}) t}, \quad t \geqslant 0, v > 0,$$

where

$$c(v) = \exp \Big(c_0 \|\boldsymbol{\theta}\| + \sum_{i \in J_e(v)} \gamma_{e,i}(v, \theta_i) v + \sum_{i \in J_s(v)} \gamma_{s,i}(v, \theta_i) v \Big).$$

Hence, the last term in (4.15) is bounded by

$$\int_0^\infty c(v) e^{\gamma_{\boldsymbol{J}(v)}(\boldsymbol{\theta}) t} dt = -\frac{1}{\gamma_{\boldsymbol{J}(v)}(\boldsymbol{\theta})} c(v) < \infty.$$

This proves the lemma. ∎

In the proof of Lemma 4.1, the condition (4.14) may be weakened as long as (4.18) holds. However, we also require the conditions (4.3) and (4.6) for $\tau_x = \tau_{B(x)}^{\text{IN}}$ to get an upper bound. In the view of these conditions, (4.14) is close to be necessary.

LEMMA 4.2. *Assume that $\boldsymbol{\theta} \in \Gamma_A^{\text{IN}}$. Then we have $\widetilde{\mathbb{E}}_{\nu_A^-}^{(\boldsymbol{\theta})}\big(f_{\boldsymbol{\theta}}(X(0)) \big) < \infty$ for $A = \{k\}$ for each $k \in \mathcal{J}$ if $\varphi_k(\boldsymbol{\theta}) < \infty$ and $\boldsymbol{\theta} \in \Theta_k$, where we recall that $\Theta_k = \big\{ \boldsymbol{\theta} \in \mathbb{R}^d; \theta_i \geqslant 0, \forall i \in \mathcal{J} \setminus \{k\} \big\} \cup \big\{ \boldsymbol{\theta} \in \mathbb{R}^d; \theta_i < 0, \forall i \in \mathcal{J} \setminus \{k\} \big\}.$*

Proof. Since $\mathbb{P}_{\nu_k^-}$ is identical with $\widetilde{\mathbb{P}}_{\nu_k^-}^{(\boldsymbol{\theta})}$ on $\mathcal{F}_0$, it is enough to show that

$$(4.19) \qquad \mathbb{E}_{\nu_k^-}\Big(f_{\boldsymbol{\theta}}(X(0)) \Big) = \mathbb{E}_{\nu_k^-}(e^{\langle \boldsymbol{\theta}, \boldsymbol{L}(0) \rangle - \langle \boldsymbol{\gamma}_e(\boldsymbol{\theta}), \boldsymbol{R}_e(0) \rangle - \langle \boldsymbol{\gamma}_s(\boldsymbol{\theta}), \boldsymbol{R}_s(0) \rangle}) < \infty.$$

We first show that $\varphi_k(\boldsymbol{\theta}) \equiv \mathbb{E}\big(e^{\langle \boldsymbol{\theta}, \boldsymbol{L} \rangle} 1(L_k = 0) \big) < \infty$ implies $\mathbb{E}_{\nu_k^-}(e^{\langle \boldsymbol{\theta}, \boldsymbol{L} \rangle}) < \infty$. To see this, let $N_{e,i}(\cdot)$ be the counting process for the exogenous arrivals at station i, and let $N_{d,j,i}(\cdot)$ be the counting process for the customers who completed service at station j and routed to station i. Then the Palm formulas for stationary point processes yield

$$(4.20) \quad \mathbb{E}_{\nu_k^-}(e^{\langle \boldsymbol{\theta}, \boldsymbol{L} \rangle}) \leqslant 1(k \in \mathcal{E}) \lambda_k \mathbb{E}_\nu \Big(\int_0^1 e^{\langle \boldsymbol{\theta}, \boldsymbol{L}(t-) \rangle} 1\big(\boldsymbol{L}(t-) \in \partial_k S_{\text{L}} \big) N_{e,k}(dt) \Big)$$

$$+ \sum_{j \in \mathcal{J} \setminus \{k\}} \alpha_j p_{jk} \mathbb{E}_\nu \Big(\int_0^1 e^{\langle \boldsymbol{\theta}, \boldsymbol{L}(t-) \rangle} 1\big(\boldsymbol{L}(t-) \in \partial_k S_{\text{L}} \big) N_{d,j,k}(dt) \Big),$$

where $\partial_k S_{\mathrm{L}} = \partial_{\{k\}} S_{\mathrm{L}}$, and the inequality becomes equality if point processes $N_{e,k}$, $N_{d,j,k}$ have no common point. Let $N_{e,i} \equiv 0$ for $i \in \mathcal{J} \setminus \mathcal{E}$ for convenience. Then

$$\max_{t \in (0,1]} \big(L_i(t) - L_i(0) \big)$$
$$\leqslant d + 1 + N_{e,i}\Big(\big(t_{e,i}(1), 1 \big] \Big) + \sum_{j \in \mathcal{J}} N_{s,j,i}\Big(\big(t_{s,j}(1), 1 \big] \Big), \quad i \in \mathcal{J},$$

where $N_{s,j,i}(\cdot)$ is the stationary counting process for the number of service completions routed to station i when the server at station j is always busy. Since $\boldsymbol{L}(0)$ is independent of $N_{e,i}\big((t_{e,i}(1), 1] \big)$ and $N_{s,j}\big((t_{s,j}(1), 1] \big)$ for $i \in \mathcal{E}, j \in \mathcal{J}$, (4.20) implies

$$\mathbb{E}_{\nu_k^-}\big(e^{\langle \boldsymbol{\theta}, \boldsymbol{L} \rangle} \big) \leqslant \varphi_k(\boldsymbol{\theta}) \big(\lambda_k + \sum_{j' \in \mathcal{J} \setminus \{k\}} \widetilde{\alpha}_{j'} p_{j'k} \big)$$
$$\times \mathbb{E}_\nu \Big(\exp \Big(\sum_{i \in \mathcal{E}} |\theta_i| \Big(d + 1 + N_{e,i}\big((0,1] \big) \Big) + \sum_{j \in \mathcal{J}} N_{s,j,i}\big((0,1] \big) \Big) \Big).$$

This proves the claim that $\mathbb{E}_{\nu_k^-}\big(e^{\langle \boldsymbol{\theta}, \boldsymbol{L} \rangle} \big) < \infty$ since $N_{e,i}\big((0,1] \big)$ and $N_{s,j}\big((0,1] \big)$ for $i \in \mathcal{E}, j \in \mathcal{J}$ are independent and have super-exponential distributions, that is, their tails are asymptotically faster than any exponential function (see, e.g., Lemma 4.1 of [18]).

We now prove (4.19). Note that its terms multiplied by $\gamma_{e,i}(\theta_i) \geqslant 0$, which is equivalent to $\theta_i \geqslant 0$, and $\gamma_{s,i}(\boldsymbol{\theta})$ for $i \in \mathcal{J} \setminus \{k\}$ can be dropped to bound the second expectation term in (4.19) because $\boldsymbol{\theta} \in \Gamma_k^{\mathrm{IN}}$. Furthermore, $R_{s,k}(0) = T_{s,k}$ under the distribution ν_k^-. Let $K_-(k, \boldsymbol{\theta}) = \big\{ i \in \mathcal{E} \setminus \{k\}; \theta_i < 0 \big\}$. Thus, it follows from the equation in (4.19) that
(4.21)
$$\mathbb{E}_{\nu_k^-}\Big(f_{\boldsymbol{\theta}}\big(X(0) \big) \Big) \leqslant \mathbb{E}_{\nu_k^-}\Big(\exp \big(\sum_{i \in \mathcal{J} \setminus \{k\}} \theta_i L_i(0) - \sum_{i \in K_-(k,\boldsymbol{\theta})} \gamma_{e,i}(\theta_i) R_{e,i}(0) \big) \Big) e^{\theta_k}.$$

Thus, (4.19) is immediate if $\theta_i \geqslant 0$, equivalently, $\gamma_{e,i}(\theta_i) \geqslant 0$, for all $i \in \mathcal{J} \setminus \{k\}$ since $\theta_k \geqslant 0$ and $\varphi_k(\boldsymbol{\theta}) < \infty$. Hence, it remains to prove (4.19) when $\theta_i < 0$ for all $i \in \mathcal{J} \setminus \{k\}$. In this case, (4.19) is obtained from the relation

$$(4.22) \qquad \mathbb{E}_{\nu_k^-}\big(e^{-W_{e,K_-(k,\boldsymbol{\theta})}(\boldsymbol{\theta}, t)} \big) < \infty,$$

where $W_{e,K_-(k,\boldsymbol{\theta})}(\boldsymbol{\theta}, t) = \sum_{i \in K_-(k,\boldsymbol{\theta})} \gamma_{e,i}(\theta_i) R_{e,i}(t)$. Let $\mathbb{E}_{e,i}$ and $\mathbb{E}_{d,j}$ represent the expectations concerning the stationary embedded distributions at exogenous arrivals at station i and at departure instants at station j, respectively. Then, they

are known as Palm distributions (e.g., see [3]), and obtained as

$$\mathbb{E}_{e,i}\Big(\exp\big(-\sum_{i\in K_-(A,\boldsymbol{\theta})} \gamma_{e,i}(\theta_i)R_{e,i}(0))\big)\Big) = \lambda_i \mathbb{E}_\nu\big(\int\limits_0^1 e^{-W_{e,K_-(A,\boldsymbol{\theta})}(\boldsymbol{\theta},t)} N_{e,i}(dt)\big),$$

$$\mathbb{E}_{d,j}\Big(\exp\big(-\sum_{i\in K_-(A,\boldsymbol{\theta})} \gamma_{e,i}(\theta_i)R_{e,i}(0))\big)\Big) = \alpha_j \mathbb{E}_\nu\big(\int\limits_0^1 e^{-W_{e,K_-(A,\boldsymbol{\theta})}(\boldsymbol{\theta},t)} N_{d,j}(dt)\big),$$

where $N_{d,j}(t) = \sum_{i\in\mathcal{J}} N_{d,j,i}(t)$. From a similar bound in (4.20), the inequality (4.22) is obtained by using

$$(4.23) \qquad \mathbb{E}_{e,i}\Big(\exp\big(-\sum_{i\in K_-(A,\boldsymbol{\theta})} \gamma_{e,i}(\theta_i)R_{e,i}(0))\big)\Big) < \infty, \quad i \in A \cap \mathcal{E},$$

$$(4.24) \qquad \mathbb{E}_{d,j}\Big(\exp\big(-\sum_{i\in K_-(A,\boldsymbol{\theta})} \gamma_{e,i}(\theta_i)R_{e,i}(0))\big)\Big) < \infty, \quad j \in \mathcal{J}.$$

Both can be proved similarly to Lemma 4.8 of [18] (see Section S2 of its supplemented version for its proof), which is originally in Lemma 4.2 of [21]. ∎

LEMMA 4.3. *For $d = 2$, $k = 1, 2$ and a compact set $B_0 \subset \mathbb{R}_+^2$, let $B_k(x) = x\mathbf{e}_k + B_0$, and let*

$$(4.25) \qquad \boldsymbol{\theta}^{(\mathrm{CP}_k)} = \arg\sup_{\boldsymbol{\theta}\in\Gamma_k^{\mathrm{IN}}} \theta_k, \quad k = 1, 2.$$

Then $\liminf_{x\to\infty} \widetilde{\mathbb{P}}_{\nu_{\{k\}}^-}^{(\boldsymbol{\theta})} (\tau_{B_k(x)}^{\mathrm{IN}} < \infty) > 0$ if $\|\boldsymbol{\theta} - \boldsymbol{\theta}^{(\mathrm{CP}_k)}\|$ is sufficiently small and $\theta_j < \theta_j^{(\mathrm{CP}_k)}$ for $j \neq k$ and $j \in \{1,2\}$.

P r o o f. For notational symmetry, we consider only the case for $k = 1$. Clearly, the lemma is obtained if $\liminf_{x\to\infty} \widetilde{\mathbb{P}}_{\nu_{\{k\}}^-}^{(\boldsymbol{\theta})} \big(\boldsymbol{L}(0) = \boldsymbol{z}, \tau_{B_k(x)}^{\mathrm{IN}} < \infty\big) > 0$. It is not hard to see that this is proved if station 1 is unstable and station 2 is stable under $\widetilde{\mathbb{P}}_{\nu_{\{k\}}^-}^{(\boldsymbol{\theta})} \big(\,\cdot\,|\boldsymbol{L}(0) = \boldsymbol{z}\big)$. By Lemma 2.9 and (3.12), this holds if

$$[\nabla\gamma(\boldsymbol{\theta})]_2 < 0, \quad \langle\nabla\gamma(\boldsymbol{\theta}), \boldsymbol{t}_1^{(\boldsymbol{\theta})}\rangle < 0.$$

The inequalities above are satisfied if $\boldsymbol{\theta}$ is chosen so that $\|\boldsymbol{\theta} - \boldsymbol{\theta}^{(\mathrm{CP}_1)}\|$ is sufficiently small, $\theta_2 < \theta_2^{(\mathrm{CP}_1)}$ and

$$(4.26) \qquad [\nabla\gamma(\boldsymbol{\theta}^{(\mathrm{CP}_1)})]_2 \leqslant 0, \quad \langle\nabla\gamma(\boldsymbol{\theta}^{(\mathrm{CP}_1)}), \boldsymbol{t}_1^{(\boldsymbol{\theta}^{(\mathrm{CP}_1)})}\rangle < 0,$$

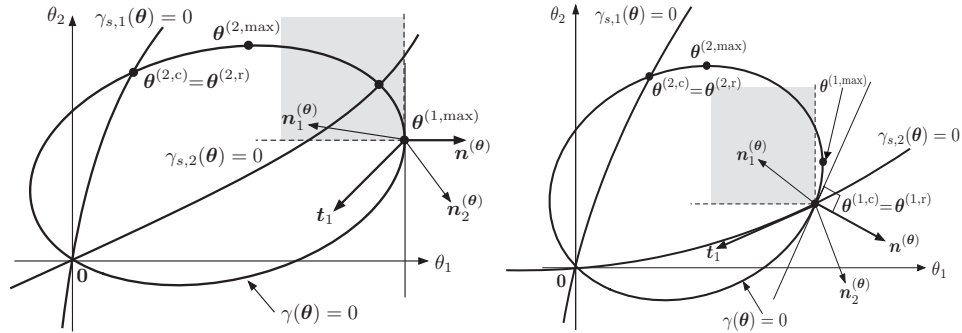where we recall that $\boldsymbol{t}_1^{(\boldsymbol{\theta})}$ is defined in Lemma 3.2. The first inequality follows

FIGURE 3. Geometric objects for $d = 2$, where $\boldsymbol{n}^{(\boldsymbol{\theta})} = \nabla\gamma(\boldsymbol{\theta})$, $\boldsymbol{n}_i^{(\boldsymbol{\theta})} = \nabla\gamma_i(\boldsymbol{\theta})$ for $i = 1, 2$.

from the convexity of Γ^{IN} and the definition (4.25) (see Figure 3). For the second inequality, let

$$\boldsymbol{\theta}^{(1,\max)} = \arg\sup_{\boldsymbol{\theta}\in\Gamma^{\text{IN}}} \theta_1.$$

If $\boldsymbol{\theta}^{(\boldsymbol{\theta}^{(\text{CP}_1)})} = \boldsymbol{\theta}^{(1,\max)}$, then the second inequality of (4.26) is immediate because $\nabla\gamma(\boldsymbol{\theta}^{\boldsymbol{\theta}^{(1,\max)}})$ is proportional to $\mathbf{e}_1 \equiv (1, 0)$ while $\boldsymbol{t}_1^{(\boldsymbol{\theta}^{(\text{CP}_1)})} < \mathbf{0}$ by Lemma 3.2. Otherwise, assume that $\boldsymbol{\theta}^{(\boldsymbol{\theta}^{(\text{CP}_1)})} \neq \boldsymbol{\theta}^{(1,\max)}$, and let f be a function from $\mathbb{R}$ to $\mathbb{R}$ such that $\theta_2 = f(\theta_1)$ is determined by $\gamma_{s,2}(\boldsymbol{\theta}) = 0$. We then observe that $f(\theta_1)$ is increasing convex in θ_1, and its derivative is smaller than that of the curve $\gamma(\boldsymbol{\theta}) = 0$ at $\boldsymbol{\theta} = \boldsymbol{\theta}^{(\text{CP}_k)}$ because $\boldsymbol{\theta}^{(\text{CP}_k)}$ is only one cross point of those two curves for $\theta_1 > 0$ and Γ^{IN} is not empty. Again the tangent vector $\boldsymbol{t}_1^{(\boldsymbol{\theta}^{(\text{CP}_1)})} < \mathbf{0}$ by Lemma 3.2, and therefore the second inequality of (4.26) must hold. ∎

LEMMA 4.4. *Under $\widetilde{\mathbb{P}}^{(\boldsymbol{\theta})}$, all stations of the GJN are weakly unstable if $\nabla\gamma(\boldsymbol{\theta}) \geqslant 0$ for $\boldsymbol{\theta} \in \mathbb{R}^d$.*

P r o o f. By (3.12) and the choice of $\boldsymbol{\theta}$, $\gamma^{(\boldsymbol{\theta})}(\mathbf{0}) \geqslant \mathbf{0}$, and therefore (c) of Lemma 2.7 proves this lemma. ∎

4.3. Proofs of the main results. In this subsection, we prove Theorems 2.1 and 2.2 and Corollary 2.1.

P r o o f o f T h e o r e m 2.1. We apply the procedure from Section 4.1.

(a) Fix $k \in \mathcal{J}$, put $B(x) = x\mathbf{e}_k + B_0$, and let $\tau_x = \tau_{B(x)}^{\text{IN}}$. Since B_0 is a compact set, (4.14) is satisfied. Hence, Steps 1–5 are verified by Lemmas 4.1 and 4.2.

(b) We first prove (2.38). To use similar arguments to those in (a), we put $B(x) = x\boldsymbol{c} + B_0$, and let $\tau_x = \tau_{B(x)}^{\text{IN}}$. We first prove, for each $A \neq \emptyset$,

$$(4.27) \qquad \limsup_{x\to\infty} \frac{1}{x}\log\mathbb{P}(\boldsymbol{L} \in x\boldsymbol{c} + B_0) \leqslant -r_A(\boldsymbol{c}).$$

We only need to verify Step 3, that is, $\mathbb{E}_{\nu_i^-}\big(f_{\boldsymbol{\theta}}\big(X(0)\big)\big) < \infty$ for all $i \in A$ because Lemma 4.2 cannot be used. We use here the assumption (A1); then it is not hard to see that, for $i \in A$, $\varphi_i(\boldsymbol{\theta}) < \infty$ implies $\mathbb{E}_{\nu_i}\big(f_{\boldsymbol{\theta}}\big(X(0)\big)\big) < \infty$. The latter finiteness implies that $\mathbb{E}_{\nu_i^-}\big(f_{\boldsymbol{\theta}}\big(X(0)\big)\big) < \infty$, as shown in the proof of Lemma 4.2. Thus, Step 3 is verified, and (4.27) is obtained. Taking the minimum of the right-hand side of (4.27) for $A \subset \mathcal{J}$ and $A \neq \emptyset$, we obtain (2.38).

We next prove (2.39). Let $\boldsymbol{\theta} = u\boldsymbol{c}$ for $u > 0$ for an arbitrarily chosen $\boldsymbol{c} \in \overleftarrow{U}_d$, and put

$$B(x) = \{\boldsymbol{z} \in \mathbb{Z}_+^d; x < \langle u\boldsymbol{c}, \boldsymbol{z} \rangle \leqslant x + 1\};$$

then (4.14) is satisfied, and therefore we can use Lemma 4.2. By (A1), Step 3 works as shown in (a). For Step 4, we put $\overline{a}_0(\boldsymbol{\theta}) = 0$ and $\overline{a}_1(\boldsymbol{\theta}) = u$; then (4.6) is satisfied. Thus, if we choose $u\boldsymbol{c} \in \Gamma_A$, all the steps work, and we have

$$\limsup_{x \to \infty} \frac{1}{x} \log \mathbb{P}(x < \langle \boldsymbol{c}, \boldsymbol{L} \rangle \leqslant x + 1) \leqslant -u.$$

This obviously implies that

$$\limsup_{x \to \infty} \frac{1}{x} \log \mathbb{P}(\langle \boldsymbol{c}, \boldsymbol{L} \rangle > x) \leqslant -u$$

for $u\boldsymbol{c} \in \overleftarrow{\Gamma}_A^{\text{IN}}$ as long as $\varphi_i(\boldsymbol{\theta}) < \infty$ for all $i \in A$. Hence, we have

$$\limsup_{x \to \infty} \frac{1}{x} \log \mathbb{P}(\langle \boldsymbol{c}, \boldsymbol{L} \rangle > x) \leqslant -m_A(\boldsymbol{c}).$$

Thus, we complete the proof by taking the minimum of the right-hand side of the above inequality over $A \subset \mathcal{J} \setminus \emptyset$. ∎

P r o o f o f T h e o r e m 2.2. We apply the lower bound procedure $1'$–$5'$.

(a) Because of symmetry, it suffices to prove for $k = 1$. Put $A = \{1\}$, and let $B(x) = x\mathbf{e}_1 + B_0$ and $\tau_x = \tau_{B(x)}^{\text{IN}}$. We choose $\boldsymbol{\theta}$ such that $\gamma(\boldsymbol{\theta}) > 0$ and $\gamma_{s,2}(\boldsymbol{\theta}) < 0$; then Step $1'$ works. Step $2'$ is obviously verified because $Y(\tau_x)$ does not decrease as x gets large. Step $3'$ is also obvious because B_0 is compact. For Step $4'$, we can take any bounded set for $U \subset \partial_A S_{\text{L}}$. Then, if we take $\boldsymbol{\theta}$ which is sufficiently close to $\boldsymbol{\theta}^{(\text{CP}_k)}$, then (4.12) holds by Lemma 4.3, while (4.13) obviously holds. This completes the procedure, and (2.40) is obtained.

(b) We restrict the initial state in a bounded set C so that $C \subset \partial_{\mathcal{J}} S_{\text{L}} \times S_{\text{R}}$ and

$$\mathbb{E}\Big(f\big(X(0)\big)1\big(X(0) \in C\big)\Big) > 0.$$

Let $c \in \text{Corn}(\overleftarrow{\Gamma}^{\text{IN}} \cap \partial\Gamma^{\text{IN}})$, which implies $\nabla g(uc) \geqslant 0$ for $uc \in \text{Corn}(\overleftarrow{\Gamma}^{\text{IN}} \cap \partial\Gamma^{\text{IN}})$ by the convexity of Γ^{IN}. Let is choose z_0 such that $x \equiv (z, y) \in C$ implies that $\max_{i \in \mathcal{J}} z_i < z_0$. We let

$$B(x) = \{z \in \mathbb{Z}_+^d; \langle c, z \rangle > x\}, \quad x > z_0,$$

and let $\tau_x = \tau_{B(x)}^{\text{IN}}$. Then, Step 2′ is obviously valid. Because the initial state is in C, we have

$$ux < \langle uc, L(\tau_x) \rangle \leqslant ux + 1.$$

Hence, the condition (4.11) in Step 3′ is satisfied for $x \geqslant z_0$. Furthermore, if we take $\theta = uc$ for the change of measure, then all the stations are weakly unstable by Lemma 4.4, which implies that

$$(4.28) \quad \mathbb{E}\Big(f\big(X(0)\big)1\big(X(0) \in C, \tau_x < \infty\big)\Big) = \mathbb{E}\Big(f\big(X(0)\big)1\big(X(0) \in C\big)\Big) > 0,$$

and therefore (4.12) is satisfied for $A = \mathcal{J}$. Thus, all the steps work well, and the proof is completed.

(c) We take the same $B(x), \tau_x$ and C as in (a). Let $\text{Corn}(a, b) = \{x \in \mathbb{R}_+^2; sa + tb, s, t \geqslant 0\}$ for $a, b \in \mathbb{R}_+^2$. For $c \in \text{Corn}(\overleftarrow{\Gamma}^{\text{IN}} \cap \partial\Gamma_1^{\text{IN}})$, we separately consider the two cases where $c \in \text{Corn}(e_1, \theta^{(\text{CP1})})$ or not. If $c \notin \text{Corn}(e_1, \theta^{(\text{CP1})})$, the asymptotic is covered by (2.41). Thus, we assume that $c \in \text{Corn}(e_1, \theta^{(\text{CP1})})$. We first choose $u > 0$ and $c \in \overrightarrow{U}_2$ such that $uc = \theta^{(\text{CP1})}$, and make the change of measure for $\theta = \theta^{(\text{CP1})}$. Then, we have (4.28) by Lemma 4.3. Hence, we have (2.42). We next let $u = \theta_1^{(\text{CP1})}$ and let $c = e_1$. In this case, we also have (2.42) by (2.40). We finally consider the case where $uc = se_1 + t\theta^{(\text{CP1})} \in \overleftarrow{\Gamma}^{\text{IN}} \cap \partial\Gamma_1^{\text{IN}}$. Let $u = \theta_1^{(\text{CP1})}/c_1$, $s = \theta_1^{(\text{CP1})}$ and $t = \theta_1^{(\text{CP1})}c_2/\theta_2^{(\text{CP1})}$; then $uc = (\theta_1^{(\text{CP1})}, \theta_1^{(\text{CP1})}c_2/c_1)$ is on $\overleftarrow{\Gamma}^{\text{IN}} \cap \partial\Gamma_1^{\text{IN}}$. Hence, we have (2.42). ∎

Proof of Corollary 2.1. (a) For $d = 2$, by Theorems 2.1 and 2.2, we have

$$(4.29) \quad \lim_{x \to \infty} \frac{1}{x} \log \mathbb{P}(L \in xe_k + B_0) = -r_{\{k\}}(e_k), \quad k = 1, 2.$$

Then, we can apply the same algorithm as in Theorem 4.1 of [15] to find $r_{\{k\}}(e_k)$, which shows that the equations (2.43) and (2.44) have a unique solution $\delta \equiv (\delta_1, \delta_2)$, and $r_{\{k\}}(e_k) = \delta_k$. This proves (2.45).

(b) The inequality (2.46) is immediate from (b) of Theorem 2.1 for $A = \mathcal{J}$. It remains to prove (2.47). We first consider the marginal distributions in the coordinate directions. By (2.39) of Theorem 2.1 for $d = 2$, it follows from $\varphi_1(0) \leqslant 1$ that

$$\limsup_{x \to \infty} \frac{1}{x} \log \mathbb{P}(\langle e_1, L \rangle > x) \leqslant -m_{\{1\}}(e_1) = -\sup\{u; ue_1 \in \overleftarrow{\Gamma}_1^{\text{IN}}\}.$$

This together with (2.42) concludes that

$$\limsup_{x\to\infty} \frac{1}{x} \log \mathbb{P}(\langle \mathbf{e}_1, \boldsymbol{L} \rangle > x) = -\sup\{u; u\mathbf{e}_1 \in \overleftarrow{\Gamma}_1^{\text{IN}}\} = -\delta_1,$$

and therefore $\varphi(\theta_1, 0)$ is finite for $\theta_1 < \delta_1$ and diverges for $\theta_1 > \delta_1$. Similarly, $\varphi(0, \theta_2)$ is finite for $\theta_2 < \delta_2$. Since $\varphi_2(\theta_1) \leqslant \varphi(\theta_1, 0)$ and $\varphi_1(\theta_2) \leqslant \varphi(\theta_2, 0)$, it follows again from (2.39) that

$$\limsup_{x\to\infty} \frac{1}{x} \log \mathbb{P}(\langle \boldsymbol{c}, \boldsymbol{L} \rangle > x) \leqslant -m_{\{1,2\}}(\boldsymbol{c})$$

$$= -\sup\{u; u\boldsymbol{c} \in \overleftarrow{\Gamma}^{\text{IN}}, \varphi_1(\theta_2), \varphi_2(\theta_1) < \infty\}$$

$$\leqslant -\sup\{u; u\boldsymbol{c} \in \overleftarrow{\Gamma}^{\text{IN}}, \theta_i < \delta_i, i = 1, 2\}$$

$$= -\sup\{u; u\boldsymbol{c} \in \mathcal{D}_2\}.$$

Thus, we got the upper bound. By (2.41) and (2.42) of Theorem 2.2, this upper bound becomes a lower bound. Hence, we have (2.47). ∎

REFERENCES

[1] S. Asmussen, *Applied Probability and Queues*, second edition, Springer, New York 2003.

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

[3] F. Baccelli and P. Brémaud, *Elements of Queueing Theory: Palm Martingale Calculus and Stochastic Recurrences*, second edition, Springer, Berlin 2003.

[4] A. Braverman, J. G. Dai, and M. Miyazawa, *Heavy traffic approximation for the stationary distribution of a generalized Jackson network: The BAR approach*, Stoch. Syst. 7 (2017), pp. 143–196.

[5] H. Chen and A. Mandelbaum, *Discrete flow networks: Bottlenecks analysis and fluid approximations*, Math. Oper. Res. 16 (1991), pp. 408–446.

[6] H. Chen and A. Mandelbaum, *Stochastic discrete flow networks: Diffusion approximation and bottlenecks*, Ann. Probab. 19 (1991), pp. 1463–1519.

[7] J. G. Dai and M. Miyazawa, *Reflecting Brownian motion in two dimensions: Exact asymptotics for the stationary distribution*, Stoch. Syst. 1 (2011), pp. 146–208.

[8] J. G. Dai and M. Miyazawa, *Stationary distribution of a two-dimensional SRBM: Geometric views and boundary measures*, Queueing Syst. 74 (2013), pp. 181–217.

[9] M. H. A. Davis, *Piecewise deterministic Markov processes: A general class of non-diffusion stochastic models*, J. Roy. Statist. Soc. Ser. B 46 (1984), pp. 353–388.

[10] P. W. Glynn and W. Whitt, *Large deviations behavior of counting processes and their inverses,* Queueing Syst. 17 (1994), pp. 107–128.

[11] J. Jacod and A. N. Shiryaev, *Limit Theorems for Stochastic Processes*, second edition, Springer, Berlin 2003.

[12] J. F. C. Kingman, *A convexity property of positive matrices*, Quart. J. Math. Oxford Ser. (2) 12 (1961), pp. 283–284.

[13] H. Kunita and T. Watanabe, *Notes on transformations of Markov processes connected with multiplicative functionals,* Mem. Fac. Sci. Kyushu Univ. Ser. A 17 (1963), pp. 181–191.

[14] K. Majewski, *Functional continuity and large deviations for the behavior of single-class queueing networks*, Queueing Syst. 61 (2009), pp. 203–241.

[15] M. Miyazawa, *Tail decay rates in double QBD processes and related reflected random walks*, Math. Oper. Res. 34 (2009), pp. 547–575.

[16] M. Miyazawa, *Light tail asymptotics in multidimensional reflecting processes for queueing networks*, TOP 19 (2011), pp. 233–299.

[17] M. Miyazawa, *A superharmonic vector for a nonnegative matrix with QBD block structure and its application to a Markov-modulated two-dimensional reflecting process*, Queueing Syst. 81 (2015), pp. 1–48.

[18] M. Miyazawa, *A unified approach for large queue asymptotics in a heterogeneous multi-server queue*, Adv. in Appl. Probab. 49 (2017), pp. 182–220. Supplemented version on arXiv (https://arxiv.org/abs/1510.01034).

[19] Z. Palmowski and T. Rolski, *A technique of the exponential change of measure for Markov processes*, Bernoulli 8 (2002), pp. 767–785.

[20] M. I. Reiman, *Open queueing networks in heavy traffic*, Math. Oper. Res. 9 (1984), pp. 441–458.

[21] J. S. Sadowsky and W. Szpankowski, *The probability of large queue lengths and waiting times in a heterogeneous multiserver queue. I: Tight limits*, Adv. in Appl. Probab. 27 (1995), pp. 532–566.

[22] W. Whitt, *Stochastic-Process Limits: An Introduction to Stochastic-Process Limits and Their Application to Queues*, Springer, New York 2002.

Masakiyo Miyazawa
Tokyo University of Science
2641 Yamazaki, Noda, Chiba 278-8510, Japan
E-mail: miyazawa@rs.tus.ac.jp