Equilibrium Behavior of Infinite-Dimensional Scaling Limits of Many-Server Stochastic Networks

by

Pooja Agarwal

Integrated BS-MS, Indian Institute of Science Education and Research

Thiruvananthapuram; India, 2014

M.Sc., Brown University; Providence, RI, 2016

A dissertation submitted in partial fulfillment of the

requirements for the degree of Doctor of Philosophy

in The Division of Applied Mathematics at Brown University

PROVIDENCE, RHODE ISLAND

October 2021

This dissertation by Pooja Agarwal is accepted in its present form
by The Division of Applied Mathematics as satisfying the
dissertation requirement for the degree of Doctor of Philosophy.

Date________________ __________________________________

Kavita Ramanan, Ph.D., Advisor

Recommended to the Graduate Council

Date________________ __________________________________

Ravi Mazumdar, Ph.D., Reader

Date________________ __________________________________

Debankur Mukherjee, Ph.D., Reader

Date________________ __________________________________

Hui Wang, Ph.D., Reader

Approved by the Graduate Council

Date________________ __________________________________

Andrew G. Campbell, Dean of the Graduate School

Vita

Education:

2014-present	Ph.D. Candidate, Applied Mathematics	Brown University
2014-2015	Masters of Science, Applied Mathematics	Brown University
2009-14	Integrated BS-MS	Indian Institute of
	(Mathematics Major)	Science Education and Research
		Thiruvananthapuram, India

Publications:

2017	P. Agarwal, U. Manna, and D. Mukherjee (2019). "Stochastic control of tidal dynamics equation with Lévy noise." Appl. Math. Optim. **79** (2019), no. 2, 327–396
Preprint	P. Agarwal, and K. Ramanan, Invariant states of the hydrodynamic equations for randomized load balancing networks
Working Paper	P. Agarwal, and K. Ramanan, Diffusion Limits of Many-Server Queues with Reneging
Working Paper	P. Agarwal, M. Simper, and R. Durrett, The q-voter model on the torus

Awards and Fellowships:

2018	Best Poster Prize, Stochastic Networks Conference, International Centre for Mathematical Sciences (ICMS), Edinburgh
2018	Stella Dafermos Award, Brown University
2016	Outstanding Poster Award, Stochastic Networks Conference, UCSD (1 of 7)

Participation in Schools and Workshops:

2020	Bernoulli-IMS One World Symposium 2020
2019	INFORMS APS Conference, University of Queensland
2019	Mathematics Research Community on Stochastic Spatial Models, American Mathematical Society
2018	51st Annual Convention of ORSI and International Conference, Indian Institute of Technology Bombay
2018	Lectures in Probability and Stochastic Processes XIII, Indian Statistical Institute Bangalore
2018	MIGSAA Summer School on Random Structures and Processes, ICMS
2018	Stochastic Networks Conference, ICMS
2018	Seminar on Stochastic Processes, Brown University
2018	Advancing Women's Impact in Mathematics Symposium, Worcester Polytechnic Institute
2017	INFORMS APS Conference, Northwestern University
2017	Women's Intellectual Network Research Symposium, Brown University
2016	Stochastic Networks Conference, University of California San Diego

Organizing Activities:

2018-present	Co-organizer of the Probability Seminar, Brown University
2018	Local Organizer, Seminar on Stochastic Processes, Brown University

Teaching Experience:

2018 Summer	Instructor, Fun With Math, STEM I for Middle School Students, Brown University
2016 Spring	Teaching Assistant, Operations Research: Probabilistic Models, Brown University
2015 Fall	Teaching Assistant, Methods of Applied Mathematics II, Brown University

Memberships:

Society for Industrial and Applied Mathematics (SIAM)

Association for Women in Mathematics (AWM)

American Mathematical Society (AMS)

Institute of Mathematical Statistics (IMS)

Applied Probability Society, INFORMS

Preface and Acknowledgments

First and foremost, I would like to thank my advisor Kavita Ramanan, for her guidance and mentorship throughout my graduate career. She has always taught with great patience and attention to detail and I am grateful for having had the opportunity to learn from her. None of this work would have been completed without your guidance.

I would also like to thank Dr. Hui Wang, Dr Ravi Mazumdar and Dr. Debankur Mukherjee for taking the time and effort to serve on my thesis committee. Also, to Dr. Haya Kaspi for collaborating and providing valuable insight for this research.

I have been fortunate enough to interact with many members of the Division of Applied Math. To the many current and previous members of the Ramanan research group, who have served as role models, colleagues, and friends: thank you for sharing your endless knowledge, entertaining and thought-provoking conversations. Karen, Kristina, Ross and Amy, thank you for your friendship and support throughout the years. Anna G. Becky, Emily and Mira, thank you for your help with organizing events for the department, and being an excellent support system. I wholeheartedly thank the amazing staff at the Division. Without their work, none of this would be possible. In particular, I thank Carly Hansen-Decelles and Candida Hall for helping me at every step.

Lastly, I want to thank my parents and brother for their love and support.

Abstract of “Equilibrium Behavior of Infinite-Dimensional Scaling Limits of Many-Server Stochastic Networks ”, by Pooja Agarwal, Ph.D., Brown University, October 2021

Stochastic networks arise in a variety of real world applications including telecommunications, service systems such as call centers, computer networks, health care services and biological systems. Leaving aside some very simple examples, it is usually infeasible to perform an exact analysis of these networks. A useful alternative is instead to identify a suitable approximation that provides insight into performance and can be shown to be accurate in a relevant asymptotic regime.

Networks in which jobs or customers are processed at nodes with multiple servers arise in many applications, and tend to be harder to analyze than networks where each node is processed only by a single server. Moreover, when the service times of jobs are generally distributed, the analysis becomes even more challenging. For example, the dimension of a Markovian state descriptor of a many-server network with general service distributions typically grows with the size of the network. Thus, a common Markovian state description that is applicable for a network of any size must necessarily be infinite-dimensional. This thesis analyzes the equilibrium behavior of two different types of scaling limits of two classes of many-server stochastic networks. In both cases, the scaling limits are infinite-dimensional and require the development of new techniques for their analysis.

The first part of the thesis considers a many-server parallel network, where each server has its own queue, to which jobs arriving with independent and identically distributed (i.i.d.) service times are routed immediately on arrival using the shortest-of-d randomized load balancing algorithm. Such models are of interest in hash tables in data switches, parallel computing, wireless networks, etc. The focus here is on the so-called hydrodynamic limit, which describes the mean behavior of the state

dynamics, as the number of servers goes to infinity. For a general class of service distributions, the hydrodynamic limit was characterized by Aghajani and Ramanan as the unique solution to a countable coupled system of deterministic measure-valued processes. This thesis shows that the hydrodynamic limit has an invariant state, which is a probability measure that describes the distribution of jobs in queues, and also provides a characterization of it in terms of a sequence of recursive fixed point problems. The latter is also shown to enable approximate computation of the invariant state, which is proved to be unique when d=2. Furthermore, under an additional finite moment condition on the service distribution and an *a priori* exponential tail decay assumption on the invariant state, it is shown that any invariant state of the hydrodynamic equations of a sub-critical network with finite mean queue length exhibits double exponential tail decay. This result takes an important step towards identifying the class of service distributions for which the power of two choices holds, beyond the case of service distributions with decreasing hazard rate considered by Bramson, Lu and Prabhakar. Numerical evidence is provided to support the analytical results.

The second model is a many-server queue with reneging often denoted as the G/GI/N+GI queue, in which there are N servers and a common queue to which jobs arrive with i.i.d. service and patience times and wait to be processed till a server becomes free. Jobs depart a system either on completion of service or by reneging from the queue when the time waiting in the queue exceeds the patience time. Under suitable assumptions, a functional central limit theorem for the state dynamics is established in the so-called Halfin-Whitt asymptotic regime, where the number of servers N goes to infinity, and the traffic intensity scales as $\lambda N - \beta\sqrt{N}$ for some constant $\beta \in (-\infty, \infty)$ and positive constant λ. The limit is characterized as the unique solution to a pair of coupled stochastic partial differential equations with a nonlinear boundary condition having a non-standard form. It is also shown that the limit process describing the total number in system is an Itô diffusion. A suitable function-valued representation of the state process plays an important role in the analysis for this model. The function-valued representation can be viewed as a more tractable reduced version of the measure-valued representation that still retain relevant information about the model, including limiting fluctuations of the queue length and virtual waiting time. In particular, our work allows the study of the overloaded or supercritical regime, where the asymptotic mean arrival rate exceeds the mean service time, which is of particular interest in applications with reneging. The work here extends previous work on both many-server queues without reneging as well as queues with reneging, which have been restricted to fluctuations of the queue length in the critical regime, where the asymptotic mean arrival rate equals the mean service time.

Contents

List of Tables

List of Figures

Chapter One

Introduction

1.1 Motivation and Context

Stochastic networks arise in a variety of real world applications ranging from telecommunications, service systems such as call centers, computer networks, health care services and biological systems. It is usually infeasible to perform an exact analysis for these networks. A useful alternative is instead to identify a suitable approximation that can be shown to be accurate in a relevant asymptotic regime. This can be carried out at various levels of accuracy, such as for example, fluid limits that approximate the mean behavior of the network, or diffusion limits that capture fluctuations of the state of the network around its mean behavior.

Networks in which there are many servers at a node arise in many applications, including cloud computing, server farms and large-scaled distributed storage systems, and tend to be harder to analyze than networks with single servers. While much analysis has focused on the case when the service distribution is exponential, statistical analysis shows that the service times of jobs typically follow more general distributions [13, 15, 19, 26, 29]. In the latter case, the analysis becomes even more challenging. This becomes clear when looking at Markovian state descriptors for service dynamics (which facilitates analysis as it allows one to use tools from Markov process theory). For example, in a many-server network with reneging, where the service and patience time distributions are exponential, the state space is shown to be finite-dimensional for both the original network state process and its scaling limit. However for a many-server network with general service distributions, the Markovian state descriptor typically grows with the size of the network. Thus, a common Markovian state description that is applicable for a network of any size must necessarily be infinite-dimensional. In the many server parallel network, where each server has its own queue, has more complex dynamics, and represented by a countable number of interacting processes. In this network, the state descriptor has a countably infinite number of interacting components. Each component here takes values in a compact interval when the service distribution is exponential and is an infinite-dimensional measure when the service times are generally distributed. Several recent works [4, 22, 23, 24, 38, 42] have used a measure-valued state descriptor to establish various limit theorems.

There are two measure-valued representations used to represent the state of the networks – age-based and residual time-based. In most of the literature, age-based representations are used in networks with head-of-the-line processing, where jobs that are waiting are served in the order in which they arrived to the server [4, 22, 23, 24]. On the other hand, residual time-based representation have been used for networks with non-head-of-the-line service like last come first served, random selection for service (RSS) [48], processor sharing (PS) [38] and Earliest-Deadline-First (EDF) [27].

This thesis focuses on studying the equilibrium properties of the infinite-dimensional scaling limit of many-server stochastic networks with first-come-first-serve (FCFS) processing using an age-based state descriptor. Specifically, two different kinds of large-scale stochastic networks in the presence of general distributions are studied: a many-server queueing model with reneging and a randomized load balancing model.

The first part of the thesis studies the equilibrium properties of the hydrodynamic limit of a randomized load balancing model with general service distributions, which was established as an infinite sequence of coupled measure-valued processes in [4]. The second part establishes the diffusion approximation for the many-server queueing model with reneging with general service and patience times as the unique solution of a coupled system of stochastic partial differential equations.

1.2 Hydrodynamic limits of a Randomized Load Balancing Network

1.2.1 Description of the Model

Randomized load balancing is an effective method that is used to improve performance in large scale networks while incurring relatively low communication overhead and low computation cost. Applications of large-scale load balancing networks have increased tremendously due to the growth in the use of server farms and computer clusters. Some examples include high-traffic web sites, high-bandwidth File Transfer Protocol sites, Domain Name System (DNS) servers, cloud computing and communication systems.

The model considered here consists of a parallel network of N servers, to which jobs with independent and identically distributed (i.i.d.) service times arrive according to a renewal process with rate λN. Upon arrival of a job, d out of the N servers are chosen independently and uniformly at random, and the job is routed to the server with the shortest queue, with ties broken uniformly at random. Each server processes jobs from its queue in a first-come first-serve (FCFS) manner and jobs leave the system on completion of service. A server never idles when there is a job in its queue. The arrival process and service times are assumed to be mutually independent, and service times of jobs have finite mean which, without loss of generality, is taken to be one. The system described above will be referred to as the

$SQ(d)$ model.

This model was first introduced and analyzed in the case of Poisson arrivals and exponential service distributions by Vvedenskaya et al. [43] and Mitzenmacher [32]. They introduced a countable state representation of the process consisting of the fraction of queues with length greater than ℓ for each integer $\ell \geq 0$ and showed that as $N \to \infty$, the state converges to a fluid limit, characterized as the unique solution to a countable system of coupled ordinary differential equations (ODEs). Under the stability condition $\lambda < 1$, for $d \geq 2$ they also obtained an analytical expression for the the unique invariant state of the fluid limit, and showed that it exhibits a doubly exponential tail decay. It was also shown by Vvedenskaya et al. [43] that the invariant state of the fluid limit is the limiting stationary distribution of the N-server system. On the other hand, when $d = 1$, the stationary queue length distribution is the same as that for an M/M/1 queue, which is well known to exhibit just an exponential tail decay. Thus, the works [43, 32] uncovered the remarkable property that a dramatic improvement in performance can be achieved by introducing just a little bit of randomness into the system, that is, even when $d = 2$, a phenomenon that has been dubbed "the power of two choices."

However, many realistic service distributions are neither exponential nor phase-type, and may not have a decreasing hazard rate. For example, statistical analyses suggest that service times follow a log-normal distribution in [13], a Gamma distribution in Automatic Teller Machines [26], or a shifted exponential distribution in [15, 29]. This motivates taking a different approach to analyzing randomized load balancing with general service distributions than that adopted in [8, 9, 10]. To do this, we build on the work of [4] that considered a very general class of service distributions and used a PDE method for analyzing randomized load balancing networks which generalizes the more classical ODE method that is valid only in the presence of exponential service distributions and can be extended to phase-type distributions. They formulated a measure-valued state representation (which generalizes the representation introduced for GI/GI/N queues in [24]) and established convergence to a hydrodynamic or fluid limit, and the works [4, 2, 1] which (under additional conditions on the service distributions) provide a reformulation of the hydrodynamic equations in terms of a countable system of coupled partial differential equations. While these papers focused on transient behavior of this network, this work characterizes the invariant state of the hydrodynamic equations derived in [4].

1.2.2 Prior Work

Until recently, most work on this model and its variants had focused on exponential service times. An extension to phase-type service distributions was first considered by Li and Lui [28], who analyzed the invariant states of a formal fluid limit, without proving convergence to the fluid limit. More general service distributions were then considered in a series of works by Bramson, Lu and Prabhakar [7, 8, 9, 10]. Under very broad assumptions and more general routing schemes, it was first shown in [7] that under the subcriticality condition $\lambda < 1$, each N-server system is ergodic with a unique stationary distribution. The works [8, 9, 10] specialize to the SQ(d) model with Poisson arrivals and service distributions with decreasing hazard rate, and directly establish convergence of steady-state distributions without first establishing a fluid limit. In particular, when $\lambda < 1$, it is shown in [9, Theorem 2.2] that the N-server stationary queue length distribution converges to the unique solution of a fixed point equation, using the cavity method from statistical physics. This equation is then analyzed to show that for power law distributions with exponent $-\alpha$, the limiting stationary distribution has a doubly exponential tail if $\alpha > d/(d-1)$, exponential tail if $\alpha = d/(d-1)$ and a power law tail if $\alpha < d/(d-1)$. But, according to the authors of [9], it would be challenging to extend their approach to establish convergence for more general service distributions (see [9, Paragraph 9, Section 1]).

A recent area of work has been to analyze the SQ(d) model where $d = d(N)$ is a function of the number of servers in the system. For systems with Poisson arrivals and exponential service times, [34, 33] proved that as $d(N) \to \infty$ the fluid limit coincides with that of the join the shortest queue policy and is independent of the growth rate of $d(N)$. For a specific function $d(N)$, Brightwell, Fairthorne and Luczak [12] showed that when the arrival rate $\lambda(N) = 1 - N^{-\alpha}$ and $d(N) = \lfloor N\beta \rfloor$, where $\alpha, \beta \in (0,1]$ with high probability as $N \to \infty$, in equilibrium, the proportion of queues with length equal to $k = \lceil \alpha/\beta \rceil$ is at least $1 - 2N - \alpha + (k-1)\beta$, and there are no longer queues.

Another area of exploration has been to analyze the SQ(d) model with Poisson arrivals where each of the N serves follow a processor sharing service discipline. In the case of generally distributed service times, Vasantam, Mukhopadhyay and Mazumdar [42] obtained a fluid limit and showed that in the limit the individual servers are statistically independent of others (propagation of chaos) and the equilibrium point of the limit is insensitive to the queue length distributions. Mukhopadhyay, Karthik and Mazumdar [35] studied a similar multi-class system with exponential service distribution, where the N servers are categorized into $M(\ll N)$ different types according to their processing capacity. In this case as well, the authors show that asymptotic independence among servers holds when M is finite and exchangeability holds only within servers of the same class. They also establish the existence and uniqueness of stationary solution of the fluid limit and show that the tail distribution of server occupancy decays doubly exponentially for each server class.

1.2.3 Our Contribution

Chapter 2 of the thesis focuses on the characterization and analysis of the hydrodynamic limit of randomized load balancing networks obtained in [4] under general service distribution. Here is a summary of the contributions.

1. Characterization of the invariant state. Under the same mild conditions required for the hydrodynamic limit in [4], we characterize the queue length distribution of the invariant state of the hydrodynamic equations as a fixed point of a certain map, and show that the fixed point is unique when $d = 2$. The methodology, which entails analysis of a system of measure-valued equations, can potentially be applied to other many-server systems with general service distributions for which measure-valued representations are useful. We show that our characterization is sufficiently explicit and computationally tractable that it yields a numerical algorithm for computing important performance measures such as finite queue length exceedance probabilities and the mean virtual waiting time in the invariant state.

2. Tail decay of the invariant state. Under an additional assumption on the tail of the service distribution and an *a priori* assumption on the tail , we show that the invariant queue length distribution exhibits a doubly exponential decay rate. Our numerics show that while the decay of the invariant queue distribution is of interest, it may not manifest itself until far into the tail, and thus it is important to be able to compute and characterize invariant finite queue length exceedance probabilities as well, which is feasible with our characterization of the invariant states.

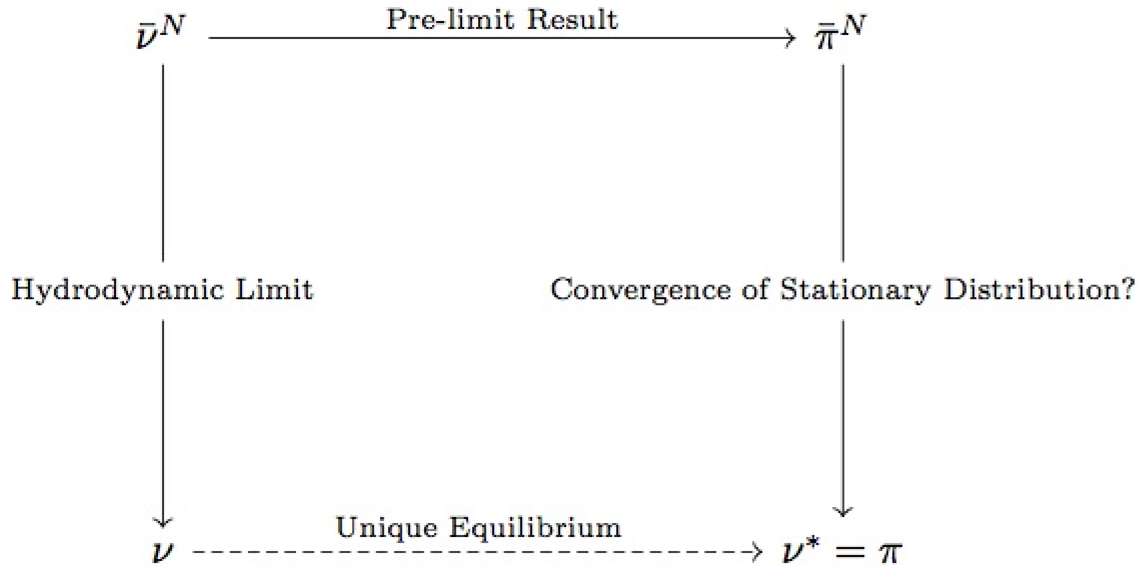


Figure 1.1: Relation between the N-server state representation $\overline{\nu}^N$, N-server stationary distribution $\overline{\pi}^N$, the hydrodynamic limit ν, invariant state of the hydrodynamic limit ν^* and the limiting stationary distribution π for the randomized load balancing network.

Our work also takes a small step towards understanding equilibrium behavior in large systems. As in the exponential case [43], we conjecture that the invariant state is the limit of the stationary distribution of the N-server network. This would involve proving the interchange of the limits in Figure 1.1. This is also motivated by the result obtained by [9], where they prove that the interchange holds when the arrival process has $\lambda < \lambda^*$, where $\lambda^* < 1/2$ is sufficiently small and the service distribution has finite second moment and a decreasing hazard rate. We provide numerical evidence to support this conjecture, however, a rigorous proof of this convergence is significantly more complicated due to it being a countable system of measure-valued equations, rather than just a system of ODEs, and is thus relegated to future work.

Another interesting area of exploration would be to extend the phase transition result shown in [10] beyond distributions with decreasing hazard rate to a more general class of service distributions.

1.3 Many-Server Queues with Reneging

1.3.1 Description of the Model

Over the last couple of decades, several applications have spurred the study of many-server models with reneging [5, 6, 19]. Specifically, in applications to telephone contact centers and (more generally) customer contact centers, the effect of customers' impatience has been shown to have a substantial impact on the performance of the system [19].

We consider a many-server queueing system in which reneging is modeled by assigning each customer a patience time. A customer in queue reneges the system without service when its waiting time (i.e., the time spent in queue since arrival) exceeds the patience time. The customers waiting in the queue enter service in an FCFS manner. In this model, customers arrive according to a renewal process with i.i.d. service requirements chosen from a general distribution with finite mean and also carry i.i.d. patience times that are drawn from another general distribution which also has a finite mean. When there are N servers and the cumulative customer arrival process is assumed to be a renewal process, this reduces to the so-called G/GI/N + GI model.

The exact analysis of a many-server queueing model with reneging when the interarrival, service and patience times are exponentially distributed was performed by Garnett, Mandelbaum and Reiman in [20]. They provided exact calculations of various steady state performance measures and their approximations in the Halfin-Whitt regime (where the number of servers N goes to infinity and the corresponding arrival rate grows as $N - \beta\sqrt{N}$ for some $\beta > 0$), both in the case of finite waiting rooms (M/M/N/B + M) and infinite waiting rooms (M/M/N + M). In the case of Poisson arrivals and exponential service but generally distributed service time (M/M/N + G), Baccelli and Hebuterne [5] obtained explicit formulae for the steady state distributions of the queue length and virtual waiting time. Several other steady state performance measures and their approximations in the Halfin-Whitt regime for this case were derived by Mandelbaum and Zetlyn [47].

1.3.2 Prior Work

The statistical analysis of call centers in Brown et al [13], shows that in many cases the service time distribution appears to follow a log-normal distribution. The patience time distribution has also been observed to not be exponential in Mandelbaum and Zetlyn [47]. When the service and patience times are generally distributed, no analytical or numerical methods are available to perform an exact analysis. A useful alternative to understand the exact system performance of the many-server queueing model with reneging is to study more tractable diffusion approximations.

For the case of exponential interarrival, service and patience time distributions, a process-level diffusion approximation was obtained by Mandelbaum, Massey and Reiman [30]. Whitt [46] subsequently generalized the result to G/M/N + M queues. Dai, He and Tezcan [16] studied G/Ph/N + GI queues using a modification of the approach introduced by Puhalskii and Reiman [39] to study diffusion approximations for the corresponding model without reneging. Reed and Tezcan [40] also proved a diffusion limit for GI/M/N + GI queues by scaling the patience time hazard rate function to obtain a refined limit process. For a G/G/N + GI queue, the diffusion approximation for the queue length and virtual waiting time have been established by Mandelbaum and Momclivic [31] in the critical regime.

A Markovian state representation for the G/GI/N + GI queues was introduced by Kang and Ramanan [22] in terms of the triple $(X^{(N)}, \nu^{(N)}, \eta^{(N)})$. The measure-valued process $\nu^{(N)}$ encodes the amount of time that customers currently receiving service have been in service. The second component $\eta^{(N)}$ is a measure-valued process that keeps track of the waiting times of customers in queue as well as potential waiting times, defined as time since entry into the system of all customers (including those who may have already entered service and possibly departed the system). The process $X^{(N)}$ denotes the total number of customers in the system. Under mild assumptions on the service and patience time distributions, Kang and Ramanan [22] established a fluid limit, characterized as the limit is the unique solution of a coupled system of deterministic measure-valued integral equations that admits an explicit representation.

Using the same state representation, Kang and Ramanan [23], proved that, under general assumptions, the fluid-scaled state process is Feller, admits a stationary distribution and is ergodic. This helps deduce important performance measures such as the equilibrium probability of a positive wait, or more generally, the equilibrium waiting time distribution. They also showed that the associated sequence of fluid-scaled stationary distributions is tight. This work also shows that (for a large class of service distributions) the sequence of fluid-scaled stationary distribution converges to the invariant state for the fluid limit when the latter is assumed to be unique. However, the fluid limit is deterministic and not refined enough to provide more accurate approximation to quantities of interest such as the probability of a positive wait. This motivates looking at stochastic fluctuations around the fluid limit and characterize the asymptotics of these fluctuations via a diffusion limit theorem.

Although this representation proved useful to study the fluid approximation, we conjecture that an associated distribution-valued diffusion-scaled limit will not be Markov on its own unless stringent assumptions are imposed on the service and patience time distribution. This conjecture is supported by the non-Markov nature of the diffusion limit in the absence of reneging as shown by Kaspi and Ramanan in [25]. In the absence of reneging, a new representation has been used by Aghajani and Ramanan [3] to obtain a Markovian diffusion limit and have also established uniqueness of the invariant distribution for the limit.

1.3.3 Our Contribution

In Chapter 3, this thesis introduces an infinite - dimensional representation for the state of the many-server queueing model with reneging and an asymptotic limit theorem is established for the fluctuation of the state process. The salient contributions are listed below.

1. A new representation. Inspired by the measure-valued representation in [22] and the infinite-dimensional representation introduced by Aghajani and Ramanan in [3], we introduce a new representation for the G/GI/N + GI queueing system. Specifically, we consider a representation $(X^{(N)}, Z_\nu^{(N)}, Z_\eta^{(N)})$ where $X^{(N)}$ is the number of customers in the N-server network, $Z_\nu^{(N)}$ and $Z_\eta^{(N)}$ are certain function-valued processes derived from the measure-valued state variables $\nu^{(N)}$ and $\eta^{(N)}$. Here $Z_\nu^{(N)}(t, r)$ represent the conditional expected number of customers that were in service at time t and have not reneged the system

by time $t+r$, given the ages of customers in service up to time t. Similarly, $Z_\eta^{(N)}(t,r)$ represent the conditional expected number of customers with potential waiting time greater than t and have not left the system by time $t+r$, given the potential waiting times of customers potentially in queue up to time t

2. SPDE characterization of the limit process. We identify the diffusion limit process of the scaled and renormalized sequence $(\widehat{X}^{(N)}, \widehat{Z}_\nu^{(N)}, \widehat{Z}_\eta^{(N)})$. Under suitable conditions on the service and patience time distributions we characterize the components $\widehat{X}$, $\widehat{Z}_\nu$ and $\widehat{Z}_\eta$, of the limit process to be the unique solution to a coupled system of stochastic equations driven by mutually independent Brownian motion and two space-time white noises. Specifically, we show that $\widehat{X}$ is an Ornstein-Uhlenbeck process with a constant diffusion coefficient, and a drift that also depends on the processes $\widehat{Z}_\nu$ and $\widehat{Z}_\eta$. The processes $\widehat{Z}_\nu$ and $\widehat{Z}_\eta$ are $\mathbb{H}^1(0,\infty)$-valued processes that satisfy a coupled system of (non-standard) stochastic partial differential equations (SPDEs) on the domain $(0,\infty)$. The components $\widehat{Z}_\nu$ and $\widehat{X}$ are further linked by a nonlinear global boundary condition: at each time $t \geq 0$, $\widehat{Z}_\nu(t,0)$ is a deterministic nonlinear function of $\widehat{X}$.

As mentioned above, it is of interest to obtain refined approximation to the N-server stationary distribution for large N. The above result takes a first step towards that goal. Indeed, we believe that the diffusion limit is an infinite-dimensional Markov process taking values in a suitable Hilbert space, with a unique stationary distribution.

Chapter Two

Invariant states of a Randomized Load-Balancing Network

2.1 Model Description

Randomized load balancing is an effective method that is used to improve performance in large scale networks while incurring relatively low communication overhead and computation costs. The model considered here consists of a parallel network of N servers, to which jobs with independent and identically distributed (i.i.d.) service times arrive according to a renewal process with rate λN. Upon arrival of a job, d out of the N servers are chosen independently and uniformly at random, and the job is routed to the server with the shortest queue, with ties broken uniformly at random. Each server processes jobs from its queue in a first-come first-serve (FCFS) manner and jobs leave the system on completion of service. A server never idles when there is a job in its queue. The arrival process and service times are assumed to be mutually independent, and service times of jobs have finite mean which, without loss of generality, is taken to be one. The system described above will be referred to as the $SQ(d)$ model.

This model was first introduced and analyzed in the case of Poisson arrivals and exponential service distributions by Vvedenskaya et al. [43] and Mitzenmacher [32]. They introduced a countable state representation of the process consisting of the fraction of queues with length greater than ℓ for each integer $\ell \geq 0$ and showed that as $N \to \infty$, the state converges to a fluid limit, characterized as the unique solution to a countable system of coupled ordinary differential equations (ODEs). Under the stability condition $\lambda < 1$, for $d \geq 2$ they also obtained an analytical expression for the the unique invariant state of the fluid limit, and showed that it exhibits a doubly exponential decay. It was also shown by Vvedenskaya et al. [43] that the stationary distributions of the N-server systems converge to this invariant state, in the limit as $N \to \infty$. On the other hand, when $d = 1$, the stationary queue length distribution is the same as that for an M/M/1 queue, which is well known to exhibit just an exponential tail decay. Thus, the works by Vvedenskaya et al. [43] and Mitzenmacher [32] uncovered the remarkable property that a dramatic improvement in performance can be achieved by introducing just a little bit of randomness into the system (i.e., even when $d = 2$), a phenomenon that has been dubbed “the power of two choices.”

Until recently, most work on this model and its variants had focused on exponential service times. An extension to phase-type service distributions was first considered by Li and Lui [28], who analyzed the invariant states of a formal fluid limit. More general service distributions were then considered in a series of works by Bramson et al. [7, 8, 9, 10]. Under very broad assumptions and more general routing schemes, it was first shown by Bramson et al. [7] that under the subcriticality condition $\lambda < 1$, each N-server system is ergodic with a unique stationary distribution. The works [8, 9, 10] consider the particular case of the SQ(d) model with Poisson arrivals and service distributions with decreasing hazard rate, and directly establish convergence of the N-server steady-state distributions without first establishing a fluid limit. In particular, when $\lambda < 1$, it is shown by Bramson et al. [9, Theorem 2.2] that the N-server stationary queue length distribution converges to the unique solution of a fixed point equation, and this equation is analyzed to get insight into the tails of the distribution. However, many realistic service distributions are neither exponential nor phase-type, and may not have a decreasing hazard rate. For example, statistical analyses suggest that service times follow a log-normal distribution in Brown et al. [13], a Gamma distribution in Automatic Teller Machines in Kolesar [26], or a shifted exponential distribution in Chen et al. [15] and Liang and Kozat [29]. However, according to the authors (see Bramson et al. [9, Paragraph 9, Section 1]), this motivates taking a new approach. We will leverage the measure-valued representation for the SQ(d) model with general service distribution introduced in [4]. Specifically, we will consider the hydrodynamic limit and study its invariant state. As mentioned in the introduction, this takes a step towards understanding the equilibrium via a different approach than that is considered in [8, 9, 10].

The rest of this chapter is organized as follows. Section 2.2 introduces the basic assumptions and the equations that characterize the hydrodynamic limit of the model. Section 2.3 states the main results. Section 2.4 presents numerical approximations of the invariant state performance measures for various distributions, and also provides evidence to support convergence of equilibrium measures to the invariant state. The proofs of Theorems 2.3.3 and 2.3.9 are given in Sections 2.5 and Section 2.6, respectively.

Throughout, we will use the following notation. Let $\mathbb{N} = \{1, 2, 3, \cdots\}$ denote the set of natural numbers and $\mathbb{Z}$ denote the set of integers. For every $E \subset \mathbb{R}$, let $\mathbb{C}_b(E)$ denote the space of continuous bounded functions on E, and $\mathcal{B}(E)$ denote the Borel σ-algebra on E. Also, for a metric space $\mathbb{X}$, let $\mathbb{D}_{\mathbb{X}}[0, \infty)$ denote the set of $\mathbb{X}$-valued functions on $[0, \infty)$ that are right continuous and have finite left limits on $(0, \infty)$, and let $\mathbb{C}_{\mathbb{X}}[0, \infty)$ be the subset of $\mathbb{X}$-valued continuous functions on $[0, \infty)$. Let $\mathbb{M}_{\leq 1}[0, a)$ denote the space of sub-probability measures on $[0, a)$ for some $a \in (0, \infty]$. Given an interval $A \subset \mathbb{R}$, let $\mathcal{B}(A)$ denote the space of Borel subsets of A. For any $V \subset \mathcal{B}[0, \infty)$, a measure μ on $[0, \infty)$ and an integrable function $f : V \mapsto \mathbb{R}$, we use the notation

$$\langle f, \mu \rangle_V := \int_V f(x)\mu(dx),$$

and omit the subscripts V when $V = [0, \infty)$. Also, let $a \wedge b$ represent $\min(a, b)$ for $a, b \in \mathbb{R}$.

2.2 Hydrodynamic Equations

In this section, we introduce the equations that characterize the hydrodynamic limit of our load balancing model, as established in [4]. Throughout, we make the following assumptions on the service distribution.

Assumption 2.2.1. The service distribution whose cumulative distribution function (cdf) is denoted by G, has density g and finite mean, which can (and will) be set to 1.

Define $\bar{G}(x) := 1-G(x)$ and let $L := \sup\{x \in [0,\infty) : \bar{G}(x) > 0\}$ be the right-end of the support of the distribution. Also, let $h : [0, L) \mapsto [0,\infty)$ denote the hazard rate function:

$$h(x) = \frac{g(x)}{\bar{G}(x)}, \quad x \in [0, L).$$

In [4] the state of an N-server randomized load balancing system is represented by the Markov process $\nu^{(N)}(t) = \{\nu_\ell^{(N)}(t)\}_{\ell\in\mathbb{N}}$, where for each ℓ, $\nu_\ell^{(N)}(t)$ is a (random) finite measure on $[0,\infty)$ that has a unit delta mass at the age (that is, the time spent in service by time t) of each job that, at time t, is in service at a queue of length no less than ℓ. For every $t \geq 0$, the scaled state $\nu^{(N)}(t)/N$ takes values in $\mathbb{S}$, where

$$\mathbb{S} := \left\{(\mu_\ell; \ell \in \mathbb{N}) \in \mathbb{M}_{\leq 1}[0, L)^{\mathbb{N}} : \langle f, \mu_\ell\rangle \geq \langle f, \mu_{\ell+1}\rangle, \forall \ell \in \mathbb{N}, f \in \mathbb{C}_b[0,\infty), f \geq 0\right\}, \tag{2.2.1}$$

where recall $\mathbb{M}_{\leq 1}[0, L)$ is the space of sub-probability measures on $[0, L)$.

For every $d \geq 2$, define

$$\mathfrak{P}_d(x,y) := \frac{x^d - y^d}{x-y} = \sum_{m=0}^{d-1} x^m \, y^{d-1-m}. \tag{2.2.2}$$

Note that $\mathfrak{P}_2(x,y) = x + y$, and for $0 \leq y \leq x$,

$$\mathfrak{P}_d(x,y) \leq dx^{d-1}. \tag{2.2.3}$$

We recall the definition of the hydrodynamic equations given in [4, Section 2.3].

Definition 2.2.2. (Hydrodynamic Equations) Given $\lambda > 0$ and $\nu(0) \in \mathbb{S}$, $\{\nu(t) = \nu_\ell(t); \ell \in \mathbb{N}, t \geq 0\}$ in $\mathbb{C}_{\mathbb{S}}[0,\infty)$ is said to solve the *hydrodynamic equations* associated with $(\lambda, \nu(0))$ if and only if for every $t \in [0,\infty)$,

$$\int_0^t \langle h, \nu_1(s)\rangle ds < \infty, \tag{2.2.4}$$

and the following equations are satisfied:

$$\langle \mathbf{1}, \nu_\ell(t)\rangle - \langle \mathbf{1}, \nu_\ell(0)\rangle = D_{\ell+1}(t) + \int_0^t \langle \mathbf{1}, \eta_\ell(s)\rangle ds - D_\ell(t), \qquad \ell \in \mathbb{N}, \tag{2.2.5}$$

and, for every $f \in \mathbb{C}_b[0,\infty)$,

$$\begin{aligned}\langle f, \nu_\ell(t)\rangle = & \left\langle f(\cdot + t)\frac{\bar{G}(\cdot + t)}{\bar{G}(\cdot)}, \nu_\ell(0)\right\rangle + \int_{[0,t]} f(t-s)\bar{G}(t-s) dD_{\ell+1}(s) \\ & + \int_0^t \left\langle f(\cdot + t - s)\frac{\bar{G}(\cdot + t - s)}{\bar{G}(\cdot)}, \eta_\ell(s)\right\rangle ds, \qquad \ell \in \mathbb{N}\end{aligned} \tag{2.2.6}$$

where

$$D_\ell(t) := \int_0^t \langle h, \nu_\ell(s)\rangle ds, \qquad \ell \in \mathbb{N}, \tag{2.2.7}$$

and

$$\eta_\ell(t) := \begin{cases} \lambda\left(1 - \langle \mathbf{1}, \nu_1(t)\rangle^d\right)\delta_0 & \text{if } \ell = 1, \\ \lambda \mathfrak{P}_d\Big(\langle \mathbf{1}, \nu_{\ell-1}(t)\rangle, \langle \mathbf{1}, \nu_\ell(t)\rangle\Big)(\nu_{\ell-1}(t) - \nu_\ell(t)) & \text{if } \ell \geq 2. \end{cases} \tag{2.2.8}$$

Remark 2.2.3. The bound (2.2.4) implies that for every $\ell \in \mathbb{N}$, the process D_ℓ in (2.2.7) is well defined.

We now briefly provide some intuition behind the form of the hydrodynamic equations. The term $\langle h, \nu_\ell(s)\rangle$ represents the limiting mean conditional scaled departure rate from queues of length at least ℓ at time s, and thus $D_\ell(t)$ given by (2.2.7) represents the limiting cumulative scaled departure rate from queues of length at least ℓ at time t. The term $\langle \mathbf{1}, \nu_\ell(t)\rangle$ represents the limiting fraction of queues of length at least ℓ at time t. Note that the scaled arrival rate of jobs to the network is given by λ, and for $\ell \geq 1$, the probability that an arriving job is routed to a queue of length $\ell - 1$ at time s is equal to $\mathfrak{P}_d(\langle \mathbf{1}, \nu_{\ell-1}(s)\rangle, \langle \mathbf{1}, \nu_\ell(s)\rangle)(\langle \mathbf{1}, \nu_{\ell-1}(s)\rangle - \langle \mathbf{1}, \nu_\ell(s)\rangle)$, with the convention $\langle \mathbf{1}, \nu_0(s)\rangle = 1$. Thus, $\langle \mathbf{1}, \eta_\ell(s)\rangle$, where η_ℓ is defined by (2.2.8), represents the scaled arrival rate at time s of jobs to queues of length $\ell - 1$. Hence, $\int_0^t \langle \mathbf{1}, \eta_\ell(s)\rangle ds$ is the total increase in the fraction of queues of length greater than or equal to ℓ due to arrivals in the interval $[0, t]$. On the other hand, $D_\ell(t) - D_{\ell+1}(t)$ represents the cumulative decrease in the fraction of queues of length at least ℓ due to service completions. The mass balance equation, (2.2.5) is a result of these observations. Lastly, the right-hand side of equation (2.2.6) describes the three terms that contribute to the measure $\nu_\ell(t)$. The first term accounts for jobs that were already in service at time 0. The second term represents the contribution due to departures from queues of length greater than or equal to $\ell + 1$ in the interval $[0, t]$. The third term represents the contribution to $\nu_\ell(t)$ due to jobs that entered service at a queue of length at least ℓ at some time $s \in (0, t]$ and are still in service at time t.

Along with Assumption 2.2.1, we also make the following assumption throughout.

Assumption 2.2.4. There exists a unique solution to the hydrodynamic equations.

Remark 2.2.5. A sufficient condition for Assumption 2.2.4, as established in [4, Theorem 2.6], is that the density g of the service distribution be bounded on every finite interval of $[0, L)$. Examples of distributions satisfying Assumptions 2.2.1 and 2.2.4 are given in Remark 2.1 of [2].

2.3 Main Results

To state our main results, we will require the following basic definition.

Definition 2.3.1 (Invariant State). Given $\lambda > 0$, $\nu^* = (\nu^*_\ell)_{\ell \in \mathbb{N}} \in \mathbb{S}$ is said to be an invariant state of the hydrodynamic equations with arrival rate λ if for every $t \geq 0$, $\nu \in \mathbb{C}_{\mathbb{S}}[0, \infty)$ defined by

$$\nu(t) = \nu^* \qquad \forall t \geq 0.$$

solves the hydrodynamic equations associated with (λ, ν^*). In this case the associated sequence $(s^*_\ell)_{\ell \in \mathbb{N}}$, defined by

$$s^*_\ell := \langle \mathbf{1}, \nu^*_\ell \rangle, \quad \ell \in \mathbb{N}, \tag{2.3.1}$$

is said to be an invariant queue length distribution (with arrival rate λ). Furthermore, any such $(s^*_\ell)_{\ell \in \mathbb{N}}$ that satisfies

$$\lim_{\ell \to \infty} s^*_\ell = 0, \tag{2.3.2}$$

will be referred to as a physical invariant queue length distribution (with arrival rate λ).

Remark 2.3.2. Note that (2.3.2) is a necessary condition for the mean of the invariant queue length distribution to be finite. Moreover, as shown in Lemma 2.5.6, the condition (2.3.2) turns out to be equivalent to requiring $s_1^* = \lambda$, which will be satisfied by any "physically relevant" invariant state, including one that arises as the limit of stationary N-server systems, which explains the nomenclature used in Definition 2.3.1.

In the sequel, we will refer to a physical invariant state as any invariant state that additionally satisfies (2.3.2). The latter condition is necessary for uniqueness even in the case of an exponential service distribution since, for example, as it is easy to verify, for any service distribution G, $\nu_\ell^*(dx) = \bar{G}(x)dx$ (and, consequently, $s_\ell^* = 1$) for every $\ell \in \mathbb{N}$, is always an invariant state of the hydrodynamic equations.

We now state our first main result on existence and characterization of physical invariant states. Its proof is given in Section 2.5.4.

Theorem 2.3.3. *Suppose Assumptions 2.2.1 and 2.2.4 hold. Also, fix $\lambda \in (0,1)$ and $d \in \mathbb{N}$. Then there exists a physical invariant state of the hydrodynamic equations with arrival rate λ if and only if there exist a sequence of measurable functions $(r_\ell)_{\ell\in\mathbb{N}}$ on $[0,\infty)$ that are continuously differentiable on $[0,L)$ such that for $x \notin [0,L)$, $r_\ell(x) = 0$ and for $x \in [0,L)$, $r_1(x) = \lambda$ and for $\ell \geq 2$,*

$$r_\ell(x) = \lambda(s_\ell^*)^d e^{-\lambda\mathfrak{P}_d(s_{\ell-1}^*, s_\ell^*)x} + \lambda\mathfrak{P}_d(s_{\ell-1}^*, s_\ell^*) \int_0^x e^{-\lambda\mathfrak{P}_d(s_{\ell-1}^*, s_\ell^*)(x-u)} r_{\ell-1}(u)du, \quad (2.3.3)$$

where for $\ell \geq 1$,

$$s_\ell^* = \int_0^L r_\ell(x)\bar{G}(x)dx. \quad (2.3.4)$$

In this case,

$$\nu_\ell^*(A) = \int_A r_\ell(x)\bar{G}(x)dx, \qquad A \in \mathcal{B}([0,L)), \quad \ell \in \mathbb{N}. \tag{2.3.5}$$

Furthermore, a physical invariant state of the hydrodynamic equations with arrival rate λ always exists, and also satisfies

$$\lambda(s_{\ell-1}^*)^d = \langle h, \nu_\ell^* \rangle = \int_0^L g(x) r_\ell(x) dx, \qquad \ell \geq 2. \tag{2.3.6}$$

Remark 2.3.4. Now, fix $\lambda \in (0,1)$ and for $d \geq 2$, define $r_1 \equiv s_1^* = \lambda$. Note that $(s_\ell^*)_{\ell \in \mathbb{N}}$ is a physical invariant queue length distribution if and only if for $\ell \geq 2$, s_ℓ^* is a fixed point of the map

$$\begin{aligned} F_\ell(s) :=& \lambda s^d \int_0^L e^{-\lambda \mathfrak{P}_d(s_{\ell-1}^*, s)x} \bar{G}(x) dx \\ &+ \lambda \mathfrak{P}_d(s_{\ell-1}^*, s) \int_0^L \left(\int_0^x e^{-\lambda \mathfrak{P}_d(s_{\ell-1}^*, s)(x-u)} r_{\ell-1}(u) du \right) \bar{G}(x) dx, \end{aligned} \tag{2.3.7}$$

on $[0, s_{\ell-1}^*]$, where for $\ell \geq 3$, the function $r_{\ell-1}$ is defined recursively via (2.3.3) in terms of the functions $r_{\ell-2}$ and quantities $s_{\ell-1}^*$ and $s_{\ell-2}^*$.

As an immediate consequence of Theorem 2.3.3, we recover the following well-known result for the exponential service distribution (see [43, Theorem 1] and [32, Lemma 2]).

Remark 2.3.5. If $\bar{G}(x) = e^{-x}$, then $g(x) = \bar{G}(x)$ and (2.3.4) and (2.3.6) show that $s_\ell^* = \lambda(s_{\ell-1}^*)^d$ for all $\ell \geq 2$. Hence, solving the recursion with $s_1^* = \lambda$, we obtain

$$s_\ell^* = \lambda^{\frac{d^\ell - 1}{d-1}}, \quad \ell \in \mathbb{N}.$$

Our next result, which concerns the tail behavior of a physical invariant state, requires some additional conditions.

Assumption 2.3.6. Fix $d \geq 2$. Suppose there exist $x_0 > 1/\lambda d$, $C_0 < \infty$, and $\beta > d/(d-1)$ such that

$$\bar{G}(x) \leq C_0 x^{-\beta}, \qquad \forall x \geq x_0. \tag{2.3.8}$$

Remark 2.3.7. This tail condition is almost equivalent to a moment condition. Specifically, a sufficient condition for the tail condition is that service distribution have finite $(\beta+1)+\varepsilon$ moment for some $\varepsilon > 0$, whereas a necessary condition is that the service distribution has a finite $(\beta+1)$ moment. Thus, it is immediate that any distribution with all moments finite, such as the Gamma, shifted exponential and lognormal, as well as the Pareto, the latter with tail parameter greater than $\beta+1$), all satisfy the conditions in Assumption 2.3.6.

Recalling that, $x_0, C_0 \in (0,\infty)$, and β are constants from Assumption 2.3.6, now define

$$C := \frac{\int_0^{x_0} \bar{G}(x)dx}{\int_{x_0}^{\infty} \bar{G}(x)dx}, \qquad C_1 := C_0(C+1). \tag{2.3.9}$$

Define

$$\hat{\beta} := \beta - \lfloor \beta \rfloor, \tag{2.3.10}$$

where $\lfloor \beta \rfloor$ denotes the largest integer less than or equal to β. Also, define $C_\beta := C_\beta(\lambda, \beta, d)$ as follows:

$$C_\beta := \begin{cases} \max\left(1, \frac{C_1(\lfloor \beta \rfloor + 1)}{\min(\hat{\beta}, 1-\hat{\beta})}(\lambda d)^{\beta-1}\right) & \text{if } \beta \notin \mathbb{N}, \\ \max\left(1, C_1(\beta+1)(\lambda d)^{\beta-1}\right) & \text{if } \beta \in \mathbb{N}. \end{cases} \tag{2.3.11}$$

We also assume exponential tail decay for the invariant queue length distribution.

Assumption 2.3.8. Fix $d \geq 2$ and C_β as defined in (2.3.11). Suppose there exist

$k > \frac{\log(3C_\beta)}{(\beta-2)(d-1)}$ and $c_k, L > 0$ such that the queue length distribution $(s^*_\ell)_{\ell\in\mathbb{N}}$ associated with any physical invariant state satisfies

$$s^*_i \le c_k e^{-ki}, \qquad \text{for all } i \ge L. \tag{2.3.12}$$

We prove the following result in Section 2.6.1.

Theorem 2.3.9. *Fix $\lambda \in (0,1)$ and $d \ge 2$ and suppose Assumptions 2.2.1–2.3.8 hold. Then, the queue length distribution $(s^*_\ell)_{\ell\in\mathbb{N}}$ associated with any physical invariant state exhibits a doubly exponential decay, in the sense that there exists a constant $n_d = n_d(\beta) > 0$ such that*

$$\liminf_{\ell\to\infty} \frac{1}{\ell} \log_d \log\left(\frac{1}{s^*_\ell}\right) \ge n_d. \tag{2.3.13}$$

We conjecture that Assumption 2.3.8 holds for the class of service distributions considered, and ongoing work is devoted to the investigation of this conjecture. If the conjecture is true, then the result in Theorem 2.3.9 significantly extends previous results on the power of two choices for specific classes of service distributions such as exponential [32, 43], phase-type [28] and Pareto with parameter $\alpha > 2$ [10]. Besides applying to more general distributions like lognormal and shifted exponential distributions that are relevant in practice, it provides a unified approach for obtaining all these results as special cases.

While the tail decay property is an interesting property, the decay property may not manifest itself until rather far into the tails and so finite (invariant) queue length exceedance probabilities are often of more practical interest. In Section 2.4, we illustrate how for $d = 2$, the characterization of the invariant states of the measure-valued hydrodynamic equations obtained in Theorem 2.3.3 allows us to compute more general invariant quantities of relevance besides the queue length, such as the virtual waiting time, which appears not to have been considered in the literature before (except in the case of exponential service times).

2.4 Numerical Results when $d = 2$

We now present numerical results, for simplicity restricting to the case $d = 2$.

2.4.1 Properties of the physical invariant state

We first obtain an alternative representation for $(r_\ell)_{\ell\in\mathbb{N}}$ that is more amenable to computation than the one given in (2.3.3)-(2.3.4). Recall the definition of a physical invariant state from Remark 2.3.2.

Lemma 2.4.1. *Suppose $\nu^* = (\nu_j^*)_{j\geq 1} \in \mathbb{S}$ is a physical invariant state of the hydrodynamic equations, with an associated sequence of measurable functions $(r_j)_{j\in\mathbb{N}}$ as in Theorem 2.3.3. Then $r_1 \equiv \lambda$ on $[0, L)$ and for $j \geq 2$,*

$$r_j(x) = \lambda + \lambda \sum_{i=2}^{j} c_{i,j} e^{-\lambda \mathfrak{P}_d(s_{i-1}^*, s_i^*)x}, \qquad \forall x \in [0, L), \tag{2.4.1}$$

where $c_{2,2} := (s_2^*)^d - 1$, and for $j \geq 3$, $c_{i,j}$ are recursively defined as follows:

$$c_{i,j} := \begin{cases} c_{i,j-1} \frac{\mathfrak{P}_d(s_{j-1}^*, s_j^*)}{\mathfrak{P}_d(s_{j-1}^*, s_j^*) - \mathfrak{P}_d(s_{i-1}^*, s_i^*)} & \text{if } 2 \leq i < j, \\ (s_j^*)^d - 1 - \sum_{k=2}^{j-1} c_{k,j} & \text{if } i = j. \end{cases} \tag{2.4.2}$$

Proof. The identity $r_1 \equiv \lambda$ on $[0, L)$ follows from Theorem 2.3.3. Substituting $\ell = 2$

in (2.3.3) with $r_1 \equiv \lambda$, we see that

$$\begin{aligned} r_2(x) &= \lambda (s_2^*)^d e^{-\lambda \mathfrak{P}_d(s_1^*, s_2^*) x} + \lambda \mathfrak{P}_d(s_1^*, s_2^*) \int_0^x e^{-\lambda \mathfrak{P}_d(s_1^*, s_2^*)(x-u)} \lambda \, du \\ &= \lambda + \lambda e^{-\lambda \mathfrak{P}_d(s_1^*, s_2^*) x} \left((s_2^*)^d - 1 \right), \end{aligned}$$

which verifies (2.4.1) for $j = 2$. Now, suppose (2.4.1) holds for some $j \geq 2$. Substituting $\ell = j + 1$ in (2.3.3) and using (2.4.1) and (2.4.2), we have

$$\begin{aligned} r_{j+1}(x) &= \lambda (s_{j+1}^*)^d e^{-\lambda \mathfrak{P}_d(s_j^*, s_{j+1}^*) x} \\ &\quad + \lambda^2 \mathfrak{P}_d(s_j^*, s_{j+1}^*) \int_0^x [1 + \sum_{i=2}^{j} c_{i,j} e^{-\lambda \mathfrak{P}_d(s_{i-1}^*, s_i^*) u}] e^{-\lambda \mathfrak{P}_d(s_j^*, s_{j+1}^*)(x-u)} du \\ &= \lambda (s_{j+1}^*)^d e^{-\lambda \mathfrak{P}_d(s_j^*, s_{j+1}^*) x} + \lambda (1 - e^{-\lambda \mathfrak{P}_d(s_j^*, s_{j+1}^*) x}) \\ &\quad + \lambda \sum_{i=2}^{j} c_{i,j} \frac{\mathfrak{P}_d(s_j^*, s_{j+1}^*)}{\mathfrak{P}_d(s_j^*, s_{j+1}^*) - \mathfrak{P}_d(s_{i-1}^*, s_i^*)} \left(e^{-\lambda \mathfrak{P}_d(s_{i-1}^*, s_i^*) x} - e^{-\lambda \mathfrak{P}_d(s_j^*, s_{j+1}^*) x} \right) \\ &= \lambda + \lambda ((s_{j+1}^*)^d - 1) e^{-\lambda \mathfrak{P}_d(s_j^*, s_{j+1}^*) x} + \lambda \sum_{i=2}^{j} c_{i,j+1} \left(e^{-\lambda \mathfrak{P}_d(s_{i-1}^*, s_i^*) x} - e^{-\lambda \mathfrak{P}_d(s_j^*, s_{j+1}^*) x} \right) \\ &= \lambda + \lambda \sum_{i=2}^{j+1} c_{i,j+1} e^{-\lambda \mathfrak{P}_d(s_{i-1}^*, s_i^*) x}, \end{aligned}$$

which shows that (2.4.1) also holds when j is replaced with $j + 1$. Thus, by the principle of mathematical induction, we are done. □

2.4.2 An algorithm for computing the invariant state

The following result is the key result that allows numerical computation of the invariant states.

Proposition 2.4.2. *When $d = 2$ and Assumptions 2.2.1 and 2.2.4 hold, then for each $\lambda \in (0, 1)$ there exists a unique physical invariant queue length distribution*

$(s_\ell^*)_{\ell \in \mathbb{N}}$ for the hydrodynamic equations with arrival rate λ. In particular, $s_1^* = \lambda$ and for each $\ell \geq 2$, s_ℓ^* is unique fixed point in $[0, s_{\ell-1}^*]$ of the function F_ℓ defined in (2.3.7).

The proof will rely on the following elementary lemma.

Lemma 2.4.3. *Let $b < \infty$ and let f be a function on $[0, b]$ such that f is continuously differentiable on $(0, b)$, $f(x) \neq 0$ for $x \in \{0, b\}$, has at least one root in $(0, b)$ and the derivative of f at every root in $(0, b)$ is negative, i.e.,*

$$f(x) = 0 \implies f'(x) < 0, \qquad \forall x \in (0, b).$$

Then there exists a unique $x \in (0, b)$ such that $f(x) = 0$.

Proof. First note that since f' is negative at all its roots, there does not exist any interval $I \subset (0, b)$ such that $f \equiv 0$ on I. Let $0 < x_1 < x_2 < b$ be two distinct roots of f. Since $f'(x_1) < 0$, $f'(x_2) < 0$ and f has a continuous derivative, there exist $y_1 > x_1$, and $y_2 < x_2$ such that $f(y_1) < 0$ and $f(y_2) > 0$. By the intermediate value theorem there exists at least one $x_* \in (x_1, x_2)$ such that $f(x_*) = 0$ and $f'(x_*) \geq 0$. This contradicts our assumption, and hence, f has a unique root. $\square$

Proof of Proposition 2.4.2. Fix $d = 2$. By definition we know $s_1^* = \lambda$ and we set $r_1 \equiv \lambda$ and note that (2.3.4) holds. Now, suppose for any $m \geq 2$, r_ℓ and s_ℓ^* satisfy (2.3.3) and (2.3.4) for $\ell = 1, 2, \cdots, m-1$. Then by Remark 2.3.4, it suffices to show that F_m has a unique fixed point in $[0, s_{m-1}^*]$, which then must be s_m^*. Define $H_m(s) = F_m(s) - s$, $s \in [0, s_{m-1}^*]$. Then, setting $d = 2$ in the expression for β_2 in

(2.2.2) and using (2.3.7),

$$H_m(s) = \lambda s^2 \int_0^L e^{-\lambda(s^*_{m-1}+s)x} \bar{G}(x)dx \\ + \lambda(s^*_{m-1}+s) \int_0^L \left(\int_0^x e^{-\lambda(s^*_{m-1}+s)(x-u)} r_{m-1}(u)du \right) \bar{G}(x)dx - s. \quad (2.4.3)$$

Note that

$$H_m(0) = \lambda s^*_{m-1} \int_0^L \left(\int_0^x e^{-\lambda s^*_{m-1}(x-u)} r_{m-1}(u)du \right) \bar{G}(x)dx > 0.$$

Also, integrating the interior integral of the second term on the right-hand side of (2.4.3) by parts, and using the fact that (2.3.3) and (2.3.4), with $\ell = m-1$, imply $r_{m-1}(0) = \lambda(s^*_{m-1})^2$ and $s^*_{m-1} = \int_0^L r_{m-1}(x)\bar{G}(x)dx$, we conclude that

$$H_m(s) = s^*_{m-1} - s - \left[\lambda(s^*_{m-1})^2 - \lambda s^2\right] \int_0^L e^{-\lambda(s^*_{m-1}+s)x} \bar{G}(x)dx \\ - \int_0^L \left(\int_0^x e^{-\lambda(s^*_{m-1}+s)(x-u)} r'_{m-1}(u)du \right) \bar{G}(x)dx.$$

Since $r'_{m-1} \geq 0$ by Lemma 2.5.5, we have $H_m(s^*_{m-1}) < 0$. Thus, $H_m(x) \neq 0$ for $x \in \{0, s^*_{m-1}\}$. On differentiating (2.4.3), we obtain

$$H'_m(s) = \int_0^L e^{-\lambda(s^*_{m-1}+s)x} \left[2\lambda s - \lambda^2 s^2 x\right] \bar{G}(x)dx \\ + \lambda \int_0^L \left(\int_0^x e^{-\lambda(s^*_{m-1}+s)(x-u)} r_{m-1}(u)du \right) \bar{G}(x)dx \\ - \int_0^L \left(\int_0^x e^{-\lambda(s^*_{m-1}+s)(x-u)} r_{m-1}(u)\lambda^2(s^*_{m-1}+s)(x-u)du \right) \bar{G}(x)dx - 1.$$

Using (2.3.7) with $d = 2$ and $\ell = m-1$ we see that

$$\lambda \int_0^L \left(\int_0^x e^{-\lambda(s^*_{m-1}+s)(x-u)} r_{m-1}(u)du \right) \bar{G}(x)dx \\ = \frac{1}{(s^*_{m-1}+s)} \left(F_m(s) - \lambda s^2 \int_0^L e^{-\lambda(s^*_{m-1}+s)x} \bar{G}(x)dx \right).$$

Now, since $s_1^* = \lambda$ and r_{m-1} satisfies (2.3.3) and therefore (2.5.13), Lemmas 2.5.5 and 2.5.6 imply the monotonicity of r_{m-1}, we have for all $u \in [0, L)$, $r_{m-1}(u) \geq r_{m-1}(0) = \lambda(s_{m-1}^*)^2$. Thus,

$$
\begin{aligned}
&\int_0^L \left(\int_0^x e^{-\lambda(s_{m-1}^*+s)(x-u)} r_{m-1}(u)\lambda^2(s_{m-1}^*+s)(x-u)du\right)\bar{G}(x)dx \\
&\quad \geq \lambda^3(s_{m-1}^*)^2(s_{m-1}^*+s)\int_0^L \left(\int_0^x e^{-\lambda(s_{m-1}^*+s)(x-u)}(x-u)du\right)\bar{G}(x)dx \\
&\quad = \lambda^3(s_{m-1}^*)^2(s_{m-1}^*+s)\int_0^L \left[\frac{1}{\lambda^2(s_{m-1}^*+s)^2} - \frac{e^{-\lambda(s_{m-1}^*+s)x}}{\lambda^2(s_{m-1}^*+s)^2} - \frac{xe^{-\lambda(s_{m-1}^*+s)x}}{\lambda(s_{m-1}^*+s)}\right]\bar{G}(x)dx \\
&\quad = \frac{\lambda(s_{m-1}^*)^2}{(s_{m-1}^*+s)}\int_0^L (1-e^{-\lambda(s_{m-1}^*+s)x})\bar{G}(x)dx - \lambda^2(s_{m-1}^*)^2\int_0^L e^{-\lambda(s_{m-1}^*+s)x}x\bar{G}(x)dx \\
&\quad = \frac{\lambda(s_{m-1}^*)^2}{(s_{m-1}^*+s)}\left(1-\int_0^L e^{-\lambda(s_{m-1}^*+s)x}\bar{G}(x)dx\right) - \lambda^2(s_{m-1}^*)^2\int_0^L e^{-\lambda(s_{m-1}^*+s)x}x\bar{G}(x)dx.
\end{aligned}
$$

Combining the last three displays and rearranging, we obtain

$$
\begin{aligned}
H_m'(s) \leq &\int_0^L e^{-\lambda(s_{m-1}^*+s)x}\left[2\lambda s - \lambda^2 s^2 x\right]\bar{G}(x)dx + \frac{F_\ell(s)}{(s_{m-1}^*+s)} - 1 \\
&- \frac{\lambda s^2}{(s_{m-1}^*+s)}\int_0^L e^{-\lambda(s_{m-1}^*+s)x}\bar{G}(x)dx - \frac{\lambda(s_{m-1}^*)^2}{(s_{m-1}^*+s)}\left(1-\int_0^L e^{-\lambda(s_{m-1}^*+s)x}\bar{G}(x)dx\right) \\
&+ \lambda^2(s_{m-1}^*)^2\int_0^L e^{-\lambda(s_{m-1}^*+s)x}x\bar{G}(x)dx \\
=&\frac{F_\ell(s)}{(s_{m-1}^*+s)} - 1 - \frac{\lambda(s_{m-1}^*)^2}{(s_{m-1}^*+s)} \\
&+ \left(2\lambda s - \frac{\lambda s^2}{(s_{m-1}^*+s)} + \frac{\lambda(s_{m-1}^*)^2}{(s_{m-1}^*+s)}\right)\int_0^L e^{-\lambda(s_{m-1}^*+s)x}\bar{G}(x)dx \\
&+ \left(\lambda^2(s_{m-1}^*)^2 - \lambda^2 s^2\right)\int_0^L e^{-\lambda(s_{m-1}^*+s)x}x\bar{G}(x)dx \\
=&\frac{(F_\ell(s)-\lambda(s_{m-1}^*)^2)}{(s_{m-1}^*+s)} - 1 + \left(2\lambda s + \frac{\lambda((s_{m-1}^*)^2-s^2)}{(s_{m-1}^*+s)}\right)\int_0^L e^{-\lambda(s_{m-1}^*+s)x}\bar{G}(x)dx \\
&+ \lambda^2((s_{m-1}^*)^2-s^2)\int_0^L e^{-\lambda(s_{m-1}^*+s)x}x\bar{G}(x)dx \\
=&\frac{(F_\ell(s)-\lambda(s_{m-1}^*)^2)}{(s_{m-1}^*+s)} - 1 + \lambda(s_{m-1}^*+s)\int_0^L e^{-\lambda(s_{m-1}^*+s)x}\bar{G}(x)dx \\
&+ \lambda^2((s_{m-1}^*)^2-s^2)\int_0^L e^{-\lambda(s_{m-1}^*+s)x}x\bar{G}(x)dx.
\end{aligned}
$$

Using Hölder's inequality, $\bar{G}^2(x) \le \bar{G}(x)$ for all $x \in [0, L)$ and the mean one condition, we have

$$\begin{aligned}\int_0^L e^{-\lambda(s^*_{m-1}+s)x}\bar{G}(x)dx &\le \left(\int_0^L e^{-2\lambda(s^*_{m-1}+s)x}dx\right)^{1/2}\left(\int_0^L \bar{G}^2(x)dx\right)^{1/2}\\ &\le \left(\int_0^L e^{-2\lambda(s^*_{m-1}+s)x}dx\right)^{1/2}\\ &\le \frac{1}{\sqrt{2\lambda(s^*_{m-1}+s)}},\end{aligned}$$

and similarly

$$\begin{aligned}\int_0^L xe^{-\lambda(s^*_{m-1}+s)x}\bar{G}(x)dx &\le \left(\int_0^L x^2e^{-2\lambda(s^*_{m-1}+s)x}dx\right)^{1/2}\left(\int_0^L \bar{G}^2(x)dx\right)^{1/2}\\ &\le \left(\int_0^L x^2e^{-2\lambda(s^*_{m-1}+s)x}dx\right)^{1/2}\\ &\le \frac{1}{2\sqrt{\lambda^3(s^*_{m-1}+s)^3}}.\end{aligned}$$

Combining the last three displays and evaluating H'_m at any fixed point $\tilde{s}^*_m$ of F_m in $[0, s^*_{m-1}]$,

$$H'_m(\tilde{s}^*_m) \le -\frac{s^*_{m-1}+\lambda(s^*_{m-1})^2}{(s^*_{m-1}+\tilde{s}^*_m)} + \left(\frac{1}{\sqrt{2}}+\frac{1}{2}\right)\sqrt{\lambda(s^*_{m-1}+\tilde{s}^*_m)} - \frac{\lambda\tilde{s}^*_m}{\sqrt{\lambda(s^*_{m-1}+\tilde{s}^*_m)}} = K_m(\tilde{s}^*_m), \tag{2.4.4}$$

where for $s \in [0, s^*_{m-1}]$,

$$K_m(s) = -\frac{s^*_{m-1}+\lambda(s^*_{m-1})^2}{(s^*_{m-1}+s)} + \left(\frac{1}{\sqrt{2}}+\frac{1}{2}\right)\sqrt{\lambda(s^*_{m-1}+s)} - \frac{\lambda s}{\sqrt{\lambda(s^*_{m-1}+s)}}.$$

Differentiating K_m we obtain

$$
\begin{aligned}
K_m'(s) =& \frac{s_{m-1}^* + \lambda(s_{m-1}^*)^2}{(s_{m-1}^*+s)^2} + \frac{1}{2}\left(\frac{1}{\sqrt{2}} + \frac{1}{2}\right)\frac{\lambda}{\sqrt{\lambda(s_{m-1}^*+s)}} \\
& - \frac{\lambda}{\sqrt{\lambda(s_{m-1}^*+s)}} + \frac{\lambda^2 s}{2\lambda(s_{m-1}^*+s)\sqrt{\lambda(s_{m-1}^*+s)}} \\
=& \frac{(s_{m-1}^* + \lambda(s_{m-1}^*)^2)}{(s_{m-1}^*+s)^2} + \frac{1}{2}\left(\frac{1}{\sqrt{2}} + \frac{1}{2}\right)\frac{\lambda}{\sqrt{\lambda(s_{m-1}^*+s)}} - \frac{2\lambda s_{m-1}^* + \lambda s}{2(s_{m-1}^*+s)\sqrt{\lambda(s_{m-1}^*+s)}} \\
=& \frac{4s_{m-1}^* + 4\lambda(s_{m-1}^*)^2 - (3s_{m-1}^*+s)\sqrt{\lambda(s_{m-1}^*+s)}}{4(s_{m-1}^*+s)^2} + \frac{1}{2\sqrt{2}}\frac{\lambda}{\sqrt{\lambda(s_{m-1}^*+s)}}.
\end{aligned}
$$

To show that $K_m'(s) \geq 0$ for $s \in [0, s_{m-1}^*]$, it suffices to show the numerator of the first term is positive. Since $a^2 > b^2$ implies $a > b$ when $a, b > 0$, this follows from

$$
\begin{aligned}
& 16(s_{m-1}^*)^2 + 16\lambda^2(s_{m-1}^*)^4 + 32\lambda(s_{m-1}^*)^3 - \left(\lambda(s_{m-1}^*+s)(9(s_{m-1}^*)^2 + s^2 + 6s_{m-1}^* s)\right) \\
& \geq 16(s_{m-1}^*)^2 + 16\lambda^2(s_{m-1}^*)^4 + 32\lambda(s_{m-1}^*)^3 - 2\lambda s_{m-1}^*(16 s_{m-1}^*) \\
& \geq 16(s_{m-1}^*)^2 + 16\lambda^2(s_{m-1}^*)^4 \\
& \geq 0.
\end{aligned}
$$

Hence, K_m is increasing on $[0, s_{m-1}^*]$. When combined with (2.4.4), this implies

$$
\begin{aligned}
H_m'(\tilde{s}_m^*) \leq K_m(\tilde{s}_m^*) \leq K_m(s_{m-1}^*) &= -\frac{1+\lambda s_{m-1}^*}{2} + \left(1 + \frac{1}{\sqrt{2}}\right)\sqrt{\lambda s_{m-1}^*} - \frac{\sqrt{\lambda s_{m-1}^*}}{\sqrt{2}} \\
&= -\frac{1}{2}\left(1 + \lambda s_{m-1}^* - 2\sqrt{\lambda s_{m-1}^*}\right) \\
&= -\frac{1}{2}(1 - \sqrt{\lambda s_{m-1}^*})^2 \\
&< 0.
\end{aligned}
$$

Hence using the definition of H_m in (2.4.3), continuous differentiability of F_m from Proposition 2.5.4 and substituting $b = s_{m-1}^*$ and $f = H_m$ in Lemma 2.4.3, we conclude that F_m has a unique fixed point in $[0, s_{m-1}^*]$. □

Now, fix $d = 2$, $\lambda \in (0,1)$ and a service distribution that satisfies Assumptions 2.2.1 and 2.2.4. Let $(s_\ell^*)_{\ell \in \mathbb{N}}$ be the invariant queue length distribution, which is uniquely defined by Proposition 2.4.2. For $\ell \geq 2$, given $\{s_i^*\}_{i \leq \ell-1}$ and $\{r_i\}_{i \leq \ell-1}$, and F_ℓ as in (2.3.4), we compute s_ℓ^* as the zero of the function $s \mapsto F_\ell(s) - s$ on the interval $[0, s_{\ell-1}^*]$ using the bisection method with an error tolerance of 10^{-12}. We then use the recursion (2.4.1) using the adaptive quadrature method with a tolerance of 10^{-12} to numerically approximate the integrals therein for the following list of distributions.

- Gamma distribution: The Gamma distribution with unit mean is parameterized by $\alpha > 0$, and

$$\bar{G}(x) = \frac{\gamma(\alpha, \alpha x)}{\Gamma(\alpha)}, \qquad x \geq 0,$$

where Γ denotes the Gamma function

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx,$$

and γ denotes the lower incomplete gamma function,

$$\gamma(a, x) = \int_0^x z^{a-1} e^{-z} dz, \qquad a > 0, x \geq 0.$$

- Weibull distribution: The Weibull distribution with unit mean is parameterized by $a > 0$, and

$$\bar{G}(x) = e^{(-x/b)^a}, \qquad x \geq 0,$$

where $b = \frac{1}{\Gamma(1+\frac{1}{a})}$.

- Lognormal distribution: The Lognormal distribution with unit mean is parameterized by $\sigma > 0$, and

$$\bar{G}(x) = \frac{1}{2} - \frac{1}{2} erf\left[\frac{\log x - \mu}{\sqrt{2}\sigma}\right], \qquad x > 0,$$

where $\mu = -\frac{\sigma^2}{2}$, and erf denotes the error function

$$erf(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-z^2} dz, \qquad x > 0.$$

- Pareto distribution: The Pareto distribution with unit mean is parameterized by $\alpha \geq 1$, and

$$\bar{G}(x) = \begin{cases} 1 & \text{if } x \in [0, x_m), \\ \left(\frac{x_m}{x}\right)^\alpha & \text{if } x \geq x_m, \end{cases}$$

where $x_m = \frac{\alpha-1}{\alpha}$.

- Burr distribution: The Burr distribution with unit mean is parameterized by $c > 0$ and $k > 1/c$ such that $kB(k - \frac{1}{c}, 1 + \frac{1}{c}) = 1$, where B denotes the beta function, i.e.,

$$B(a,b) = \int_0^1 t^{(a-1)}(1-t)^{(b-1)} dt, \qquad a, b > 0,$$

and

$$\bar{G}(x) = (1 + x^c)^{-k}, \qquad x \geq 0.$$

The resulting numerical estimates for $\lambda = 0.5$ are tabulated in Table 2.1, where s_ℓ^* is set to zero if the numerical estimate for s_ℓ^* is below the precision error, i.e., $s_\ell^* \leq 10^{-12}$. Continuing the numerical computation for $6 \leq \ell \leq 8$ for the distributions in Table 2.1, the only non-zero values are $s_6^* = 2.7 \times 10^{-6}$, and $s_7^* = 1.8147 \times 10^{-11}$ for the Weibull distribution with $a = 0.5$, and $s_6^* = 0.0231$, $s_7^* = 0.0182$ and $s_8^* = 0.015$ for the Pareto distribution with $\alpha = 1.5$. Even for small values of ℓ the decay of

	s_2^*	s_3^*	s_4^*	s_5^*
Exponential	0.125	0.0078	3.0518×10^{-5}	4.6566×10^{-10}
Gamma ($\alpha = 3$)	0.0871	0.0022	1.1892×10^{-6}	3.629×10^{-11}
Weibull ($a = 2$)	0.0838	0.0018	7.1889×10^{-7}	0
Weibull ($a = 0.5$)	0.2425	0.0872	0.0166	0.0008
Lognormal ($\sigma = \frac{1}{3}$)	0.0738	0.0012	2.4229×10^{-7}	2.985×10^{-11}
Pareto ($\alpha = 3$)	0.0812	0.0024	1.2208×10^{-5}	0
Pareto ($\alpha = 1.5$)	0.1820	0.0797	0.0460	0.0311
Burr ($c = 2$, $k = 1.5$)	0.1117	0.0076	9.8831×10^{-5}	1.1969×10^{-7}

Table 2.1: Numerical estimates for s_ℓ^*, $2 \leq \ell \leq 5$, for various distributions.

s_ℓ^* appears to be faster for lighter-tailed as opposed to heavier-tailed distributions (such as Weibull with $\alpha = 0.5$, and Pareto with $\alpha = 1.5$). In Section 2.4.3, we show that these approximations to the invariant state agree well with the approximations to the limiting stationary distribution obtained via simulations.

To understand the rate of decay of the invariant queue length distribution with respect to ℓ, Table 2.2 tabulates the function $H(\ell) = \log_2(\log(1/s_\ell^*))/\ell$ for the above distributions, where the entry NaN denotes that $H(\ell)$ could not be computed due to loss of precision while estimating s_ℓ^*. For service distributions exhibiting a doubly exponential tail decay, by Theorem 2.3.9 $H(\ell)$ should converge to (or be lower bounded by) a constant as $\ell \to \infty$. For example, it is easy to check that for the exponential service distribution $H(\ell)$ converges to 1 as $\ell \to \infty$. This assertion cannot be inferred from Table 2.2, thus showing that the decay rate does not manifest itself till far into the tail. Thus, from a practical point of view, this underscores the importance of being able to compute s_ℓ^* for finite ℓ, as our approach allows.

	$H(2)$	$H(3)$	$H(4)$	$H(5)$
Exponential	0.5281	0.7595	0.8445	0.8851
Gamma ($\alpha = 3$)	0.6436	0.8724	0.9425	0.9688
Weibull ($a = 2$)	0.6549	0.8859	0.9556	NaN
Weibull ($a = 0.5$)	0.2513	0.4289	0.5085	0.5650
Lognormal ($\sigma = \frac{1}{3}$)	0.6911	0.9182	0.9821	0.9198
Pareto ($\alpha = 3$)	0.6640	0.8629	0.8878	NaN
Pareto ($\alpha = 1.5$)	0.3843	0.4462	0.4057	0.3590
Burr ($c = 2$, $k = 1.5$)	0.5562	0.7621	0.8013	0.7989

Table 2.2: Computations of $H(\ell) = \frac{\log_2(\log(\frac{1}{s_\ell^*}))}{\ell}$ for s_ℓ^*, $2 \leq \ell \leq 5$.

Due to the measure-valued representation, it is also possible to calculate the invariant states of other important performance measures in addition to the queue length. For example, consider the virtual waiting time at any time t, defined to be the amount of time that a virtual job hypothetically arriving at time t would have to wait before entering service. Using [2, Theorem 4.5], it can be easily deduced that when $d = 2$, the invariant mean virtual waiting time, W^* is given by

$$W^* = \sum_{\ell \geq 2} Z_\ell(0)^2 + \sum_{\ell \geq 1} [Z_\ell(0) + Z_{\ell+1}(0)] \int_0^\infty [Z_\ell(r) + Z_{\ell+1}(r)] dr, \tag{2.4.5}$$

where $Z_\ell(r) := \langle (\bar{G}(\cdot + r)/\bar{G}(\cdot)), \nu_\ell^* \rangle$. We now compute the invariant mean virtual waiting time by using the following approximation to (2.4.5).

$$W^* \sim \sum_{\ell=2}^{L_0} Z_\ell(0)^2 + \sum_{\ell=1}^{L_0 - 1} [Z_\ell(0) + Z_{\ell+1}(0)] \sum_{j=0}^{\lfloor R_0/\delta \rfloor} [Z_\ell(r_j) + Z_{\ell+1}(r_j)] \delta,$$

where $L_0 = 6$, $R_0 = 20$, $\delta = 0.003$ and $r_j = j\delta$, with a tolerance of 10^{-12}.

	Exponential	Pareto ($\alpha = 3$)	Pareto ($\alpha = 2.5$)	Pareto ($\alpha = 2$)	Pareto ($\alpha = 1.75$)	Pareto ($\alpha = 1.5$)
W^*	0.4246	0.1673	0.1905	0.2449	0.3042	0.4287

Table 2.3: Invariant mean virtual waiting time for Exponential and Pareto distributions.

	Weibull ($a = 2$)	Weibull ($a = 1.5$)	Weibull ($a = 1$)	Weibull ($a = 0.5$)	Gamma ($\alpha = 3$)
W^*	0.1716	0.1966	0.2661	0.6781	0.1789

Table 2.4: Invariant mean virtual waiting time for Weibull and Gamma distributions.

The numerics show that for a given distribution, the invariant mean virtual waiting time increases with increase in the heaviness of the tail of the distribution. While this may appear intuitive, it is not completely obvious, since it is in contrast to transient time results obtained in [2, Section 5.2.2], where the relaxation time (defined to be the time it takes for the network backlog to decrease to the extent that the mean virtual waiting time reaches half of its initial value) is smaller in the heavy-tailed case when compared to the light-tailed case.

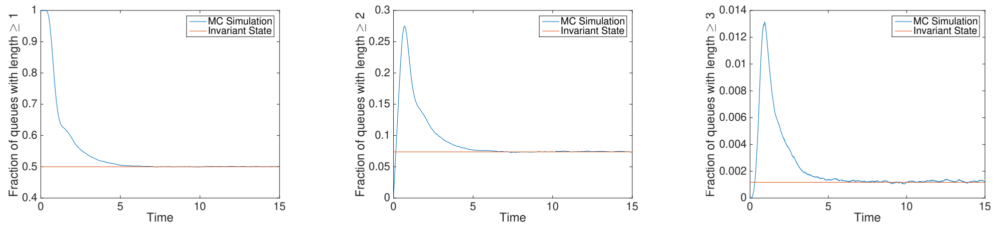
2.4.3 Potential convergence of the N-server equilibria to the invariant state

As mentioned in Section 1.2.2, for service distributions with decreasing hazard rate function, it was shown in [9] that the sequence of stationary distributions π_N of N-server systems subject to our load balancing algorithm converges to a limit π^* and that in the case of power law distributions with tail parameter β, this limit has a doubly exponential tail if and only if $\beta > d/(d-1)$. They use a different approach not involving the study of dynamic behavior (or the hydrodynamic limit).

An alternative approach is to characterize the hydrodynamic limit and show that the unique invariant state of the hydrodynamic equations coincides with the limiting stationary distribution. The results of this chapter, along with that of [4] and [9] can be combined to show that this is true in the case of service distributions with decreasing hazard rate. We conjecture that the unique physical invariant state coincides with the limiting stationary distribution even for more general service distributions. Below, we provide numerical evidence to support this conjecture.

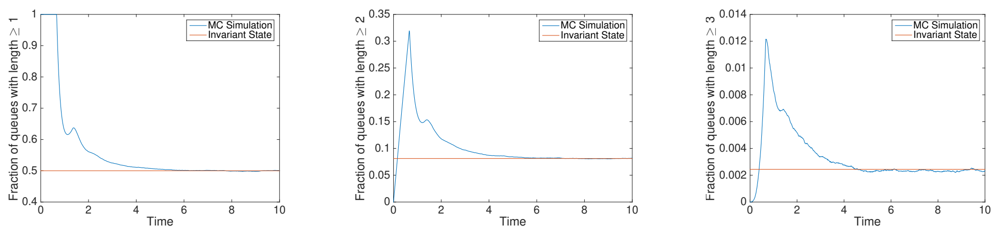
We only present results for the lognormal and Pareto distributions. Note that, while the Pareto distribution has a decreasing hazard rate function, and so one expects agreement of the two quantities due to [4, 9], it is significant that the lognormal distribution still shows agreement despite not having a monotonic hazard rate function. Similar trends are observed for a variety of other service distributions and arrival rates. Figures 2.1 and 2.2 below present a comparison, for $\ell = 1, 2, 3$, of s^*_ℓ with estimates of the probability that a typical queue has length no less than ℓ at time t, using Monte Carlo simulations of an N-server network with $N = 600$ and $\lambda = 0.5$ for the time interval $t \in [0, 15]$ for the Lognormal distribution, and $t \in [0, 10]$ for the Pareto distribution, using 600 realizations, with the initial condition that at time $t = 0$, every server has exactly one job with age 0. We then compute the probability that a typical queue has length $\geq \ell$ at time t, and compare this quantity with the invariant state s^*_ℓ obtained in Table 2.1. For large ℓ (e.g., $\ell \geq 4$) it becomes harder to use simulations to get accurate estimates of the equilibrium probability of the queue exceeding ℓ; whereas, the recursions can still be used to get an approximation to s^*_ℓ. Figure 2.1 confirms the results for the Lognormal distribution and Figure 2.2

for the Pareto distribution and provides support for the conjecture (similar trends are observed for a variety of initial conditions and service distributions.)



(a) $\ell = 1$ (b) $\ell = 2$ (c) $\ell = 3$

Figure 2.1: Comparison of the probability of the queue exceeding ℓ obtained from MC simulation with the numerical approximation to the invariant state s_ℓ^* for the Lognormal distribution with $\sigma = 1/3$



(a) $\ell = 1$ (b) $\ell = 2$ (c) $\ell = 3$

Figure 2.2: Comparison of the probability of the queue exceeding ℓ obtained from MC simulation with the numerical approximation to the invariant states$_\ell^*$ for the Pareto distribution with $\alpha = 3$

2.5 Proof of Theorem 2.3.3

Let $\nu^* = (\nu_\ell^*)_{\ell\in\mathbb{N}}$ be a physical invariant state of the hydrodynamic equations in the sense of Remark 2.3.2. Then (by definition) $\nu(t) = (\nu_\ell^*)_{\ell\in\mathbb{N}}$, $t > 0$, is a solution to the hydrodynamic equations with initial condition ν^*, and condition (2.3.2) holds with $(s_\ell^*)_{\ell\in\mathbb{N}}$ defined as in (2.3.1). Define the associated invariant quantities

$$\eta_\ell^* := \begin{cases} \lambda(1 - \langle \mathbf{1}, \nu_1^* \rangle^d)\delta_0 & \text{if } \ell = 1, \\ \lambda \mathfrak{P}_d\left(\langle \mathbf{1}, \nu_{\ell-1}^*\rangle, \langle \mathbf{1}, \nu_\ell^* \rangle\right)(\nu_{\ell-1}^* - \nu_\ell^*) & \text{if } \ell \geq 2. \end{cases} \tag{2.5.1}$$

From (2.2.7), it follows that for every $t \geq 0$, the departure process with initial state ν^* satisfies

$$D_\ell(t) = \langle h, \nu_\ell^* \rangle t, \qquad \forall \ell \in \mathbb{N}, \tag{2.5.2}$$

and using (2.2.8), the measure-valued routing process is given by $\eta_\ell(t) = \eta_\ell^*$ for all $t \geq 0$.

The proof of Theorem 2.3.3, which is presented in Section 2.5.4, relies on several preliminary results. First, in Section 2.5.1, we obtain a bound on the departure rate in an invariant state, in Section 2.5.2 we obtain a useful characterization of an invariant state (see Proposition 2.5.4), and in Section 2.5.3 we establish existence of a physical invariant state.

2.5.1 The departure rate in an invariant state

Lemma 2.5.1. *Suppose $\nu^* = (\nu_\ell^*)_{\ell \in \mathbb{N}}$ is an invariant state of the corresponding hydrodynamic equations, and let $s_\ell^* = \langle \mathbf{1}, \nu_\ell^* \rangle$ be the corresponding queue length distribution. Then, for every $\ell \geq 1$,*

$$\langle h, \nu_{\ell+1}^* \rangle = \lambda (s_\ell^*)^d - \lambda + \langle h, \nu_1^* \rangle. \tag{2.5.3}$$

Moreover if (2.3.2) holds, that is, $s_\ell^* \to 0$ as $\ell \to \infty$, then

$$\langle h, \nu_1^* \rangle \geq \lambda. \tag{2.5.4}$$

Proof. Let $\nu^* = (\nu_\ell^*)_{\ell \in \mathbb{N}}$ be an invariant state of the hydrodynamic equations. Then, for every $\ell \geq 1$, (2.2.4) implies $\langle h, \nu_\ell^* \rangle < \infty$, and (2.2.5), (2.5.1) and (2.5.2) imply that

$$\langle h, \nu_2^* \rangle - \langle h, \nu_1^* \rangle = -\lambda(1 - (s_1^*)^d), \tag{2.5.5}$$

and for $\ell \geq 2$,

$$\langle h, \nu_{\ell+1}^* \rangle - \langle h, \nu_\ell^* \rangle = -\lambda \mathfrak{P}_d \Big(s_{\ell-1}^*, s_\ell^* \Big)(s_{\ell-1}^* - s_\ell^*) = -\lambda[(s_{\ell-1}^*)^d - (s_\ell^*)^d], \tag{2.5.6}$$

where the last equality invokes (2.2.2). Thus, defining $s_0^* := 1$, we see that for $\ell \geq 1$,

$$\begin{aligned} \langle h, \nu_{\ell+1}^* \rangle &= \sum_{i=1}^{\ell} \big(\langle h, \nu_{i+1}^* \rangle - \langle h, \nu_i^* \rangle \big) + \langle h, \nu_1^* \rangle \\ &= -\lambda \sum_{i=1}^{\ell} \big((s_{i-1}^*)^d - (s_i^*)^d \big) + \langle h, \nu_1^* \rangle \\ &= \lambda (s_\ell^*)^d - \lambda + \langle h, \nu_1^* \rangle, \end{aligned} \tag{2.5.7}$$

which proves (2.5.3). Furthermore, since $\langle h, \nu_{\ell+1}^* \rangle \geq 0$ for all ℓ, and $\lim_{\ell \to \infty} s_\ell^* = 0$, (2.5.7) implies (2.5.4). □

2.5.2 Characterization of invariant states

We begin by establishing the absolute continuity of the invariant state.

Lemma 2.5.2. *Suppose $\nu^* = (\nu_\ell^*)_{\ell \in \mathbb{N}} \in \mathbb{S}$ is an invariant state of the hydrodynamic equations and $(s_\ell^*)_{\ell \in \mathbb{N}}$ is the associated invariant queue length distribution as defined in (2.3.1). Then for every $\ell \in \mathbb{N}$, ν_ℓ^* is absolutely continuous with respect to Lebesgue measure on $[0, \infty)$.*

We first make a simple observation that will be used in the proof of Lemma 2.5.2.

Lemma 2.5.3. *If $(\nu_\ell)_{\ell \in \mathbb{N}}$ is a solution to the hydrodynamic equations with initial* *condition $\nu(0)$, then (2.2.6) holds for all Lebesgue integrable functions f on $[0,\infty)$.*

Proof. We show below that (2.2.6) holds for indicator functions. Indeed, fix $0 \le a \le b < \infty$ consider a sequence $(\varepsilon_n)_{n\ge 1}$ with $\varepsilon_n < 1$ for each $n \ge 1$ and $\varepsilon_n \downarrow 0$. For each $n \ge 1$, consider $f_n \in \mathbb{C}_b[0,\infty)$ defined by

$$f_n(x) := \begin{cases} 1 & x \in [a,b], \\ \text{linear} & x \in [a-\varepsilon_n, a) \cup (b, b+\varepsilon_n], \\ 0 & \text{otherwise.} \end{cases}$$

Then $f_n \to 1_{[a,b]}$ pointwise as $n \to \infty$. Using (2.2.6) with $f = f_n$ and the dominated convergence theorem, we have

$$\begin{aligned} \lim_{n\to\infty} \langle f_n, \nu_\ell(t)\rangle = \lim_{n\to\infty} \Bigg[&\langle f_n(\cdot + t)\frac{\bar{G}(\cdot+t)}{\bar{G}(\cdot)}, \nu_\ell(0)\rangle + \int_{[0,t]} f_n(t-s)\bar{G}(t-s)dD_{\ell+1}(s) \\ &+ \int_0^t \langle f_n(\cdot + t - s)\frac{\bar{G}(\cdot + t - s)}{\bar{G}(\cdot)}, \eta_\ell(s)\rangle ds \Bigg] \\ =&\langle 1_{[a,b]}(\cdot + t)\frac{\bar{G}(\cdot+t)}{\bar{G}(\cdot)}, \nu_\ell(0)\rangle + \int_{[0,t]} 1_{[a,b]}(t-s)\bar{G}(t-s)dD_{\ell+1}(s) \\ &+ \int_0^t \langle 1_{[a,b]}(\cdot + t - s)\frac{\bar{G}(\cdot + t - s)}{\bar{G}(\cdot)}, \eta_\ell(s)\rangle ds \end{aligned}$$

By linearity, it then follows that (2.2.6) holds for all simple functions. Since any Lebesgue integrable function can be represented as a monotone limit of simple functions, another application of the DCT shows that (2.2.6) is satisfied whenever f is a Lebesgue integrable function. □

In what follows, ν^* is an invariant state of the hydrodynamic equations.

Proof of Lemma 2.5.2. For $0 \le a < b < \infty$, substituting $\nu = \nu^*$ and $f = 1_{[a,b]}$ in

(2.2.6), and using (2.5.2), we have for each $t > 0$,

$$\begin{aligned}
\nu_\ell^*[a,b] =& \int_0^L 1_{[a,b]}(x+t)\frac{\bar{G}(x+t)}{\bar{G}(x)}\nu_\ell^*(dx) + \int_{[0,t]} 1_{[a,b]}(t-s)\bar{G}(t-s)\langle h, \nu_{\ell+1}^*\rangle ds \\
&+ \int_0^t \int_0^L 1_{[a,b]}(x+t-s)\frac{\bar{G}(x+t-s)}{\bar{G}(x)}\eta_\ell^*(dx)ds \\
=& \int_{(a-t)^+\wedge L}^{(b-t)^+\wedge L} \frac{\bar{G}(x+t)}{\bar{G}(x)}\nu_\ell^*(dx) + \langle h, \nu_{\ell+1}^*\rangle \int_{(t-b)^+}^{(t-a)^+} \bar{G}(t-s)ds \\
&+ \int_0^t \int_{(a-(t-s))^+\wedge L}^{(b-(t-s))^+\wedge L} \frac{\bar{G}(x+t-s)}{\bar{G}(x)}\eta_\ell^*(dx)ds \\
\leq& \nu_\ell^*[(a-t)^+\wedge L, (b-t)^+\wedge L] + \langle h, \nu_{\ell+1}^*\rangle(b-a) \\
&+ \int_0^t \int_{(a-(t-s))^+\wedge L}^{(b-(t-s))^+\wedge L} \frac{\bar{G}(x+t-s)}{\bar{G}(x)}\eta_\ell^*(dx)ds. \qquad (2.5.8)
\end{aligned}$$

The relation (2.5.1) implies that for $\ell = 1$,

$$\int_0^t \int_{(a-(t-s))^+\wedge L}^{(b-(t-s))^+\wedge L} \frac{\bar{G}(x+t-s)}{\bar{G}(x)}\eta_1^*(dx)ds \leq \int_{(t-b)^+}^{(t-a)+} \bar{G}(t-s)\lambda(1-\langle \mathbf{1}, \nu_1^*\rangle^d)ds \leq \lambda(b-a). \tag{2.5.9}$$

Claim: Given $0 \leq a < b < \infty$,

$$\nu_\ell^*[a,b] \leq A_\ell(b-a) \qquad \ell \geq 1, \tag{2.5.10}$$

where

$$A_\ell := \begin{cases} \langle h, \nu_2^*\rangle + \lambda & \ell = 1, \\ \langle h, \nu_{\ell+1}^*\rangle + d\lambda t A_{\ell-1} & \ell \geq 2. \end{cases} \tag{2.5.11}$$

The claim for $\ell = 1$ follows on using (2.5.9) and (2.5.8), with $\ell = 1$, and sending $t \to \infty$.

Now, suppose (2.5.10) holds for some $\ell - 1$ with $\ell \geq 2$. For $0 \leq s \leq t < \infty$, using (2.5.1), the inequality $\mathfrak{P}_d(x,y) \leq d\ \forall x, y \geq 0$, and (2.5.10), with ℓ replaced by $\ell - 1$, we have

$$\begin{aligned}
&\int_0^t \int_{(a-(t-s))^+\wedge L}^{(b-(t-s))^+\wedge L} \frac{\bar{G}(x+t-s)}{\bar{G}(x)} \eta_\ell^*(dx)ds \\
&\quad \leq \int_0^t \int_{(a-(t-s))^+\wedge L}^{(b-(t-s))^+\wedge L} \frac{\bar{G}(x+t-s)}{\bar{G}(x)} d\lambda(\nu_{\ell-1}^*(dx) - \nu_\ell^*(dx))ds \\
&\quad \leq d\lambda \int_0^t \nu_{\ell-1}^*((a-(t-s))^+ \wedge L, (b-(t-s))^+ \wedge L)ds \\
&\quad \leq d\lambda t A_{\ell-1}(b-a).
\end{aligned}$$

When combined with (2.5.8) and(2.5.11), and sending $t \to \infty$, the above equation yields (2.5.10). Given $\varepsilon_1 > 0$, define $\delta_1(\varepsilon_1) := \varepsilon_1/2(\langle h, \nu_2^* \rangle + \lambda)$. Let $E \subset [0, \infty)$ be a Lebesgue measurable set with $\mu(E) < \delta_1(\varepsilon_1)$, where μ denotes Lebesgue measure. Since (by, e.g., [18, Lemma 1.17])

$$\mu(E) = \inf \left\{ \sum_{i\in\mathbb{N}} \mu((a_i, b_i)) : E \subset \bigcup_{i\in\mathbb{N}} (a_i, b_i) \right\},$$

there exists $((a_i, a_i + \Delta_i))_{i\in\mathbb{N}}$ such that $E \subset \bigcup_{i\in\mathbb{N}}(a_i, a_i + \Delta_i)$ and $\sum_{i\in\mathbb{N}} \mu((a_i, a_i + \Delta_i)) = \sum_{i\in\mathbb{N}} \Delta_i < 2\delta_1(\varepsilon_1)$. Since $\langle h, \nu_2^* \rangle \geq 0$ by Lemma 2.5.1, using (2.5.10) and (2.5.11) with $\ell = 1$, we obtain for all i,

$$\nu_1^*(E) \leq \sum_{i\in\mathbb{N}} \nu_1^*(a_i, a_i + \Delta_i) \leq \sum_{i\in\mathbb{N}} (\langle h, \nu_2^* \rangle + \lambda)\Delta_i = 2(\langle h, \nu_2^* \rangle + \lambda)\delta_1(\varepsilon_1) = \varepsilon_1.$$

Likewise, fix $\ell \geq 2$. Given $\varepsilon_\ell > 0$, define $\delta_\ell(\varepsilon_\ell) = \varepsilon_\ell/2n_\ell$. Now suppose $E \subset [0, \infty)$ is Lebesgue measurable with $\mu(E) < \delta_\ell$. Then there exist $((a_i, a_i + \Delta_i))_{i\in\mathbb{N}}$ such that $E \subset \bigcup_{i\in\mathbb{N}}(a_i, a_i + \Delta_i)$ and $\sum_{i\in\mathbb{N}} \mu((a_i, a_i + \Delta_i)) = \sum_{i\in\mathbb{N}} \Delta_i < 2\delta_\ell$. Using (2.5.10), we see that,

$$\sum_{i\in\mathbb{N}} \nu_\ell^*[a_i, a_i + \Delta_i] \leq \sum_{i\in\mathbb{N}} A_\ell \Delta_i.$$

Hence, using $\langle h, \nu_\ell^* \rangle \geq 0$ for all $\ell \geq 1$ from Lemma 2.5.1, we have,

$$\nu_\ell^*(E) \leq \sum_{i \in \mathbb{N}} \nu_\ell^*(a_i, a_i + \Delta_i) \leq \sum_{i \in \mathbb{N}} A_\ell \Delta_i < 2A_\ell \delta_\ell(\varepsilon_\ell) = \varepsilon_\ell.$$

Thus, we have shown that for every $\ell \geq 1$, given $\varepsilon_\ell > 0$, $\exists \delta_\ell(\varepsilon_\ell) > 0$ such that $\mu(E) < \delta_\ell(\varepsilon_\ell) \implies \nu_\ell^*(E) < \varepsilon_\ell$. This proves ν_ℓ^* is absolutely continuous with respect to μ. $\square$

Proposition 2.5.4. *Suppose $\nu^* = (\nu_\ell^*)_{\ell \in \mathbb{N}}$ is an invariant state of the hydrodynamic equations, and s_ℓ^* is defined in terms of ν_ℓ^* via (2.3.1) for each $\ell \geq 1$. Then there exists a sequence $(r_\ell)_{\ell \in \mathbb{N}}$, of measurable functions on $[0, \infty)$ that admit a version that is continuously differentiable on $[0, L)$, such that $r_\ell(x) = 0$ for $x \notin [0, L)$, (2.3.4) and (2.3.5) hold,*

$$r_1 \equiv s_1^*, \qquad \text{on } [0, L), \tag{2.5.12}$$

and for $\ell \geq 2$, and $x \in [0, L)$,

$$r_\ell(x) = (\lambda (s_\ell^*)^d + s_1^* - \lambda) e^{-\lambda \mathfrak{P}_d(s_{\ell-1}^*, s_\ell^*) x} + \lambda \mathfrak{P}_d(s_{\ell-1}^*, s_\ell^*) \int_0^x e^{-\lambda \mathfrak{P}_d(s_{\ell-1}^*, s_\ell^*)(x-u)} r_{\ell-1}(u) du. \tag{2.5.13}$$

Proof. Since each ν_ℓ^* is absolutely continuous with respect to Lebesgue measure by Lemma 2.5.2, and $\bar{G}(x) > 0$ for $x \in [0, L)$, there exists $(r_\ell)_{\ell \in \mathbb{N}}$, a sequence of measurable functions on $[0, L)$ such that $\frac{d\nu_\ell}{dx}(x) = r_\ell(x) \bar{G}(x)$. Hence, $r_\ell(x) = \mathbb{I}_{[0,L)}(x) \frac{1}{\bar{G}(x)} \frac{d\nu_\ell}{dx}(x)$ clearly satisfies (2.3.5), and since $s_\ell^* = \langle \mathbf{1}, \nu_\ell^* \rangle$, this immediately implies (2.3.4). Now using $\nu(t) = \nu^*$ in (2.2.6), and for $\ell = 1$, combining (2.5.1), (2.5.2), and (2.5.5), we have for every $f \in \mathbb{C}_b[0,\infty)$ and $t \geq 0$,

$$\begin{aligned}\int_0^L f(x)\bar{G}(x)r_1(x)dx &= \int_0^L f(x+t)\bar{G}(x+t)r_1(x)dx + \langle h, \nu_2^* \rangle \int_0^t f(s)\bar{G}(s)ds \\ &\quad + \lambda\left(1-(s_1^*)^d\right)\int_0^t f(s)\bar{G}(s)ds \\ &= \int_0^L f(x+t)\bar{G}(x+t)r_1(x)dx + \langle h, \nu_1^* \rangle \int_0^t f(s)\bar{G}(s)ds.\end{aligned}$$

Recalling that $h = g/\bar{G}$, this implies that

$$\int_0^L f(x)\bar{G}(x)\left[r_1(x) - \mathbb{I}_{\{x\geq t\}}r_1(x-t) - \mathbb{I}_{\{x<t\}}\left(\int_0^L g(s)r_1(s)\,ds\right)\right]dx = 0.$$

A standard monotone class argument shows that this holds for every bounded measurable function f. Hence, the above equation holds if and only if for a.e. $(x,t) \in [0,L) \times [0,\infty)$,

$$r_1(x) = \mathbb{I}_{\{x\geq t\}}r_1(x-t) + \mathbb{I}_{\{x<t\}}\left(\int_0^L g(s)r_1(s)\,ds\right).$$

It is easy to see that this implies that r_1 is a.s. equal to the constant $\int_0^L g(s)r_1(s)ds$, and with some abuse of notation, we denote this constant again by r_1. In turn, since both $\int_{[0,L)} g(x)dx = 1$ and $\int_{[0,L)} \bar{G}(x)dx = 1$, this implies that

$$\langle h, \nu_1^* \rangle = r_1 \int_{[0,L)} h(x)\bar{G}(x)dx = r_1 = r_1 \int_{[0,L)} \bar{G}(x)dx = s_1^*, \tag{2.5.14}$$

which proves (2.5.12).

Similarly, for $\ell \geq 2$, setting $\nu_\ell(0) = \nu_\ell^*$ in (2.2.6), and using the identity $\nu_\ell(t) = \nu_\ell^*$ for any t along with (2.3.5), (2.5.1) and (2.5.2), we have for every $f \in \mathbb{C}_b[0,\infty)$, and

$t \geq 0$,

$$\int_0^L f(x)\bar{G}(x)r_\ell(x)dx = \int_0^L f(x+t)\bar{G}(x+t)r_\ell(x)dx + \langle h, \nu^*_{\ell+1}\rangle \int_0^t f(s)\bar{G}(s)ds \\ + \lambda\mathfrak{P}_d\Big(s^*_{\ell-1}, s^*_\ell\Big) \int_0^t \left(\int_0^L f(x+s)\bar{G}(x+s)\left(r_{\ell-1}(x) - r_\ell(x)\right)dx\right)ds. \tag{2.5.15}$$

Since r_ℓ is integrable with respect to $\bar{G}(x)dx$ by definition, we can use Fubini's theorem to change the order of integration in the last term on the right-hand side of (2.5.15), and apply the identity $\bar{G}(x) = 0$ for $x > L$, to obtain

$$\begin{aligned}&\int_0^t \left(\int_{s\wedge L}^L f(x)\bar{G}(x)\left(r_{\ell-1}(x-s) - r_\ell(x-s)\right)dx\right)ds \\ &= \int_0^{t\wedge L} f(x)\bar{G}(x)\left(\int_0^x \left(r_{\ell-1}(x-s) - r_\ell(x-s)\right)ds\right)dx \\ &\quad + \int_{t\wedge L}^L f(x)\bar{G}(x)\left(\int_0^t \left(r_{\ell-1}(x-s) - r_\ell(x-s)\right)ds\right)dx,\end{aligned}$$

where we use the notation $a \wedge b$ to denote the minimum of a and b. Thus, we can rewrite (2.5.15) as

$$\int_0^L f(x)\bar{G}(x)\left[r_\ell(x) - K^{(1)}_\ell(x,t)\mathbb{I}_{\{x\geq t\wedge L\}} - K^{(2)}_\ell(x)\mathbb{I}_{\{x<t\wedge L\}}\right]dx = 0, \tag{2.5.16}$$

where, for $(x,t) \in [0, L) \times [0, \infty)$,

$$\begin{aligned}K^{(1)}_\ell(x,t) &:= r_\ell(x-t) + \lambda\mathfrak{P}_d\Big(s^*_{\ell-1}, s^*_\ell\Big)\int_0^t \left(r_{\ell-1}(x-s) - r_\ell(x-s)\right)ds, \\ K^{(2)}_\ell(x) &:= \langle h, \nu^*_{\ell+1}\rangle + \lambda\mathfrak{P}_d\Big(s^*_{\ell-1}, s^*_\ell\Big)\int_0^x \left(r_{\ell-1}(u) - r_\ell(u)\right)du.\end{aligned}$$

Now, (2.5.16) clearly holds if and only if for every $t \geq 0$, $r_\ell(x) = K^{(1)}_\ell(x,t)\mathbb{I}_{\{x\geq t\wedge L\}} +$

$K_\ell^{(2)}(x)\mathbb{I}_{\{x<t\wedge L\}}$ for a.e. $x \in [0, L)$, which in turn holds if and only if for every $t \geq 0$,

$$r_\ell(x) = K_\ell^{(1)}(x,0) = K_\ell^{(1)}(x,t), \quad \text{for a.e.} \quad x \in [t\wedge L, L), \tag{2.5.17}$$

and

$$r_\ell(x) = K_\ell^{(2)}(x), \quad \text{for a.e.} \quad x < t \wedge L. \tag{2.5.18}$$

We will first consider the implications of the equality in (2.5.18). Define

$$I_\ell(x) := \int_0^x r_\ell(u)du,$$

and

$$B_\ell(x) := \Big[\langle h, \nu_\ell^*\rangle - \lambda\mathfrak{P}_d\Big(s_{\ell-1}^*, s_\ell^*\Big)(s_{\ell-1}^* - s_\ell^*)\Big] + \lambda\mathfrak{P}_d\Big(s_{\ell-1}^*, s_\ell^*\Big)I_{\ell-1}(x). \tag{2.5.19}$$

Note that B_ℓ is continuous on $[0, L)$, and using the definition of $K_\ell^{(2)}$ and (2.5.6), we can rewrite (2.5.18) as

$$\frac{dI_\ell}{dx}(x) + \lambda\mathfrak{P}_d\Big(s_{\ell-1}^*, s_\ell^*\Big)I_\ell(x) = B_\ell(x), \tag{2.5.20}$$

with boundary condition $I_\ell(0) = 0$. Solving this linear ordinary differential equation, we see that

$$I_\ell(x) = e^{-\lambda\mathfrak{P}_d(s_{\ell-1}^*, s_\ell^*)x}\int_0^x e^{\lambda\mathfrak{P}_d(s_{\ell-1}^*, s_\ell^*)u}B_\ell(u)du.$$

Combining this with (2.5.20) and recalling that $r_\ell = dI_\ell/dx$, we obtain

$$r_\ell(x) = B_\ell(x) - \lambda\mathfrak{P}_d\Big(s_{\ell-1}^*, s_\ell^*\Big)\int_0^x e^{-\lambda\mathfrak{P}_d(s_{\ell-1}^*, s_\ell^*)(x-u)}B_\ell(u)du, \tag{2.5.21}$$

which in particular shows that r_ℓ is continuous. Also, to further simplify the right-hand side, note that (2.5.19) shows that B_ℓ is differentiable with continuous derivative $\lambda\mathfrak{P}_d(s^*_{\ell-1}, s^*_\ell)r_{\ell-1}$. An integration by parts then yields

$$\begin{aligned}
&\lambda\mathfrak{P}_d(s^*_{\ell-1}, s^*_\ell)\int_0^x e^{\lambda\mathfrak{P}_d(s^*_{\ell-1},s^*_\ell)u}B_\ell(u)du \\
=&e^{\lambda\mathfrak{P}_d(s^*_{\ell-1},s^*_\ell)x}B_\ell(x) - B_\ell(0) - \lambda\mathfrak{P}_d(s^*_{\ell-1}, s^*_\ell)\int_0^x e^{\lambda\mathfrak{P}_d(s^*_{\ell-1},s^*_\ell)u}r_{\ell-1}(u)du.
\end{aligned}$$

Substituting this back into (2.5.21), and using (2.5.19), (2.2.2), (2.5.3) and (2.5.14) we obtain

$$\begin{aligned}
r_\ell(x) =&e^{-\lambda\mathfrak{P}_d(s^*_{\ell-1},s^*_\ell)x}\left[\langle h, \nu^*_\ell\rangle - \lambda\mathfrak{P}_d(s^*_{\ell-1}, s^*_\ell)(s^*_{\ell-1} - s^*_\ell)\right] \\
&+ \lambda e^{-\lambda\mathfrak{P}_d(s^*_{\ell-1},s^*_\ell)x}\mathfrak{P}_d(s^*_{\ell-1}, s^*_\ell)\int_0^x e^{\lambda\mathfrak{P}_d(s^*_{\ell-1},s^*_\ell)u}r_{\ell-1}(u)du \\
=&(\lambda(s^*_\ell)^d + s^*_1 - \lambda)e^{-\lambda\mathfrak{P}_d(s^*_{\ell-1},s^*_\ell)x} + \lambda\mathfrak{P}_d(s^*_{\ell-1}, s^*_\ell)\int_0^x e^{-\lambda\mathfrak{P}_d(s^*_{\ell-1},s^*_\ell)(x-u)}r_{\ell-1}(u)du,
\end{aligned}$$

which coincides with the identity in (2.5.13).

To complete the proof, it only remains to check that with this definition of r_ℓ, (2.5.17) is also satisfied, and that each r_ℓ is continuously differentiable. First, note that for any $t \geq 0$ and $x \in [t \wedge L, L)$, we have

$$\int_0^t r_\ell(x-s)\,ds = \int_{x-t}^x r_\ell(u)du = I_\ell(x) - I_\ell(x-t).$$

When combined with (2.5.20) both as is and with x replaced by $x-t$, and the relation $r_\ell = \frac{dI_\ell}{dx}$, this yields

$$\lambda\mathfrak{P}_d(s^*_{\ell-1}, s^*_\ell)\int_0^t r_\ell(x-s)\,ds = B_\ell(x) - B_\ell(x-t) - r_\ell(x) + r_\ell(x-t),$$

and when combined with (2.5.19), this implies

$$\begin{aligned}\lambda\mathfrak{P}_d(s^*_{\ell-1}, s^*_\ell)\int_0^t r_{\ell-1}(x-s)ds &= \lambda\mathfrak{P}_d(s^*_{\ell-1}, s^*_\ell)(I_{\ell-1}(x) - I_{\ell-1}(x-t)) \\ &= B_\ell(x) - B_\ell(x-t).\end{aligned}$$

Together, the last two displays imply that

$$\lambda\mathfrak{P}_d(s^*_{\ell-1}, s^*_\ell)\int_0^t (r_{\ell-1}(x-s) - r_\ell(x-s))\, ds = r_\ell(x) - r_\ell(x-t),$$

which, given the definition of $K^{(1)}_\ell$, proves that (2.5.17) is also satisfied.

Since r_1 is a constant function, it is clearly continuously differentiable and $r_1'(x) = 0$ for all $x \in [0, L)$. Given that $r_{\ell-1}$ is continuously differentiable on $[0, L)$, the continuous differentiability of r_ℓ on $[0, L)$ is an immediate consequence of (2.5.13). The continuously differentiable property of $(r_\ell)_{\ell\in\mathbb{N}}$ on $[0, L)$ then follows from the principle of mathematical induction.

□

2.5.3 A criterion for the existence of a physical invariant state

In this section, we use the characterization obtained in Proposition 2.5.4 to show that an invariant state always exists (see Proposition 2.5.7). The proof of existence relies on a few preliminary results. The first is a monotonicity property of the sequence of functions $(r_\ell)_{\ell\in\mathbb{N}}$ that satisfy (2.5.13).

Lemma 2.5.5. *Given an invariant state $(\nu^*_\ell)_{\ell\in\mathbb{N}}$, let $(r_\ell)_{\ell\in\mathbb{N}}$ be the associated measurable functions on $[0, \infty)$ that are continuously differentiable on $[0, L)$, as described* *in Proposition 2.5.4. Then, for any* $\ell \geq 1$,

$$r_{\ell+1}(x) \leq r_\ell(x) \qquad \textit{and} \qquad r'_\ell(x) \geq 0, \qquad \textit{for } x \in (0, L)$$

Proof. The facts that ν^* lies in $\mathbb{S}$ and ν^* satisfies relation (2.3.5), which was established in Proposition 2.5.4, imply that $\int_A r_{\ell+1}(x)\bar{G}(x)dx \leq \int_A r_\ell(x)\bar{G}(x)dx$ for all $A \in \mathcal{B}[0, \infty)$ and $\ell \geq 1$. Since, by Assumption 2.2.1 and Proposition 2.5.4, $r_{\ell+1}\bar{G}$ and $r_\ell\bar{G}$ are continuously differentiable on $[0, L)$, (a straightforward generalization of) Proposition 2.23 of [18] shows that $r_{\ell+1} \leq r_\ell$ on $[0, L)$, $\ell \geq 1$.

To show $r'_\ell(x) \geq 0$, again using the differentiability of r_ℓ, and differentiating both sides of (2.5.13), we obtain that for $\ell \geq 2$ and $x \in (0, L)$,

$$\begin{aligned} r'_\ell(x) = & -\lambda(\lambda(s^*_\ell)^d + s^*_1 - \lambda)\mathfrak{P}_d(s^*_{\ell-1}, s^*_\ell)e^{-\lambda\mathfrak{P}_d(s^*_{\ell-1}, s^*_\ell)x} \\ & - (\lambda\mathfrak{P}_d(s^*_{\ell-1}, s^*_\ell))^2 \int_0^x e^{-\lambda\mathfrak{P}_d(s^*_{\ell-1}, s^*_\ell)(x-u)} r_{\ell-1}(u)du + \lambda\mathfrak{P}_d(s^*_{\ell-1}, s^*_\ell)r_{\ell-1}(x). \end{aligned}$$

Using (2.5.13) to replace the second term on the right-hand side above, it follows that for $\ell \geq 2$ and $x \in (0, L)$,

$$\begin{aligned} r'_\ell(x) = & -\lambda(\lambda(s^*_\ell)^d + s^*_1 - \lambda)\mathfrak{P}_d(s^*_{\ell-1}, s^*_\ell)e^{-\lambda\mathfrak{P}_d(s^*_{\ell-1}, s^*_\ell)x} \\ & + \lambda\mathfrak{P}_d(s^*_{\ell-1}, s^*_\ell)(\lambda(s^*_\ell)^d + s^*_1 - \lambda)e^{-\lambda\mathfrak{P}_d(s^*_{\ell-1}, s^*_\ell)x} \\ & - \lambda\mathfrak{P}_d(s^*_{\ell-1}, s^*_\ell)r_\ell(x) + \lambda\mathfrak{P}_d(s^*_{\ell-1}, s^*_\ell)r_{\ell-1}(x) \\ = & \lambda\mathfrak{P}_d(s^*_{\ell-1}, s^*_\ell)(r_{\ell-1}(x) - r_\ell(x)). \end{aligned}$$

Using the identity $r_{\ell-1} \geq r_\ell$ on $[0, L)$ proved above, this implies $r'_\ell \geq 0$ on $(0, L)$ for $\ell \geq 2$, as desired. For $\ell = 1$, $r_1(x) = s^*_1$ for all $x \in [0.L)$ and so $r'_1(x) = 0$ for all $x \in (0, L)$ which completes the proof. $\square$

We now consider physical invariant states, namely those that satisfy condition (2.3.2).

Lemma 2.5.6. *Let $\nu^* = (\nu^*_\ell)_{\ell\in\mathbb{N}}$ be a physical invariant state of the hydrodynamic equations, that is, for which the associated $(s^*_\ell)_{\ell\in\mathbb{N}}$ satisfies $\lim_{\ell\to\infty} s^*_\ell = 0$, with s^*_ℓ as in (2.3.1). Then $\langle h, \nu^*_1\rangle = s^*_1 = \lambda$.*

Proof. Fix an invariant state, and let $(r_\ell)_{\ell\in\mathbb{N}}$ be the associated sequence of functions described in Proposition 2.5.4. Since (2.3.5) and (2.5.12) imply $\langle h, \nu^*_1\rangle = s^*_1 \int_{[0,L)} g(x)dx = s^*_1$, it remains to show that $\langle h, \nu^*_1\rangle = \lambda$. By Lemma 2.5.5 for every $x \in [0, L)$, $r_\ell(x)$ is monotonically decreasing in ℓ, and is bounded below by 0. Together with (2.5.12), this shows $r_\ell(x) \le r_1 = s^*_1$, for all $x \in [0, L)$, $\ell \in \mathbb{N}$, and the limit $r_\infty(x) = \lim_{\ell\to\infty} r_\ell(x)$ exists for all $x \in [0, L)$. Since $\bar{G}$ is integrable, the dominated convergence theorem implies $\lim_{\ell\to\infty} \int_0^L r_\ell(x)\bar{G}(x)dx = \int_0^L r_\infty(x)\bar{G}(x)dx$. On the other hand, (2.3.4) and the tail condition $\lim_{\ell\to\infty} s^*_\ell = 0$ of (2.3.2) show that $\lim_{\ell\to\infty} \int_0^L r_\ell(x)\bar{G}(x)dx = \lim_{\ell\to\infty} s^*_\ell = 0$. When combined, these limits imply $\int_0^L r_\infty(x)\bar{G}(x)dx = 0$. Since $\bar{G}(x) > 0$ and $r_\infty(x) \ge 0$ for $x \in [0, L)$, this implies $r_\infty = 0$ a.e.. Likewise, using $h = g/\bar{G}$, (2.3.5), the integrability of g, and the dominated convergence theorem, one has

$$\lim_{\ell\to\infty} \langle h, \nu^*_\ell\rangle = \lim_{\ell\to\infty} \int_0^L r_\ell(x)g(x)dx = \int_0^L r_\infty(x)g(x)dx = 0.$$

Thus, taking the limit in equation (2.5.3) as $\ell \to \infty$, and using the fact that $s^*_\ell \to 0$, we see that

$$\lim_{\ell\to\infty} \langle h, \nu^*_\ell\rangle = -\lambda + \langle h, \nu^*_1\rangle.$$

Comparing the last two equations, we conclude that $\langle h, \nu^*_1\rangle = \lambda$. ∎

Let $\mathbb{C}_b^1[0,\infty)$ denote the space of bounded functions on $[0,\infty)$ that are continuously differentiable on $[0,L)$. Fix $d \geq 2$, $\lambda \in (0,1]$ and consider the associated map $\bar{F} = F^{(d)} : \mathbb{C}_b^1[0,\infty) \times (0,1) \mapsto \mathbb{R}_+$ defined as follows: for $r \in \mathbb{C}_b^1[0,\infty)$ define

$$\bar{F}(r,s) = \lambda s^d \int_0^L e^{-\lambda \mathfrak{P}_d(\bar{s},s)x} \bar{G}(x)dx + \lambda \mathfrak{P}_d\Big(\bar{s},s\Big) \int_0^L \left(\int_0^x e^{-\lambda \mathfrak{P}_d(\bar{s},s)(x-u)} r(u)du \right) \bar{G}(x)dx, \tag{2.5.22}$$

where $\bar{s} = \bar{s}(r)$ is defined by

$$\bar{s}(r) := \int_0^L r(x)\bar{G}(x)dx, \tag{2.5.23}$$

Then the following criterion for existence of a physical invariant state is an immediate consequence of Proposition 2.5.4 and Lemma 2.5.6.

Proposition 2.5.7. *Suppose that whenever either $r \equiv \lambda$ or when $r \in \mathbb{C}_b^1[0,\infty)$ is non-constant and*

$$r(0) = \lambda \left(\int_0^L r(x)\bar{G}(x)dx \right)^d, \tag{2.5.24}$$

then $s \mapsto \bar{F}(r,s)$ has at least one fixed point on $(0,\bar{s}(r))$. Then there exists a physical invariant state for the hydrodynamic equations with arrival rate λ.

Proof. Assume the supposition of the proposition holds. By Proposition 2.5.4 and Lemma 2.5.6, any physical invariant state $(\nu_\ell^*)_{\ell\in\mathbb{N}}$ and the corresponding physical invariant queue length distribution must satisfy $s_1^* = \lambda$ and $r_1 \equiv \lambda$. On the other hand, (2.5.3) and Lemma 2.5.6 imply that $\langle h, \nu_{\ell+1}^* \rangle = \lambda (s_\ell^*)^d$. When substituted into (2.5.13) and (2.3.5), this implies (2.3.3) and (2.3.6) must hold. Thus, to show existence of an invariant state it suffices to find sequences of continuously differentiable functions $(r_\ell)_{\ell\geq 2}$ and positive constants $(s_\ell^*)_{\ell\geq 2}$ that satisfy (2.3.3) and (2.3.4). For $\ell = 2$, define $F_2(s) = \bar{F}(r,s)$, with $r \equiv s_1^* = \lambda$, and let s_2^* be a fixed point of F_2 in $(0,\lambda]$, which exists by the claim, and let r_2 be as defined in terms of s_2^*

via (2.3.3). Since $r_1 \equiv \lambda$ and the exponential function is non-negative and continuously differentiable, it is easy to see that r_2 is non-negative and $r_2 \in \mathbb{C}_b^1[0,\infty)$. Then s_2^* clearly satisfies (2.3.4) due to (2.5.23), and there is a unique s_2^* satisfying (2.3.4) with r_2 satisfying (2.3.3) if and only if the fixed point of F_2 is unique. Now, suppose that for some $j \geq 2$, there exist non-negative, bounded continuously differentiable functions $(r_\ell)_{2\leq\ell\leq j}$ and constants $(s_\ell^*)_{2\leq\ell\leq j}$ that satisfy (2.3.3) and (2.3.4) for $2 \leq \ell \leq j$. Then (2.3.3) and (2.3.4) imply that $r = r_j$ satisfies (2.5.24). Therefore, setting $F_{j+1}(\cdot) = \bar{F}(r_j, \cdot)$, the claim above shows that F_{j+1} has a fixed point that we denote by s_{j+1}^*, and let r_{j+1} be as defined in (2.3.3), and note that then r_{j+1} is non-negative and $r_{j+1} \in \mathbb{C}_b^1[0,\infty)$ by definition. Also note that (2.3.4) holds on account of (2.5.23). Thus, we have constructed $(r_\ell)_{2\leq\ell\leq j+1}$ and $(s_\ell^*)_{2\leq\ell\leq j+1}$ that satisfy (2.3.3) and (2.3.4) for all $\ell \leq j+1$, and the assertion of the proposition follows by induction. □

2.5.4 Proof of Theorem 2.3.3

Proof of Theorem 2.3.3. To establish Theorem 2.3.3 it suffices to verify the supposition of Proposition 2.5.7. To see why the latter is true, fix r that satisfies the stated conditions, and define $J(s) := \bar{F}(r,s) - s$. Note that then we need to show that J has at least one zero on $(0, \bar{s}]$, for $\bar{s} = \int_0^L r(x)\bar{G}(x)dx$. Note that by (2.2.2), $\mathfrak{P}_d(\bar{s}, s) > 0$ for all $s \in [0,1]$ whenever $\bar{s} > 0$. Since r is not identically zero and $\bar{G}$ is strictly positive on $[0, L)$, (2.5.22) implies that $\bar{F}(r, 0) > 0$ and hence, $J(0) > 0$. Since J is continuous, by the intermediate value theorem, to show J has a zero on $(0, \bar{s}]$, it suffices to show that $J(\bar{s}) \leq 0$. Note that a simple integration by parts on the interior of the second term on the right-hand side of (2.5.22) yields

$$\bar{F}(r,s) = \int_0^L r(x)\bar{G}(x)dx - \left[r(0) - \lambda s^d\right] \int_0^L e^{-\lambda \mathfrak{P}_d(\bar{s},s)x}\bar{G}(x)dx$$
$$- \int_0^L \left(\int_0^x e^{-\lambda \mathfrak{P}_d(\bar{s},s)(x-u)} r'(u)du \right) \bar{G}(x)dx. \tag{2.5.25}$$

If $r \equiv \lambda$, then by (2.5.23), $\bar{s} = \lambda$ and (2.5.25) shows that $J(\bar{s}) = \bar{F}(r,\bar{s}) - \bar{s} = -\lambda(1-\lambda^d)\int_0^L e^{-\lambda\mathfrak{P}_d(\lambda,\lambda)x}\bar{G}(x)dx$. Thus, $J(\bar{s}) \leq 0$ (with equality holding only if $\lambda = 1$) and so J has a zero on $(0,\bar{s})$ if $\lambda < 1$ and at $\bar{s}$ if $\lambda = 1$. On the other hand, if r satisfies (2.5.24), then (2.5.25) and (2.5.23) show that

$$J(\bar{s}) = -\int_0^L \left(\int_0^x e^{-\lambda \mathfrak{P}_d(\bar{s},s)(x-u)} r'(u)du \right) \bar{G}(x)dx < 0,$$

where the last inequality also uses the fact that r is non-constant. This proves the first assertion of Theorem 2.3.3. Also, in view of (2.3.5), the right-hand side of (2.3.6) is equal to $\langle h, \nu_\ell^* \rangle$. Therefore, (2.3.5) follows from Proposition 2.5.4 and Lemma 2.5.6. This completes the proof of Theorem 2.3.3. □

2.6 Tail Decay for the Invariant State

Throughout this section we fix $d \geq 2$, $\lambda \in (0,1)$, suppose Assumptions 2.2.1-2.3.8 hold and let $\beta > d/(d-1)$, $x_0 > 1/\lambda d$ and $C_0 < \infty$ be as in Assumption 2.3.6. Also, let $(s_\ell^*)_{\ell\in\mathbb{N}}$ be the queue length distribution associated with a physical invariant state $(\nu_\ell^*)_{\ell\in\mathbb{N}}$ of the hydrodynamic equations. Note that the analysis here does not explicitly require uniqueness of the physical invariant state, and only relies on the characterization stated in Theorem 2.3.3 (and proved in Section 2.5).

2.6.1 A reduction

In this section we show that the proof of Theorem 2.3.9 can be reduced to establishing certain key estimates stated in Proposition 2.6.1 below. Throughout this section, given $m \in \mathbb{N}$, $n \in \mathbb{N}$ and $x_i \in \mathbb{R}$ for all i, we will use the convention that

$$\prod_{i=m}^{n} x_i = 1 \quad \text{ if } m > n.$$

We will also use the convention $s_i^* = 1$ for all $i \leq 0$.

Proposition 2.6.1. *Suppose Assumptions 2.2.1 - 2.3.6 hold. Let $(s_\ell^*)_{\ell\in\mathbb{N}}$ be the queue length distribution associated with a physical invariant state of the hydrodynamic equations. If β is not an integer, then there exists $1 \leq C_\beta < \infty$ such that for all $\ell \geq 2$,*

$$s_\ell^* \leq \lambda(3C_\beta)^{\ell-1}\left(\prod_{i=\ell-\lfloor\beta\rfloor+1}^{\ell-1}(s_i^*)^{d-1}\right)(s^*_{\ell-\lfloor\beta\rfloor})^{\hat{\beta}(d-1)}. \tag{2.6.1}$$

If β is an integer, then there exists $\delta \in (0,1)$, and $1 \leq C_\beta < \infty$ such that for all $\ell \geq 2$,

$$s_\ell^* \leq \lambda(3C_\beta)^{\ell-1}\left(\prod_{i=\ell-\beta+2}^{\ell-1}(s_i^*)^{d-1}\right)(s^*_{\ell-\beta+1})^{(1-\delta)(d-1)}. \tag{2.6.2}$$

The proof of Proposition 2.6.1, which involves a careful analysis of the equations characterizing the invariant state, is lengthy and deferred to Section 2.6.3. We now use this result to prove Theorem 2.3.9 in two steps. First, in Lemma 2.6.2, we use the assumption on the tail decay in 2.3.8 to obtain a technical result. We then use Lemma 2.6.2 with a result on inhomogeneous recursions (Lemma 2.6.3 below) to show that the decay rate is at least doubly exponential.

Lemma 2.6.2. *Suppose Assumptions 2.2.1 – 2.3.8 hold. Let $(s_\ell^*)_{\ell\in\mathbb{N}}$ be the queue length distribution associated with a physical invariant state of the hydrodynamic* *equations and let $C_\beta \in [1,\infty)$ and $\delta \in (0,1)$ be as in Proposition 2.6.1. If β is not an integer, then there exists $M(\beta) < \infty$ such that for all $\ell \geq M(\beta)$,*

$$\lambda(3C_\beta)^{\ell-1}\left(\prod_{i=\ell-\lfloor\beta\rfloor+1}^{\ell-1}(s_i^*)^{d-1}\right)(s_{\ell-\lfloor\beta\rfloor}^*)^{\hat{\beta}(d-1)} < 1. \tag{2.6.3}$$

If β is an integer, then there exists $M(\beta) < \infty$ such that for all $\ell \geq M(\beta)$,

$$\lambda(3C_\beta)^{\ell-1}\left(\prod_{i=\ell-\beta+2}^{\ell-1}(s_i^*)^{d-1}\right)(s_{\ell-\beta+1}^*)^{(1-\delta)(d-1)} < 1. \tag{2.6.4}$$

Proof. We consider two cases:

Case 1: β is not an integer. Define $c = k(d-1)(\beta-1) - \log(3C_\beta)$. Note that $c > 0$ by Assumption 2.3.8. Using the fact that $i \to s_i^*$ is non-decreasing, this implies

$$\begin{aligned}
&\lambda(3C_\beta)^{\ell-1}\left(\prod_{i=\ell-\lfloor\beta\rfloor+1}^{\ell-1}(s_i^*)^{d-1}\right)(s_{\ell-\lfloor\beta\rfloor}^*)^{\hat{\beta}(d-1)}\\
\leq&\lambda(3C_\beta)^{\ell-1}(s_{\ell-\lfloor\beta\rfloor}^*)^{(\beta-1)(d-1)}\\
\leq&\lambda\frac{c_k^{(\beta-1)(d-1)}}{3C_\beta}e^{k\lfloor\beta\rfloor(\beta-1)(d-1)}e^{\ell(\log(3C_\beta)-k(\beta-1)(d-1))}\\
=.&\lambda\frac{c_k^{(\beta-1)(d-1)}}{3C_\beta}e^{k\lfloor\beta\rfloor(\beta-1)(d-1)}e^{-c\ell}.
\end{aligned}$$

Since $\lim_{\ell\to\infty} e^{-c\ell} = 0$, there exists $M(\beta) < \infty$ such that (2.6.3) holds for all $\ell \geq M(\beta)$.

Case 2: β is an integer. In this case the proof of (2.6.4) follows the same argument as in the non-integral case along with the observation that $\beta - 1 - \delta > \beta - 2$ for any $\delta \in (0,1)$. $\square$

Lemma 2.6.3. *Suppose that c_1, c_2 and $(R_\ell)_{\ell\in\mathbb{Z}}$ are positive numbers. For every*

$\eta \in (0,1)$ and $j \in \mathbb{N}$ such that $(d-1)(j+\eta-1) > 1$, and

$$R_\ell = c_1 - (\ell-1)c_2 + (d-1)\left(\sum_{i=\ell-j+1}^{\ell-1} R_i + \eta R_{\ell-j}\right) \quad \textit{for } \ell \geq 1, \tag{2.6.5}$$

there exists $n_d = n_d(j,\eta) \in (0,\infty)$ such that

$$\lim_{\ell\to\infty} \frac{1}{\ell} \log_d R_\ell = n_d.$$

Proof. Subtracting $R_{\ell-1}$ from R_ℓ and using (2.6.5), we see that for $\ell \geq 2$,

$$R_\ell = -c_2 + R_{\ell-1} + (d-1)\left(R_{\ell-1} - (1-\eta)R_{\ell-j} - \eta R_{\ell-j-1}\right).$$

Subtracting $R_{\ell-1}$ from R_ℓ again, using the above equation, and rearranging, we have for $\ell \geq 3$,

$$R_\ell = (d+1)R_{\ell-1} - dR_{\ell-2} + (d-1)\left(-(1-\eta)R_{\ell-j} + (1-2\eta)R_{\ell-j-1} + \eta R_{\ell-j-2}\right).$$

Define $F = F_{j,\eta} : \mathbb{R} \mapsto \mathbb{R}$ by

$$F(x) := 1 - (d+1)x + dx^2 - (d-1)(-(1-\eta)x^j + (1-2\eta)x^{j+1} + \eta x^{j+2}). \tag{2.6.6}$$

Noting that 0 is not a root of F, let $1/\gamma_i$ be the k distinct roots of F, with $k \leq j+2$, and let d_i denotes the multiplicity of the root $1/\gamma_i$. Then $F(x) = \sum_{i=1}^k (1-\gamma_i x)^{d_i}$. Since $j \geq 1$ and $\eta(d-1) \neq 0$, by Theorem 4.1.1 of [41],

$$R_\ell = \sum_{i=1}^k \phi_i(\ell)\gamma_i^\ell, \qquad \forall \ell \geq 0, \tag{2.6.7}$$

where $\phi_i(\ell)$ is a polynomial in ℓ of degree at least d_i.

A direct inspection of (2.6.6) shows that $F(1)=0$, the root $\gamma_1=1$ has degree $d_1=2$, and $F(x)=(1-x)^2P(x)$, where

$$P(x)=\left(1-(d-1)\sum_{i=1}^{j-1}x^i-(d-1)\eta x^j\right).$$

By Descartes' rule of signs, there exists exactly one positive root of $P(\cdot)$. Let us denote it by $1/\gamma_2$. Note that $P(0)=1$ and $P(1)<0$ since $(d-1)(j+\eta-1)>1$ by assumption. Thus, $1/\gamma_2\in(0,1)$ or equivalently $\gamma_2>1$. Using Rouche's theorem (see [36, Exercise 237, page 321]) it is easy to show that $P(\cdot)$ has no roots other than $1/\gamma_2$ in the disc $\{z:|z|\leq 1/\gamma_2\}$. Hence, $|\gamma_i|<\gamma_2$ for $i\geq 3$. In view of the representation for R_ℓ in (2.6.7), this implies

$$\lim_{\ell\to\infty}\frac{\log_d R_\ell}{\ell}=\log_d\gamma_2.$$

Setting $n_d=\log_d\gamma_2$ proves the claim. □

Proof of Theorem 2.3.9. Fix a physical invariant state $(\nu_\ell^*)_{\ell\in\mathbb{N}}$ of the hydrodynamic equations, and let $(s_\ell^*)_{\ell\in\mathbb{N}}$ be the associated queue length distribution. We consider two cases:

Case 1: β is not an integer. Let $C_\beta\in[0,\infty)$ be as in Proposition 2.6.1. In this case, using the monotonicity of the logarithm function along with (2.6.1), we see that for $\ell\geq 2$

$$\log\left(\frac{1}{s_\ell^*}\right)\geq-\log\lambda-(\ell-1)\log(3C_\beta)+(d-1)\left(\sum_{i=\ell-\lfloor\beta\rfloor+1}^{\ell-1}\log\left(\frac{1}{s_i^*}\right)+\hat{\beta}\log\left(\frac{1}{s_{\ell-\lfloor\beta\rfloor}^*}\right)\right).$$

Let $M=M(\beta)<\infty$ where $M(\beta)$ is as in (2.6.3) of Lemma 2.6.2. Set $Q_i:=-\log(s_i^*)$

for $\{M-\lfloor\beta\rfloor,\cdots,M-1\}$, and recursively define

$$Q_\ell := -\log\lambda - (\ell-1)\log(3C_\beta) + (d-1)\left(\sum_{i=\ell-\lfloor\beta\rfloor+1}^{\ell-1} Q_i + \hat{\beta}Q_{\ell-\lfloor\beta\rfloor}\right), \qquad \forall \ell \geq M.$$

Then, applying Lemma 2.6.3 with $j=\lfloor\beta\rfloor$, $\eta=\hat{\beta}$ (noting that $(d-1)(\beta-1)>1$), $c_1=-\log\lambda>0$, $c_2=\log(3C_\beta)>0$ and $R_\ell := Q_{M+\ell-1}$ for $\ell\geq-\lfloor\beta\rfloor+1$, we conclude there exists $n_d=n_d(\beta,\delta)\in(0,\infty)$ such that

$$\lim_{\ell\to\infty}\frac{\log Q_\ell}{\ell} = n_d.$$

Since $\log(1/s_i^*)\geq Q_i$ for all $i\geq M$, this proves (2.3.13), in this case.

Case 2: β is an integer. Let $C_\beta\in[0,\infty)$ and $\delta\in(0,1)$ be as in Proposition 2.6.1. Then (2.6.2) implies that for every $\ell\geq 2$,

$$\log\left(\frac{1}{s_\ell^*}\right) \geq -\log\lambda - (\ell-1)\log(3C_\beta) + (d-1)\left(\sum_{i=\ell-\beta+2}^{\ell-1}\log\left(\frac{1}{s_i^*}\right) + (1-\delta)\log\left(\frac{1}{s_{\ell-\beta+1}^*}\right)\right).$$

Let $M=M(\beta)<\infty$ where $M(\beta)$ is as in (2.6.4) of Lemma 2.6.2. Now, set $Q_i := -\log(s_i^*)$ for $i\in\{M-\beta+1,\cdots,M-1\}$ and recursively define

$$Q_\ell = -\log\lambda - (\ell-1)\log(3C_\beta) + (d-1)\left(\sum_{i=\ell-\beta+2}^{\ell-1} Q_i + (1-\delta)Q_{\ell-\beta+1}\right), \qquad \forall\ell\geq M.$$

Since β is an integer, $(d-1)(\beta-1)>1$ implies $\beta\geq 2$. Since $(1-\delta)\in(0,1)$, $\beta-1\geq 1$, $d\geq 2$, we have $(d-1)(\beta-1-\delta)>1$. Thus, applying Lemma 2.6.3 with $j=\beta-1$, $\eta=1-\delta$ and $R_\ell=Q_{M+\ell-1}$ for $\ell\geq-\beta$, we conclude there exists $n_d=n_d(\beta,\delta)\in(0,\infty)$ such that

$$\lim_{\ell\to\infty}\frac{\log Q_\ell}{\ell} = n_d.$$

Since $\log(1/s_i^*) \geq Q_i$ for all $i \geq M$, this proves (2.3.13), for Case 2 as well. □

Remark 2.6.4. The estimates for $(s_\ell^*)_{\ell\in\mathbb{N}}$ in (2.6.1) and (2.6.2) are analogous to the estimates for corresponding probabilities obtained in (3.5) and (3.6) of Proposition 3.2 in [10]. One may expect that the proof of the tail decay property in Theorem 2.3.9 could therefore be deduced from Proposition 2.6.1 by simply referring to the argument used in [10] to deduce their tail decay result (in Theorem 1.1 of [10]) from Proposition 3.2 therein, which only involves comparison with a homogeneous linear recursion (see Proposition 3.3 of [10]). However, we could not quite resolve the argument in [10], and, under the *a priori exponential tail decay assumption on the invariant queue length distribution, instead provide a self-contained proof that entails a two-step argument presented above, that involves comparison with the slightly more complicated inhomogeneous recursion analyzed in Lemma 2.6.3.*

2.6.2 Preliminary Estimates on the Density of the Physical Invariant State

In this section we obtain preliminary estimates that are used to prove Proposition 2.6.1 in Section 2.6.3.

Lemma 2.6.5. *Suppose Assumptions 2.2.1 and 2.2.4 hold. For all $x \in [0,\infty)$, the inequality*

$$r_2(x) \leq \lambda(s_2^*)^d + \lambda(1 - e^{-\lambda d(s_1^*)^{d-1}x}), \tag{2.6.8}$$

holds, and for $\ell \geq 3$,

$$r_\ell(x) \leq \lambda(s_\ell^*)^d + \lambda(s_{\ell-1}^*)^d + \lambda\sum_{j=2}^{\ell-2}(s_j^*)^d\prod_{i=j+1}^{\ell-1}(1 - e^{-\lambda d(s_i^*)^{d-1}x}) + \lambda\prod_{i=1}^{\ell-1}(1 - e^{-\lambda d(s_i^*)^{d-1}x}), \tag{2.6.9}$$

where the third term on the right-hand side of (2.6.9) *is treated as zero if* $\ell = 3$.

Proof. Since $(s^*_\ell)_{\ell \in \mathbb{N}}$ is the queue length distribution associated with a physical invariant state of the hydrodynamic equations, by Theorem 2.3.3, $r_1 = s^*_1 = \lambda$, and by convention $s^*_0 = 1$. Hence, (2.6.8) follows from (2.3.3) with $\ell = 2$, and the inequality $\mathfrak{P}_d(s^*_1, s^*_2) \le d(s^*_1)^{d-1}$, which holds by (2.2.3). For the proof of (2.6.9) we will make repeated use of the fact that the function $x \mapsto 1 - e^{-cx}$ is increasing for any $c > 0$. Using (2.3.3) again, but now with $\ell = 3$, and using the inequality (2.6.8), we obtain

$$\begin{aligned}
r_3(x) \le & \lambda(s^*_3)^d e^{-\lambda \mathfrak{P}_d(s^*_2, s^*_3)x} + \lambda^2 (s^*_2)^d \mathfrak{P}_d(s^*_2, s^*_3) \int_0^x e^{-\lambda \mathfrak{P}_d(s^*_2,s^*_3)(x-u)} du \\
& + \lambda^2 \mathfrak{P}_d(s^*_2, s^*_3) \int_0^x e^{-\lambda \mathfrak{P}_d(s^*_2,s^*_3)(x-u)} (1 - e^{-\lambda d (s^*_1)^{d-1} u}) du \\
\le & \lambda(s^*_3)^d + \lambda(s^*_2)^d (1 - e^{-\lambda \mathfrak{P}_d(s^*_2,s^*_3)x}) + \lambda^2 (1 - e^{-\lambda d(s^*_1)^{d-1} x}) \mathfrak{P}_d(s^*_2, s^*_3) \int_0^x e^{-\lambda \mathfrak{P}_d(s^*_2,s^*_3)(x-u)} du \\
\le & \lambda(s^*_3)^d + (\lambda(s^*_2)^d + \lambda(1 - e^{-\lambda d(s^*_1)^{d-1}x}))(1 - e^{-\lambda \mathfrak{P}_d(s^*_2,s^*_3)x}).
\end{aligned}$$

Together with the inequality $\mathfrak{P}_d(s^*_2, s^*_3) \le d(s^*_2)^{d-1}$ from (2.2.3), this implies,

$$r_3(x) \le \lambda(s^*_3)^d + \lambda(s^*_2)^d + \lambda(1 - e^{-\lambda d(s^*_1)^{d-1}x})(1 - e^{-\lambda d(s^*_2)^d x}),$$

which proves (2.6.9) for $\ell = 3$.

Now, suppose (2.6.9) holds for some $\ell \ge 3$. Then using (2.3.3) with ℓ replaced by $\ell + 1$ and (2.6.9), we have

$$\begin{aligned}
r_{\ell+1}(x) = & \lambda(s^*_{\ell+1})^d e^{-\lambda \mathfrak{P}(s^*_\ell, s^*_{\ell+1})x} + \lambda \mathfrak{P}_d(s^*_\ell, s^*_{\ell+1}) \int_0^x e^{-\lambda \mathfrak{P}(s^*_\ell, s^*_{\ell+1})(x-u)} r_\ell(u) du \\
\le & \lambda(s^*_{\ell+1})^d + \lambda^2 \left[(s^*_\ell)^d + (s^*_{\ell-1})^d \right] \mathfrak{P}_d(s^*_\ell, s^*_{\ell+1}) \int_0^x e^{-\lambda \mathfrak{P}_d(s^*_\ell, s^*_{\ell+1})(x-u)} du \\
& + \lambda^2 \sum_{j=2}^{\ell-2} \left[(s^*_j)^d \mathfrak{P}_d(s^*_\ell, s^*_{\ell+1}) \int_0^x e^{-\lambda \mathfrak{P}_d(s^*_\ell, s^*_{\ell+1})(x-u)} \prod_{i=j+1}^{\ell-1} (1 - e^{-\lambda d(s^*_i)^{d-1} u}) du \right]
\end{aligned}$$

$$
\begin{aligned}
&+ \lambda^2 \mathfrak{P}_d(s_\ell^*, s_{\ell+1}^*) \int_0^x e^{-\lambda \mathfrak{P}_d(s_\ell^*, s_{\ell+1}^*)(x-u)} \prod_{i=1}^{\ell-1} (1 - e^{-\lambda d (s_i^*)^{d-1} u}) du \\
\leq& \lambda (s_{\ell+1}^*)^d + \lambda (s_\ell^*)^d (1 - e^{-\lambda \mathfrak{P}_d(s_\ell^*, s_{\ell+1}^*) x}) + \lambda (s_{\ell-1}^*)^d (1 - e^{-\lambda \mathfrak{P}_d(s_\ell^*, s_{\ell+1}^*) x}) \\
&+ \lambda^2 \sum_{j=2}^{\ell-2} \left[(s_j^*)^d \prod_{i=j+1}^{\ell-1} (1 - e^{-\lambda d (s_i^*)^{d-1} x}) \mathfrak{P}_d(s_\ell^*, s_{\ell+1}^*) \int_0^x e^{-\lambda \mathfrak{P}_d(s_\ell^*, s_{\ell+1}^*)(x-u)} du \right] \\
&+ \lambda^2 \prod_{i=1}^{\ell-1} (1 - e^{-\lambda d (s_i^*)^{d-1} x}) \mathfrak{P}_d(s_\ell^*, s_{\ell+1}^*) \int_0^x e^{-\lambda \mathfrak{P}_d(s_\ell^*, s_{\ell+1}^*)(x-u)} du.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
r_{\ell+1}(x) \leq& \lambda (s_{\ell+1}^*)^d + \lambda (s_\ell^*)^d + \lambda (s_{\ell-1}^*)^d (1 - e^{-\lambda \mathfrak{P}_d(s_\ell^*, s_{\ell+1}^*) x}) \\
&+ \lambda \sum_{j=2}^{\ell-2} (s_j^*)^d \prod_{i=j+1}^{\ell-1} (1 - e^{-\lambda d (s_i^*)^{d-1} x}) (1 - e^{-\lambda \mathfrak{P}_d(s_\ell^*, s_{\ell+1}^*) x}) \\
&+ \lambda \prod_{i=1}^{\ell-1} (1 - e^{-\lambda d (s_i^*)^{d-1} x}) (1 - e^{-\lambda \mathfrak{P}_d(s_\ell^*, s_{\ell+1}^*) x}).
\end{aligned}
$$

Using the inequality $\mathfrak{P}_d(s_\ell^*, s_{\ell+1}^*) \leq d(s_\ell^*)^{d-1}$ from (2.2.3), we then obtain

$$
r_{\ell+1}(x) \leq \lambda (s_{\ell+1}^*)^d + \lambda (s_\ell^*)^d + \lambda \sum_{j=2}^{\ell-1} (s_j^*)^d \prod_{i=j+1}^{\ell} (1 - e^{-\lambda d (s_i^*)^{d-1} x}) + \lambda \prod_{i=1}^{\ell} (1 - e^{-\lambda d (s_i^*)^{d-1} x}),
$$

which shows that (2.6.9) holds with ℓ replaced by $\ell + 1$. Thus, by the principle of mathematical induction, (2.6.9) holds for all $\ell \geq 3$. □

We now state a result that is an immediate consequence of Lemma 2.6.5 and the elementary inequalities $s_\ell^* \leq s_{\ell-1}^*$ and $1 - e^{-x} \leq 1 \wedge x$ for $x \geq 0$.

Corollary 2.6.6. *Suppose Assumptions 2.2.1 and 2.2.4 hold. For $x \in [0, \infty)$,*

$$
r_2(x) \leq \lambda (s_1^*)^d + \lambda (1 \wedge \lambda d (s_1^*)^{d-1} x), \tag{2.6.10}
$$

and for $\ell \geq 3$,

$$r_\ell(x) \leq 2\lambda(s^*_{\ell-1})^d + \lambda\sum_{j=2}^{\ell-2}(s^*_j)^d \prod_{i=j+1}^{\ell-1}(1\wedge \lambda d(s^*_i)^{d-1}x) + \lambda\prod_{i=1}^{\ell-1}(1\wedge\lambda d(s^*_i)^{d-1}x), \tag{2.6.11}$$

where the second term on the right-hand side of (2.6.11) *is treated as zero if* $\ell = 3$.

We now make an observation that will be used repeatedly in the next section. Recall the definition of C from (2.3.9) and note that $C \in (0,\infty)$. Fix f to be a non-decreasing measurable function $f : [0,\infty) \mapsto [0,\infty)$. Then clearly

$$\frac{\int_0^{x_0} f(x)\bar{G}(x)dx}{\int_{x_0}^{\infty} f(x)\bar{G}(x)dx} \leq \frac{f(x_0)\int_0^{x_0}\bar{G}(x)dx}{f(x_0)\int_{x_0}^{\infty}\bar{G}(x)dx} = C.$$

Using the above inequality, definition of C_1 from (2.3.9), (2.3.8), and rescaling, we see that

$$\begin{aligned}\int_0^\infty f(x)\bar{G}(x)dx = \int_0^{x_0} f(x)\bar{G}(x)dx + \int_{x_0}^\infty f(x)\bar{G}(x)dx \leq& (C+1)\int_{x_0}^\infty f(x)\bar{G}(x)dx \\ \leq& (C+1)C_0\int_{x_0}^\infty f(x)x^{-\beta}dx \\ =& \frac{C_1}{x_0^{\beta-1}}\int_1^\infty f(x_0y)y^{-\beta}dy.\end{aligned} \tag{2.6.12}$$

Recall that $s^*_i = 1$ for all $i \leq 0$, and define

$$m_i := \frac{1}{\lambda d(s^*_i)^{d-1}x_0}, \qquad i \in \mathbb{Z}. \tag{2.6.13}$$

Fix $d \geq 2$, $\lambda < 1$ and let $x_0 > 1/\lambda d$, $\beta > d/(d-1)$ be as in Assumption 2.3.6, and let $C_\beta = C_\beta(\lambda,\beta,d)$ be as defined in (2.3.11). Then, recalling $s^*_\ell = \int_0^L r_\ell(x)\bar{G}(x)dx$ from (2.3.4), integrating the inequality (2.6.11) with respect to $\bar{G}(x)dx$ and using the identity $\int_0^\infty \bar{G}(x)dx = 1$, it follows that for $\ell \geq 3$

$$\begin{aligned} s^*_\ell \leq & 2\lambda(s^*_{\ell-1})^d + \int_0^\infty \lambda \sum_{j=2}^{\ell-2} (s^*_j)^d \prod_{i=j+1}^{\ell-1} (1 \wedge \lambda d(s^*_i)^{d-1}x)\bar{G}(x)dx \\ & + \int_0^\infty \lambda \prod_{i=1}^{\ell-1} (1 \wedge \lambda d(s^*_i)^{d-1}x)\bar{G}(x)dx. \end{aligned}$$

Applying (2.6.12) with $f(x) = \prod_{i=j+1}^{\ell-1}(1 \wedge \lambda d(s^*_i)^{d-1}x)$ for $j \in \{0, 2, \cdots, \ell-1\}$, and using (2.6.13), this implies

$$s^*_\ell \leq 2\lambda(s^*_{\ell-1})^d + \frac{\lambda C_1}{x_0^{\beta-1}} \sum_{j=2}^{\ell-2} (s^*_j)^d \psi_{\ell,j} + \frac{\lambda C_1}{x_0^{\beta-1}} \psi_{\ell,\ell-1}, \tag{2.6.14}$$

where , for $\ell \geq 2$,

$$\psi_{\ell,j} := \int_1^\infty \prod_{i=j+1}^{\ell-1} (1 \wedge m_i^{-1} y)dy, \quad j \in \mathbb{Z}. \tag{2.6.15}$$

The proof of Proposition 2.6.1 will proceed by bounding the right-hand side of (2.6.14) by (2.6.1) or (2.6.2), depending on whether β is or is not an integer, using an inductive argument and the estimates in the following lemma. In what follows, recall the convention that $s^*_i = 1$ for all $i \leq 0$.

Lemma 2.6.7. *Suppose $\ell \geq 2$ and $\psi_{\ell,j}$, $j \in \{0, 2, \cdots, \ell-2\}$ is as in (2.6.15). If $j \leq \ell - 2$ and $\beta \in (d/d-1, \infty) \setminus \mathbb{N}$, we have*

$$\frac{C_1}{x_0^{\beta-1}} \psi_{\ell,j} \leq \begin{cases} C_\beta \left(\prod_{i=\ell-\lfloor\beta\rfloor+1}^{\ell-1} (s^*_i)^{d-1} \right) (s^*_{\ell-\lfloor\beta\rfloor})^{\hat{\beta}(d-1)} & \text{if } j = \ell - \lfloor\beta\rfloor - 1, \\ C_\beta \left(\prod_{i=j+1}^{\ell-1} (s^*_i)^{d-1} \right) & \text{if } j \geq \ell - \lfloor\beta\rfloor, \end{cases} \tag{2.6.16}$$

whereas if $j \leq \ell - 2$ and $\beta \in (d/(d-1), \infty) \cap \mathbb{N}$, then there exists $\delta \in (0, 1)$ such *that*

$$\frac{C_1}{x_0^{\beta-1}}\psi_{\ell,j} \leq \begin{cases} C_\beta\left(\prod_{i=\ell-\beta+2}^{\ell-1}(s_i^*)^{d-1}\right)(s_{\ell-\beta+1}^*)^{(1-\delta)(d-1)} & \textit{if } j=\ell-\beta, \\ C_\beta\left(\prod_{i=j+1}^{\ell-1}(s_i^*)^{d-1}\right) & \textit{if } j\geq \ell-\beta+1. \end{cases} \tag{2.6.17}$$

Moreover, we also have

$$\frac{C_1}{x_0^{\beta-1}}\psi_{\ell,0} \leq \begin{cases} C_\beta\left(\prod_{i=\ell-\lfloor\beta\rfloor+1}^{\ell-1}(s_i^*)^{d-1}\right)(s_{\ell-\lfloor\beta\rfloor}^*)^{\hat{\beta}(d-1)} & \textit{if } \beta\in\left(\frac{d}{d-1},\infty\right)\setminus\mathbb{N}, \\ C_\beta\left(\prod_{i=\ell-\beta+2}^{\ell-1}(s_i^*)^{d-1}\right)(s_{\ell-\beta+1}^*)^{(1-\delta)(d-1)} & \textit{if } \beta\in\left(\frac{d}{d-1},\infty\right)\cap\mathbb{N}. \end{cases} \tag{2.6.18}$$

The proof of Lemma 2.6.7 is somewhat involved, and hence relegated to Section 2.6.4.

2.6.3 Proof of Proposition 2.6.1

First note that the convention $s_i^*=1$ for $i\leq 0$ and the assumption $\sum_{\ell\geq 1}s_\ell^*<\infty$ imply $s_1^*=\lambda$ by Theorem 2.3.3. Suppose β is not an integer. Using (2.3.4), (2.6.10) of Corollary 2.6.6, the fact that $\int_0^\infty \bar{G}(x)dx=1$ and (2.6.15), we have

$$s_2^* \leq \lambda(s_1^*)^d + \int_0^\infty \lambda(1\wedge\lambda d(s_1^*)^{d-1}x)\bar{G}(x)dx \leq \lambda(s_1^*)^d + \frac{\lambda C_1}{x_0^{\beta-1}}\psi_{2,0},$$

where the second inequality uses (2.6.12) with $f(x)=\lambda(1\wedge m_1^{-1}x)$, with m_1 as defined in (2.6.13), and (2.6.15). Applying (2.6.18) of Lemma 2.6.7, with $\ell=2$, and using $C_\beta > 1$ and $s_i^* \le 1$ for all $i \in \mathbb{Z}$, we obtain

$$s_2^* \le \lambda (s_1^*)^d + \lambda C_\beta \left(\prod_{i=3-\lfloor\beta\rfloor}^{1} (s_i^*)^{d-1} \right) (s_{2-\lfloor\beta\rfloor}^*)^{\hat\beta(d-1)} \le 2\lambda C_\beta \left(\prod_{i=3-\lfloor\beta\rfloor}^{1} (s_i^*)^{d-1} \right) (s_{2-\lfloor\beta\rfloor}^*)^{\hat\beta(d-1)}, \tag{2.6.19}$$

which shows that (2.6.1) holds for $\ell = 2$.

Similarly, using (2.3.4) and (2.6.11) of Corollary 2.6.6, with $\ell = 3$, and (2.6.12), with $f(x) = \lambda \prod_{i=1}^{2} (1 \wedge m_i^{-1} x)$ and (2.6.15), we have

$$s_3^* \le 2\lambda (s_2^*)^d + \frac{C_1}{x_0^{\beta-1}} \int_1^\infty \lambda \prod_{i=1}^{2} (1 \wedge m_i^{-1} y) y^{-\beta} dy \le 2\lambda (s_2^*)^d + \frac{\lambda C_1}{x_0^{\beta-1}} \psi_{3,0}. \tag{2.6.20}$$

We now claim that for any $k \ge 2$ and $L_k < \infty$,

$$s_k^* \le L_k \left(\prod_{i=k-\lfloor\beta\rfloor+1}^{k-1} (s_i^*)^{d-1} \right) (s_{k-\lfloor\beta\rfloor}^*)^{\hat\beta(d-1)} \implies (s_k^*)^d \le L_k \left(\prod_{i=k-\lfloor\beta\rfloor+2}^{k} (s_i^*)^{d-1} \right) (s_{k-\lfloor\beta\rfloor+1}^*)^{\hat\beta(d-1)}. \tag{2.6.21}$$

Indeed, this follows from the fact that for $k \ge 2$

$$\begin{aligned} (s_k^*)^d = (s_k^*)^{d-1} s_k^* \le& (s_k^*)^{d-1} L_k \left(\prod_{i=k-\lfloor\beta\rfloor+1}^{k-1} (s_i^*)^{d-1} \right) (s_{k-\lfloor\beta\rfloor}^*)^{\hat\beta(d-1)} \\ \le& L_k \left(\prod_{i=k-\lfloor\beta\rfloor+2}^{k} (s_i^*)^{d-1} \right) (s_{k-\lfloor\beta\rfloor+1}^*)^{\hat\beta(d-1)}, \end{aligned}$$

and the fact that $s_{k-\lfloor\beta\rfloor}^* \le 1$ and $s_{k-\lfloor\beta\rfloor+1}^* \le (s_{k-\lfloor\beta\rfloor+1}^*)^{\hat\beta}$. Together, (2.6.18), (2.6.19), (2.6.20) and (2.6.21), with $k = 2$ and $L_2 = 2C_\beta$, yield

$$s_3^* \le 5\lambda C_\beta \left(\prod_{i=4-\lfloor\beta\rfloor}^{2} (s_i^*)^{d-1} \right) (s_{3-\lfloor\beta\rfloor}^*)^{\hat\beta(d-1)}. \tag{2.6.22}$$

We now use induction to obtain estimates for all $\ell \geq 2$. Define $K_2 := 2$, $K_3 := 5$, and

$$K_\ell := 2K_{\ell-1} + \sum_{j=2}^{\ell-2} K_j + 1 \qquad \forall \ell \geq 4. \tag{2.6.23}$$

Moreover, suppose for some $\ell > 3$,

$$s_j^* \leq \lambda K_j (C_\beta)^{j-2} \left(\prod_{i=j-\lfloor\beta\rfloor+1}^{j-1} (s_i^*)^{d-1} \right) (s_{j-\lfloor\beta\rfloor}^*)^{\hat\beta(d-1)}, \qquad j \in \{3, \cdots \ell-1\}. \tag{2.6.24}$$

To upper bound the first term on the right-hand side of (2.6.14), we combine (2.6.24), with $j = \ell - 1$, and (2.6.21), with $k = \ell - 1$ and $L_{\ell-1} = K_{\ell-1} C_\beta^{\ell-3}$, to conclude that

$$2\lambda (s_{\ell-1}^*)^d \leq 2\lambda K_{\ell-1} C_\beta^{\ell-3} \left(\prod_{i=\ell-\lfloor\beta\rfloor+1}^{\ell-1} (s_i^*)^{d-1} \right) (s_{\ell-\lfloor\beta\rfloor}^*)^{\hat\beta(d-1)}. \tag{2.6.25}$$

Next, to estimate the last term on the right-hand side of (2.6.14), combine (2.6.15) with (2.6.18) of Lemma 2.6.7, to obtain

$$\frac{C_1}{x_0^{\beta-1}} \psi_{\ell,0} \leq C_\beta \left(\prod_{i=\ell-\lfloor\beta\rfloor+1}^{\ell-1} (s_i^*)^{d-1} \right) (s_{\ell-\lfloor\beta\rfloor}^*)^{\hat\beta(d-1)}. \tag{2.6.26}$$

We now identify respective upper bounds for each of the summands in the middle term on the right-hand side of (2.6.14). First, note that when $2 \leq j \leq \ell - \lfloor\beta\rfloor - 1$, (2.6.16) and the inequalities $s_i^* \leq 1$ for all i, and $C_\beta \geq 1$ imply

$$\frac{C_1}{x_0^{\beta-1}} (s_j^*)^d \psi_{\ell,j} \leq \frac{C_1}{x_0^{\beta-1}} (s_j^*)^d \psi_{\ell,\ell-\lfloor\beta\rfloor-1} \leq C_\beta^{\ell-2} \left(\prod_{i=\ell-\lfloor\beta\rfloor+1}^{\ell-1} (s_i^*)^{d-1} \right) (s_{\ell-\lfloor\beta\rfloor}^*)^{\hat\beta(d-1)}. \tag{2.6.27}$$

On the other hand, if $\ell - \lfloor\beta\rfloor \leq j \leq \ell - 2$, using the fact that $j + 1 < \ell$ and $s_i^* \leq 1$

for all i, as well as (2.6.24) and (2.6.21), with $k = j$ and $L_k = K_j C_\beta^{j-2}$, one obtains

$$\frac{C_1}{x_0^{\beta-1}}(s_j^*)^d \psi_{\ell,j} \le K_j C_\beta^{j-1} \left(\prod_{i=\ell-\lfloor\beta\rfloor+1}^{\ell-1} (s_i^*)^{d-1} \right) (s_{\ell-\lfloor\beta\rfloor}^*)^{\hat{\beta}(d-1)}. \tag{2.6.28}$$

We now observe that $K_\ell \le 3^{\ell-1}$. Since $K_2 = 2$, $K_3 = 5$ and for $\ell > 3$, by definition

$$\begin{aligned} K_{\ell-1} = 2K_{\ell-2} + \sum_{j=2}^{\ell-3} K_j + 1 &= K_{\ell-2} + \sum_{j=2}^{\ell-2} K_j + 1 \\ &= K_{\ell-2} + K_\ell - 2K_{\ell-1} \\ &\ge K_\ell - 2K_{\ell-1}, \end{aligned}$$

which implies $K_\ell \le 3^{\ell-1}$ for all $\ell \ge 2$. Combining this with (2.6.14), (2.6.25)-(2.6.28), the definition (2.6.23) and the inequality $C_\beta \ge 1$, we conclude that

$$\begin{aligned} s_\ell^* &\le \left(2\lambda K_{\ell-1} C_\beta^{\ell-3} + \lambda C_\beta^{\ell-2} \sum_{j=2}^{\ell-2} K_j + \lambda C_\beta \right) \left(\prod_{i=\ell-\lfloor\beta\rfloor+1}^{\ell-1} (s_i^*)^{d-1} \right) (s_{\ell-\lfloor\beta\rfloor}^*)^{\hat{\beta}(d-1)} \\ &\le \lambda K_\ell C_\beta^{\ell-2} \left(\prod_{i=\ell-\lfloor\beta\rfloor+1}^{\ell-1} (s_i^*)^{d-1} \right) (s_{\ell-\lfloor\beta\rfloor}^*)^{\hat{\beta}(d-1)}. \end{aligned}$$

This shows that (2.6.24) also holds when $\ell - 1$ is replaced by ℓ. By the principle of induction, (2.6.24) holds for all $\ell \ge 2$. Since $K_\ell \le 3^{\ell-1}$ and $C_\beta > 1$, this proves (2.6.1) of Proposition 2.6.1.

If β is an integer, the proof for (2.6.2) uses the same argument as in the non-integral case except that $\hat{\beta}$ is replaced with $1 - \delta$, and $i - \lfloor\beta\rfloor$ with $i - \beta + 1$ for all i, and the upper bounds in (2.6.17) of Lemma 2.6.7 are used in place of the bounds in (2.6.16). Thus, we omit the details.

2.6.4 Proof of Key Estimates

To complete the proof of Proposition 2.6.1, in this section we present the proof of Lemma 2.6.7. We start with a preliminary result in Lemma 2.6.8 that will be used in the proof. Recall the definition of m_i from (2.6.13) and note that $m_0 < 1$ and the sequence $\{m_i\}_{i\in\mathbb{Z}}$ is non-decreasing. Define

$$\kappa := \min\{i \in \mathbb{N} : m_i \geq 1\}, \tag{2.6.29}$$

and note that $\kappa < \infty$ since $s_i^* \to 0$ as $i \to \infty$; see (2.3.2). As always, $\beta \in (d/(d-1), \infty)$.

Lemma 2.6.8. *Fix $\ell \in \mathbb{N}$. Let $j' \leq \ell - 2$. Then for any $z > 0$,*

$$\int_z^\infty \prod_{i=j'+1}^{\ell-1} (1 \wedge m_i^{-1} y) y^{-\beta} dy \leq \frac{z^{-(\beta-1)}}{(\beta - 1)}. \tag{2.6.30}$$

Moreover, if $n \in \mathbb{N}$ is such that $j' \leq n - 2$ and $\ell - \lfloor\beta\rfloor + 2 \leq n \leq \ell - 1$, then

$$\int_{m_{n-1}}^{m_n} \prod_{i=j'+1}^{\ell-1} (1 \wedge m_i^{-1} y) y^{-\beta} dy \leq \frac{(\lambda d x_0)^{\beta-1}}{(\beta - \ell + n - 1)} \left(\prod_{i=n}^{\ell-1} (s_i^*)^{d-1}\right) (s_{n-1}^*)^{(\beta-\ell+n-1)(d-1)}. \tag{2.6.31}$$

If $\beta \in (d/(d-1), \infty) \setminus \mathbb{N}$ and $\ell - \lfloor\beta\rfloor < j' + 1 \leq \kappa \leq \ell - 1$, then

$$\int_1^{m_\kappa} \prod_{i=j'+1}^{\ell-1} (1 \wedge m_i^{-1} y) y^{-\beta} dy \leq \frac{(\lambda d x_0)^{\beta-\hat{\beta}-1}}{\hat{\beta}} \left(\prod_{i=j'+1}^{\ell-1} (s_i^*)^{d-1}\right). \tag{2.6.32}$$

Proof. Fix $j' \leq \ell - 2$. The relation (2.6.30) follows immediately since $1 \wedge x \leq 1$, and

$\beta > 1$ imply

$$\int_z^\infty \prod_{i=j'+1}^{\ell-1} (1 \wedge m_i^{-1} y) y^{-\beta} dy \le \int_z^\infty y^{-\beta} dy = \frac{z^{-(\beta-1)}}{(\beta-1)}.$$

Next, fix $j', \ell, n \in \mathbb{N}$ with $j' \le n-2$ and $\ell - \lfloor \beta \rfloor + 2 \le n \le \ell - 1$. Then, for $y \in [m_{n-1}, m_n)$, the fact that m_i is increasing in i implies $m_i^{-1} y \ge 1$ for all $i \le n-1$ and $m_i^{-1} y \le 1$ for all $i \ge n$. Hence, the definition of m_{n-1} in (2.6.13) and the relation $\ell - \beta - n + 1 < 0$ yield (2.6.31) as follows:

$$\begin{aligned}
\int_{m_{n-1}}^{m_n} \prod_{i=j'+1}^{\ell-1} (1 \wedge m_i^{-1} y) y^{-\beta} dy &= \int_{m_{n-1}}^{m_n} (\lambda d x_0)^{\ell-n} \left(\prod_{i=n}^{\ell-1} (s_i^*)^{d-1} \right) y^{-\beta+\ell-n} dy \\
&\le \frac{(\lambda d x_0)^{\ell-n}}{(\beta-\ell+n-1)} \left(\prod_{i=n}^{\ell-1} (s_i^*)^{d-1} \right) (\lambda d (s_{n-1}^*)^{d-1} x_0)^{\beta-\ell+n-1} \\
&= \frac{(\lambda d x_0)^{\beta-1}}{(\beta-\ell+n-1)} \left(\prod_{i=n}^{\ell-1} (s_i^*)^{d-1} \right) (s_{n-1}^*)^{(\beta-\ell+n-1)(d-1)}.
\end{aligned}$$

Next, let $\ell - \lfloor \beta \rfloor < j'+1 \le \kappa \le \ell - 1$. Note that $1 \wedge m_i^{-1} y \le m_i^{-1} y$ for all $y \in [1, m_\kappa)$ and $i \ge 0$. Hence, combining this with the definition of m_i in (2.6.13), and the relation $\beta - \lfloor \beta \rfloor + 1 = \hat{\beta} + 1 > 0$, we obtain (2.6.32) as follows:

$$\begin{aligned}
\int_1^{m_\kappa} \prod_{i=j'+1}^{\ell-1} (1 \wedge m_i^{-1} y) y^{-\beta} dy &\le \int_1^{m_\kappa} \left(\prod_{i=j'+1}^{\ell-1} m_i^{-1} y \right) y^{-\beta+\ell-j'-1} dy \\
&= (\lambda d x_0)^{\ell-j'-1} \left(\prod_{i=j'+1}^{\ell-1} (s_i^*)^{d-1} \right) \int_1^{m_\kappa} y^{-\beta+\ell-j'-1} dy \\
&\le (\lambda d x_0)^{\lfloor \beta \rfloor - 1} \left(\prod_{i=j'+1}^{\ell-1} (s_i^*)^{d-1} \right) \int_1^{m_\kappa} y^{-\beta+\lfloor \beta \rfloor - 1} dy \\
&\le \frac{(\lambda d x_0)^{\beta-\hat{\beta}-1}}{\hat{\beta}} \left(\prod_{i=j'+1}^{\ell-1} (s_i^*)^{d-1} \right),
\end{aligned}$$

where the third step uses the inequalities $\ell - \lfloor \beta \rfloor < j'+1$ and $\lambda d x_0 \ge 1$. $\square$

Recall the definition of C_1 from (2.3.9), C_β from (2.3.11) and ψ from (2.6.15).

Proof of Lemma 2.6.7. Fix $\ell \in \mathbb{N}$, $\ell \geq 2$ and $\beta \in (d/(d-1), \infty)$. We first show that (2.6.16) and (2.6.17) imply (2.6.18). First, suppose $\beta \notin \mathbb{N}$. Recalling the definition of $\psi_{\ell,j}$ from (2.6.15), we now claim (and justify below) that

$$\frac{C_1}{x_0^{\beta-1}} \int_1^\infty \prod_{i=1}^{\ell-1}(1 \wedge m_i^{-1} y) y^{-\beta} dy \leq \frac{C_1}{x_0^{\beta-1}} \int_1^\infty \prod_{i=\ell-\lfloor\beta\rfloor}^{\ell-1}(1 \wedge m_i^{-1} y) y^{-\beta} dy = \frac{C_1}{x_0^{\beta-1}} \psi_{\ell,\ell-\lfloor\beta\rfloor-1}. \tag{2.6.33}$$

Indeed, note that if $i \leq 0$, then $1 \wedge m_i^{-1} = 1 \wedge \lambda d x_0 (s_i^*)^{d-1} = 1$ since $\lambda d x_0 > 1$, and $s_i^* = 1$ for $i \leq 0$. This immediately implies (2.6.33) when $\ell - \lfloor\beta\rfloor \leq 1$. On the other hand, if $\ell - \lfloor\beta\rfloor > 1$, then the trivial inequality $1 \wedge x \leq 1$ and (2.6.15) imply (2.6.33). Noting that $\lambda d x_0 \geq 1$ implies $C_1/x_0^{\beta-1}$ is bounded above by the constant C_β defined in (2.3.11), the last two displays imply (2.6.18). If $\beta \in \mathbb{N}$, (2.6.17) can be shown to imply (2.6.18) using the same argument, but with $\ell - \lfloor\beta\rfloor$ replaced by $\ell - \beta + 1$.

We now turn to the proofs of (2.6.16) and (2.6.17). Recall the definition of κ in (2.6.29).

Case 1: Suppose $\kappa \geq \ell$ and $\ell - \lfloor\beta\rfloor - 1 \leq j \leq \ell - 2$. Then, using the definition of $\psi_{\ell,j}$ in (2.6.15) and (since $j \leq \ell - 2$) applying (2.6.30) with $z = 1$ and $j' = j$, as well as (2.3.11), this implies

$$\frac{C_1}{x_0^{\beta-1}} \psi_{\ell,j} = \frac{C_1}{x_0^{\beta-1}} \int_1^\infty \prod_{i=j+1}^{\ell-1}(1 \wedge m_i^{-1} y) y^{-\beta} dy \leq \frac{C_1}{(\beta-1) x_0^{\beta-1}} \leq \frac{C_\beta}{(\lambda d x_0)^{\beta-1}}.$$

Combining this with the observation that $m_i^{-1} = \lambda d x_0 (s_i^*)^{d-1} \geq 1$ for all $i \leq \ell - 1$

due to (2.6.29) and the case assumption, we have, for any $r_{j,\beta} \geq 1$,

$$\frac{C_1}{x_0^{\beta-1}}\psi_{\ell,j} \leq \frac{C_\beta}{(\lambda d x_0)^{\beta-1}}\left(\prod_{i=j+1}^{\ell-1} m_i^{-1}\right) r_{j,\beta} = C_\beta(\lambda d x_0)^{-(j-\ell+\beta)}\left(\prod_{i=j+1}^{\ell-1}(s_i^*)^{d-1}\right) r_{j,\beta}. \tag{2.6.34}$$

Since $\lambda d x_0 > 1$, choosing $r_{j,\beta} = 1$, we obtain

$$\frac{C_1}{x_0^{\beta-1}}\psi_{\ell,j} \leq C_\beta\left(\prod_{i=j+1}^{\ell-1}(s_i^*)^{d-1}\right) \qquad \text{if } j \geq \ell-\beta.$$

This immediately implies (2.6.16) when $\beta \notin \mathbb{N}$ and $j \geq \ell - \lfloor\beta\rfloor$, and (2.6.17) when $\beta \in \mathbb{N}$ and $j \geq \ell-\beta+1$ and also, noting that $s^*_{\ell-\beta+1} \leq 1$ implies $s^*_{\ell-\beta+1} \leq (s^*_{\ell-\beta+1})^{(1-\delta)}$ for any $\delta > 0$, when $j = \ell-\beta$. When $\beta \notin \mathbb{N}$, since $j \geq \ell-\beta$ implies $j \geq \ell - \lfloor\beta\rfloor$, (2.6.34) also proves (2.6.16) in that case. The remaining case when $\beta \notin \mathbb{N}$ and $j = \ell - \lfloor\beta\rfloor - 1$ can be deduced similarly from (2.6.34) by setting $r_{j,\beta} = m_{\ell-\lfloor\beta\rfloor}^{-(1-\hat{\beta})} = (\lambda d x_0)^{-(1-\hat{\beta})}(s^*_{\ell-\lfloor\beta\rfloor})^{-(1-\hat{\beta})(d-1)}$ therein.

<u>**Case 2**</u>: Suppose $\kappa < \ell$ and $\ell - \lfloor\beta\rfloor - 1 \leq j < \kappa - 1$. We now look at the partition $\pi_2 = \{[1, m_\kappa), [m_\kappa, m_{\kappa+1}), \cdots, [m_{\ell-1}, \infty)\}$ of the interval $[1,\infty)$. Note that the

$$\text{number of intervals in } \pi_2 = \ell - \kappa + 1 \leq \lfloor\beta\rfloor - 1 + 1 < \lfloor\beta\rfloor + 1. \tag{2.6.35}$$

Since $j < \ell - 2$, using (2.6.30) with $z = m_{\ell-1} = (\lambda d (s^*_{\ell-1})^{d-1} x_0)^{-1}$ and $j' = j$, we have

$$\frac{C_1}{x_0^{\beta-1}}\int_{m_{\ell-1}}^{\infty}\prod_{i=j+1}^{\ell-1}(1 \wedge m_i^{-1}y)y^{-\beta}dy \leq \frac{C_1}{(\beta-1)}(\lambda d)^{\beta-1}(s^*_{\ell-1})^{(d-1)(\beta-1)}. \tag{2.6.36}$$

Next, since $\ell - \lfloor\beta\rfloor \leq j+1 < \kappa$ implies $\ell - \lfloor\beta\rfloor + 1 \leq \kappa$, we have for $n \in \{\kappa+1, \kappa+2, \cdots, \ell-1\}$, $n \geq \kappa+1 \geq \ell - \lfloor\beta\rfloor + 2$ and $n \geq \kappa + 1 > j+2$. Since

$\ell - \lfloor \beta \rfloor + 2 \leq n \leq \ell - 1$ and $j \leq n-2$, we use (2.6.31) with $j' = j$ to obtain

$$\frac{C_1}{x_0^{\beta-1}} \int_{m_{n-1}}^{m_n} \prod_{i=j+1}^{\ell-1} (1 \wedge m_i^{-1} y) y^{-\beta} dy \leq \frac{C_1}{(\beta - \ell + n - 1)} (\lambda d)^{\beta-1} \sigma_n^{d-1}, \tag{2.6.37}$$

where

$$\sigma_n := \left(\prod_{i=n}^{\ell-1} s_i^* \right) (s_{n-1}^*)^{(\beta-\ell+n-1)}. \tag{2.6.38}$$

Now, for any $n \in \{\kappa+1, \cdots, \ell-1\}$ the case assumption, which in particular implies $j < n-2$, the inequality $s_{n-1}^* \leq 1$, and the fact that $i \mapsto s_i^*$ is non-increasing imply that $(s_{\ell-1}^*)^{\beta-1} \leq \sigma_n$ and $\sigma_n \leq \sigma_n (s_{n-1}^*)^{\ell-j-\beta} = \left(\prod_{i=n}^{\ell-1} s_i^* \right) (s_{n-1}^*)^{n-1-j} \leq \prod_{i=j+1}^{\ell-1} s_i^*$, and hence, that

$$\max_{n \in \{\kappa+1, \cdots, \ell-1\}} \left((s_{\ell-1}^*)^{\beta-1}, \sigma_n, \prod_{i=j+1}^{\ell-1} s_i^* \right) = \prod_{i=j+1}^{\ell-1} s_i^*, \tag{2.6.39}$$

<u>Case 2A</u>: Suppose $\beta \notin \mathbb{N}$. Restricting j so that $\ell - \lfloor \beta \rfloor - 1 < j < \kappa - 1$, and using (2.6.32) with $j' = j$, we have for any $r_{j,\beta} \geq 1$,

$$\frac{C_1}{x_0^{\beta-1}} \int_1^{m_\kappa} \prod_{i=j+1}^{\ell-1} (1 \wedge m_i^{-1} y) y^{-\beta} dy \leq \frac{C_1}{\hat{\beta} x_0^{\beta-1}} (\lambda d x_0)^{\beta-\hat{\beta}-1} \left(\prod_{i=j+1}^{\ell-1} (s_i^*)^{d-1} \right) r_{j,\beta}. \tag{2.6.40}$$

Since $\hat{\beta} > 0$ and $\lambda d x_0 \geq 1$, choosing $r_{j,\beta} = 1$, we obtain

$$\frac{C_1}{x_0^{\beta-1}} \int_1^{m_\kappa} \prod_{i=j+1}^{\ell-1} (1 \wedge m_i^{-1} y) y^{-\beta} dy \leq \frac{C_1}{\hat{\beta}} (\lambda d)^{\beta-1} \left(\prod_{i=j+1}^{\ell-1} (s_i^*)^{d-1} \right). \tag{2.6.41}$$

Combining (2.6.35) – (2.6.38), (2.6.41) and (2.6.39) with the fact that

$$\min_{n \in \{\kappa+1, \cdots, \ell-1\}} (\beta - 1, \beta - \ell + n - 1, \hat{\beta}) \geq \min(\hat{\beta}, 1 - \hat{\beta}), \tag{2.6.42}$$

and the definitions of C_β and $\psi_{\ell,j}$ in (2.3.11) and (2.6.15), respectively, we obtain (2.6.16) whenever $j \geq \ell - \lfloor \beta \rfloor$.

Setting $j = \ell - \lfloor \beta \rfloor$ and $r_{j,\beta} = m_{\ell-\lfloor\beta\rfloor}^{-\hat{\beta}} = (\lambda d x_0)^{\hat{\beta}} (s^*_{\ell-\lfloor\beta\rfloor})^{\hat{\beta}(d-1)}$ in (2.6.40), we obtain

$$\begin{aligned}\frac{C_1}{x_0^{\beta-1}} \int_1^{m_\kappa} \prod_{i=\ell-\lfloor\beta\rfloor}^{\ell-1} (1 \wedge m_i^{-1} y) y^{-\beta} dy &\leq \frac{C_1}{x_0^{\beta-1}} \int_1^{m_\kappa} \prod_{i=\ell-\lfloor\beta\rfloor+1}^{\ell-1} (1 \wedge m_i^{-1} y) y^{-\beta} dy \\ &\leq \frac{C_1}{\hat{\beta}} (\lambda d)^{\beta-1} \left(\prod_{i=\ell-\lfloor\beta\rfloor+1}^{\ell-1} (s_i^*)^{d-1} \right) (s^*_{\ell-\lfloor\beta\rfloor})^{\hat{\beta}(d-1)}.\end{aligned} \tag{2.6.43}$$

Now, noting that $\ell - \lfloor\beta\rfloor - 1 < n - 2$ and $i \mapsto s_i^*$ is non-increasing, we conclude that

$$\max_{n \in \{\kappa+1, \cdots, \ell-1\}} \left((s^*_{\ell-1})^{\beta-1}, \sigma_n, \left(\prod_{i=\ell-\lfloor\beta\rfloor+1}^{\ell-1} s_i^* \right) (s^*_{\ell-\lfloor\beta\rfloor})^{\hat{\beta}} \right) = \left(\prod_{i=\ell-\lfloor\beta\rfloor+1}^{\ell-1} s_i^* \right) (s^*_{\ell-\lfloor\beta\rfloor})^{\hat{\beta}}, \tag{2.6.44}$$

Combining (2.6.35) – (2.6.38), (2.6.42) – (2.6.44) and the definitions of C_β and $\psi_{\ell,j}$ in (2.3.11) and (2.6.15), respectively, we obtain (2.6.16) when $j = \ell - \lfloor\beta\rfloor - 1$.

Case 2B: Suppose $\beta \in \mathbb{N}$ and recall $\ell - \beta \leq j + 1 < \kappa$. Note that $y \in [1, m_\kappa)$ and the fact that $i \mapsto m_i$ is increasing implies $m_i^{-1} y \geq 1$ for all $i \leq \kappa - 1$ and $m_i^{-1} y \leq 1$ for all $i \geq \kappa$. Then, substituting the expression for m_i^{-1} in (2.6.13), and $m_{j+1}^{-1} \geq 1$, we obtain

$$\begin{aligned}&\frac{C_1}{x_0^{\beta-1}} \int_1^{m_\kappa} \prod_{i=j+1}^{\ell-1} (1 \wedge m_i^{-1} y) y^{-\beta} dy \\ &\quad \leq \frac{C_1}{x_0^{\beta-1}} \int_1^{m_\kappa} \prod_{i=\kappa}^{\ell-1} (1 \wedge m_i^{-1} y) y^{-\beta} dy \\ &\quad \leq \frac{C_1}{x_0^{\beta-1}} (\lambda d x_0)^{\ell-\kappa} \int_1^{m_\kappa} \left(\prod_{i=\kappa}^{\ell-1} (s_i^*)^{d-1} \right) y^{-\beta+\ell-\kappa} dy\end{aligned}$$

$$
\begin{aligned}
&\le \frac{C_1}{x_0^{\beta-1}}(\lambda d x_0)^{\ell-\kappa}\left(\prod_{i=\kappa}^{\ell-1}(s_i^*)^{d-1}\right)\int_1^{m_\kappa}\left(\prod_{i=j+2}^{\kappa-1} m_i^{-1}y\right) m_{j+1}^{-1}y^{-\beta+\ell-\kappa}dy\\
&= \frac{C_1}{x_0^{\beta-1}}(\lambda d x_0)^{\ell-j-1}\left(\prod_{i=j+1}^{\ell-1}(s_i^*)^{d-1}\right)\int_1^{m_\kappa} y^{-\beta+\ell-j-2}dy\\
&\le \frac{C_1}{x_0^{\beta-1}}(\lambda d x_0)^{\beta-1}\left(\prod_{i=j+1}^{\ell-1}(s_i^*)^{d-1}\right)\int_1^{m_\kappa} y^{-2}dy\\
&\le C_1(\lambda d x_0)^{\beta-1}\left(\prod_{i=j+1}^{\ell-1}(s_i^*)^{d-1}\right). \qquad (2.6.45)
\end{aligned}
$$

Combining (2.6.35) – (2.6.39) and (2.6.45) with the fact that

$$\min_{n\in\{\kappa+1,\cdots,\ell-1\}}(\beta-1,\beta-\ell+n-1,1)=1, \qquad (2.6.46)$$

the observation that $s^*_{\ell-\beta+1}\le(s^*_{\ell-\beta+1})^{1-\delta}$ for any $\delta\in(0,1)$ and the definitions of C_β and $\psi_{\ell,j}$ in (2.3.11) and (2.6.15), respectively, we obtain (2.6.17).

Case 3: Suppose $\kappa<\ell$, $j\ge\ell-\lfloor\beta\rfloor-1$ and $\kappa-1\le j\le\ell-2$. Using (2.6.29), this implies $m_i\ge1$ for all $i\ge j+1$. We now look at the partition $\pi_3=\{[1,m_{j+1}),[m_{j+1},m_{j+2}),\cdots,[m_{\ell-1},\infty)$ of the interval $[1,\infty)$. Note that

$$\text{number of intervals in } \pi_3=\ell-j\le\lfloor\beta\rfloor+1. \qquad (2.6.47)$$

Since $j+1\le\ell-1$, using (2.6.30) with $z=m_{\ell-1}=(\lambda d(s^*_{\ell-1})^{d-1}x_0)^{-1}$ and $j'=j$, we have

$$\frac{C_1}{x_0^{\beta-1}}\int_{m_{\ell-1}}^{\infty}\prod_{i=\ell-\lfloor\beta\rfloor}^{\ell-1}(1\wedge m_i^{-1}y)y^{-\beta}dy\le\frac{C_1}{(\beta-1)}(\lambda d)^{\beta-1}(s^*_{\ell-1})^{(d-1)(\beta-1)}. \qquad (2.6.48)$$

For the interval $[1,m_{j+1})$ note that $y\in[1,m_{j+1})$ and the fact that m_i is increasing in i implies $m_i^{-1}y \leq 1$ for all $i \geq j+1$. Hence, we have

$$\frac{C_1}{x_0^{\beta-1}} \int_1^{m_{j+1}} \prod_{i=j+1}^{\ell-1} (1 \wedge m_i^{-1}y) y^{-\beta} dy = \frac{C_1}{x_0^{\beta-1}} (\lambda d x_0)^{\ell-j-1} \left(\prod_{i=j+1}^{\ell-1} (s_i^*)^{d-1} \right) \int_1^{m_{j+1}} y^{-\beta+\ell-j-1} dy. \tag{2.6.49}$$

Additionally, if $j \geq \ell - \lfloor \beta \rfloor$, note that for $n \in \{j+2, j+3, \cdots, \ell-1\}$, $n \geq j+2 \geq \ell - \lfloor \beta \rfloor + 2$. Since $\ell - \lfloor \beta \rfloor + 2 \leq n \leq \ell - 1$, we use (2.6.31) with $j' = j$ and the definition of σ_n in (2.6.38) to obtain

$$\frac{C_1}{x_0^{\beta-1}} \int_{m_{n-1}}^{m_n} \prod_{i=j+1}^{\ell-1} (1 \wedge m_i^{-1}y) y^{-\beta} dy \leq \frac{C_1}{(\beta - \ell + n - 1)} (\lambda d)^{\beta-1} \sigma_n^{d-1}. \tag{2.6.50}$$

Case 3A: Suppose $\beta \notin \mathbb{N}$. First, let $j = \ell - \lfloor \beta \rfloor - 1$. Since $\beta - \hat{\beta} = \lfloor \beta \rfloor$ by (2.3.10), (2.6.49) simplifies to

$$\begin{aligned} &\frac{C_1}{x_0^{\beta-1}} \int_1^{m_{\ell-\lfloor \beta \rfloor}} \prod_{i=\ell-\lfloor \beta \rfloor}^{\ell-1} (1 \wedge m_i^{-1}y) y^{-\beta} dy \\ &= \frac{C_1}{x_0^{\beta-1}} (\lambda d x_0)^{\beta-\hat{\beta}} \left(\prod_{i=\ell-\lfloor \beta \rfloor}^{\ell-1} (s_i^*)^{d-1} \right) \int_1^{m_{\ell-\lfloor \beta \rfloor}} y^{-\hat{\beta}} dy \\ &\leq \frac{C_1}{(1-\hat{\beta}) x_0^{\beta-1}} (\lambda d x_0)^{\beta-\hat{\beta}} \left(\prod_{i=\ell-\lfloor \beta \rfloor}^{\ell-1} (s_i^*)^{d-1} \right) (\lambda d (s_{\ell-\lfloor \beta \rfloor}^*)^{d-1} x_0)^{\hat{\beta}-1} \\ &= \frac{C_1}{(1-\hat{\beta})} (\lambda d)^{\beta-1} \left(\prod_{i=\ell-\lfloor \beta \rfloor+1}^{\ell-1} (s_i^*)^{d-1} \right) (s_{\ell-\lfloor \beta \rfloor}^*)^{\hat{\beta}(d-1)}. \end{aligned} \tag{2.6.51}$$

Note that for $n \in \{\ell - \lfloor \beta \rfloor + 2, \ell - \lfloor \beta \rfloor + 3, \cdots, \ell - 1\}$, $n \geq \ell - \lfloor \beta \rfloor + 2$. Since $\ell - \lfloor \beta \rfloor + 2 \leq n \leq \ell - 1$ and, we use (2.6.31), with $j' = \ell - \lfloor \beta \rfloor - 1$ and the definition of σ_n in (2.6.38) to obtain

$$\frac{C_1}{x_0^{\beta-1}} \int_{m_{n-1}}^{m_n} \prod_{i=\ell-\lfloor \beta \rfloor}^{\ell-1} (1 \wedge m_i^{-1}y) y^{-\beta} dy \leq \frac{C_1}{(\beta - \ell + n - 1)} (\lambda d)^{\beta-1} \sigma_n^{d-1}. \tag{2.6.52}$$

Also, note that $y \in [m_{\ell-\lfloor\beta\rfloor}, m_{\ell-\lfloor\beta\rfloor+1})$ and the fact that $i \mapsto m_i$ is increasing implies $m_i^{-1}y \geq 1$ for all $i \leq \ell - \lfloor\beta\rfloor$ and $m_i^{-1}y \leq 1$ for all $i \geq \ell - \lfloor\beta\rfloor + 1$. Hence, using (2.6.13), we obtain

$$\begin{aligned}
&\frac{C_1}{x_0^{\beta-1}} \int_{m_{\ell-\lfloor\beta\rfloor}}^{m_{\ell-\lfloor\beta\rfloor+1}} \prod_{i=\ell-\lfloor\beta\rfloor}^{\ell-1} (1 \wedge m_i^{-1}y) y^{-\beta} dy \\
&\quad= \frac{C_1}{x_0^{\beta-1}} \int_{m_{\ell-\lfloor\beta\rfloor}}^{m_{\ell-\lfloor\beta\rfloor+1}} (\lambda d x_0)^{\lfloor\beta\rfloor-1} \left(\prod_{i=\ell-\lfloor\beta\rfloor+1}^{\ell-1} (s_i^*)^{d-1} \right) y^{-\hat{\beta}-1} dy \\
&\quad\leq \frac{C_1 (\lambda d x_0)^{\beta-\hat{\beta}-1}}{\hat{\beta} x_0^{\beta-1}} \left(\prod_{i=\ell-\lfloor\beta\rfloor+1}^{\ell-1} (s_i^*)^{d-1} \right) (\lambda d (s_{\ell-\lfloor\beta\rfloor}^*)^{d-1} x_0)^{\hat{\beta}} \\
&\quad= \frac{C_1 (\lambda d)^{\beta-1}}{\hat{\beta}} \left(\prod_{i=\ell-\lfloor\beta\rfloor+1}^{\ell-1} (s_i^*)^{d-1} \right) (s_{\ell-\lfloor\beta\rfloor}^*)^{\hat{\beta}(d-1)}. \qquad (2.6.53)
\end{aligned}$$

Combining (2.6.47), (2.6.48), (2.6.51) – (2.6.53), (2.6.44), (2.6.42) (since $\kappa \leq \ell - \lfloor\beta\rfloor$) and the definitions of C_β and $\psi_{\ell,j}$ in (2.3.11) and (2.6.15), respectively, we obtain (2.6.16) when $j = \ell - \lfloor\beta\rfloor - 1$.

Now, let $j \geq \ell - \lfloor\beta\rfloor$. Using (2.6.13), (2.6.49) simplifies to

$$\begin{aligned}
\frac{C_1}{x_0^{\beta-1}} \int_1^{m_{j+1}} \prod_{i=j+1}^{\ell-1} (1 \wedge m_i^{-1}y) y^{-\beta} dy &\leq \frac{C_1}{x_0^{\beta-1}} (\lambda d x_0)^{\lfloor\beta\rfloor-1} \left(\prod_{i=j+1}^{\ell-1} (s_i^*)^{d-1} \right) \int_1^{m_{j+1}} y^{-\hat{\beta}-1} dy \\
&\leq \frac{C_1}{\hat{\beta}} (\lambda d)^{\beta-1} \left(\prod_{i=j+1}^{\ell-1} (s_i^*)^{d-1} \right). \qquad (2.6.54)
\end{aligned}$$

Combining (2.6.47), (2.6.48), (2.6.50), (2.6.54), (2.6.42), (2.6.39) (since $\kappa \leq j + 1$) and the definitions of C_β and $\psi_{\ell,j}$ in (2.3.11) and (2.6.15), respectively, we obtain (2.6.16) when $j \geq \ell - \lfloor\beta\rfloor$.

Case 3B: Suppose $\beta \in \mathbb{N}$. First, let $j = \ell - \beta$. Using (2.6.13) and the fact that there exists $\delta \in (0,1)$ such that $\log x \le x^\delta$ for all $x \ge 1$, (2.6.49) simplifies to

$$\begin{aligned}
&\frac{C_1}{x_0^{\beta-1}} \int_1^{m_{\ell-\beta+1}} \prod_{i=\ell-\beta+1}^{\ell-1} (1 \wedge m_i^{-1} y) y^{-\beta} dy \\
&\quad \le \frac{C_1}{x_0^{\beta-1}} (\lambda d x_0)^{\beta-1} \left(\prod_{i=\ell-\beta+1}^{\ell-1} (s_i^*)^{d-1} \right) \log\left(\frac{1}{\lambda d (s_{\ell-\beta+1}^*)^{d-1} x_0}\right) \\
&\quad \le C_1 (\lambda d)^{\beta-1} \left(\prod_{i=\ell-\beta+2}^{\ell-1} (s_i^*)^{d-1} \right) (s_{\ell-\beta+1}^*)^{(1-\delta)(d-1)}. \qquad (2.6.55)
\end{aligned}$$

Now, noting that $\ell - \beta \le n-2$, $s_{n-1}^* \le 1$, $\delta \in (0,1)$ and $i \mapsto s_i^*$ is non-increasing, we conclude that

$$\max_{n \in \{(\ell-\beta+2, \cdots, \ell-1\}} \left((s_{\ell-1}^*)^{\beta-1}, \sigma_n, \left(\prod_{i=\ell-\beta+2}^{\ell-1} s_i^* \right) (s_{\ell-\beta+1}^*)^{1-\delta} \right) = \left(\prod_{i=\ell-\beta+2}^{\ell-1} s_i^* \right) (s_{\ell-\beta+1}^*)^{1-\delta}. \qquad (2.6.56)$$

Combining (2.6.47), (2.6.48), (2.6.50), (2.6.55), (2.6.56), (2.6.46) (since $\kappa \le \ell-\beta+1$) and the definitions of C_β and $\psi_{\ell,j}$ in (2.3.11) and (2.6.15), respectively, we obtain (2.6.17) when $j = \ell - \beta$.

On the other hand, if $j \ge \ell - \beta + 1$, then (2.6.49) simplifies to

$$\begin{aligned}
\frac{C_1}{x_0^{\beta-1}} \int_1^{m_{j+1}} \prod_{i=j+1}^{\ell-1} (1 \wedge m_i^{-1} y) y^{-\beta} dy &\le \frac{C_1}{x_0^{\beta-1}} (\lambda d x_0)^{\beta-1} \left(\prod_{i=j+1}^{\ell-1} (s_i^*)^{d-1} \right) \int_1^{m_{j+1}} y^{-2} dy \\
&\le C_1 (\lambda d)^{\beta-1} \left(\prod_{i=j+1}^{\ell-1} (s_i^*)^{d-1} \right). \qquad (2.6.57)
\end{aligned}$$

Combining (2.6.47), (2.6.48), (2.6.50), (2.6.57), (2.6.39), (2.6.46) (since $\kappa \le j+1$) and the definitions of C_β and $\psi_{\ell,j}$ in (2.3.11) and (2.6.15), respectively, we obtain (2.6.17) when $j \ge \ell - \beta + 1$. $\square$

Chapter Three

Diffusion Approximations for Many-Server Queues with Reneging

In this Chapter, we consider a G/GI/N+GI model, a system with N servers and a single common queue in which arriving customers are served in a non-idling, First-Come-First-Serve (FCFS) manner, i.e., a newly arriving customer immediately enters service if there are any idle servers or, if all servers are busy, then the customer joins the back of the queue. The customer at the head of the queue (if one is present) enters service as soon as a server becomes free, with our results not being sensitive to the exact mechanism used to assign an arriving customer to an idle server, as long as the non-idling condition is met, that is, there is never simultaneously a positive queue and an idle server, is satisfied. It is assumed that customers are impatient, and that a customer reneges from the queue as soon as the amount of time they have spent in queue reaches their patience time. However, customers do not renege once they have entered service.

For this model, the the measure-valued fluid limit as shown by Kang and Ramanan [22] is deterministic and not refined enough to provide more accurate approximation to quantities of interest such as the probability of a positive wait. This motivates looking at stochastic fluctuations around the fluid limit and characterize the asymptotics of these fluctuations via a diffusion limit theorem. However, as shown by Kaspi and Ramanan [25], the distribution-valued diffusion limit is non-Markovian in the absence of reneging. Hence, we conjecture that an associated distribution-valued diffusion-scaled limit will not be Markov on its own unless stringent assumptions are imposed on the service and patience time distribution. To overcome this, we introduce a new representation $(\widehat{X}^{(N)}, \widehat{Z}_\nu^{(N)}, \widehat{Z}_\eta^{(N)})$ for the state of the system, where $\widehat{X}^{(N)}$ denotes the scaled fluctuation in the number of customers in the N-server network, $\widehat{Z}_\nu^{(N)}$ and $\widehat{Z}_\eta^{(N)}$ are function-valued processes that, respectively, contain information about the ages and patience times of customers in the system. This state representation is presented in Section 3.5.

First, in Section 3.1, we introduce some common terminology that we be used throughout. In Section 3.2 we describe the G/GI/N + GI queueing model in detail, and state our assumptions on service time distribution, patience time distribution and arrival process. We then describe the fluid limit for our model in Section 3.3, leveraging off the corresponding fluid limit for a measure-valued state representation obtained in [22]. We introduce a family of martingales in Section 3.4 that are obtained as compensated departure processes and compensated reneging processes. The main results for this model are stated in Section 3.6 with the proofs deferred to Section 3.10. The proofs rely on auxiliary results obtained Sections 3.7 - 3.9.

3.1 Notation and Terminology

The following notation will be used throughout the chapter. $\mathbb{Z}$ is the set of integers, $\mathbb{N}$ is the set of strictly positive integers, $\mathbb{R}$ is set of real numbers, $\mathbb{R}_+$ the set of non-negative real numbers and $\mathbb{Z}_+$ is the set of non-negative integers. For $a, b \in \mathbb{R}$, $a \vee b$ denotes the maximum of a and b, $a \wedge b$ the minimum of a and b and the short-hand a^+ is used for $a \vee 0$. Given $A \subset \mathbb{R}$ and $a \in \mathbb{R}$, $A - a$ equals the set $\{x \in \mathbb{R} : x + a \in A\}$ and $\mathbb{I}_B$ denotes the indicator function of the set B (that is, $\mathbb{I}_B(x) = 1$ if $x \in B$ and $\mathbb{I}_B(x) = 0$ otherwise).

3.1.1 Function and Measure Spaces

Given any metric space E, let $\mathcal{B}(E)$ denote the Borel sets of E. Also, let $\mathcal{C}_b(E)$ and $\mathcal{C}_c(E)$ denote, respectively, the space of bounded, continuous functions and the space of continuous real-valued functions with compact support defined on E, while $\mathcal{C}^1(E)$

is the space of real-valued, once continuously differentiable functions on E, and $\mathcal{C}_c^1(E)$ is the subspace of functions in $\mathcal{C}^1(E)$ that have compact support. The subspace of functions in $\mathcal{C}^1(E)$ that, together with their first derivatives, are bounded, will be denoted by $\mathcal{C}_b^1(E)$. The constant functions $f \equiv 1$ and $f \equiv 0$ will be represented by the symbols $\mathbf{1}$ and $\mathbf{0}$, respectively. Given any càdlàg, real-valued function φ defined on $[0,\infty)$, we define $\|\varphi\|_T := \sup_{s\in[0,T]} |\varphi(s)|$ for every $T < \infty$, and let $\|\varphi\|_\infty := \sup_{s\in[0,\infty)} |\varphi(s)|$, which could possibly take the value ∞. In addition, the support of a function φ is denoted by $\mathrm{supp}(\varphi)$. Given a nondecreasing function f on $[0,\infty)$, f^{-1} denotes the inverse function of f in the sense that

$$f^{-1}(y) = \inf\{x \geq 0 : f(x) \geq y\}. \tag{3.1.1}$$

For each differentiable function f defined on $\mathbb{R}$, f' denotes the first derivative of f. For each function $f(t,x)$ defined on $\mathbb{R}\times\mathbb{R}^n$, f_t denotes the partial derivative of f with respect to t and f_x denotes the partial derivative of f with respect to x.

Let $\mathbb{L}^1(0,\infty)$, $\mathbb{L}^1_{\mathrm{loc}}(0,\infty)$, and $\mathbb{L}^2(0,\infty)$, denote, respectively, the spaces of integrable, locally integrable, and square integrable functions on $(0,\infty)$ with their corresponding standard norms. The space $\mathbb{L}^2(0,\infty)$ is a Hilbert space with the inner product

$$\langle f, g\rangle_{\mathbb{L}^2} = \int_0^\infty f(x)g(x)dx.$$

The space $\mathbb{H}^1(0,\infty)$ denotes the space of square integrable functions f on $(0,\infty)$ whose weak derivative f' exists and is also square integrable. $\mathbb{H}^1(0,\infty)$ equipped with the inner product

$$\langle f, g\rangle_{\mathbb{H}^1} = \langle f, g\rangle_{\mathbb{L}^2} + \langle f', g'\rangle_{\mathbb{L}^2}, \tag{3.1.2}$$

and the corresponding norm $\|f\|_{\mathbb{H}^1} = \left(\|f\|^2_{\mathbb{L}^2(0,\infty)} + \|f'\|^2_{\mathbb{L}^2(0,\infty)}\right)^{\frac{1}{2}}$, is a separable

Banach space, and hence, a Polish space (see, e.g., [11, Proposition 8.1 on pg 203])

The space of Radon measures on a metric space E, endowed with the Borel σ-algebra, is denoted by $\mathcal{M}(E)$, while $\mathcal{M}_F(E)$, $\mathcal{M}_1(E)$ and $\mathcal{M}_{\leq 1}(E)$ are, respectively, the subspaces of finite, probability and sub-probability measures in $\mathcal{M}(E)$. Also, given $B < \infty$, $\mathcal{M}_B(E) \subset \mathcal{M}_F(E)$ denotes the space of measures μ in $\mathcal{M}_F(E)$ such that $|\mu(E)| \leq B$. Recall that a Radon measure is one that assigns finite measure to every relatively compact subset of $\mathbb{R}_+$. The space $\mathcal{M}(E)$ is equipped with the vague topology, i.e., a sequence of measures $\{\mu_n\}$ in $\mathcal{M}(E)$ is said to converge to μ in the vague topology (denoted $\mu_n \xrightarrow{v} \mu$) if and only if for every $\varphi \in \mathcal{C}_c(E)$,

$$\int_E \varphi(x)\, \mu_n(dx) \to \int_E \varphi(x)\, \mu(dx) \qquad \text{as } n \to \infty. \tag{3.1.3}$$

By identifying a Radon measure $\mu \in \mathcal{M}(E)$ with the mapping on $\mathcal{C}_c(E)$ defined by

$$\varphi \mapsto \int_E \varphi(x)\, \mu(dx),$$

one can equivalently define a Radon measure on E as a linear mapping from $\mathcal{C}_c(E)$ into $\mathbb{R}$ such that for every compact set $\mathcal{K} \subset E$, there exists $L_{\mathcal{K}} < \infty$ such that

$$\left| \int_E \varphi(x)\, \mu(dx) \right| \leq L_{\mathcal{K}} \, \|\varphi\|_\infty \qquad \forall \varphi \in \mathcal{C}_c(E) \text{ with } \mathrm{supp}(\varphi) \subset \mathcal{K}.$$

On $\mathcal{M}_F(E)$, we will also consider the weak topology, i.e., a sequence $\{\mu_n\}$ in $\mathcal{M}_F(E)$ is said to converge weakly to μ (denoted $\mu_n \xrightarrow{w} \mu$) if and only if (3.1.3) holds for every $\varphi \in \mathcal{C}_b(E)$. As is well-known, $\mathcal{M}(E)$ and $\mathcal{M}_F(E)$, endowed with the vague and weak topologies, respectively, are Polish spaces. The symbol δ_x will be used to denote the measure with unit mass at the point x and, by some abuse of notation, we will use $\mathbf{0}$ to denote the identically zero Radon measure on E. When E is an interval, say

$[0, H)$, for notational conciseness, we will often write $\mathcal{M}[0, H)$ instead of $\mathcal{M}([0, H))$. For any finite measure μ on $[0, H)$, we define

$$F^{\mu}(x) := \mu[0, x], \quad x \in [0, H). \tag{3.1.4}$$

We say a measure μ is continuous at x if and only if $\mu(\{x\}) = 0$.

For any Borel measurable function $f : [0, H) \to \mathbb{R}$ that is integrable with respect to $\xi \in \mathcal{M}[0, H)$, we often use the short-hand notation

$$\langle f, \xi \rangle := \int_{[0,H)} f(x)\, \xi(dx).$$

Also, for ease of notation, given $\xi \in \mathcal{M}[0, H)$ and an interval $(a, b) \subset [0, H)$, we will use $\xi(a, b)$ and $\xi(a)$ to denote $\xi((a, b))$ and $\xi(\{a\})$, respectively.

3.1.2 Measure-valued Stochastic Processes

Given a Polish space $\mathcal{H}$, we denote by $\mathcal{D}_{\mathcal{H}}[0, T]$ (respectively, $\mathcal{D}_{\mathcal{H}}[0, \infty)$) the space of $\mathcal{H}$-valued, càdlàg functions on $[0, T]$ (respectively, $[0, \infty)$), and we endow this space with the usual Skorokhod J_1-topology [37]. Then $\mathcal{D}_{\mathcal{H}}[0, T]$ and $\mathcal{D}_{\mathcal{H}}[0, \infty)$ are also Polish spaces (see [37]). A sequence $\{X_n\}$ of càdlàg, $\mathcal{H}$-valued processes, with X_n defined on the probability space $(\Omega_n, \mathcal{F}_n, \mathbb{P}_n)$, is said to converge in distribution to a càdlàg $\mathcal{H}$-valued process X defined on $(\Omega, \mathcal{F}, \mathbb{P})$ if, for every bounded, continuous functional $F : \mathcal{D}_{\mathcal{H}}[0, \infty) \to \mathbb{R}$, we have

$$\lim_{n\to\infty} \mathbb{E}_n \left[F(X_n)\right] = \mathbb{E}\left[F(X)\right],$$

where $\mathbb{E}_n$ and $\mathbb{E}$ are the expectation operators with respect to the probability measures $\mathbb{P}_n$ and $\mathbb{P}$, respectively. Convergence in distribution of X_n to X will be denoted by $X_n \Rightarrow X$. Let $\mathcal{D}^0_{\mathbb{R}}[0,\infty)$ be the subset of functions $f \in \mathcal{D}_{\mathbb{R}}[0,\infty)$ with $f(0) = 0$, and $\mathcal{I}_{\mathbb{R}_+}[0,\infty)$ be the subset of non-decreasing functions $f \in \mathcal{D}_{\mathbb{R}_+}[0,\infty)$ with $f(0) = 0$.

3.2 Description of Model and State Dynamics

In Section 3.2.1 we describe the basic model and the model primitives. In Section 3.2.2 we introduce the state descriptor, some auxiliary processes, and equations that describe the dynamics of the state. Finally, in Section 3.2.3 (see Theorem 3.2.1), we provide a succinct characterization of the state dynamics. This characterization motivates the form of the fluid equations, which are introduced in Section 3.3.1.

3.2.1 Model Description and Primitive Data

As introduced in the beginning of the chapter, we consider a G/GI/N+GI model, a system with N servers and a single common queue in which arriving customers are served in a non-idling, FCFS manner. The customer at the head of the queue (if one is present) enters service as soon as a server becomes free. In this model, a customer reneges from the queue as soon as the amount of time they have spent in queue reaches their patience time. However, customers do not renege once they have entered service. The patience times of customers are given by an i.i.d. sequence, $\{r_i,\ i \in \mathbb{Z}\}$, with common cumulative distribution function G^r on $[0,\infty]$, while the service requirements of customers are given by another i.i.d. sequence, $\{v_i,\ i \in \mathbb{Z}\}$, with common cumulative distribution function G^s on $[0,\infty)$. For $i \in \mathbb{N}$, r_i and v_i represent, respectively, the patience time and the service requirement of the ith customer to enter the system after time zero, while $\{r_i,\ i \in -\mathbb{N} \cup \{0\}\}$ and $\{v_i,\ i \in -\mathbb{N} \cup \{0\}\}$ represent, respectively, the patience times and the service requirements of customers that arrived prior to time zero (if such customers exist), ordered according to their arrival times (prior to time zero). We assume that G^s has density g^s and G^r, restricted on $[0,\infty)$, has density g^r. This implies, in particular, that $G^r(0+) = G^s(0+) = 0$. Let

$$\begin{aligned} H^r &:= \sup\{x \in [0,\infty) : g^r(x) > 0\}, \\ H^s &:= \sup\{x \in [0,\infty) : g^s(x) > 0\} \end{aligned}$$

denote the right ends of the supports of g^r and g^s, respectively. The superscript (N) will be used to refer to quantities associated with the system with N servers.

Let $E^{(N)}$ denote the cumulative arrival process, with $E^{(N)}(t)$ representing the total number of customers that arrive into the system with N servers in the time interval $[0,t]$. Also, consider the càdlàg, real-valued process $\alpha_E^{(N)}$ defined by $\alpha_E^{(N)}(s) = s$ if $E^{(N)}(s) = 0$ and, if $E^{(N)}(s) > 0$, then

$$\alpha_E^{(N)}(s) := s - \sup\left\{u < s :\ E^{(N)}(u) < E^{(N)}(s)\right\}, \tag{3.2.1}$$

which denotes the time elapsed since the last arrival. If $E^{(N)}$ is a renewal process, then $\alpha_E^{(N)}$ is simply the backward recurrence time process. Also, let $\mathcal{E}_0^{(N)}$ be an a.s. $\mathbb{Z}_+$-valued random variable that represents the number of customers that entered the system prior to time zero. This random variable does not play an important role in the analysis, but is used for bookkeeping purposes to keep track of the indices of customers.

The following mild assumptions on $E^{(N)}$ will be imposed throughout, without explicit mention:

- $E^{(N)}$ is a non-decreasing, pure jump process with $E^{(N)}(0) = 0$ and a.s., for $t \in [0, \infty)$, $E^{(N)}(t) < \infty$ and $E^{(N)}(t) - E^{(N)}(t-) \in \{0, 1\}$;
- The process $\alpha_E^{(N)}$ is Markovian with respect to its own natural filtration (this holds, for example, when $E^{(N)}$ is a renewal process);
- The cumulative arrival process $E^{(N)}$, the sequence of service requirements $\{v_j, j \in \mathbb{Z}\}$, and the sequence of patience times $\{r_j, j \in \mathbb{Z}\}$ are independent;

The assumption on the jump size of $E^{(N)}$ is not crucial, and is imposed mainly for convenience. On the other hand, the assumed independence of the service and patience times is a genuine restriction. It would be of interest to consider the case of correlated service and patience times.

3.2.2 State Descriptor and Dynamical Equations

As explained in the paper on the functional law of large numbers for this model [22], the state of the system is represented by the vector of processes $(\alpha_E^{(N)}, X^{(N)}, \nu^{(N)}, \eta^{(N)})$, where $\alpha_E^{(N)}$ determines the cumulative arrival process via (3.2.1), $X^{(N)}$ represents the total number of customers in system with N servers, $\eta^{(N)}$ denotes the "potential queue measure" process which keeps track of the waiting times of customers in queue, and $\nu^{(N)}$ denotes the "age measure" process which encodes the amount of time that the customers currently receiving service have been in service.

3.2.2.1 Description of Queue Dynamics

The potential waiting time process $w_j^{(N)}$ of customer j is (for every realization) defined to be the piecewise linear function on $[0,\infty)$ that is identically zero till the customer enters the system, then increases linearly, representing the amount of time elapsed since entering the system, and then remains constant (equal to the patience time) once the time elapsed exceeds the patience time. More precisely, for $j \in \mathbb{N}$, if $\zeta_j^{(N)}$ is the time at which the jth customer arrives into the system after time 0, then for $j \in \mathbb{N}$ $\zeta_j^{(N)} = (E^{(N)})^{-1}(j) := \inf\{t > 0 : E^{(N)}(t) = j\}$ and

$$w_j^{(N)}(t) = \begin{cases} \left[t - \zeta_j^{(N)}\right] \vee 0 & \text{if } t - \zeta_j^{(N)} < r_j, \\ r_j & \text{otherwise.} \end{cases} \tag{3.2.2}$$

For $j \in -\mathbb{N} \cup \{0\}$, $w_j^{(N)}$ represents the potential waiting time process of the jth customer who entered the system before time zero (if such a customer exists). For $t \in [0,\infty)$, let $\eta_t^{(N)}$ be the non-negative Borel measure on $[0, H^r)$ that has a unit mass at the potential waiting time of each customer that has entered the system by time t and whose potential waiting time has not yet reached its patience time. Recall that δ_x represents the Dirac mass at x. The potential queue measure $\eta_t^{(N)}$ can be written in the form

$$\eta_t^{(N)} = \sum_{j=-\mathcal{E}_0^{(N)}+1}^{E^{(N)}(t)} \delta_{w_j^{(N)}(t)} \mathbb{I}_{\{w_j^{(N)}(t) < r_j\}} = \sum_{j=-\mathcal{E}_0^{(N)}+1}^{E^{(N)}(t)} \delta_{w_j^{(N)}(t)} \mathbb{I}_{\left\{\frac{dw_j^{(N)}}{dt}(t+) > 0\right\}}, \tag{3.2.3}$$

where the last equality holds because at any time t, the potential waiting time process of any customer has a right derivative that is positive if and only if the customer has entered the system and the customer's potential waiting time has not yet reached its patience time.

For $t \in [0, \infty)$, let $Q^{(N)}(t)$ be the number of customers waiting in queue at time t. Due to the non-idling condition, the queue length process is then given by

$$Q^{(N)}(t) = [X^{(N)}(t) - N]^{+}. \tag{3.2.4}$$

Moreover, since the head-of-the-line customer is the customer in queue with the longest waiting time, the quantity

$$\chi^{(N)}(t) := \inf \left\{ x > 0 : \ \eta_t^{(N)}[0, x] \geq Q^{(N)}(t) \right\} = \left(F^{\eta_t^{(N)}} \right)^{-1} (Q^{(N)}(t)) \tag{3.2.5}$$

represents the waiting time of the head-of-the-line customer in the queue at time t. Here, recall from (3.1.4) that $F^{\eta_t^{(N)}}$ is the c.d.f. of the measure $\eta_t^{(N)}$ and $(F^{\eta_t^{(N)}})^{-1}$ represents its inverse, as defined in (3.1.1). Since this is an FCFS system, any mass in $\eta_t^{(N)}$ that lies to the right of $\chi^{(N)}(t)$ represent customers that have already entered service by time t. Therefore, the queue length process $Q^{(N)}$ admits the following alternative representation in terms of $\chi^{(N)}$ and $\eta^{(N)}$:

$$\begin{aligned} Q^{(N)}(t) &= \sum_{j=-\mathcal{E}_0^{(N)}+1}^{E^{(N)}(t)} \mathbb{I}_{\{w_j^{(N)}(t) \leq \chi^{(N)}(t),\ w_j^{(N)}(t) < r_j\}} \\ &= \eta_t^{(N)}[0, \chi^{(N)}(t)]. \end{aligned} \tag{3.2.6}$$

3.2.2.2 Description of Service Dynamics

Analogous to the potential waiting process $w_j^{(N)}$, the age process $a_j^{(N)}$ associated with customer j is (for every realization) defined to be the piecewise linear function on $[0, \infty)$ that equals 0 till the customer enters service, then increases linearly while the customer is in service (representing the amount of time elapsed since entering service) and is then constant (equal to the total service requirement v_j) after the customer completes service and departs the system.

Now, for $t \in [0, \infty)$, let $\nu_t^{(N)}$ be the discrete non-negative Borel measure on $[0, H^s)$ that has a unit mass at the age of each of the customers in service at time t. Then, in a fashion analogous to (3.2.3), the age measure $\nu_t^{(N)}$ can be explicitly represented as

$$\nu_t^{(N)} = \sum_{j=-\mathcal{E}_0^{(N)}+1}^{E^{(N)}(t)} \delta_{a_j^{(N)}(t)} \mathbb{I}_{\left\{\frac{da_j^{(N)}}{dt}(t+)>0\right\}}. \tag{3.2.7}$$

3.2.2.3 Auxiliary Processes

We now introduce certain auxiliary processes that will be useful for the study of the evolution of the system.

- The cumulative reneging process $R^{(N)}$, where $R^{(N)}(t)$ is the cumulative number of customers that have reneged from the system in the time interval $[0, t]$;
- the cumulative potential reneging process $S^{(N)}$, where $S^{(N)}(t)$ represents the cumulative number of customers whose potential waiting times have reached their patience times in the interval $[0, t]$;
- the cumulative departure process $D^{(N)}$, where $D^{(N)}(t)$ is the cumulative number of customers that have departed the system after completion of service in the interval $[0, t]$;
- the process $K^{(N)}$, where $K^{(N)}(t)$ represents the cumulative number of customers that have entered service in the interval $[0, t]$.

Now, $\langle \mathbf{1}, \nu_t^{(N)} \rangle = \nu_t^{(N)}[0, \infty)$ represents the total number of customers in service at time t. Therefore, mass balances on the total number of customers in the system, the number of customers waiting in the "potential queue", and the number of customers in service show that

$$X^{(N)}(0) + E^{(N)} = X^{(N)} + D^{(N)} + R^{(N)}, \tag{3.2.8}$$

$$\langle \mathbf{1}, \eta_0^{(N)} \rangle + E^{(N)} = \langle \mathbf{1}, \eta^{(N)} \rangle + S^{(N)}, \tag{3.2.9}$$

and

$$\langle \mathbf{1}, \nu_0^{(N)} \rangle + K^{(N)} = \langle \mathbf{1}, \nu^{(N)} \rangle + D^{(N)}. \tag{3.2.10}$$

In addition, it is also clear that

$$X^{(N)} = \langle \mathbf{1}, \nu^{(N)} \rangle + Q^{(N)}. \tag{3.2.11}$$

Combining (3.2.8), (3.2.10) and (3.2.11), we obtain the following mass balance for the number of customers in queue:

$$Q^{(N)}(0) + E^{(N)} = Q^{(N)} + R^{(N)} + K^{(N)}. \tag{3.2.12}$$

Furthermore, the non-idling condition takes the form

$$N - \langle \mathbf{1}, \nu^{(N)} \rangle = [N - X^{(N)}]^+.$$

Indeed, note that this ensures that when $X^{(N)}(t) < N$, the number in the system is equal to the number in service, and so there is no queue, while if $X^{(N)}(t) > N$, there is a positive queue and $\langle \mathbf{1}, \nu_t^{(N)} \rangle = N$, indicating that there are no idle servers.

3.2.2.4 Filtration

The total number of customers in service at time t is given by $\langle \mathbf{1}, \nu_t^{(N)} \rangle = \nu_t^{(N)}[0, H^s)$ and is bounded above by N. In addition, from (3.2.9) it follows that

$$\langle \mathbf{1}, \eta_t^{(N)} \rangle = \eta_t^{(N)}[0, H^r) \leq E^{(N)}(t) + \langle \mathbf{1}, \eta_0^{(N)} \rangle \leq E^{(N)}(t) + \mathcal{E}_0^{(N)}$$

which is a.s. finite by assumption. Therefore, for every $t \in [0, \infty)$, a.s., $\nu_t^{(N)} \in \mathcal{M}_F[0, H^s)$ and $\eta_t^{(N)} \in \mathcal{M}_F[0, H^r)$. Hence, the state descriptor $(\alpha_E^{(N)}, X^{(N)}, \nu^{(N)}, \eta^{(N)})$ takes values in $\mathbb{R}_+ \times \mathbb{Z}_+ \times \mathcal{M}_F[0, H^s) \times \mathcal{M}_F[0, H^r)$. For purely technical purposes we will find it convenient to also introduce the additional "station process" $s^{(N)} := (s_j^{(N)}, j \in \mathbb{Z})$, defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$. For each $t \in [0, \infty)$, if customer j has already entered service by time t, then $s_j^{(N)}(t)$ is equal to the index $i \in \{1, \ldots, N\}$ of the station at which customer j receives/received service and $s_j^{(N)}(t) := 0$ otherwise. For $t \in [0, \infty)$, let $\tilde{\mathcal{F}}_t^{(N)}$ be the σ-algebra generated by

$$\left\{ \mathcal{E}_0^{(N)}, X^{(N)}(0), \alpha_E^{(N)}(s), w_j^{(N)}(s), a_j^{(N)}(s), s_j^{(N)}, j \in \{-\mathcal{E}_0^{(N)} + 1, \ldots, 0\} \cup \mathbb{N}, s \in [0, t] \right\}$$

and let $\{\mathcal{F}_t^{(N)}\}$ denote the associated right-continuous filtration, completed with respect to $\mathbb{P}$. In [22] , an explicit construction of the state descriptor and auxiliary processes is provided, which shows in particular that the state descriptor $(\alpha_E^{(N)}, X^{(N)}, \nu^{(N)}, \eta^{(N)})$ and auxiliary processes are càdlàg. Moreover, in [22, Lemma A.1], it is proved that the state process $V^{(N)} := (\alpha_E^{(N)}, X^{(N)}, \nu^{(N)}, \eta^{(N)})$ and the processes $E^{(N)}$, $Q^{(N)}$, $S^{(N)}$, $R^{(N)}$, $D^{(N)}$ and $K^{(N)}$ are all $\mathcal{F}_t^{(N)}$-adapted, and in [22, Lemma B.1], it is shown that $(V^{(N)}, \mathcal{F}_t^{(N)})$ is a strong Markov process.

3.2.3 A Succinct Characterization of the Dynamics

The main result of this section is Theorem 3.2.1, which provides equations that more succinctly characterize the dynamics of the state $(\alpha_E^{(N)}, X^{(N)}, \nu^{(N)}, \eta^{(N)})$ described in Section 3.2.2. First, we introduce some more notation that is required to state the result.

For any measurable function φ on $[0, H^s) \times \mathbb{R}_+$, consider the process $D_\varphi^{(N)}$ that takes values in $\mathbb{R}$, and is given by

$$D_\varphi^{(N)}(t) := \sum_{s\in[0,t]} \sum_{j=-\mathcal{E}_0^{(N)}+1}^{E^{(N)}(t)} \mathbb{I}_{\left\{\frac{da_j^{(N)}}{dt}(s-)>0,\ \frac{da_j^{(N)}}{dt}(s+)=0\right\}} \varphi(a_j^{(N)}(s-), s) \tag{3.2.13}$$

for $t \in [0,\infty)$. It follows immediately from (3.2.13) and the right continuity of the filtration $\{\mathcal{F}_t^{(N)}\}$ that $D_\varphi^{(N)}$ is $\{\mathcal{F}_t^{(N)}\}$-adapted. Also, comparing (3.2.13) with the definition of the process $D^{(N)}$, it is clear that when φ is the constant function $\mathbf{1}$, $D_\mathbf{1}^{(N)}$ is exactly the cumulative departure process $D^{(N)}$, i.e.,

$$D_\mathbf{1}^{(N)} = D^{(N)}. \tag{3.2.14}$$

In an exactly analogous fashion, for any measurable function ψ on $[0, H^r) \times \mathbb{R}_+$, consider the process $S_\psi^{(N)}$ that takes values in $\mathbb{R}$, and is given by

$$S_\psi^{(N)}(t) := \sum_{s\in[0,t]} \sum_{j=-\mathcal{E}_0^{(N)}+1}^{E^{(N)}(t)} \mathbb{I}_{\left\{\frac{dw_j^{(N)}}{dt}(s-)>0,\ \frac{dw_j^{(N)}}{dt}(s+)=0\right\}} \psi(w_j^{(N)}(s-), s). \tag{3.2.15}$$

It follows immediately from (3.2.15) and the right continuity of the filtration $\{\mathcal{F}_t^{(N)}\}$ that for $t \in [0,\infty)$, $S_\psi^{(N)}$ is $\{\mathcal{F}_t^{(N)}\}$-adapted. Moreover, $S_\mathbf{1}^{(N)}$ is clearly equal to the cumulative potential reneging process $S^{(N)}$, i.e.,

$$S_{\mathbf{1}}^{(N)} = S^{(N)}. \tag{3.2.16}$$

In addition, using (3.2.8), (3.2.11) and the non-negativity of $Q^{(N)}$, $R^{(N)}$ and $\langle \mathbf{1}, \nu^{(N)} \rangle$, it follows that for any $t \in [0, \infty)$ and bounded, measurable φ,

$$\mathbb{E}\left[\left|D_{\varphi}^{(N)}(t)\right|\right] \leq \|\varphi\|_{\infty} \mathbb{E}\left[X^{(N)}(0) + E^{(N)}(t)\right] < \infty \tag{3.2.17}$$

and likewise, for each $t \in [0, \infty)$ and bounded measurable ψ, (3.2.9) shows that

$$\mathbb{E}\left[\left|S_{\psi}^{(N)}(t)\right|\right] \leq \|\psi\|_{\infty} \mathbb{E}\left[\langle \mathbf{1}, \eta_0^{(N)} \rangle + E^{(N)}(t)\right] < \infty. \tag{3.2.18}$$

Next, comparing the definition of the process $R^{(N)}$ with (3.2.15), it is clear that the cumulative reneging process $R^{(N)}$ satisfies

$$R^{(N)}(t) = S_{\theta^{(N)}}^{(N)}(t), \qquad t \geq 0, \tag{3.2.19}$$

where $\theta^{(N)}$ is given by

$$\theta^{(N)}(x, s) = \mathbb{I}_{[x,\infty)}(\chi^{(N)}(s-)), \qquad x \in \mathbb{R},\ s \geq 0. \tag{3.2.20}$$

The following result was established in [22, Theorem 2.1].

Theorem 3.2.1. *The processes $(E^{(N)}, X^{(N)}, \nu^{(N)}, \eta^{(N)})$ a.s. satisfy the following cou-*

pled set of equations: for $\varphi \in \mathcal{C}_c^1([0,H^s)\times\mathbb{R}_+)$ *and* $t\in[0,\infty)$,

$$\begin{aligned}\left\langle \varphi(\cdot,t),\nu_t^{(N)}\right\rangle &= \left\langle \varphi(\cdot,0),\nu_0^{(N)}\right\rangle + \int_0^t \left\langle \varphi_x(\cdot,s)+\varphi_s(\cdot,s),\nu_s^{(N)}\right\rangle\, ds \\ &\quad -D_\varphi^{(N)}(t) + \int_{[0,t]} \varphi(0,s)dK^{(N)}(s),\end{aligned} \tag{3.2.21}$$

for $\psi \in \mathcal{C}_c^1([0,H^r)\times\mathbb{R}_+)$ *and* $t\in[0,\infty)$,

$$\begin{aligned}\left\langle \psi(\cdot,t),\eta_t^{(N)}\right\rangle &= \left\langle \psi(\cdot,0),\eta_0^{(N)}\right\rangle + \int_0^t \left\langle \psi_x(\cdot,s)+\psi_s(\cdot,s),\eta_s^{(N)}\right\rangle\, ds \\ &\quad -S_\psi^{(N)}(t) + \int_{[0,t]} \psi(0,s)dE^{(N)}(s),\end{aligned} \tag{3.2.22}$$

$$X^{(N)}(t) = X^{(N)}(0)+E^{(N)}(t)-D_{\mathbf{1}}^{(N)}(t)-R^{(N)}(t), \tag{3.2.23}$$

$$N-\left\langle \mathbf{1},\nu_t^{(N)}\right\rangle = [N-X^{(N)}(t)]^+, \tag{3.2.24}$$

where $K^{(N)}$ *satisfies (3.2.10),* $R^{(N)}$ *satisfies (3.2.19) and* $D_\varphi^{(N)}$ *and* $S_\psi^{(N)}$ *are the processes defined in (3.2.13) and (3.2.15), respectively.*

3.3 Fluid Limits

Next, consider the following scaled versions of the basic processes described in Section 3.2. For each $N\in\mathbb{N}$, the scaled version of the processes $I = X,E,D,K,Q,R,S$ is given by

$$\overline{I}^{(N)} := \frac{I^{(N)}}{N}. \tag{3.3.1}$$

Similarly, the scaled versions of the measures ν and η are given by

$$\overline{\nu}_t^{(N)}(B) := \frac{\nu_t^{(N)}(B)}{N}, \qquad \overline{\eta}_t^{(N)}(B) := \frac{\eta_t^{(N)}(B)}{N}, \tag{3.3.2}$$

for $t \in [0,\infty)$ and any Borel subset B of $\mathbb{R}_+$.

Recall that $\mathcal{I}_{\mathbb{R}_+}[0,\infty)$ is the subset of non-decreasing functions $f \in \mathcal{D}_{\mathbb{R}_+}[0,\infty)$ with $f(0)=0$, $H^s = \sup\{x \in [0,\infty) : g^s(x) > 0\}$ and $H^r = \sup\{x \in [0,\infty) : g^r(x) > 0\}$. Define

$$\mathcal{S}_0 := \left\{ \begin{array}{c} (e,x,\nu,\eta) \in \mathcal{I}_{\mathbb{R}_+}[0,\infty) \times \mathbb{R}_+ \times \mathcal{M}_F[0,H^s) \times \mathcal{M}_F[0,H^r) : \\ 1 - \langle \mathbf{1}, \nu \rangle = [1-x]^+ \end{array} \right\}. \tag{3.3.3}$$

$\mathcal{S}_0$ serves as the space of possible input data for the fluid equations. We impose some natural assumptions on the sequence of initial conditions $(\overline{E}^{(N)}, \overline{X}^{(N)}(0), \overline{\nu}_0^{(N)}, \overline{\eta}_0^{(N)})$.

Assumption 3.3.1. **(Initial conditions)** There exists an $\mathcal{S}_0$-valued random variable $(\overline{E}, \overline{X}(0), \overline{\nu}_0, \overline{\eta}_0)$ such that, as $N \to \infty$, the following limits hold:

1. $\overline{E}^{(N)} \to \overline{E}$ in $\mathcal{D}_{\mathbb{R}_+}[0,\infty)$ $\mathbb{P}$-a.s., and $\mathbb{E}\left[\overline{E}^{(N)}(t)\right] \to \mathbb{E}\left[\overline{E}(t)\right] < \infty$ for every $t \in [0,\infty)$;
2. $\overline{X}^{(N)}(0) \to \overline{X}(0)$ in $\mathbb{R}_+$ $\mathbb{P}$-a.s.;
3. $\overline{\nu}_0^{(N)} \xrightarrow{w} \overline{\nu}_0$ in $\mathcal{M}_F[0,H^s)$;
4. $\overline{\eta}_0^{(N)} \xrightarrow{w} \overline{\eta}_0$ in $\mathcal{M}_F[0,H^r)$, and $\mathbb{E}\left[\langle 1, \overline{\eta}_0^{(N)} \rangle\right] \to \mathbb{E}[\langle 1, \overline{\eta}_0 \rangle] < \infty$.

The next assumption imposes some regularity conditions on $\overline{\eta}_0$ and $\overline{E}$.

Assumption 3.3.2. For each $t \geq 0$, if $\overline{\eta}_0(\{t\}) > 0$ then $\overline{\eta}_0(t, t+\varepsilon) > 0$ for every $\varepsilon > 0$ and if $\overline{E}(t) - \overline{E}(t-) > 0$, then $\overline{E}(t-) - \overline{E}(t-\varepsilon) > 0$ for every $\varepsilon > 0$.

Remark 3.3.3. Assumption 3.3.2 is trivially satisfied if $\overline{\eta}_0$ and $\overline{E}$ are continuous, that is, $\overline{\eta}_0(\{t\}) = 0$ for all $t \geq 0$ and the function $\overline{E}$ is continuous.

In order to state our last assumption, define the hazard rate functions of G^r and G^s in the usual manner:

$$h^r(x) := \frac{g^r(x)}{1 - G^r(x)}, \qquad x \in [0, H^r), \tag{3.3.4}$$

$$h^s(x) := \frac{g^s(x)}{1 - G^s(x)}, \qquad x \in [0, H^s). \tag{3.3.5}$$

It is easy to verify that h^r and h^s are locally integrable on $[0, H^r)$ and $[0, H^s)$, respectively.

Assumption 3.3.4. There exists $L^s < H^s$ such that h^s is either bounded or lower-semicontinuous on (L^s, H^s), and likewise, there exists $L^r < H^r$ such that h^r is either bounded or lower-semicontinuous on (L^r, H^r).

3.3.1 Fluid Equations

We now introduce the fluid equations.

Definition 3.3.5. **(Fluid Equations)** *The càdlàg function* $(\overline{X}, \overline{\nu}, \overline{\eta})$ *defined on* $[0, \infty)$ *and taking values in* $\mathbb{R}_+ \times \mathcal{M}_F[0, H^s) \times \mathcal{M}_F[0, H^r)$ *is said to solve the* fluid equations *associated with* $(\overline{E}, \overline{X}(0), \overline{\nu}_0, \overline{\eta}_0) \in \mathcal{S}_0$ *and the hazard rate functions* h^r *and* h^s *if and only if for every* $t \in [0, \infty)$,

$$\int_0^t \langle h^r, \overline{\eta}_s \rangle \, ds < \infty, \qquad \int_0^t \langle h^s, \overline{\nu}_s \rangle \, ds < \infty, \tag{3.3.6}$$

and the following relations are satisfied: for every $\varphi \in \mathcal{C}_c^1([0, H^s) \times \mathbb{R}_+)$,

$$\begin{aligned} \langle \varphi(\cdot, t), \overline{\nu}_t \rangle &= \langle \varphi(\cdot, 0), \overline{\nu}_0 \rangle + \int_0^t \langle \varphi_s(\cdot, s), \overline{\nu}_s \rangle \, ds + \int_0^t \langle \varphi_x(\cdot, s), \overline{\nu}_s \rangle \, ds \\ &\quad - \int_0^t \langle h^s(\cdot) \varphi(\cdot, s), \overline{\nu}_s \rangle \, ds + \int_0^t \varphi(0, s) \, d\overline{K}(s), \end{aligned} \tag{3.3.7}$$

where

$$\overline{K}(t) = \langle \mathbf{1}, \overline{\nu}_t \rangle - \langle \mathbf{1}, \overline{\nu}_0 \rangle + \int_0^t \langle h^s, \overline{\nu}_s \rangle \, ds; \tag{3.3.8}$$

for every $\psi \in \mathcal{C}_c^1([0, H^r) \times \mathbb{R}_+)$

$$\begin{aligned} \langle \psi(\cdot, t), \overline{\eta}_t \rangle &= \langle \psi(\cdot, 0), \overline{\eta}_0 \rangle + \int_0^t \langle \psi_s(\cdot, s), \overline{\eta}_s \rangle \, ds + \int_0^t \langle \psi_x(\cdot, s), \overline{\eta}_s \rangle \, ds \\ &\quad - \int_0^t \langle h^r(\cdot)\psi(\cdot, s), \overline{\eta}_s \rangle \, ds + \int_0^t \psi(0, s) \, d\overline{E}(s); \end{aligned} \tag{3.3.9}$$

$$\overline{Q}(t) = \overline{X}(t) - \langle \mathbf{1}, \overline{\nu}_t \rangle; \tag{3.3.10}$$

$$\overline{Q}(t) \leq \langle \mathbf{1}, \overline{\eta}_t \rangle; \tag{3.3.11}$$

$$\overline{R}(t) = \int_0^t \left(\int_0^{\overline{Q}(s)} h^r((F^{\overline{\eta}_s})^{-1}(y)) dy \right) ds, \tag{3.3.12}$$

where we recall that $F^{\overline{\eta}_t}(x) = \overline{\eta}_t[0, x]$;

$$\overline{X}(t) = \overline{X}(0) + \overline{E}(t) - \int_0^t \langle h^s, \overline{\nu}_s \rangle \, ds - \overline{R}(t); \tag{3.3.13}$$

and

$$1 - \langle \mathbf{1}, \overline{\nu}_t \rangle = [1 - \overline{X}(t)]^+. \tag{3.3.14}$$

3.3.2 Results on the Fluid Limit

The following result was established in [22, Theorem 3.5].

Theorem 3.3.6. *Given any* $(\overline{E}, \overline{X}(0), \overline{\nu}_0, \overline{\eta}_0) \in \mathcal{S}_0$, *there exists at most one solution* $(\overline{X}, \overline{\nu}, \overline{\eta})$ *to the associated fluid equations (3.3.6)–(3.3.14). Moreover, if* $\overline{\nu}$ *and* $\overline{\eta}$ *satisfy (3.3.6), then* $(\overline{X}, \overline{\nu}, \overline{\eta})$ *is a solution to the fluid equations if and only if, for* *every* $f \in \mathcal{C}_b(\mathbb{R}_+)$,

$$\begin{aligned}\int_{[0,H^r)} f(x)\,\overline{\eta}_t(dx) &= \int_{[0,H^r)} f(x+t)\frac{1-G^r(x+t)}{1-G^r(x)}\,\overline{\eta}_0(dx) \\ &\quad + \int_{[0,t]} f(t-s)(1-G^r(t-s))\,d\overline{E}(s),\end{aligned} \tag{3.3.15}$$

$$\begin{aligned}\int_{[0,H^s)} f(x)\,\overline{\nu}_t(dx) &= \int_{[0,H^s)} f(x+t)\frac{1-G^s(x+t)}{1-G^s(x)}\,\overline{\nu}_0(dx) \\ &\quad + \int_{[0,t]} f(t-s)(1-G^s(t-s))\,d\overline{K}(s),\end{aligned} \tag{3.3.16}$$

where

$$\overline{K}(t) = [\overline{X}(0)-1]^+ - [\overline{X}(t)-1]^+ + \overline{E}(t) - \int_0^t \left(\int_0^{[\overline{X}(s)-1]^+} h^r\left(\left(F^{\overline{\eta}_s}\right)^{-1}(y)\right) dy \right) ds \tag{3.3.17}$$

and for all $t \in [0,\infty)$, $\overline{X}$ *satisfies* $[\overline{X}(t)-1]^+ \leq \langle \mathbf{1}, \overline{\eta}_t \rangle$, *the non-idling condition (3.3.14) and*

$$\overline{X}(t) = \overline{X}(0) + \overline{E}(t) - \int_0^t \langle h^s, \overline{\nu}_s \rangle\, ds - \int_0^t \left(\int_0^{[\overline{X}(s)-1]^+} h^r\left(\left(F^{\overline{\eta}_s}\right)^{-1}(y)\right) dy \right) ds. \tag{3.3.18}$$

Moreover, $\overline{K}$ *also satisfies*

$$\begin{aligned}\overline{K}(t) &= \langle \mathbf{1}, \overline{\nu}_t \rangle - \langle \mathbf{1}, \overline{\nu}_0 \rangle + \int_{[0,H^s)} \frac{G^s(x+t-s) - G^s(x)}{1-G^s(x)}\,\overline{\nu}_0(dx) \\ &\quad + \int_0^t \left(\langle \mathbf{1}, \overline{\nu}_{t-s} \rangle - \langle \mathbf{1}, \overline{\nu}_0 \rangle + \int_{[0,H^s)} \frac{G^s(x+t-s) - G^s(x)}{1-G^s(x)}\,\overline{\nu}_0(dx) \right) u^s(s)\, ds,\end{aligned} \tag{3.3.19}$$

where u^s *is the density of the renewal function* U^s *associated with* G^s *(*u^s *exists since* G^s *is assumed to have a density).*

Next, we state the main result of [22, Theorem 3.6] which shows that, under fairly general conditions, a solution to the fluid equations exists and is the functional law of large numbers limit, as $N \to \infty$, of the N-server system with reneging.

Theorem 3.3.7. *Suppose that Assumptions 3.3.1–3.3.4 hold, and let* $(\overline{E}, \overline{X}(0), \overline{\nu}_0, \overline{\eta}_0) \in \mathcal{S}_0$ *be the limiting initial condition. Then there exists a unique solution* $(\overline{X}, \overline{\nu}, \overline{\eta})$ *to the associated fluid equations, and the sequence* $(\overline{X}^{(N)}, \overline{\nu}^{(N)}, \overline{\eta}^{(N)})$ *converges weakly, as* $N \to \infty$*, to* $(\overline{X}, \overline{\nu}, \overline{\eta})$*.*

In the fluid system, for any $u > t$ the total mass of customers in queue at time u that arrived before time t equals $\overline{Q}(u) - \overline{\eta}_u[0, u-t]$, and the ages of these (fluid) customers lie in the interval $(u - t, \overline{\chi}(u)]$, where

$$\overline{\chi}(u) := (F^{\overline{\eta}_u})^{-1}(\overline{Q}(u)). \tag{3.3.20}$$

Let ν_* and η_* be the probability measures defined as follows:

$$\nu_*[0, x) := \int_0^x (1 - G^s(y))dy, \qquad x \in [0, H^s), \tag{3.3.21}$$

and

$$\eta_*[0, x) := \int_0^x (1 - G^r(y))dy, \qquad x \in [0, H^r). \tag{3.3.22}$$

For $\lambda \geq 1$, define the set B_λ as follows:

$$B_\lambda := \left\{ x \in [1, \infty) : G^r((F^{\lambda\eta_*})^{-1}((x-1)^+)) = \frac{\lambda - 1}{\lambda} \right\}. \tag{3.3.23}$$

Let I_λ be the set of states defined by

$$I_\lambda := \begin{cases} \{(\lambda, \lambda\nu_*, \lambda\eta_*)\}, & \text{if } \lambda < 1, \\ \{(x_*, \nu_*, \lambda\eta_*) : x_* \in B_\lambda\}, & \text{if } \lambda \geq 1. \end{cases} \tag{3.3.24}$$

Suppose I_λ satisfies following assumption.

Assumption 3.3.8. The set I_λ has a single element.

Under the above assumption, it was proved in [23, Theorem 5.5] that I_λ is the invariant manifold of the fluid limit associated with arrival rate λ is defined as $\overline{E}(t) = \lambda t$(as is explained later in Remark 3.6.2). Since I_λ has a single element, the set B_λ is a singleton, which in turn implies that the invariant queue length, Q_*, is unique. Thus, $(F^{\lambda\eta_*})^{-1}(Q_*)$ is a unique constant. Define

$$\chi_* = (F^{\lambda\eta_*})^{-1}(Q_*) \tag{3.3.25}$$

3.4 A Family of Martingales

In Section 3.4.1, we identify the compensators (with respect to the filtration $\mathcal{F}_t^{(N)}$) of the cumulative departure, potential reneging and (actual) reneging processes.

3.4.1 Compensators

Here we summarize some relevant results from [22]. For any bounded measurable function φ on $[0, H^s) \times \mathbb{R}_+$, consider the sequence $\{A^{(N)}_{\varphi,\nu}\}$ of processes given by

$$A^{(N)}_{\varphi,\nu}(t) := \int_0^t \left(\int_{[0,H^s)} \varphi(x,s) h^s(x)\, \nu_s^{(N)}(dx) \right) ds, \qquad t \in [0,\infty). \tag{3.4.1}$$

Likewise, for any bounded measurable function φ on $[0, H^r) \times \mathbb{R}_+$ and $N \in \mathbb{N}$, let

$$A^{(N)}_{\varphi,\eta}(t) := \int_0^t \left(\int_{[0,H^r)} \varphi(x,s) h^r(x)\, \eta_s^{(N)}(dx) \right) ds, \qquad t \in [0,\infty). \tag{3.4.2}$$

In [22, Proposition 5.1] stated below, it was shown that $A^{(N)}_{\varphi,\nu}$ (respectively, $A^{(N)}_{\varphi,\eta}$) is the $\mathcal{F}^{(N)}_t$-compensator of the associated "φ-weighted" cumulative departure process $D^{(N)}_\varphi$ (respectively, $S^{(N)}_\varphi$).

Proposition 3.4.1. The following properties hold.

1. For every bounded measurable function φ on $[0, H^s) \times \mathbb{R}_+$, the process $M^{(N)}_{\varphi,\nu}$ defined by

$$M^{(N)}_{\varphi,\nu} := D^{(N)}_\varphi - A^{(N)}_{\varphi,\nu} \tag{3.4.3}$$

is a local $\mathcal{F}^{(N)}_t$-martingale. Moreover, for every $N \in \mathbb{N}$, $t \in [0, \infty)$ and $m \in [0, H^s)$,

$$|A^{(N)}_{\varphi,\nu}(t)| \le \|\varphi\|_\infty \left(X^{(N)}(0) + E^{(N)}(t)\right) \left(\int_0^m h^s(x)\, dx\right) < \infty \tag{3.4.4}$$

for every $\varphi \in \mathcal{C}_c([0, H^s) \times \mathbb{R}_+)$ with $\text{supp}(\varphi) \subset [0, m] \times \mathbb{R}_+$. In addition, the quadratic variation process $\langle \overline{M}^{(N)}_{\varphi,\nu} \rangle$ of the scaled process $\overline{M}^{(N)}_{\varphi,\nu} := M^{(N)}_{\varphi,\nu}/N$ satisfies

$$\lim_{N\to\infty} \mathbb{E}\left[\langle \overline{M}^{(N)}_{\varphi,\nu} \rangle(t)\right] = 0; \qquad \overline{M}^{(N)}_{\varphi,\nu} \Rightarrow \mathbf{0} \text{ as } N \to \infty. \tag{3.4.5}$$

2. Furthermore, properties (3.4.3)–(3.4.5) also hold with D, a_j, ν, H^s and h^s, respectively, replaced by S, w_j, η, H^r and h^r.

We now reproduce Remark 5.2 [22].

Remark 3.4.2. *It is easy to see that Lemmas 5.6 and 5.8 of [24] continue to be valid in the presence of abandonments. Indeed, the proofs of Lemmas 5.6 and 5.8 of [24] only depend on Assumption 1 and Corollary 5.5 therein (since, as shown in Lemma 5.12 of [24], the additional conditions (5.32) and (5.33) of Lemma 5.8 of [24] can be* *derived from Assumption 1), which correspond to Assumption 3.3.1 and Proposition 3.4.1 of this chapter. In addition, due to the parallels between the dynamics of $\nu^{(N)}$ and $\eta^{(N)}$ (see Remark 22 [22]), the analogs of the results in Lemmas 5.6 and 5.8, with $D^{(N)}, \nu^{(N)}, G^s$ and H^s, respectively, replaced by $S^{(N)}, \eta^{(N)}, G^r$ and H^r, also hold. In this case, even though $\eta_0^{(N)}$ is (unlike $\nu_0^{(N)}$) not necessarily a sub-probability measure, the verification of the conditions analogous to (5.32) and (5.33) of Lemma 5.8 in [24] can still be carried out in the same manner since Assumption 3.3.1 implies that the sequence $\{\langle \mathbf{1}, \eta_0^{(N)} \rangle\}$ is tight. Moreover, even though G^r is allowed to have a mass at ∞, the proofs of Lemmas 5.6 and 5.8 are still valid, with the renewal function U^s now replaced by the function $U^r(\cdot) = \int_0^{\cdot} \sum_{n=0}^{\infty} (g^r)^{*n}(s)\, ds$, where $(g^r)^{*n}$ is the nth convolution of g^r on $[0, \infty)$.*

Now, note from (3.2.19) that $R^{(N)} = S^{(N)}_{\theta^{(N)}}$, where $\theta^{(N)}$ is defined by (3.2.20). In view of the fact that $A^{(N)}_{\varphi,\eta}$ is the compensator for $S^{(N)}_{\varphi}$, it was shown in [22, Lemma 5.4] (stated below) that the compensator of $R^{(N)}$ is equal to $A^{(N)}_{\theta^{(N)},\eta}$, where

$$A^{(N)}_{\theta^{(N)},\eta}(t) := \int_0^t \left(\int_{[0,H^r)} \mathbb{I}_{[0,\chi^{(N)}(s-)]}(x) h^r(x)\, \eta_s^{(N)}(dx) \right) ds, \qquad t \in [0, \infty). \tag{3.4.6}$$

Lemma 3.4.3. *For every $N \in \mathbb{N}$, the process $M^{(N)}_{\theta^{(N)},\eta}$ defined by*

$$M^{(N)}_{\theta^{(N)},\eta} := R^{(N)} - A^{(N)}_{\theta^{(N)},\eta} \tag{3.4.7}$$

is a local $\mathcal{F}_t^{(N)}$-martingale. In addition, as $N \to \infty$,

$$\lim_{N\to\infty} \mathbb{E}\left[\langle \overline{M}^{(N)}_{\theta^{(N)},\eta} \rangle(t) \right] = 0, \quad \overline{M}^{(N)}_{\psi,\eta} \Rightarrow \mathbf{0} \quad \text{and} \quad \overline{M}^{(N)}_{\theta^{(N)},\eta} \Rightarrow \mathbf{0}. \tag{3.4.8}$$

In what follows, recall that $F^{\eta_t^{(N)}}(x) = \eta_t^{(N)}[0, x]$ and its inverse $(F^{\eta_t^{(N)}})^{-1}$ is as defined in (3.1.1). The following result was established in [22, Proposition 5.5].

Proposition 3.4.4. For each $N \in \mathbb{N}$, $t \geq 0$ and measurable function h on $[0, H^r)$,

$$\int_{[0,H^r)} \mathbb{I}_{[0,\chi^{(N)}(t-)]}(x)h(x)\eta_t^{(N)}(dx) = \int_0^{Q^{(N)}(t)+\iota^{(N)}(t)} h((F^{\eta_t^{(N)}})^{-1}(y))dy, \tag{3.4.9}$$

where

$$\iota^{(N)}(t) := \begin{cases} 0 & \text{if } (\chi^{(N)}(t-) - \chi^{(N)}(t))(K^{(N)}(t) - K^{(N)}(t-)) = 0, \\ 1 & \text{if } (\chi^{(N)}(t-) - \chi^{(N)}(t))(K^{(N)}(t) - K^{(N)}(t-)) > 0. \end{cases} \tag{3.4.10}$$

The proposition above shows an alternative, more convenient, representation for $A^{(N)}_{\theta^{(N)},\eta}$, or more generally, for processes of the form $A^{(N)}_{\theta^{(N)},\eta}$, but with h^r replaced by an arbitrary measurable function h. From the discussion below [22, Theorem 7.1] we have $(\overline{A}^{(N)}_{\cdot,\nu}, \overline{A}^{(N)}_{\cdot,\eta}) \to (\overline{A}_{\cdot,\nu}, \overline{A}_{\cdot,\eta})$ $\mathbb{P}$-a.s, where for any bounded measurable function φ on $[0, H^s) \times \mathbb{R}_+$,

$$\overline{A}_{\varphi,\nu}(t) := \int_0^t \left(\int_{[0,H^s)} \varphi(x,s)h^s(x)\,\overline{\nu}_s(dx) \right) ds, \qquad t \in [0,\infty). \tag{3.4.11}$$

and for any bounded measurable function φ on $[0, H^r) \times \mathbb{R}_+$ and

$$\overline{A}_{\varphi,\eta}(t) := \int_0^t \left(\int_{[0,H^r)} \varphi(x,s)h^r(x)\,\overline{\eta}_s(dx) \right) ds, \qquad t \in [0,\infty). \tag{3.4.12}$$

3.4.2 Certain martingale measures and their stochastic integrals

From the proof of Proposition 3.4.1 (which uses the proof of Lemma 5.9 of [24]) it follows that for any bounded and measurable φ on $[0, H^s) \times [0, \infty)$, the $\{\mathcal{F}^{(N)}_t\}$-

predictable quadratic variation of $M^{(N)}_{\varphi,\nu}$ takes the form

$$\langle M^{(N)}_{\varphi,\nu}\rangle = A^{(N)}_{\varphi,\nu}(t) \tag{3.4.13}$$

$$= \int_0^t \left(\int_{[0,H^s)} \varphi^2(x,s) h^s(x) \nu^{(N)}_s(dx) \right) ds, \qquad t \in [0,\infty), \tag{3.4.14}$$

and for any bounded and measurable ψ on $[0,H^r) \times [0,\infty)$, the $\{\mathcal{F}^{(N)}_t\}$- predictable quadratic variation of $M^{(N)}_{\psi,\eta}$ takes the form

$$\langle M^{(N)}_{\psi,\eta}\rangle = A^{(N)}_{\psi,\eta}(t) \tag{3.4.15}$$

$$= \int_0^t \left(\int_{[0,H^r)} \psi^2(x,s) h^r(x) \eta^{(N)}_s(dx) \right) ds, \qquad t \in [0,\infty), \tag{3.4.16}$$

Now, for $B \in \mathcal{B}[0,H^s)$ and $t \in [0,\infty)$, define,

$$\mathcal{M}^{(N)}_{t,\nu}(B) := M^{(N)}_{1_B,\nu}(t) = D^{(N)}_{1_B}(t) - A^{(N)}_{1_B,\nu}(t). \tag{3.4.17}$$

Similarly for $\tilde{B} \in \mathcal{B}[0,H^r)$ and $t \in [0,\infty)$, define,

$$\mathcal{M}^{(N)}_{t,\eta}(\tilde{B}) := M^{(N)}_{1_{\tilde{B}},\eta}(t) = S^{(N)}_{1_{\tilde{B}}}(t) - A^{(N)}_{1_{\tilde{B}},\eta}(t). \tag{3.4.18}$$

Let $\mathcal{B}_0[0,H^s)$ and $\mathcal{B}_0[0,H^r)$ denote the algebra generated by the intervals $[0,x]$, $x \in [0,H^s)$ and $[0,y]$, $y \in [0,H^r)$ respectively. It is easy to verify that $M^{(N)}_\nu = \{\mathcal{M}^{(N)}_{t,\nu}(B), \mathcal{F}^{(N)}_t, t \geq 0, B \in \mathcal{B}_0[0,H^s)\}$ and $M^{(N)}_\eta = \{\mathcal{M}^{(N)}_{t,\eta}(B), \mathcal{F}^{(N)}_t, t \geq 0, B \in \mathcal{B}_0[0,H^r)\}$ are martingale measures (the proof is similar to the one in Lemma A.1 of [25]). We now show that $\mathcal{M}^{(N)}_\nu$ is in fact an orthogonal martingale measure (see page 288 of Walsh [44] for a definition). Essentially, the orthogonality property holds because almost surely, no two departures or abandonments occur at the same time. In Lemma 3.4.5 below, we first state a slight generalization of this latter property, which will also be used to prove the orthogonality property of $\mathcal{M}^{(N)}_\eta$. Define the jump of a function f at time t by $\Delta f(t) = f(t) - f(t-)$. We now prove the following lemmas.

Lemma 3.4.5. *For every* $N \in \mathbb{N}, \mathbb{P}$- *a.s.*

$$\Delta D^{(N)}(t) \leq 1, \qquad t \in [0, \infty), \tag{3.4.19}$$

$$\Delta S^{(N)}(t) \leq 1, \qquad t \in [0, \infty). \tag{3.4.20}$$

Proof. The proof follows by the strong Markov property of the N-server state process from Lemma B.1 [22] and independence of the service times and patience times of the customers, using the same idea as in the proof of equation (4.6) of Lemma 4.2 [25]. □

Before the orthogonality property, we now prove another lemma that will be used in proving an asymptotic independence result in Section 3.7

Lemma 3.4.6. *As* $N \to \infty$

$$\frac{1}{N}\mathbb{E}\left[\sum_{s\leq t} \Delta D^{(N)}(s)\Delta S^{(N)}(s)\right] \to 0. \tag{3.4.21}$$

Proof. Let $D^{(N),r}(s)$ denote the cumulative number of departures during $(r, r+s]$ of customers that entered service at or before time r, and let $D^{(N)+,r}(s)$ be the cumulative number of departures during $(r, r+s]$ of customers that entered service after time r.

For $\delta > 0$, and $k = 0, 1, 2, \cdots$, we have

$$\begin{aligned}\sum_{s\in(k\delta,(k+1)\delta]} \Delta D^{(N)}(s)\Delta S^{(N)}(s) &= \sum_{s\in(0,\delta]} \Delta D^{(N),k\delta}(s)\Delta S^{(N)}(k\delta+s) \\ &\quad + \sum_{s\in(0,\delta]} \Delta D^{(N)+,k\delta}(s)\Delta S^{(N)}(k\delta+s). \end{aligned} \tag{3.4.22}$$

Claim: $\sum_{s\in(0,\infty]} \Delta D^{(N),r}(s)\Delta S^{(N)}(r+s) = 0, \qquad r \in [0,\infty)$

Define $J^{(N),r}$ to be the jump times of $D^{(N),r}$ in $[0,\infty)$ and let $\mathcal{G}_t^{(N)} = \mathcal{F}_{r+t}^{(N)}$. Using the same idea as in the proof of Proposition 3.4.1 (see Proposition 5.1(1)[22] for proof), $\{\Delta D^{(N),r}(t)\}$ has a continuous $\{\mathcal{G}_t^{(N)}\}$ compensator given by

$$\int_0^t \left(\int_{[0,H^s)} \frac{g^s(x+s)}{1-G^s(x)}\nu_r^{(N)}(dx)\right) ds, \qquad t \in [0,\infty),$$

and hence has no fixed jump times, that is, $\mathbb{P}(t \in J^{(N),r}|\mathcal{F}_r^{(N)}) = 0$ for every $t \in [0,\infty)$. Moreover, due to the assumption of independence of the service times and the patience times, $\{S^{(N)}(r+t) - S^{(N)}(r)\}_{t\geq 0}$ and $\{D^{(N),r}(t)\}_{t\geq 0}$ are conditionally independent, given $\mathcal{F}_r^{(N)}$.

Let $\mathcal{T} = \{\bar{t} = (t_1, \cdots, t_m, ...) \in [0,\infty)^\infty : 0 \leq t_1 \leq t_2 \leq \cdots\}$, and let $\overline{T}^{(N),r}$ denote the random $\mathcal{T}$-valued sequence of times after r at which $S^{(N)}$ has a jump. Moreover, let $\mu^{(N)}$ denote the conditional probability distribution of $\overline{T}^{(N),r}$, given $\mathcal{F}_r^{(N)}$. For any $\bar{t} \in \mathcal{T}$, using the fact that $0 \leq \Delta S^{(N)}(t) \leq 1$ from Lemma 3.4.5, the conditional independence of $\{D^{(N),r}(t), t \geq 0\}$ from $\overline{T}^{(N),r}$ given $\mathcal{F}_r^{(N)}$ and the property established above, we have

$$\begin{aligned}
\mathbb{E}[\sum_{s\in(0,\infty]} \Delta D^{(N),r}(s)\Delta S^{(N)}(r+s)|\mathcal{F}_r^{(N)}, \overline{T}^{(N),r} = \overline{t}] &= \sum_{t\in\overline{T}^{(N)}} \mathbb{E}[\Delta D^{(N),r}(t)|\mathcal{F}_r^{(N)}, \overline{T}^{(N),r} = \overline{t}] \\
&= \sum_{t\in\overline{T}^{(N)}} \mathbb{E}[\Delta D^{(N),r}(t)|\mathcal{F}_r^{(N)}] \\
&= \sum_{t\in\overline{T}^{(N)}} \mathbb{P}(t\in J^{(N),r}|\mathcal{F}_r^{(N)}) = 0.
\end{aligned}$$

Integrating the left-hand side above with respect to the conditional distribution $\mu^{(N)}$ and then taking expectation, it follows that

$$\mathbb{E}[\sum_{s\in(0,\infty]} \Delta D^{(N),r}(s)\Delta S^{(N)}(r+s)] = 0.$$

Since the term inside is non-negative, the claim is proved.

Hence the first term in the right-hand side of (3.4.22) is 0. Since $\Delta S^{(N)}(s)\le 1$ by Lemma 3.4.5, the second term can be bounded as follows,

$$\begin{aligned}
\sum_{s\in(k\delta,(k+1)\delta]} \Delta D^{(N)}(s)\Delta S^{(N)}(s) &\le \sum_{s\in(0,\delta]} \Delta D^{(N)+,k\delta}(s) \\
&\le \sum_{j=K^{(N)}(k\delta)+1}^{K^{(N)}((k+1)\delta)} 1_{\{v_j\le\delta\}}.
\end{aligned}$$

Summing the above equation over $k=0,1,\cdots,\lfloor t/\delta\rfloor$ and dividing by N, we obtain

$$\begin{aligned}
\frac{1}{N}\mathbb{E}[\sum_{s\le t} \Delta D^{(N)}(s)\Delta S^{(N)}(s)] &\le \mathbb{E}\left[\int_0^{t+\delta} 1_{\{\nu_{K^{(N)}(s)}\le\delta\}} dK^{(N)}(s)\right] \\
&\le \mathbb{E}\left[\overline{D}^{(N)}_{1_{(0,\delta]}}(t+2\delta)\right] \\
&= \mathbb{E}\left[\overline{A}^{(N)}_{1_{(0,\delta]},\nu}(t+2\delta)\right].
\end{aligned}$$

On both sides taking the limit supremum as $N\to\infty$, and then the limit as $\delta\downarrow 0$, we obtain

$$\limsup_{N\to\infty} \frac{1}{N}\mathbb{E}[\sum_{s\le t} \Delta D^{(N)}(s)\Delta S^{(N)}(s)] \le \lim_{\delta\downarrow 0} \limsup_{N\to\infty} \mathbb{E}\left[\overline{A}^{(N)}_{1_{(0,\delta]},\nu}(t+2\delta)\right] = 0,$$

where the last equality follows from (3.4.4). As $\Delta D^{(N)}(s), \Delta S^{(N)}(s) \ge 0$, we are done. □

Corollary 3.4.7. 1. For each $N \in \mathbb{N}$, the martingale measure

$$M^{(N)}_\nu = \{\mathcal{M}^{(N)}_{t,\nu}(B), \mathcal{F}^{(N)}_t, t \ge 0, B \in \mathcal{B}_0[0, H^s)\},$$

is orthogonal and has covariance functional

$$\langle \mathcal{M}^{(N)}_\nu(B), \mathcal{M}^{(N)}_\nu(\tilde{B})\rangle_t = A^{(N)}_{1_{B\cap\tilde{B}},\nu}(t) = \int_0^t \left(\int_{B\cap\tilde{B}} h^s(x)\nu^{(N)}_s(dx)\right) ds, \quad (3.4.23)$$

for $B, \tilde{B} \in \mathcal{B}_0[0, H^s)$.

2. For each $N \in \mathbb{N}$, the martingale measure

$$M^{(N)}_\eta = \{\mathcal{M}^{(N)}_{t,\eta}(B), \mathcal{F}^{(N)}_t, t \ge 0, B \in \mathcal{B}_0[0, H^r)\},$$

is orthogonal and has covariance functional

$$\langle \mathcal{M}^{(N)}_\eta(B), \mathcal{M}^{(N)}_\eta(\tilde{B})\rangle_t = A^{(N)}_{1_{B\cap\tilde{B}},\eta}(t) = \int_0^t \left(\int_{B\cap\tilde{B}} h^r(x)\eta^{(N)}_s(dx)\right) ds, \quad (3.4.24)$$

for $B, \tilde{B} \in \mathcal{B}_0[0, H^r)$.

Proof. 1. Using (3.4.19), the proof is analogous to the proof of [24, Corollary 4.3].

2. Using (3.4.20), the proof is analogous to the proof of [24, Corollary 4.3].

□

The orthogonality property established in Corollary 3.4.7 allows us to define stochastic integrals with respect to the martingale measures $\mathcal{M}_{\nu}^{(N)}$ and $\mathcal{M}_{\eta}^{(N)}$. The stochastic integral is defined for a large class of so-called predictable integrands satisfying a suitable integrability property (see page 292 of Walsh [44]) which, since $\mathbb{E}[A_{\mathbf{1},\nu}^{(N)}(T)], \mathbb{E}[A_{\mathbf{1},\eta}^{(N)}(T)] < \infty$ by Remark 3.4.2 and Lemma 5.6 of [24], and $\nu^{(N)}$ and $\eta^{(N)}$ are finite non-negative measures, includes the class of deterministic functions in $\mathcal{C}_b([0,H^s)\times[0,\infty))$ and $\mathcal{C}_b([0,H^r)\times[0,\infty))$ respectively. Moreover, by Theorem 2.5 on page 295 of Walsh [44], it follows that for all $\varphi \in \mathcal{C}_b([0,H^s)\times[0,\infty))$,the stochastic integral $\{\mathcal{M}_{t,\nu}^{(N)}(\varphi)(B), \{\mathcal{F}^{(N)}\}; t \geq 0, B \in \mathcal{B}[0,H^s)\}$ of φ with respect to $\mathcal{M}_{\nu}^{(N)}$ is a càdlàg orthogonal martingale measure with covariance functional

$$\langle \mathcal{M}_{\nu}^{(N)}(\varphi)(B), \mathcal{M}_{\nu}^{(N)}(\tilde{\varphi})(\tilde{B})\rangle_t = \int_0^t \left(\int_{B\cap\tilde{B}} \varphi(x,s)\tilde{\varphi}(x,s)h^s(x)\nu_s^{(N)}(dx) \right) ds \tag{3.4.25}$$

for bounded, continuous φ, $\tilde{\varphi}$ and $B, \tilde{B} \in \mathcal{B}_0[0,H^s)$. Similarly for all $\psi \in \mathcal{C}_b([0,H^r)\times[0,\infty))$,the stochastic integral $\{\mathcal{M}_{t,\eta}^{(N)}(\psi)(B), \{\mathcal{F}^{(N)}\}; t \geq 0, B \in \mathcal{B}[0,H^r)\}$ of ψ with respect to $\mathcal{M}_{\eta}^{(N)}$ is a càdlàg orthogonal martingale measure with covariance functional

$$\langle \mathcal{M}_{\eta}^{(N)}(\psi)(B), \mathcal{M}_{\eta}^{(N)}(\tilde{\psi})(\tilde{B})\rangle_t = \int_0^t \left(\int_{B\cap\tilde{B}} \psi(x,s)\tilde{\psi}(x,s)h^r(x)\eta_s^{(N)}(dx) \right) ds \tag{3.4.26}$$

for bounded, continuous ψ, $\tilde{\psi}$ and $B, \tilde{B} \in \mathcal{B}_0[0,H^r)$. When $B = [0,H^s)$, we will drop the dependence on B and simply write

$$\mathcal{M}_{\nu}^{(N)}(\varphi) = \mathcal{M}_{\nu}^{(N)}(\varphi)([0,H^s)),$$

and when $B = [0,H^r)$, we similarly write

$$\mathcal{M}_{\eta}^{(N)}(\psi) = \mathcal{M}_{\eta}^{(N)}(\psi)([0,H^r)),$$

Remark 3.4.8. For $\varphi \in \mathcal{C}_b([0,H^s)\times[0,\infty))$, the stochastic integral $\mathcal{M}_\nu^{(N)}(\varphi)$ admits a càdlàg version. Indeed, the càdlàg martingale $M_{\varphi,\nu}^{(N)}$ defined in (3.4.3) is a version of the stochastic integral $\mathcal{M}_\nu^{(N)}(\varphi)$. Also, for $\psi \in \mathcal{C}_b([0,H^r)\times[0,\infty))$, the stochastic integral $\mathcal{M}_\eta^{(N)}(\psi)$ admits a càdlàg version, and the càdlàg martingale $M_{\psi,\eta}^{(N)}$ defined as the analogue of (3.4.3) in Proposition 3.4.1 is a version of the stochastic integral $\mathcal{M}_\eta^{(N)}(\psi)$.

Let $\tilde{\mathbf{0}}$ denote the zero Radon measure. It was shown in Proposition 3.4.1 that

$$\overline{\mathcal{M}}_\nu^{(N)} := \frac{\mathcal{M}_\nu^{(N)}}{N} \to \overline{\mathcal{M}}_\nu \equiv \tilde{\mathbf{0}},$$

in the space $\mathcal{D}_{\mathcal{M}_F[0,H^s)}[0,\infty)$, and

$$\overline{\mathcal{M}}_\eta^{(N)} := \frac{\mathcal{M}_\eta^{(N)}}{N} \to \overline{\mathcal{M}}_\eta \equiv \tilde{\mathbf{0}},$$

in the space $\mathcal{D}_{\mathcal{M}_F[0,H^r)}[0,\infty)$. Now, let $\widehat{M}_\nu^{(N)}$ and $\widehat{\mathcal{M}}_\eta^{(N)}$ be the diffusion-scaled version of the processes

$$\widehat{M}_\nu^{(N)} := \frac{\widehat{\mathcal{M}}_\nu^{(N)}}{\sqrt{N}}, \qquad \widehat{\mathcal{M}}_\eta^{(N)} := \frac{\widehat{\mathcal{M}}_\eta^{(N)}}{\sqrt{N}}. \tag{3.4.27}$$

It is clear from the above discussion that each $\widehat{\mathcal{M}}_\nu^{(N)}$ is an orthogonal martingale measure with covariance functional $\int_0^t \left(\int_{B\cap\tilde{B}} h^s(x)\overline{\nu}_s^{(N)}(dx)\right) ds$, and each $\widehat{\mathcal{M}}_\eta^{(N)}$ is an orthogonal martingale measure with covariance functional $\int_0^t \left(\int_{B\cap\tilde{B}} h^r(x)\overline{\eta}_s^{(N)}(dx)\right) ds$. Also for any φ,ψ in suitable classes of functions that include the spaces $\mathcal{C}_b([0,H^s)\times[0,\infty))$ and $\mathcal{C}_b([0,H^r)\times[0,\infty))$ respectively, the stochastic integrals $\widehat{\mathcal{M}}_\nu^{(N)}(\varphi)$ and $\widehat{\mathcal{M}}_\eta^{(N)}(\psi)$ are well-defined càdlàg, orthogonal $\{\mathcal{F}_t^{(N)}\}$ martingale measures. Moreover, for every $\varphi,\tilde{\varphi} \in \mathcal{C}_b([0,H^s)\times[0,\infty))$ and for every $\psi,\tilde{\psi} \in \mathcal{C}_b([0,H^r)\times[0,\infty))$

and $t \in [0, \infty)$,

$$\langle \widehat{\mathcal{M}}_{\nu}^{(N)}(\varphi), \widehat{\mathcal{M}}_{\nu}^{(N)}(\tilde{\varphi}) \rangle_t = \int_0^t \left(\int_{[0,H^s)} \varphi(x,s)\tilde{\varphi}(x,s) h^s(x) \overline{\nu}_s^{(N)}(dx) \right) ds, \tag{3.4.28}$$

and

$$\langle \widehat{\mathcal{M}}_{\eta}^{(N)}(\psi), \widehat{\mathcal{M}}_{\eta}^{(N)}(\tilde{\psi}) \rangle_t = \int_0^t \left(\int_{[0,H^r)} \psi(x,s)\tilde{\psi}(x,s) h^r(x) \overline{\eta}_s^{(N)}(dx) \right) ds. \tag{3.4.29}$$

3.4.3 Some associated stochastic convolution integrals

We now define stochastic convolution integrals with respect to the martingale measures $\mathcal{M}_{\nu}^{(N)}$ and $\mathcal{M}_{\eta}^{(N)}$. For $N \in \mathbb{N}$, $\varphi \in \mathcal{C}_b([0, H^s) \times [0, \infty))$ and $t \in [0, \infty)$, define

$$\mathcal{H}_{t,\nu}^{(N)}(\varphi) := \int\int_{[0,H^s)\times[0,t]} \varphi(x+t-s, s) \frac{1 - G^s(x+t-s)}{1 - G^s(x)} \widehat{M}_{\nu}^{(N)}(dx, ds), \tag{3.4.30}$$

and for $N \in \mathbb{N}$, $\psi \in \mathcal{C}_b([0, H^r) \times [0, \infty))$ and $t \in [0, \infty)$, define

$$\mathcal{H}_{t,\eta}^{(N)}(\psi) := \int\int_{[0,H^r)\times[0,t]} 1_{[0,\chi_*]}(x) \varphi(x+t-s, s) \frac{1 - G^r(x+t-s)}{1 - G^r(x)} \widehat{M}_{\eta}^{(N)}(dx, ds). \tag{3.4.31}$$

For each $t \in [0, \infty)$, the stochastic integrals with respect to $\mathcal{M}_{\nu}^{(N)}$ and $\mathcal{M}_{\eta}^{(N)}$ in (3.4.30) and (3.4.31) are well defined because $\mathcal{M}_{\nu}^{(N)}$ and $\mathcal{M}_{\eta}^{(N)}$ are orthogonal martingale measures and the functions $(x, s) \mapsto \varphi(x+t-s, s)(1-G^s(x+t-s))/(1-G^s(x))$ and $(x, s) \mapsto 1_{[0,\chi_*]}(x)\psi(x+t-s, s)(1-G^r(x+t-s))/(1-G^r(x))$ lie in $\mathcal{C}_b([0, H^s) \times [0, \infty))$ and $\mathcal{C}_b([0, \chi_*] \times [0, \infty))$ respectively for all $\varphi \in \mathcal{C}_b([0, H^s) \times [0, \infty))$ and $\psi \in \mathcal{C}_b([0, H^r) \times [0, \infty))$. The scaled version of these quantities are then defined in the obvious manner: for $N \in \mathbb{N}$, $\varphi \in \mathcal{C}_b([0,H^s) \times [0,\infty))$ and $t \in [0,\infty)$, let

$$\begin{aligned}\widehat{\mathcal{H}}^{(N)}_{t,\nu}(\varphi) &:= \frac{\mathcal{H}^{(N)}_{t,\nu}(\varphi)}{\sqrt{N}} \\ &= \int\int_{[0,H^s)\times[0,t]} \varphi(x+t-s,s)\frac{1-G^s(x+t-s)}{1-G^s(x)}\widehat{\mathcal{M}}^{(N)}_{\nu}(dx,ds), \end{aligned} \tag{3.4.32}$$

and for $N \in \mathbb{N}$, $\psi \in \mathcal{C}_b([0,H^r) \times [0,\infty))$ and $t \in [0,\infty)$, let

$$\begin{aligned}\widehat{\mathcal{H}}^{(N)}_{t,\eta}(\psi) &:= \frac{\mathcal{H}^{(N)}_{t}(\psi)}{\sqrt{N}} \\ &= \int\int_{[0,H^s)\times[0,t]} 1_{[0,\chi_*]}(x)\psi(x+t-s,s)\frac{1-G^r(x+t-s)}{1-G^r(x)}\widehat{\mathcal{M}}^{(N)}_{\eta}(dx,ds), \end{aligned} \tag{3.4.33}$$

3.4.4 Related limit quantities

We now define some additional quantities, which we subsequently show (in Propositions 3.7.5 and 3.10.9) to be limits of the sequences $\{\widehat{\mathcal{M}}^{(N)}_{\nu}\}_{N\in\mathbb{N}}$, $\{\widehat{\mathcal{M}}^{(N)}_{\eta}\}_{N\in\mathbb{N}}$, $\{\widehat{\mathcal{H}}^{(N)}_{\nu}\}_{N\in\mathbb{N}}$ and $\{\widehat{\mathcal{H}}^{(N)}_{\eta}\}_{N\in\mathbb{N}}$. Let $\widehat{\mathcal{M}}_{\nu} = \{\widehat{\mathcal{M}}_{t,\nu}(B), B \in \mathcal{B}_0[0,H^s), t \in [0,\infty)\}$ be a continuous martingale measure with the deterministic covariance functional

$$\langle \widehat{\mathcal{M}}_{\nu}(B), \widehat{\mathcal{M}}_{\nu}(\tilde{B})\rangle_t = \int_0^t \left(\int_{[0,H^s)} 1_{B\cap\tilde{B}}(x)(\lambda\wedge 1)g^s(x)(dx)\right) ds, \tag{3.4.34}$$

and $\widehat{\mathcal{M}}_{\eta} = \{\widehat{\mathcal{M}}_{t,\eta}(B), B \in \mathcal{B}_0[0,H^r), t \in [0,\infty)\}$ be a continuous martingale measure with the deterministic covariance functional

$$\langle \widehat{\mathcal{M}}_{\eta}(B), \widehat{\mathcal{M}}_{\eta}(\tilde{B})\rangle_t = \int_0^t \left(\int_{[0,H^r)} 1_{B\cap\tilde{B}}(x)\lambda g^r(x)(dx)\right) ds, \tag{3.4.35}$$

for $t \in [0, \infty)$. Thus, both $\widehat{\mathcal{M}}_\nu$ and $\widehat{\mathcal{M}}_\eta$ are white noises. Let $\mathcal{C}_{\widehat{\mathcal{M}}_\nu}$ denote the subset of continuous functions on $[0, H^s) \times [0, \infty)$ that satisfies

$$\int_0^t \left(\int_{[0,H^s)} \varphi^2(x, s)(\lambda \wedge 1) g^s(x)(dx) \right) ds < \infty, t \in [0, \infty). \tag{3.4.36}$$

Similarly let $\mathcal{C}_{\widehat{\mathcal{M}}_\eta}$ denote the subset of continuous functions on $[0, H^r) \times [0, \infty)$ that satisfies

$$\int_0^t \left(\int_{[0,H^r)} \psi^2(x, s)\lambda g^r(x)(dx) \right) ds < \infty, t \in [0, \infty). \tag{3.4.37}$$

Note that $\mathcal{C}_{\widehat{\mathcal{M}}_\nu}$ and $\mathcal{C}_{\widehat{\mathcal{M}}_\eta}$ include,in particular,the spaces $\mathcal{C}_b([0, H^s) \times [0, \infty))$ and $\mathcal{C}_b([0, H^r) \times [0, \infty))$ respectively.

For any $\varphi \in \mathcal{C}_{\widehat{\mathcal{M}}_\nu}$ and $t \in [0, \infty)$, the stochastic integral of φ with respect to $\widehat{\mathcal{M}}_\nu$ on $[0, H^s) \times [0, t]$, denoted by

$$\widehat{\mathcal{M}}_{t,\nu}(\varphi) := \int\int_{[0,H^s)\times[0,t]} \varphi(x, s)\widehat{\mathcal{M}}_\nu(dx, ds), \tag{3.4.38}$$

is well defined. Similarly, for any $\psi \in \mathcal{C}_{\widehat{\mathcal{M}}_\eta}$ and $t \in [0, \infty)$, the stochastic integral of ψ with respect to $\widehat{\mathcal{M}}_\eta$ on $[0, H^r) \times [0, t]$, denoted by

$$\widehat{\mathcal{M}}_{t,\eta}(\psi) := \int\int_{[0,H^r)\times[0,t]} \psi(x, s)\widehat{\mathcal{M}}_\eta(dx, ds) \tag{3.4.39}$$

is well defined. In fact, for such φ, ψ, $\widehat{\mathcal{M}}_\nu(\varphi) = \{\widehat{\mathcal{M}}_{t,\nu}(\varphi), t \geq 0\}$ and $\widehat{\mathcal{M}}_\eta(\psi) = \{\widehat{\mathcal{M}}_{t,\eta}(\psi), t \geq 0\}$ are càdlàg, orthogonal martingale measure; see page 292 of Walsh [44] for the definition. Moreover, because $\widehat{\mathcal{M}}_\nu$ and $\widehat{\mathcal{M}}_\eta$ are continuous martingale measures, $\widehat{\mathcal{M}}_\nu(\varphi)$ and $\widehat{\mathcal{M}}_\eta(\psi)$ have versions as a continuous real-valued process.

Next, for $t \in [0, \infty)$ and $f \in \mathcal{C}_b[0, H^s)$, let $\widehat{\mathcal{H}}_{t,\nu}(f)$ be the random variable given by the following convolution integral:

$$\widehat{\mathcal{H}}_{t,\nu}(f) := \int\int_{[0,H^s)\times[0,t]} f(x+t-s)\frac{1-G^s(x+t-s)}{1-G^s(x)}\widehat{\mathcal{M}}_\nu(dx,ds). \tag{3.4.40}$$

Similarly for $g \in \mathcal{C}_b[0,H^r)$, let $\widehat{\mathcal{H}}_{t,\eta}(g)$ be the random variable given by the following convolution integral:

$$\widehat{\mathcal{H}}_{t,\eta}(g) := \int\int_{[0,H^r)\times[0,t]} 1_{[0,\chi_*]}(x)g^r(x+t-s)\frac{1-G^r(x+t-s)}{1-G^r(x)}\widehat{\mathcal{M}}_\eta(dx,ds). \tag{3.4.41}$$

In order to express the convolution integrals in a more succinct fashion, consider the family of operators $\{\Psi^s_t, t \geq 0\}$ defined, for $t > 0$ and $(x,s) \in [0,H^s)\times[0,\infty)$, by

$$(\Psi^s_t f)(x,s) := f(x+(t-s)^+)\frac{1-G^s(x+(t-s)^+)}{1-G^s(x)}, \tag{3.4.42}$$

for bounded and measurable functions f on $[0,H^s)$, and $\{\Psi^r_t, t \geq 0\}$ defined, for $t > 0$ and $(x,s) \in [0,H^r)\times[0,\infty)$, by

$$(\Psi^r_t g)(x,s) := g(x+(t-s)^+)\frac{1-G^r(x+(t-s)^+)}{1-G^r(x)}, \tag{3.4.43}$$

for bounded and measurable functions g on $[0,H^r)$, where recall $(t-s)^+ = \max(t-s,0)$. The operators Ψ^s_t and Ψ^r_t map the space of bounded measurable functions on $[0,\infty)$ to the space of bounded measurable functions on $[0,H^s)\times[0,\infty)$ and $[0,H^s)\times[0,\infty)$ respectively. We also have

$$\sup_{t\in[0,\infty)} \|\Psi^s_t f\|_\infty \leq \|f\|_\infty, \qquad \sup_{t\in[0,\infty)} \|\Psi^r_t g\|_\infty \leq \|g\|_\infty. \tag{3.4.44}$$

The processes $\widehat{\mathcal{H}}_\nu, \widehat{\mathcal{H}}_\eta, \widehat{\mathcal{H}}_\nu^{(N)}, \widehat{\mathcal{H}}_\eta^{(N)}$, respectively, can then be rewritten as

$$\widehat{\mathcal{H}}_{t,\nu}(f) = \widehat{\mathcal{M}}_{t,\nu}(\Psi_t^s f), \qquad \widehat{\mathcal{H}}_{t,\nu}^{(N)}(f) = \widehat{\mathcal{M}}_{t,\nu}^{(N)}(\Psi_t^s f), \qquad t \geq 0, \tag{3.4.45}$$

and

$$\widehat{\mathcal{H}}_{t,\eta}(f) = \widehat{\mathcal{M}}_{t,\eta}(1_{[0,\chi_*]}\Psi_t^r f), \qquad \widehat{\mathcal{H}}_{t,\eta}^{(N)}(f) = \widehat{\mathcal{M}}_{t,\eta}^{(N)}(1_{[0,\chi_*]}\Psi_t^r f), \qquad t \geq 0. \tag{3.4.46}$$

3.5 Central Limit Scaling

In this section we restrict the initial conditions of the N-server system so that $(\overline{X}^{(N)}, \overline{\nu}^{(N)}, \overline{\eta}^{(N)})$ converges weakly to $(\overline{X}, \overline{\nu}, \overline{\eta})$, where $(\overline{X}, \overline{\nu}, \overline{\eta}) = I_\lambda$, with I_λ defined in (3.3.24).

In what follows, we will introduce the following notation for conciseness: for $N \in \mathbb{N}$, we set

$$A_D^{(N)} := A_{\mathbf{1},\nu}^{(N)}, \qquad A_R^{(N)} := A_{\theta^{(N)},\eta}^{(N)} \tag{3.5.1}$$

$$\mathcal{M}(\mathbb{R}_+)^{(N)} := M_{\mathbf{1},\nu}^{(N)}, \qquad M_R^{(N)} := M_{\theta^{(N)},\eta}^{(N)}. \tag{3.5.2}$$

Note that by Proposition 3.4.1 and Lemma 3.4.3, $\overline{M}_D^{(N)} \Rightarrow 0$ and $\overline{M}_R^{(N)} \Rightarrow 0$, and hence set

$$\overline{M}_D \equiv 0, \qquad \overline{M}_R \equiv 0. \tag{3.5.3}$$

Using the definition (3.4.27), define $\widehat{\mathcal{M}}_D^{(N)} = \widehat{\mathcal{M}}_\nu^{(N)}(\mathbf{1})$, and $\widehat{\mathcal{M}}_R^{(N)} = \widehat{\mathcal{M}}_\eta^{(N)}(\theta^{(N)})$ with $\theta^{(N)}$ defined in (3.2.20). Now, for $Y = D, E, S, R, X, Q, A_D, A_R, \nu, \eta$ with $\overline{\nu} =$

$(\lambda \wedge 1)\nu_*$ and $\overline{\eta} = \lambda\eta_*$, we define the diffusion scaled version of the process as follows:

$$\widehat{Y}^{(N)}(t) = \sqrt{N}\left(\overline{Y}^{(N)}(t) - \overline{Y}(t)\right). \tag{3.5.4}$$

For $r \geq 0$ define the functions

$$\alpha^r(\cdot) := \frac{1 - G^s(\cdot + r)}{1 - G^s(\cdot)}, \qquad \text{and} \qquad \beta^r(\cdot) := \frac{1 - G^r(\cdot + r)}{1 - G^r(\cdot)}. \tag{3.5.5}$$

Using the function defined above, and recalling the definition of χ_* from (3.3.25), define the processes

$$\widehat{Z}^{(N)}_\nu(t,r) := \langle \alpha^r, \widehat{\nu}^{(N)}_t \rangle, \qquad \text{and} \qquad \widehat{Z}^{(N)}_\eta(t,r) := \langle 1_{[0,\chi_*]}\beta^r, \widehat{\eta}^{(N)}_t \rangle. \tag{3.5.6}$$

Using the definition above it is easy to see that

$$\partial_r \widehat{Z}^{(N)}_\nu(t,r) = \langle \partial_r \alpha^r, \widehat{\nu}^{(N)}_t \rangle, \qquad \text{and} \qquad \partial_r \widehat{Z}^{(N)}_\eta(t,r) = \langle 1_{[0,\chi_*]}\partial_r \beta^r, \widehat{\eta}^{(N)}_t \rangle. \tag{3.5.7}$$

3.6 Main Results

The main results of this chapter are stated in Section 3.6.2. They rely on some basic assumptions and the definition of a certain map (called the diffusion many-server reneging map). Before stating the main results, we state the assumptions required.

3.6.1 Assumptions

We have the following assumptions on the arrival process.

Assumption 3.6.1. There exist constants $\lambda, \sigma^2 \in (0,\infty)$ and $\beta \in \mathbb{R}$ such that for every $N \in \mathbb{N}$, $E^{(N)}$ is a renewal process with i.i.d. inter-renewal times $\{\xi^{(N)}\}_{j\in\mathbb{N}}$ that have mean $1/\lambda^{(N)}$ and variance $(\sigma^2/\lambda)/(\lambda^{(N)})^2$, where

$$\lambda^{(N)} = \lambda N - \beta\sqrt{N}, \tag{3.6.1}$$

and the following Lindeberg condition holds: for every $\varepsilon > 0$,

$$\lim_{N\to\infty} N^2 \mathbb{E}[(\xi_1^{(N)})^2 \mathbf{1}_{\{\xi_1^{(N)}\sqrt{N}>\varepsilon\}}] = 0.$$

Remark 3.6.2. Let λ, β and σ be as defined in Assumption 3.6.1. Then, given a standard Brownian motion B, the process $\widehat{E}$ given by

$$\widehat{E}(t) := \sigma B(t) - \beta t, \qquad t \in [0,\infty). \tag{3.6.2}$$

is a well-defined diffusion and therefore a semimartingale, with $\sigma B(t)$, $t \geq 0$ being the local martingale and βt, the finite variation process in the decomposition. If Assumption 3.6.1 holds, then it is easy to see that $\overline{E}$ in Assumption 3.3.1 is given by $\overline{E}(t) = \lambda t$.

We impose the following assumptions on the distributions and the initial conditions.

Assumption 3.6.3. The function G^s satisfies the following properties:

a. The function h^s is uniformly bounded, that is $H_1^s := \sup_{x\in[0,\infty)} h^s(x) < \infty$.

b. The derivative g^s is continuously differentiable with derivative $(g^s)'$, and the function

$$h_2^s(x) := \frac{(g^s)'(x)}{1 - G^s(x)}, \qquad x \in [0, \infty),$$

is uniformly bounded, that is, $H_2^s := \sup_{x\in[0,\infty)} |h_2^s(x)| < \infty$.

Assumption 3.6.4. For some $\epsilon > 0$, $\overline{G^s}(x) = \mathcal{O}(x^{-2-\epsilon})$ as $x \to \infty$, that is, there exists a finite positive constant c_s such that $\overline{G^s}(x) \le c_s x^{-2-\epsilon}$ for all sufficiently large x. In other words, the service time distribution has a finite $(2 + \epsilon)$ moment.

Assumption 3.6.5. The function G^r satisfies the following properties:

a. The function h^r is uniformly bounded, that is $H_1^r := \sup_{x\in[0,\infty)} h^r(x) < \infty$.

b. The derivative g^r is continuously differentiable with derivative $(g^r)'$, and the function

$$h_2^r(x) := \frac{g^{r'}(x)}{1 - G^r(x)}, \qquad x \in [0, \infty),$$

is uniformly bounded, that is, $H_2^r := \sup_{x\in[0,\infty)} |h_2^r(x)| < \infty$.

Assumption 3.6.6. For some $\epsilon > 0$, $\overline{G^r}(x) = \mathcal{O}(x^{-2-\epsilon})$ as $x \to \infty$, that is, there exists a finite positive constant c_r such that $\overline{G^r}(x) \le c_r x^{-2-\epsilon}$ for all sufficiently large x. In other words, the patience time distribution has a finite $(2 + \epsilon)$ moment.

Assumption 3.6.7. There exists an $\mathbb{R}$-valued random variable $\widehat{X}(0)$ and families of random variables $\{\widehat{Z}_\nu(0, r),\ r \ge 0\}$ and $\{\widehat{Z}_\eta(0, r), r \ge 0\}$ all defined on a common probability space, such that:

(a) $\{\widehat{Z}_\nu(0, t), t \ge 0\}$, and $\{\widehat{Z}_\eta(0, t), t \ge 0\}$ admit versions as a continuous $\mathbb{R}$-valued process. Almost surely, $t \mapsto \widehat{Z}_\nu(0, t)$ and $t \mapsto \widehat{Z}_\eta(0, t)$ are measurable functions on $[0, \infty)$;

(b) As $N \to \infty$, $(\widehat{X}^{(N)}(0), \widehat{Z}_\nu^{(N)}(0,\cdot), \widehat{Z}_\eta^{(N)}(0,\cdot)) \Rightarrow (\widehat{X}(0), \widehat{Z}_\nu(0,\cdot), \widehat{Z}_\eta(0,\cdot))$ in $\mathbb{R} \times \mathbb{H}^1(0,\infty)^2$;

Remark 3.6.8. Since $\widehat{\nu}_0^{(N)}$ is a signed measure with finite total mass bounded by $\sqrt{N}$, using the norm inequality $\|f\|_\infty \le 2\|f\|_{\mathbb{H}^1}$ it is easy to see that almost surely for $r \ge 0$,

$$|\widehat{Z}_\nu^{(N)}(0,r)| \le 2\sqrt{N}\|\alpha^r\|_\infty \le 4\sqrt{N}\|\alpha^r\|_{\mathbb{H}^1}.$$

3.6.2 Statement of Main Results

We introduce the diffusion many-server reneging map below will be used to characterize the limit process. Let $B^+[0,\infty)$ denote the space of non-negative bounded functions on $[0,\infty)$.

Definition 3.6.9 (Diffusion many-server reneging equations). Let $\overline{X} \in [0,\infty)$ be fixed, and let g be a probability density function. Given $(E, X(0), W, Y) \in \mathcal{D}^0_{\mathbb{R}}[0,\infty) \times \mathbb{R} \times \mathcal{D}_{\mathbb{R}}[0,\infty) \times B^+[0,\infty)$, $(K, X, Z, Q) \in \mathcal{D}^0_{\mathbb{R}}[0,\infty) \times \mathcal{D}_{\mathbb{R}}[0,\infty)^3$ is said to solve the diffusion many-server reneging equations (DMSRE) associated with $\overline{X}$ and $(E, X(0), W, Y)$ if for any $T < \infty$, $\sup_{s\in[0,T]} |X(s)| < \infty$, and for $t \in [0,\infty)$,

$$Z(t) = W(t) + K(t) - \int_0^t g(t-s)K(s)ds, \tag{3.6.3}$$

$$K(t) = E(t) + X(0) - X(t) + Z(t) - Z(0) - \int_0^t Q(s)Y(s)ds, \tag{3.6.4}$$

and

$$Z(t) = \begin{cases} X(t) & \text{if } \overline{X} < 1, \\ X(t) \wedge 0 & \text{if } \overline{X} = 1, \\ 0 & \text{if } \overline{X} > 1. \end{cases} \tag{3.6.5}$$

$$Q(t) = \begin{cases} 0 & \text{if } \overline{X} < 1, \\ X^+(t) & \text{if } \overline{X} = 1, \\ X(t) & \text{if } \overline{X} > 1. \end{cases} \tag{3.6.6}$$

When a solution to the DMSRE exists, let Λ denote the corresponding "diffusion many-server reneging" mapping (associated with $\overline{X}$) that takes $(E, X(0), W, Y) \in \mathcal{D}^0_{\mathbb{R}}[0,\infty) \times \mathbb{R} \times \mathcal{D}_{\mathbb{R}}[0,\infty) \times B^+[0,\infty)$ to the corresponding solution (K, X, Z, Q) of the DMSRE.

Note that this definition automatically requires that $E(0) = K(0) = 0$ and $Z(0) = W(0)$.

Theorems 3.6.10 and 3.6.11 below identify the limit of the sequence $(\widehat{X}^{(N)}, \widehat{Q}^{(N)}, \widehat{Z}^{(N)}_\nu, \widehat{Z}^{(N)}_\eta)$. We now introduce some limiting processes required for the statement of the theorems.

Let f and g be absolutely continuous functions on $[0, H^r)$ and $[0, H^r)$ respectively. Define

$$\widehat{Z}_\eta(t,r) = \widehat{Z}_\eta(0, t+r) - \widehat{\mathcal{H}}_{t,\eta}(\beta^r) + \widehat{\mathcal{E}}_t(\beta^r), \tag{3.6.7}$$

where

$$\widehat{\mathcal{E}}_t(f) := f(0)\widehat{E}(t) - \int_{[0,t]} \widehat{E}(s) \frac{d((1-G^r(t-s))f(t-s))}{ds} ds, \qquad t \in [0,\infty), \tag{3.6.8}$$

and

$$\widehat{Z}_\nu(t,r) = \widehat{Z}_\nu(0, t+r) - \widehat{\mathcal{H}}_{t,\nu}(\alpha^r) + \widehat{\mathcal{K}}_t(\alpha^r), \tag{3.6.9}$$

where

$$\widehat{\mathcal{K}}_t(g) := g(0)\widehat{K}(t) - \int_{[0,t]} \widehat{K}(s)\frac{d((1-G^s(t-s))g(t-s))}{ds}ds, \qquad t \in [0,\infty). \tag{3.6.10}$$

Let $\{e_k; k \in \mathbb{N}\}$ be a sequence of functions with compact support in $\mathbb{H}^1(0,\infty)$ that form a countable orthonormal basis for the Hilbert space $\mathbb{H}^1(0,\infty)$ with inner product denoted by $\langle \cdot, \cdot \rangle_{\mathbb{H}^1}$, and let e_k' denote the weak derivative of e_k. Theorems 3.6.10 and 3.6.11 show convergence of finite-dimensional projections of $\widehat{Z}_\eta^{(N)}$ and $(\widehat{X}^{(N)}, \widehat{Q}^{(N)}, \widehat{Z}_\nu^{(N)})$ to the finite-dimensional projections of $\widehat{Z}_\eta$ and $(\widehat{X}, \widehat{Q}, \widehat{Z}_\nu)$ respectively.

Theorem 3.6.10. *Suppose the Assumptions 3.6.1, 3.6.5,3.6.6 and 3.6.7 hold. and $\widehat{Z}_\eta$ satisfies (3.6.7). Then for every $t \geq 0$ and $k \in \mathbb{N}$, as $N \to \infty$,*

$$\left(\langle \widehat{Z}_\eta^{(N)}(t,\cdot), e_1\rangle_{\mathbb{H}^1}, \cdots, \langle \widehat{Z}_\eta^{(N)}(t,\cdot), e_k\rangle_{\mathbb{H}^1}\right) \Rightarrow (\langle \widehat{Z}_\eta(t,\cdot), e_1\rangle_{\mathbb{H}^1}, \cdots, \langle \widehat{Z}_\eta(t,\cdot), e_k\rangle_{\mathbb{H}^1}). \tag{3.6.11}$$

Theorem 3.6.11. *Suppose Assumptions 3.6.1 - 3.6.7 hold and let $(\widehat{K}, \widehat{X}, \widehat{Z}_\nu(\cdot,0), \widehat{Q})$ be the solution to DMSRE defined in Definition 3.6.9 with $\overline{X}$ and g replaced by x_* and g^s respectively, i.e., $(\widehat{K}, \widehat{X}, \widehat{Z}_\nu(\cdot,0), \widehat{Q}) = \Lambda(\widehat{E} - \widehat{\mathcal{M}}_R + \int_0^t \partial_r \widehat{Z}_\eta(s,0)ds - \int_0^t \widehat{Z}_\eta(s,0)h^r(\chi_*)ds, \widehat{X}(0), \widehat{Z}_\nu(0,\cdot) - \widehat{\mathcal{H}}_\nu(\mathbf{1}), h^r(\chi_*))$ and $\widehat{Z}_\nu$ satisfies (3.6.9). Then for every $t \geq 0$ and $k \in \mathbb{N}$, as $N \to \infty$,*

$$\begin{aligned}&\left(\widehat{X}_t^{(N)}, \widehat{Q}_t^{(N)}, \langle \widehat{Z}_\nu^{(N)}(t,\cdot), e_1\rangle_{\mathbb{H}^1}, \cdots, \langle \widehat{Z}_\nu^{(N)}(t,\cdot), e_k\rangle_{\mathbb{H}^1}\right) \\ \Rightarrow &(\widehat{X}_t, \widehat{Q}_t, \langle \widehat{Z}_\nu(t,\cdot), e_1\rangle_{\mathbb{H}^1}, \cdots, \langle \widehat{Z}_\nu(t,\cdot), e_k\rangle_{\mathbb{H}^1}).\end{aligned} \tag{3.6.12}$$

We now show that the limit processes described in Theorems 3.6.10 and 3.6.11 can alternatively be characterized as the unique solution to a stochastic partial differential equation (SPDE).

Define the space $\mathbb{Y}$ as

$$\mathbb{Y} := \left\{ (x,f,g) \in \mathbb{R} \times \mathbb{H}^1(0,\infty) \times \mathbb{H}^1(0,\infty) : f(0) = \begin{cases} x & \text{if } \lambda < 1 \\ x \wedge 0 & \text{if } \lambda = 1 \\ 0 & \text{if } \lambda > 1. \end{cases} \right\}. \tag{3.6.13}$$

and the $\mathbb{Y}$-valued random element $Y(t) = (\widehat{X}(t), \widehat{Z}_\nu(t,\cdot), \widehat{Z}_\eta(t,\cdot)$ for $t \geq 0$.

Definition 3.6.12. (Diffusion SPDE) Let $\overline{X} \in [0,\infty)$ be fixed. Given a $\mathbb{Y}$-valued random element $Y(0)$, a $\mathbb{Y}$-valued stochastic process $Y = \{(\widehat{X}(t), \widehat{Z}_\nu(t,\cdot), \widehat{Z}_\eta(t,\cdot); t \geq 0\}$ is said to be a solution of the diffusion SPDE if

$$\begin{aligned}\widehat{X}(t) =& \widehat{X}(0) + \int_0^t \theta(s)dW(s) - \beta t + \int_0^t \partial_r \widehat{Z}_\nu(s,0)ds + \int_0^t \partial_r \widehat{Z}_\eta(s,0)ds \\ &- \int_0^t \widehat{Q}(s)h^r(\chi_*)ds - \int_0^t \widehat{Z}_\eta(s,0)h^r(\chi_*)ds,\end{aligned} \tag{3.6.14}$$

where

$$\widehat{Q}(t) = \begin{cases} 0 & \text{if } \overline{X} < 1, \\ \widehat{X}^+(t) & \text{if } \overline{X} = 1, \\ \widehat{X}(t) & \text{if } \overline{X} > 1, \end{cases} \tag{3.6.15}$$

and

$$\widehat{Z}_\nu(t,r) = \widehat{Z}_\nu(0,r) + \int_0^t \partial_r \widehat{Z}_\nu(s,r)ds - \widehat{\mathcal{M}}_{t,\nu}(\alpha^r) + (1 - G^s(r))\widehat{K}(t), \tag{3.6.16}$$

$$\widehat{Z}_\eta(t,r) = \widehat{Z}_\eta(0,r) + \int_0^t \partial_r \widehat{Z}_\eta(s,r)ds - \widehat{\mathcal{M}}_{t,\eta}(1_{[0,\chi_*]}\beta^r) + (1 - G^r(r))\widehat{E}(t), \tag{3.6.17}$$

where

$$\widehat{E}(t) = \sigma B(t) - \beta t$$

and $\widehat{K}$ satisfies

$$\widehat{K}(t) = \begin{cases} \widehat{E}(t) - \widehat{R}(t) & \text{if } \overline{X} < 1 \\ \widehat{E}(t) - \widehat{R}(t) - \int_0^t \widehat{X}^+(s) h^r(\chi_*) ds + (\widehat{X}(0) - \widehat{X}(t)) \vee 0 & \text{if } \overline{X} = 1 \\ \widehat{E}(t) - \widehat{R}(t) - \int_0^t \widehat{X}(s) h^r(\chi_*) ds + \widehat{X}(0) - \widehat{X}(t) & \text{if } \overline{X} > 1 \end{cases} \tag{3.6.18}$$

where

$$\widehat{R}(t) = \widehat{\mathcal{M}}_R(t) - \int_0^t \partial_r \widehat{Z}_\eta(s,0) ds + \int_0^t \widehat{Z}_\eta(s,0) h^r(\chi_*) ds.$$

Theorem 3.6.13. *Suppose Assumptions 3.6.3 - 3.6.7 are satisfied, and given a $\mathbb{Y}$-valued random element $Y(0)$, let $Y = \{Y(t); t \geq 0\}$, where $Y(t) = (\widehat{X}(t), \widehat{Z}_\nu(t,\cdot), \widehat{Z}_\eta(t,\cdot))$ for $t \geq 0$ and $(\widehat{K}, \widehat{X}, \widehat{Z}_\nu(\cdot,0), \widehat{Q}) := \Lambda(\widehat{E} - \widehat{\mathcal{M}}_R + \int_0^t \partial_r \widehat{Z}_\eta(s,0)ds - \int_0^t \widehat{Z}_\eta(s,0) h^r(\chi_*) ds, \widehat{X}(0), \widehat{Z}_\nu(0,\cdot) - \widehat{\mathcal{H}}_\nu(\mathbf{1}), h^r(\chi_*))$. Also suppose equations (3.6.7) and (3.6.9) hold. Then Y is a solution of the diffusion SPDE.*

Theorem 3.6.14. *Suppose Assumptions 3.6.3, 3.6.4, 3.6.5 and 3.6.6 hold. Then, for every $\mathbb{Y}$-valued random element $Y(0) = (\widehat{X}_0, \widehat{Z}_\nu(0,\cdot), \widehat{Z}_\eta(0,\cdot))$, there is at most one solution of the equations (3.6.14)-(3.6.17) with initial condition $Y(0)$.*

3.7 An Asymptotic Independence Result

Recall from Remark 3.4.2 that U^s and U^r are the renewal functions associated with the service and patience distributions G^s and G^r respectively. We first state two propositions whose proof follow along the lines of [25, Lemma 8.1] using (3.4.4) and its analogue for $A_{\psi,\eta^{(N)}}$ from Proposition 3.4.1, along with Remark 3.4.2.

Lemma 3.7.1. *Fix $T < \infty$. For every $N \in \mathbb{N}$ and positive integer k,*

$$\mathbb{E}[(\overline{A}^{(N)}_{\mathbf{1},\nu}(T))^k] = \mathbb{E}\left[\left(\int_0^T \int_{[0,H^s)} h^s(x)\overline{\nu}^{(N)}_s(dx)ds\right)^k\right] \leq k!(U^s(T))^k. \qquad (3.7.1)$$

Moreover, there exists $\overline{C}^s(T) < \infty$ such that for every positive integer k and measurable function φ on $[0, H^s) \times [0, T]$,

$$\sup_{N\in\mathbb{N}} \mathbb{E}[(\overline{A}^{(N)}_{\varphi,\nu}(T))^k] \leq k!(\overline{C}^s(T))^k \left(\int_{[0,H^s)} \varphi^*(x)h^s(x)dx\right)^k,$$

where $\varphi^*(x) := \sup_{s\in[0,T]} |\varphi(x, s)|$. Furthermore, if Assumptions 3.3.1 and 3.3.4 hold, then

$$(\overline{A}_{\mathbf{1},\nu}(T))^k = \left(\int_0^T \int_{[0,H^s)} h^s(x)(\lambda \wedge 1)\nu_*(dx)ds\right)^k \leq k!(U^s(T))^k. \qquad (3.7.2)$$

Lemma 3.7.2. *Fix $T < \infty$. For every $N \in \mathbb{N}$ and positive integer k,*

$$\mathbb{E}[(\overline{A}^{(N)}_{\mathbf{1},\eta}(T))^k] = \mathbb{E}\left[\left(\int_0^T \int_{[0,H^r)} h^r(x)\overline{\eta}^{(N)}_s(dx)ds\right)^k\right] \leq k!(U^r(T))^k. \qquad (3.7.3)$$

Moreover, there exists $\overline{C}^r(T) < \infty$ such that for every positive integer k and measurable function ψ on $[0, H^r) \times [0, T]$,

$$\sup_{N\in\mathbb{N}} \mathbb{E}[(\overline{A}^{(N)}_{\psi,\eta}(T))^k] \leq k!(\overline{C}^r(T))^k \left(\int_{[0,H^r)} \psi^* h^r(x)dx\right)^k,$$

where $\psi^*(x) := \sup_{s\in[0,T]} |\psi(x, s)|$. Furthermore, if Assumptions 3.3.1 and 3.3.4 hold, then

$$(\overline{A}_{\mathbf{1},\eta}(T))^k = \left(\int_0^T \int_{[0,H^r)} h^r(x)\lambda\eta_*(dx)ds\right)^k \leq k!(U^r(T))^k. \qquad (3.7.4)$$

The proofs of Lemmas 3.7.3 and 3.7.4 below follow along the lines of the proof of [25, Lemma 8.2] using Lemmas 3.7.1 and 3.7.2 respectively.

Lemma 3.7.3. *For every integer k, there exists a universal constant $C^s_k < \infty$ such that for every $\varphi \in \mathcal{C}_b([0, H^s) \times [0, \infty))$ and $T < \infty$,*

$$\mathbb{E}[\sup_{s\in[0,T]} |\widehat{\mathcal{M}}^{(N)}_{s,\nu}(\varphi)|^k] \leq C^s_k \|\varphi\|^k_\infty \left[\left(\frac{k}{2}\right)!(U^s(T))^{k/2} + \frac{1}{N^{k/2}}\right], \qquad N \in \mathbb{N}, \quad (3.7.5)$$

$$\mathbb{E}[\sup_{s\in[0,T]} |\widehat{\mathcal{M}}_{s,\nu}(\varphi)|^k] \leq C^s_k \left(\frac{k}{2}\right)!(U^s(T))^{k/2}\|\varphi\|^k_\infty, \quad (3.7.6)$$

and for $0 \leq s \leq t$,

$$\mathbb{E}[|\widehat{\mathcal{M}}_{t,\nu}(\varphi) - \widehat{\mathcal{M}}_{s,\nu}(\varphi)|^r] \leq C^s_k(\overline{A}_{\varphi^2,\nu}(t) - \overline{A}_{\varphi^2,\nu}(s))^{k/2}. \quad (3.7.7)$$

Lemma 3.7.4. *For every integer k, there exists a universal constant $C^r_k < \infty$ such that for every $\psi \in \mathcal{C}_b([0, H^r) \times [0, \infty))$ and $T < \infty$,*

$$\mathbb{E}[\sup_{s\in[0,T]} |\widehat{\mathcal{M}}^{(N)}_{s,\eta}(\psi)|^k] \leq C^r_k \|\psi\|^k_\infty \left[\left(\frac{k}{2}\right)!(U^r(T))^{k/2} + \frac{1}{N^{k/2}}\right], \qquad N \in \mathbb{N}, \quad (3.7.8)$$

$$\mathbb{E}[\sup_{s\in[0,T]} |\widehat{\mathcal{M}}_{s,\eta}(\psi)|^k] \leq C^r_k \left(\frac{k}{2}\right)!(U^r(T))^{k/2}\|\psi\|^k_\infty, \quad (3.7.9)$$

and for $0 \leq s \leq t$,

$$\mathbb{E}[|\widehat{\mathcal{M}}_{t,\eta}(\psi) - \widehat{\mathcal{M}}_{s,\eta}(\psi)|^r] \leq C^r_k(\overline{A}_{\psi^2,\eta}(t) - \overline{A}_{\psi^2,\eta}(s))^{k/2}. \quad (3.7.10)$$

We now identify the limit of the sequence of martingale measures $\{\widehat{\mathcal{M}}^{(N)}_\nu, \widehat{\mathcal{M}}^{(N)}_\eta\}_{N\in\mathbb{N}}$ and also show that the sequence is asymptotically independent of the sequences of centered arrival processes and initial conditions.

Before stating the proposition, define

$$\widehat{L}^{(N)}(t) := \frac{1}{\sqrt{N}} \sum_{j=2}^{E^{(N)}(t)+1} \left(1 - \lambda^{(N)} \xi_j^{(N)}\right). \tag{3.7.11}$$

Note that by Assumption 3.6.1 and equation 8.10[25],

$$\widehat{E}^{(N)} = \widehat{L}^{(N)}(t) + \widehat{\gamma}^{(N)}(t) - \beta t, \tag{3.7.12}$$

where

$$\widehat{\gamma}^{(N)}(t) = \frac{1}{\sqrt{N}} \left(\sum_{j=2}^{E^{(N)}(t)+1} \lambda^{(N)} \xi_j^{(N)} - \lambda^{(N)} t \right),$$

and it was shown in [39, p. 30, Lemma A.1 and (5.15)] (see also [25, Section 8.2, p. 190]) that

$$\widehat{\gamma}^{(N)} \Rightarrow 0. \tag{3.7.13}$$

Now, note that by (3.2.8), (3.4.3), (3.4.7), (3.4.1), (3.5.1) and(3.5.2), we have

$$X^{(N)}(t) = X^{(N)}(0) + E^{(N)}(t) - \mathcal{M}(\mathbb{R}_+)^{(N)}(t) - \int_0^t \langle h^s, \nu_s^{(N)} \rangle \, ds - M_R^{(N)}(t) - A_R^{(N)}(t).$$

Combining this with (3.3.13), (3.5.3), (3.5.4), (3.5.7) and (3.7.12) we see that

$$\begin{aligned}
\widehat{X}^{(N)}(t) &= \widehat{X}^{(N)}(0) + \widehat{L}^{(N)}(t) + \widehat{\gamma}^{(N)} - \beta t - \widehat{\mathcal{M}}_D^{(N)}(t) - \widehat{\mathcal{M}}_R^{(N)}(t) \\
&\quad - \int_0^t \langle h^s, \widehat{\nu}_s^{(N)} \rangle ds - \sqrt{N} \left[\overline{A}_R^{(N)}(t) - \overline{R}(t) \right]. \\
&= \widehat{X}^{(N)}(0) + \widehat{L}^{(N)}(t) + \widehat{\gamma}^{(N)} - \beta t - \widehat{\mathcal{M}}_D^{(N)}(t) - \widehat{\mathcal{M}}_R^{(N)}(t) \qquad (3.7.14) \\
&\quad + \int_0^t \partial_r \widehat{Z}_\nu^{(N)}(s,0) ds - \sqrt{N} \left[\overline{A}_R^{(N)}(t) - \overline{R}(t) \right].
\end{aligned}$$

Proposition 3.7.5. Suppose Assumptions 3.3.1, 3.3.4, 3.6.1 and 3.6.7 hold. Then for every $f \in \mathcal{C}_b([0, H^s) \times \mathbb{R})$, $\widehat{\mathcal{M}}_\nu^{(N)}(f) \Rightarrow \widehat{\mathcal{M}}_\nu(f)$ in $\mathcal{D}_\mathbb{R}[0, \infty)$ and for every $g \in$

$\mathcal{C}_b([0,H^r)\times\mathbb{R})$, $\widehat{\mathcal{M}}^{(N)}_\eta(g)\Rightarrow\widehat{\mathcal{M}}_\eta(g)$ in $\mathcal{D}_\mathbb{R}[0,\infty)$. Moreover, as $N\to\infty$,

$$\begin{aligned}&(\widehat{L}^{(N)},\widehat{X}^{(N)}(0),\widehat{Z}^{(N)}_\nu(0,\cdot),\widehat{Z}^{(N)}_\eta(0,\cdot),\widehat{\mathcal{M}}^{(N)}_\nu(f),\widehat{\mathcal{M}}^{(N)}_\eta(g))\\ \Rightarrow&(\sigma B,\widehat{X}(0),\widehat{Z}_\nu(0,\cdot),\widehat{Z}_\eta(0,\cdot),\widehat{\mathcal{M}}_\nu(f),\widehat{\mathcal{M}}_\eta(g)),\end{aligned}\tag{3.7.15}$$

in $\mathcal{D}_\mathbb{R}[0,\infty)\times\mathbb{R}\times\mathbb{H}^1(0,\infty)^2\times\mathcal{D}_\mathbb{R}[0,\infty)^2$.

Proof. Let $\{[\widehat{L}^{(N)}]_t\}$, $\{[\widehat{\mathcal{M}}^{(N)}_\nu(f)]_t\}$ and $\{[\widehat{\mathcal{M}}^{(N)}_\eta(g)]_t\}$ represent the $\{\mathcal{F}^{(N)}_t\}$-optional quadratic variation processes associated with $\{\widehat{L}^{(N)},t\geq 0\}$, $\{\widehat{\mathcal{M}}^{(N)}_{t,\nu}(f),t\geq 0\}$ and $\{\widehat{\mathcal{M}}^{(N)}_{t,\eta}(g),t\geq 0\}$ respectively.

Claim 1: For $t>0$, as $N\to\infty,[\widehat{L}^{(N)}]_t\to\sigma^2t$, $[\widehat{\mathcal{M}}^{(N)}_\nu(f)]_t\to\overline{A}_{f^2,\nu}$ and $[\widehat{\mathcal{M}}^{(N)}_\eta(g)]_t\to\overline{A}_{g^2,\eta}$ in probability.

Using the same idea as in the proof of Claim 1 of Proposition 8.4 [25] and Assumption 3.6.1 $[\widehat{L}^{(N)}]_t\to\sigma^2t$.

Note that $\widehat{\mathcal{M}}^{(N)}_\nu(f)$ and $\widehat{\mathcal{M}}^{(N)}_\eta(g)$ are both compensated sums of jumps, and hence local martingales of finite variation. Thus, they are purely discontinuous martingales (Lemma 4.14(b) of Chapter I of [21]) and hence (by Theorem 4.52 of Chapter 1 of [21]), their $\{\mathcal{F}^{(N)}_t\}$ optional quadratic variation at time t satisfy

$$[\widehat{\mathcal{M}}^{(N)}_\nu(f)]_t=\sum_{s\leq t}\left(\Delta\widehat{\mathcal{M}}^{(N)}_{s,\nu}(f)\right)^2=\overline{D}^{(N)}_{f^2}(t),$$

and

$$[\widehat{\mathcal{M}}^{(N)}_\eta(g)]_t=\sum_{s\leq t}\left(\Delta\widehat{\mathcal{M}}^{(N)}_{s,\eta}(g)\right)^2=\overline{S}^{(N)}_{g^2}(t),$$

where the last equality follows because the jumps of $\mathcal{M}^{(N)}_\nu(f)$ and $\mathcal{M}^{(N)}_\eta(g)$ coincide with those of $\overline{D}^{(N)}_f$ and $\overline{S}^{(N)}_g$ respectively. By Proposition 3.4.1 and the law of large numbers results in [22](specifically the discussion below Theorem 7.1 along with equations (3.4.11) and (3.4.12)), $\overline{D}^{(N)}_{f^2}(t) \to \overline{A}_{f^2,\nu}(t)$ and $\overline{S}^{(N)}_{g^2}(t) \to \overline{A}_{g^2,\eta}(t)$ in probability.

Claim 2: For $t > 0$, $[\widehat{L}^{(N)}, \widehat{\mathcal{M}}^{(N)}_{\nu}(f)]_t \to 0, [\widehat{L}^{(N)}, \widehat{\mathcal{M}}^{(N)}_{\eta}(g)]_t \to 0$ in probability as $N \to \infty$.

The proof for this claim follows the same idea as in the proof of Claim 2 in Proposition 8.4 [25] by using the first two equations in the proof of Lemma 6.3 [22], i.e.,

$$\sup_N \mathbb{E}[\overline{D}^{(N)}_f(t)] < \infty, \qquad \sup_N \mathbb{E}[\overline{S}^{(N)}_g(t)] < \infty,$$

for all $t \in [0, \infty)$.

Claim 3: For every $t > 0$, $[\widehat{\mathcal{M}}^{(N)}_{\nu}(f), \widehat{\mathcal{M}}^{(N)}_{\eta}(g)]_t \to 0$ in probability as $N \to \infty$
The $\{\mathcal{F}^{(N)}_t\}$ optional quadratic covariation is given by

$$[\widehat{\mathcal{M}}^{(N)}_{\nu}(f), \widehat{\mathcal{M}}^{(N)}_{\eta}(g)]_t = \frac{1}{N}\sum_{s \le t} \Delta D^{(N)}_f(s) \Delta S^{(N)}_g(s),$$

since the jumps of $\widehat{\mathcal{M}}^{(N)}_{\nu}(f)$ and $\widehat{\mathcal{M}}^{(N)}_{\eta}(g)$ coincide with those of $\overline{D}^{(N)}_f$ and $\overline{S}^{(N)}_g$ respectively. Using Markov's inequality for any $\epsilon > 0$

$$\begin{aligned}
\mathbb{P}([\widehat{\mathcal{M}}^{(N)}_{\nu}(f), \widehat{\mathcal{M}}^{(N)}_{\eta}(g)]_t > \epsilon) &\le \frac{\mathbb{E}[[\widehat{\mathcal{M}}^{(N)}_{\nu}(f), \widehat{\mathcal{M}}^{(N)}_{\eta}(g)]_t]}{\epsilon} \\
&= \frac{\mathbb{E}[\sum_{s \le t} \Delta D^{(N)}_f(s) \Delta S^{(N)}_g(s)]}{N\epsilon} \\
&\le \|f\|_\infty \|g\|_\infty \frac{\mathbb{E}[\sum_{s \le t} \Delta D^{(N)}(s) \Delta S^{(N)}(s)]}{N\epsilon}.
\end{aligned}$$

Taking limit as $N \to \infty$ and using Lemma 3.4.6,

$$\lim_{N \to \infty} \mathbb{P}([\widehat{M}^{(N)}_{f,\nu}, \widehat{M}^{(N)}_{g,\eta}]_t > \epsilon) = 0.$$

Claim 4: The jumps of $(\widehat{L}^{(N)}, \widehat{\mathcal{M}}_\nu^{(N)}(f), \widehat{\mathcal{M}}_\eta^{(N)}(g))$ are asymptotically negligible and

$$(\widehat{L}^{(N)}, \widehat{\mathcal{M}}_\nu^{(N)}(f), \widehat{\mathcal{M}}_\eta^{(N)}(g)) \Rightarrow (\sigma B, \widehat{\mathcal{M}}_\nu(f), \widehat{\mathcal{M}}_\eta(g)) \qquad \text{as } N \to \infty. \tag{3.7.16}$$

The size of the jumps of $\widehat{E}^{(N)}$ and $\widehat{L}^{(N)}$ converge to zero as $N \to \infty$ because $\widehat{E}^{(N)}$ is a counting process with unit jumps and $\sup_{t \le T} |\widehat{\gamma}^{(N)}(t)| \to 0$ in probability. Also by Lemma 3.4.5 and the continuity of $A_{f,\nu}^{(N)}$ and $A_{g,\eta}^{(N)}$, the sizes of the jumps of $\widehat{\mathcal{M}}_\nu^{(N)}(f)$ and $\widehat{\mathcal{M}}_\eta^{(N)}(g)$ are uniformly bounded by $\|f\|_\infty/\sqrt{N}$ and $\|g\|_\infty/\sqrt{N}$ respectively, so they converge to zero in probability as well. Because $\{(\widehat{L}^{(N)}, \widehat{\mathcal{M}}_\nu^{(N)}(f), \widehat{\mathcal{M}}_\eta^{(N)}(g))\}$ is a sequence of martingales starting at zero, applying the martingale central limit theorem (see Theorem 1.4 on page 339[17]), we see that (3.7.16) holds, where the conditions of the theorem are verified by Claims 1,2,3 and the first assertion of Claim 4, hence the second part of the Claim follows from the observation that $\sigma^2\cdot$, $\overline{A}_{f^2,\nu}(t)$ and $\overline{A}_{g^2,\eta}(t)$ are variance processes of the independent centered Gaussian processes σB, $\widehat{\mathcal{M}}_\nu(f)$ and $\widehat{\mathcal{M}}_\eta(g)$ respectively.

For the next two claims, we introduce some standard Markov process notation associated with the Markov process $(\alpha_E^{(N)}, X^{(N)}, \nu^{(N)}, \eta^{(N)}) = \{(\alpha_E^{(N)}(t), X^{(N)}(t), \nu_t^{(N)}, \eta_t^{(N)}),$ $t \ge 0\}$. Let $\mathcal{G}_t^{(N),0}$ be the σ-algebra generated by $\{(\alpha_E^{(N)}(s), X^{(N)}(s), \nu_s^{(N)}, \eta_s^{(N)}), s \in [0,t]\}$, and let $\mathcal{G}_\infty^{(N),0} = \sigma(\cup_{t\ge 0}\mathcal{G}_t^{(N),0})$. For $(r,k,\mu,\gamma) \in [0,\infty) \times \mathbb{N} \times \mathcal{M}_F[0,H^s) \times \mathcal{M}_F[0,H^r)$, let $\mathbb{P}_{r,k,\mu,\gamma}^{(N)}$ be the law of the Markov process $(\alpha_E^{(N)}, X^{(N)}, \nu^{(N)}, \eta^{(N)})$ with initial condition $(\alpha_E^{(N)}(0), X^{(N)}(0), \nu_0^{(N)}, \eta_0^{(N)}) = (r,k,\mu,\gamma)$. Also, let $\{\tilde{\mathcal{G}}_t^{(N)}, t \ge 0\}$ be the usual augmentation of the filtration $\{\mathcal{G}_t^{(N),0}, t \ge 0\}$ and let $\mathcal{G}_t^{(N)} = \cap_{s>t}\mathcal{G}_s^{(N),0}, t \ge 0$, be the associated right-continuous filtration.

Claim 5: For every $N \in \mathbb{N}$, the process $\{\widehat{L}_t^{(N)}, \widehat{\mathcal{M}}_{t,\nu}^{(N)}(f), \widehat{\mathcal{M}}_{t,\eta}^{(N)}(g), t \ge 0\}$ is $\{\mathcal{G}^{(N)}\}$-adapted. Moreover, let $\{r^{(N)}, k^{(N)}, \mu^{(N)}, \gamma^{(N)}\}_{N\in\mathbb{N}} \subset [0,\infty)\times\mathbb{N}\times\mathcal{M}_F[0,H^s)\times$

$\mathcal{M}_F[0,H^r)$ be a deterministic sequence that satisfies $r^{(N)} \to 0$, $k^{(N)} \to \overline{X}(0)$, $\frac{\mu^{(N)}}{N} \Rightarrow (\lambda \wedge 1)\nu_*$, $\frac{\gamma^{(N)}}{N} \Rightarrow \lambda\eta_*$, where $(\overline{X}(0),\ (\lambda \wedge 1)\nu_*,\ \lambda\eta_*) \in I_\lambda$ are the invariant manifold limit of the fluid-scaled initial conditions defined in (3.3.21)-(3.3.24) provided Assumption 3.3.8 holds. Then, given any bounded and continuous functional F on $\mathcal{D}_{\mathbb{R}^3}[0,\infty)$,

$$\lim_{N\to\infty} \mathbb{E}^{(N)}_{(r^{(N)},k^{(N)},\mu^{(N)},\gamma^{(N)})}[F(\widehat{L}^{(N)}, \widehat{\mathcal{M}}^{(N)}_\nu(f), \widehat{\mathcal{M}}^{(N)}_\eta(g))] = \mathbb{E}[F(\sigma B, \widehat{\mathcal{M}}_\nu(f), \widehat{\mathcal{M}}_\nu(f))]. \tag{3.7.17}$$

Fix $t \in [0,\infty)$. It is easy to see from the definitions in Section 3.2.1, Assumption 3.6.1 and (3.7.12) that $E^{(N)}(t)$, $\xi^{(N)}_j$, $j \le E^{(N)}(t)$, and hence $\widehat{L}^{(N)}(t)$, are all measurable with respect to $\sigma(\alpha^{(N)}_E(s), s \in [0,t]) \subset \mathcal{G}^{(N)}_t$. In a similar fashion, for $f \in \mathcal{C}_b([0,H^s) \times \mathbb{R})$, it follows from the definitions given in (3.4.1) and (3.2.13) that $A^{(N)}_{f,\nu}$ and $D^{(N)}_f$ are measurable with respect to $\sigma(\nu^{(N)}_s, s \in [0,t)) \subseteq \mathcal{G}^{(N)}_t$. Thus, from (3.4.3) and (3.5.4) it follows that $\widehat{\mathcal{M}}^{(N)}_{t,\nu}(f)$ is $\mathcal{G}^{(N)}_t$-measurable. Similarly, for $g \in \mathcal{C}_b([0,H^r)\times\mathbb{R})$, it follows from the definition given in (3.4.2) and the last claim in Proposition 3.2.15 that $A^{(N)}_{g,\eta}$ and $S^{(N)}_g$ are measurable with respect to $\sigma(\eta^{(N)}_s, s \in [0,t)) \subseteq \mathcal{G}^{(N)}_t$. Thus from (3.4.1) and (3.5.4) it follows that $\widehat{\mathcal{M}}^{(N)}_{t,\eta}(g)$ is also $\mathcal{G}^{(N)}_t$-measurable. Hence the first assertion of the claim is proved. Since $\widehat{L}^{(N)}$, $\widehat{\mathcal{M}}^{(N)}_\nu(f)$ and $\widehat{\mathcal{M}}^{(N)}_\eta(g)$ are adapted with respect to $\{\mathcal{G}^{(N)}_t\}$, one can use the Markov property and repeat the argument used in the proof of Claim 4, to establish (3.7.17).

Claim 6. The asymptotic independence property holds, that is, for real-valued, bounded continuous functions F_1 on $\mathcal{D}_{\mathbb{R}^3}[0,\infty)$ and F_2 on $\mathbb{R}\times\mathbb{H}^1(0,\infty)^2$,

$$\begin{aligned}\lim_{N\to\infty} &\mathbb{E}[F_1(\widehat{L}^{(N)}, \widehat{\mathcal{M}}^{(N)}_\nu(f), \widehat{\mathcal{M}}^{(N)}_\eta(g))F_2(\widehat{X}^{(N)}(0), \widehat{Z}^{(N)}_\nu(0,\cdot), \widehat{Z}^{(N)}_\eta(0,\cdot))] \\ &= \mathbb{E}[F_1(\sigma B, \widehat{\mathcal{M}}_\nu(f), \widehat{\mathcal{M}}_\eta(g))]\mathbb{E}[F_2(\widehat{X}(0), \widehat{Z}_\nu(0,\cdot), \widehat{Z}_\eta(0,\cdot))] \\ &= \mathbb{E}[F_1(\sigma B, \widehat{\mathcal{M}}_\nu(f), \widehat{\mathcal{M}}_\eta(g))F_2(\widehat{X}(0), \widehat{Z}_\nu(0,\cdot), \widehat{Z}_\eta(0,\cdot))].\end{aligned} \tag{3.7.18}$$

Let F_1 be a continuous functional on $\mathcal{D}_{\mathbb{R}^3}[0,\infty)$ and F_2 a continuous functional on $\mathbb{R}\times\mathbb{H}^1(0,\infty)^2$. Then

$$\begin{aligned}&\lim_{N\to\infty}\mathbb{E}[F_1(\widehat{L}^{(N)},\widehat{\mathcal{M}}^{(N)}_\nu(f),\widehat{\mathcal{M}}^{(N)}_\eta(g))F_2(\widehat{X}^{(N)}(0),\widehat{Z}^{(N)}_\nu(0,\cdot),\widehat{Z}^{(N)}_\eta(0,\cdot))]\\&\quad=\lim_{N\to\infty}\mathbb{E}[\mathbb{E}[F_1(\widehat{L}^{(N)},\widehat{\mathcal{M}}^{(N)}_\nu(f),\widehat{\mathcal{M}}^{(N)}_\eta(g))F_2(\widehat{X}^{(N)}(0),\widehat{Z}^{(N)}_\nu(0,\cdot),\widehat{Z}^{(N)}_\eta(0,\cdot))|\mathcal{G}^{(N)}_0]].\end{aligned}\tag{3.7.19}$$

Note that, $\widehat{X}^{(N)}(0)$, $\widehat{Z}^{(N)}_\nu(0,\cdot)$ and $\widehat{Z}^{(N)}_\eta(0,\cdot)$ are measurable functions of $X^{(N)}(0)$, $\nu^{(N)}_0$ and $\eta^{(N)}_0$ because the corresponding fluid limit quantities are almost surely deterministic. Therefore, $F_2(\widehat{X}^{(N)}(0),\widehat{Z}^{(N)}_\nu(0,\cdot),\widehat{Z}^{(N)}_\eta(0,\cdot))$ is $\mathcal{G}^{(N)}_0$-measurable. Since $\{(\widehat{L}^{(N)}(t),\widehat{\mathcal{M}}^{(N)}_{t,\nu}(f),\widehat{\mathcal{M}}^{(N)}_{t,\eta}(g)),t\geq 0\}$ is $\{\mathcal{G}^{(N)}_t\}$-adapted by Claim 5, the Markov property shows that

$$\begin{aligned}&\mathbb{E}[\mathbb{E}[F_1(\widehat{L}^{(N)},\widehat{\mathcal{M}}^{(N)}_\nu(f),\widehat{\mathcal{M}}^{(N)}_\eta(g))F_2(\widehat{X}^{(N)}(0),\widehat{Z}^{(N)}_\nu(0,\cdot),\widehat{Z}^{(N)}_\eta(0,\cdot))|\mathcal{G}^{(N)}_0]]\\&=\mathbb{E}[F_2(\widehat{X}^{(N)}(0),\widehat{Z}^{(N)}_\nu(0,\cdot),\widehat{Z}^{(N)}_\eta(0,\cdot))\mathbb{E}^{(N)}_{(\alpha^{(N)}_E,X^{(N)}(0),\nu^{(N)}_0,\eta^{(N)}_0)}[F_1(\widehat{L}^{(N)},\widehat{\mathcal{M}}^{(N)}_\nu(f),\widehat{\mathcal{M}}^{(N)}_\eta(g))]].\end{aligned}\tag{3.7.20}$$

Since $(R^{(N)}_E,\overline{X}^{(N)}(0),\overline{\nu}^{(N)}_0,\overline{\eta}^{(N)}_0)\to(0,\overline{X}(0),(\lambda\wedge 1)\nu_*,\lambda\eta_*)$ $\mathbb{P}$- almost surely by Assumption 3.3.1, the conditions of Claim 5 are satisfied and it follows from (3.7.17) that $\mathbb{P}$-almost surely,

$$\lim_{N\to\infty}\mathbb{E}^{(N)}_{(\alpha^{(N)}_E,X^{(N)}(0),\nu^{(N)}_0,\eta^{(N)}_0)}[F_1(\widehat{L}^{(N)},\widehat{\mathcal{M}}^{(N)}_\nu(f),\widehat{\mathcal{M}}^{(N)}_\eta(g))]=\mathbb{E}[F_1(\sigma B,\widehat{\mathcal{M}}_\nu(f),\widehat{\mathcal{M}}_\eta(g))].$$

Since F_1 and F_2 are bounded (say by $\overline{C}$), the bounded convergence theorem implies that

$$\begin{aligned}
&\lim_{N\to\infty} |\mathbb{E}[F_2(\widehat{X}^{(N)}(0), \widehat{Z}^{(N)}_\nu(0,\cdot), \widehat{Z}^{(N)}_\eta(0,\cdot))\mathbb{E}^{(N)}_{(\alpha_E^{(N)}, X^{(N)}(0), \nu_0^{(N)}, \eta_0)}[F_1(\widehat{L}^{(N)}, \widehat{\mathcal{M}}^{(N)}_\nu(f), \widehat{\mathcal{M}}^{(N)}_\eta(g))]] \\
&\qquad -\mathbb{E}[F_2(\widehat{X}^{(N)}(0), \widehat{Z}^{(N)}_\nu(0,\cdot), \widehat{Z}^{(N)}_\eta(0,\cdot))\mathbb{E}[F_1(\sigma B, \widehat{\mathcal{M}}_\nu(f), \widehat{\mathcal{M}}_\eta(g))]]| \\
&\le \overline{C} \lim_{N\to\infty} \mathbb{E}[|\mathbb{E}^{(N)}_{(\alpha_E^{(N)}, X^{(N)}(0), \nu_0^{(N)}, \eta_0)}[F_1(\widehat{L}^{(N)}, \widehat{\mathcal{M}}^{(N)}_\nu(f), \widehat{\mathcal{M}}^{(N)}_\eta(g))] - \mathbb{E}[F_1(\sigma B, \widehat{\mathcal{M}}_\nu(f), \widehat{\mathcal{M}}_\eta(g))]|] \\
&= 0. \qquad (3.7.21)
\end{aligned}$$

On the other hand, since F_2 is bounded and continuous, by Assumption 3.6.7,

$$\lim_{N\to\infty} \mathbb{E}[F_2(\widehat{X}^{(N)}(0), \widehat{Z}^{(N)}_\nu(0,\cdot), \widehat{Z}^{(N)}_\eta(0,\cdot))] = \mathbb{E}[F_2(\widehat{X}(0), \widehat{Z}_\nu(0,\cdot), \widehat{Z}_\eta(0,\cdot))]. \quad (3.7.22)$$

Combining (3.7.19)-(3.7.22), we obtain the first equality in (3.7.18). The second equality of (3.7.18) follows from the first due to the fact that, by construction, $(\sigma B, \widehat{\mathcal{M}}_\nu(f), \widehat{\mathcal{M}}_\eta(g))$ depend only on the fluid initial conditions $\overline{x}_0$, $(\lambda \wedge 1)\nu_*$, $\lambda\eta_*$, which are $\mathbb{P}$ almost surely deterministic.

Given (3.7.12), an application of the continuous mapping theorem then shows that for every $f \in \mathcal{C}_b([0,H^s)\times\mathbb{R})$ and $g \in \mathcal{C}_b([0,H^r)\times\mathbb{R})$, $(\widehat{L}^{(N)}, \widehat{\mathcal{M}}^{(N)}_\nu(f), \widehat{\mathcal{M}}^{(N)}_\eta(g)) \Rightarrow (\sigma B, \widehat{\mathcal{M}}_\nu(f), \widehat{\mathcal{M}}_\eta(g))$ in $\mathcal{D}_\mathbb{R}[0,\infty)^3$. Similarly, the continuous mapping theorem and (3.7.18) imply that for every $f \in \mathcal{C}_b([0,H^s)\times\mathbb{R})$ and $g \in \mathcal{C}_b([0,H^r)\times\mathbb{R})$, as $N\to\infty$,

$$\begin{aligned}
&(\widehat{L}^{(N)}, \widehat{X}^{(N)}, \widehat{Z}^{(N)}_\nu(0,\cdot), \widehat{Z}^{(N)}_\eta(0,\cdot), \widehat{\mathcal{M}}^{(N)}_\nu(f), \widehat{\mathcal{M}}^{(N)}_\eta(g)) \\
&\qquad \Rightarrow (\sigma B, \widehat{X}, \widehat{Z}_\nu(0,\cdot), \widehat{Z}_\eta(0,\cdot), \widehat{\mathcal{M}}_\nu(f), \widehat{\mathcal{M}}_\eta(g))
\end{aligned}$$

□

Corollary 3.7.6. The processes $\widehat{L}^{(N)}$, $\widehat{\mathcal{M}}^{(N)}_D$ and $\widehat{\mathcal{M}}^{(N)}_R$ are asymptotically independent, as $N \to \infty$. Moreover,

$$\widehat{L}^{(N)}(t) - \widehat{\mathcal{M}}_D^{(N)}(t) - \widehat{\mathcal{M}}_R^{(N)}(t) \Rightarrow \int_0^t \theta(s)dB(s), \tag{3.7.23}$$

where

$$\theta(t) = \sqrt{\sigma^2 + \langle h^s, (\lambda \wedge 1)\nu_* \rangle + \left(\int_0^{Q_*} h^r \left((F^{\lambda\eta_*})^{-1}(y)\right) dy\right)}.$$

Proof. First note that by Lemma 3.4.3, Proposition 3.4.4, and Proposition 7.2 [22],

$$[\widehat{\mathcal{M}}_R^{(N)}]_t \to \int_0^t \left(\int_0^{Q_*} h^r \left((F^{\lambda\eta_*})^{-1}(y)\right) dy\right) ds, \qquad \text{as} \qquad N \to \infty.$$

Substituting $g = \theta^{(N)}$ and $f = \mathbf{1}$ in Claim 3 of Proposition 3.7.5, we see that,

$$[\widehat{\mathcal{M}}_D^{(N)}, \widehat{\mathcal{M}}_R^{(N)}]_t \to 0,$$

$$[\widehat{L}^{(N)}, \widehat{\mathcal{M}}_D^{(N)}]_t \to 0,$$

and

$$[\widehat{L}^{(N)}, \widehat{\mathcal{M}}_R^{(N)}]_t \to 0,$$

in probability as $N \to \infty$. Now, using the same argument as in Claim 4 of Proposition 3.7.5, proves (3.7.23). □

3.8 N-server dynamics

In this section we provide a characterization of the centered state process introduced in (3.5.4) and (3.5.6).

Proposition 3.8.1. For each $N \in \mathbb{N}$, the process $(\widehat{X}^{(N)}, \widehat{Z}_\nu^{(N)}, \widehat{Z}_\eta^{(N)})$ almost surely satisfies the following coupled set of equations:

$$\widehat{Z}_{\nu}^{(N)}(t,r) = \widehat{Z}_{\nu}^{(N)}(0,r) + \int_0^t \partial_r \widehat{Z}_{\nu}^{(N)}(s,r)ds - \widehat{\mathcal{M}}_{t,\nu}^{(N)}(\alpha^r) + (1-G^s(r))\widehat{K}^{(N)}(t), \tag{3.8.1}$$

$$\widehat{Z}_{\eta}^{(N)}(t,r) = \widehat{Z}_{\eta}^{(N)}(0,r) + \int_0^t \partial_r \widehat{Z}_{\eta}^{(N)}(s,r)ds - \widehat{\mathcal{M}}_{t,\eta}^{(N)}(1_{[0,\chi*]}\beta^r) + (1-G^r(r))\widehat{E}^{(N)}(t), \tag{3.8.2}$$

where $K^{(N)}$ satisfies

$$\widehat{K}^{(N)}(t) = \widehat{Z}_{\nu}^{(N)}(t,0) - \widehat{Z}_{\nu}^{(N)}(0,0) + \widehat{\mathcal{M}}_{t,\nu}^{(N)}(\mathbf{1}) + \sqrt{N}\left(\overline{A}_{\mathbf{1},\nu}^{(N)}(t) - \int_0^t \langle h^s, \overline{\nu}_s\rangle\, ds\right) \tag{3.8.3}$$

$$\begin{aligned} &= \widehat{Z}_{\nu}^{(N)}(t,0) - \widehat{Z}_{\nu}^{(N)}(0,0) + \widehat{X}^{(N)}(0) - \widehat{X}^{(N)}(t) + \widehat{E}^{(N)}(t) \\ &\quad - \widehat{\mathcal{M}}_{R}^{(N)}(t) - \sqrt{N}(\overline{A}_{R}^{(N)}(t) - \overline{R}(t)), \end{aligned} \tag{3.8.4}$$

and there exists $\Omega^* \in \mathcal{F}$ with $\mathbb{P}(\Omega^*) = 1$ such that for every $T < \infty$ and $\omega \in \Omega^*$ there exists $N_0 = N_0(\omega, t) < \infty$ such that for all $N \geq N_0$

$$\widehat{Z}_{\nu}^{(N)}(t,0) = \begin{cases} \widehat{X}^{(N)} & \text{if } \overline{X}(t) < 1, \\ \widehat{X}^{(N)}(t) \wedge 0 & \text{if } \overline{X}(t) = 1, \\ 0 & \text{if } \overline{X}(t) > 1. \end{cases} \tag{3.8.5}$$

Proof. Equation (3.8.1) is obtained by substituting $\varphi(x,t) = \frac{1-G^s(x+r)}{1-G^s(x)}$ in equations (3.2.21) and (3.3.7) and then dividing (3.2.21) by N, subtracting the corresponding side of (3.3.7) from it and multiplying the resulting quantities by $\sqrt{N}$ and using the fact that $M_{\alpha^r,\nu}^{(N)}$ is indistinguishable from $\mathcal{M}_{\alpha^r,\nu}^{(N)}$. In an analogous manner (3.8.2) is obtained by using a standard approximation argument and substituting $\psi(x,t) = 1_{[0,\chi*]}(x)\frac{1-G^r(x+r)}{1-G^r(x)}$ in equations (3.2.22) and (3.3.9) and then dividing (3.2.22) by N, subtracting the corresponding side of (3.3.9) from it and multiplying the resulting quantities by $\sqrt{N}$ and using the fact that $M^{(N)}_{1_{[0,\chi*]}\beta^r,\eta}$ is indistinguishable from $\mathcal{M}^{(N)}_{1_{[0,\chi*]}\beta^r,\eta}$. Rearrange (3.2.10), divide it by N and subtract (3.3.8) from it and multiply the resulting quantities by $\sqrt{N}$ to obtain

$$\widehat{K}^{(N)}(t) = \widehat{Z}^{(N)}_\nu(t,0) - \widehat{Z}^{(N)}_\nu(0,0) + \sqrt{N}\left(\overline{D}^{(N)}(t) - \int_0^t \langle h^s, \overline{\nu}_s\rangle\, ds\right).$$

The first equality in equation (3.8.3) is obtained by using the definition (3.4.3) of $\widehat{\mathcal{M}}^{(N)}_\nu(\mathbf{1})$ and (3.5.3) in the above equation. The second equality in equation (3.8.3) is obtained by substituting $D^{(N)}(t)$ from (3.2.8) and $\int_0^t \langle h^s, \overline{\nu}_s\rangle ds$ from (3.3.13) and using (3.4.7) in the above equation. The final equation (3.8.5) follows from Proposition 6.2 [25] and the corresponding Remark 6.3 [25] using (3.2.24) and (3.3.14). □

For $N \in \mathbb{N}$ and continuous f, define

$$\widehat{\mathcal{K}}^{(N)}_t(f) := \int_{[0,t]} (1 - G^s(t-s))f(t-s)d\widehat{K}^{(N)}(s), \qquad t \in [0,\infty). \tag{3.8.6}$$

By applying integration by parts to the right-hand side of (3.8.6), and using the fact that $\widehat{K}^{(N)}$ has at most a countable number of discontinuities, we see that for all absolutely continuous functions f defined on $[0, H^s)$

$$\widehat{\mathcal{K}}^{(N)}_t(f) = f(0)\widehat{K}^{(N)}(t) - \int_{[0,t]} \widehat{K}^{(N)}(s)\frac{d((1-G^s(t-s))f(t-s))}{ds}ds, \qquad t \in [0,\infty), \tag{3.8.7}$$

Similarly given a continuous function f let us define

$$\widehat{\mathcal{E}}^{(N)}_t(f) := \int_{[0,t]} (1 - G^r(t-s))f(t-s)d\widehat{E}^{(N)}(s), \qquad t \in [0,\infty). \tag{3.8.8}$$

By applying integration by parts to the right-hand side of (3.8.8), and using the fact that $\widehat{E}^{(N)}$ has at most a countable number of discontinuities, we see that for all absolutely continuous functions f defined on $[0, H^r)$

$$\widehat{\mathcal{E}}_t^{(N)}(f) = f(0)\widehat{E}^{(N)}(t) - \int_{[0,t]} \widehat{E}^{(N)}(s)\frac{d((1-G^r(t-s))f(t-s))}{ds}ds, \qquad t \in [0,\infty). \tag{3.8.9}$$

We now state the representation result,

Proposition 3.8.2. For every $N \in \mathbb{N}$ and $r, t \in [0,\infty)$,

$$\widehat{Z}_\nu^{(N)}(t,r) = \widehat{Z}_\nu^{(N)}(0,t+r) - \widehat{\mathcal{H}}_{t,\nu}^{(N)}(\alpha^r) + \widehat{\mathcal{K}}_t^{(N)}(\alpha^r), \tag{3.8.10}$$

and

$$\widehat{Z}_\eta^{(N)}(t,r) = \widehat{Z}_\eta^{(N)}(0,t+r) - \widehat{\mathcal{H}}_{t,\eta}^{(N)}(\beta^r) + \widehat{\mathcal{E}}_t^{(N)}(\beta^r), \tag{3.8.11}$$

Proof. The proof follows using the same arguments as in the proof of (3.8.1) and (3.8.2) but with $\varphi(x,s) = \frac{1-G^s(x+t-s+r)}{1-G^s(x)}$ and $\psi(x,s) = 1_{[0,\chi*]}(x)\frac{1-G^r(x+t-s+r)}{1-G^r(x)}$ respectively. □

Proposition 3.8.3. For every $N \in \mathbb{N}$, and $r, t \in [0,\infty)$,

$$\begin{aligned}\partial_r \widehat{Z}_\nu^{(N)}(t,r) =& \partial_r \widehat{Z}_\nu^{(N)}(0,t+r) + \widehat{\mathcal{M}}_{t,\nu}^{(N)}(\Psi_{t+r}^s h^s) \\ &- g^s(r)\widehat{K}^{(N)}(t) - \int_0^t (g^s)'(t-s+r)\widehat{K}^{(N)}(s)ds,\end{aligned} \tag{3.8.12}$$

and

$$\begin{aligned}\partial_r \widehat{Z}_\eta^{(N)}(t,r) =& \partial_r \widehat{Z}_\eta^{(N)}(0,t+r) - \widehat{\mathcal{M}}_{t,\eta}^{(N)}(1_{[0,\chi*]}\Psi_{t+r}^r h^r) \\ &- g^r(r)\widehat{E}^{(N)}(t) - \int_0^t (g^r)'(t-s+r)\widehat{E}^{(N)}(s)ds,\end{aligned} \tag{3.8.13}$$

Proof. The proof follows using the same arguments as in the proof of (3.8.1) and (3.8.2) but with $\varphi(x,s) = \frac{g^s(x+t-s+r)}{1-G^s(x)}$ and $\psi(x,s) = 1_{[0,\chi_*]}(x)\frac{g^r(x+t-s+r)}{1-G^r(x)}$ respectively, using the definition in (3.5.7) and noting that $\varphi(x,s) = (\Psi^s_{t+r}h^s)(x,s)$ and $\psi(x,s) = (\Psi^r_{t+r}h^r)(x,s)$. □

Before proceeding further let us look at a simplified representation for $\sqrt{N}[\overline{A}^{(N)}_R(t) - \overline{R}(t)]$. By (3.5.1),

$$\sqrt{N}\left[\overline{A}^{(N)}_R(t) - \overline{R}(t)\right] = \sqrt{N}\left[\overline{A}^{(N)}_{\theta^{(N)},\eta}(t) - \overline{R}(t)\right].$$

Using (3.4.6), (3.4.9) and (3.3.12), the above expression is equal to

$$\sqrt{N}\left[\int_0^t \left[\int_0^{\overline{Q}^{(N)}(s)+(\iota^{(N)}(s)/N)} h^r((F^{\overline{\eta}^{(N)}_s})^{-1}(y))dy - \int_0^{\overline{Q}(s)} h^r\left((F^{\overline{\eta}_s})^{-1}(y))dy\right)\right]ds\right], \tag{3.8.14}$$

where $\overline{Q}(s) = Q_*$ and $\overline{\eta}_s = \lambda\eta_*$ for all $s \geq 0$. Since $\overline{\eta}_s$ is a right continuous measure for all $s \in [0,\infty)$, using the first equation in assumption (3.3.6), we apply the change of variables formula to get,

$$\int_0^t \int_0^{\overline{Q}(s)} h^r\left((F^{\overline{\eta}_s})^{-1}(y)\right)dyds = \int_0^t \int_0^{(F^{\overline{\eta}_s})^{-1}(\overline{Q}(s))} h^r(y)d\overline{\eta}_s(y)ds.$$

Using change of variable again

$$\begin{aligned}
&\int_0^{\overline{Q}^{(N)}(s)+(\iota^{(N)}(s)/N)} h^r((F^{\overline{\eta}^{(N)}_s})^{-1}(y))dy \\
&\quad= \int_0^{\overline{Q}^{(N)}(s)} h^r((F^{\overline{\eta}^{(N)}_s})^{-1}(y))dy + \int_{\overline{Q}^{(N)}(s)}^{\overline{Q}^{(N)}(s)+(\iota^{(N)}(s)/N)} h^r((F^{\overline{\eta}^{(N)}_s})^{-1}(y))dy \\
&\quad= \int_0^{(F^{\overline{\eta}^{(N)}_s})^{-1}(\overline{Q}^{(N)}(s))} h^r(y)d\overline{\eta}^{(N)}_s(y) + \int_{\overline{Q}^{(N)}(s)}^{\overline{Q}^{(N)}(s)+(\iota^{(N)}(s)/N)} h^r((F^{\overline{\eta}^{(N)}_s})^{-1}(y))dy.
\end{aligned}$$

Substituting both the equations above in (3.8.14), using the definition of $\partial_r \widehat{Z}_\eta^{(N)}$ from (3.5.7) and rearranging we have

$$\sqrt{N}\left[\int_0^t \left[\int_0^{(F^{\overline{\eta}_s^{(N)}})^{-1}(\overline{Q}^{(N)}(s)+(\iota^{(N)}(s)/N))} h^r(y)d\overline{\eta}_s^{(N)}(y) - \int_0^{\overline{Q}(s)} h^r\left((F^{\overline{\eta}_s})^{-1}(y)\right)dy\right] ds\right]$$
$$= -\int_0^t \partial_r \widehat{Z}_\eta^{(N)}(s,0)ds + \sqrt{N}\int_0^t \int_{(F^{\overline{\eta}_s})^{-1}(\overline{Q}(s))}^{(F^{\overline{\eta}_s^{(N)}})^{-1}(\overline{Q}^{(N)}(s))} h^r(y)d\overline{\eta}_s^{(N)}(y)ds$$
$$+ \sqrt{N}\int_0^t \int_{\overline{Q}^{(N)}(s)}^{\overline{Q}^{(N)}(s)+(\iota^{(N)}(s)/N)} h^r((F^{\overline{\eta}_s^{(N)}})^{-1}(y))dyds. \tag{3.8.15}$$

Using mean value theorem, there exists a $C^{(N)}(s) \in (0, \iota^{(N)}(s)/N)$ such that

$$\sqrt{N}\int_0^t \int_{\overline{Q}^{(N)}(s)}^{\overline{Q}^{(N)}(s)+(\iota^{(N)}(s)/N)} h^r((F^{\overline{\eta}_s^{(N)}})^{-1}(y))dyds$$
$$= \frac{1}{\sqrt{N}}\int_0^t \iota^{(N)}(s)h^r((F^{\overline{\eta}_s^{(N)}})^{-1}(\overline{Q}^{(N)}(s)+C^{(N)}(s)))ds.$$

Since $\iota^{(N)} = 0$ a.s. for a fixed N, we see that the above equation reduces to

$$\sqrt{N}\int_0^t \int_{\overline{Q}^{(N)}(s)}^{\overline{Q}^{(N)}(s)+(\iota^{(N)}(s)/N)} h^r((F^{\overline{\eta}_s^{(N)}})^{-1}(y))dyds = 0. \tag{3.8.16}$$

Using the change of variable formula

$$\sqrt{N}\textstyle\int_0^t \displaystyle\int_{(F^{\overline{\eta}_s})^{-1}(\overline{Q}(s))}^{(F^{\overline{\eta}_s^{(N)}})^{-1}(\overline{Q}^{(N)}(s))} h^r(y)d\overline{\eta}_s^{(N)}(y)ds$$
$$= \sqrt{N}\int_0^t \int_{F^{\overline{\eta}_s^{(N)}}\left(\left(F^{\overline{\eta}_s}\right)^{-1}(\overline{Q}(s))\right)}^{\overline{Q}^{(N)}(s)} h^r\left((F^{\overline{\eta}_s^{(N)}})^{-1}(x)\right)dxds.$$

Therefore using the above equation and (3.3.25)

$$
\begin{aligned}
&\sqrt{N}\int_0^t \int_{(F^{\overline{\eta}_s})^{-1}(\overline{Q}(s))}^{(F^{\overline{\eta}_s^{(N)}})^{-1}(\overline{Q}^{(N)}(s))} h^r(y)d\overline{\eta}_s^{(N)}(y)ds \\
&\quad= \sqrt{N}\int_0^t \int_{\overline{\eta}_s^{(N)}[0,\chi_*]}^{\overline{Q}^{(N)}(s)} h^r\left((F^{\overline{\eta}_s^{(N)}})^{-1}(x)\right)dxds \\
&\quad= \sqrt{N}\int_0^t \int_{\overline{Q}(s)}^{\overline{Q}^{(N)}(s)} h^r\left((F^{\overline{\eta}_s^{(N)}})^{-1}(x)\right)dxds + \sqrt{N}\int_0^t \int_{\overline{\eta}_s^{(N)}[0,\chi_*]}^{\overline{Q}(s)} h^r\left((F^{\overline{\eta}_s^{(N)}})^{-1})(x)\right)dxds.
\end{aligned}
$$

Using mean value theorem, there exists a

$$
C_1^{(N)}(s) \in (0, \overline{Q}^{(N)}(s) - \overline{Q}(s)), \tag{3.8.17}
$$

and

$$
C_2^{(N)}(s) \in (0, \overline{Q}(s) - \overline{\eta}_s^{(N)}[0, \overline{\chi}(s)]), \tag{3.8.18}
$$

such that

$$
\begin{aligned}
&\sqrt{N}\int_0^t \int_{\overline{Q}(s)}^{\overline{Q}^{(N)}(s)} h^r\left((F^{\overline{\eta}_s^{(N)}})^{-1}(x)\right)dxds + \sqrt{N}\int_0^t \int_{\overline{\eta}_s^{(N)}[0,\chi_*]}^{\overline{Q}(s)} h^r\left((F^{\overline{\eta}_s^{(N)}})^{-1})(x)\right)dxds \\
&\;=\sqrt{N}\int_0^t (\overline{Q}^{(N)}(s) - \overline{Q}(s))h^r\left((F^{\overline{\eta}_s^{(N)}})^{-1}(\overline{Q}(s) + C_1^{(N)}(s))\right)ds \\
&\quad+ \sqrt{N}\int_0^t \left(\overline{Q}(s) - \overline{\eta}_s^{(N)}[0,\chi_*]\right) h^r\left((F^{\overline{\eta}_s^{(N)}})^{-1})(\overline{\eta}_s^{(N)}[0,\chi_*] + C_2^{(N)}(s))\right)ds \\
&\;= \int_0^t \widehat{Q}^{(N)}(s)h^r\left((F^{\overline{\eta}_s^{(N)}})^{-1}(\overline{Q}(s) + C_1^{(N)}(s))\right)ds \\
&\quad+ \int_0^t \widehat{Z}_\eta^{(N)}(s,0)h^r\left((F^{\overline{\eta}_s^{(N)}})^{-1})(\overline{\eta}_s^{(N)}[0,\chi_*] + C_2^{(N)}(s))\right)ds.
\end{aligned} \tag{3.8.19}
$$

Hence using (3.8.15) - (3.8.19), we obtain

$$
\begin{aligned}
\sqrt{N}\left[\overline{A}_R^{(N)}(t) - \overline{R}(t)\right] = &-\int_0^t \partial_r \widehat{Z}_\eta^{(N)}(s,0)ds + \int_0^t \widehat{Q}^{(N)}(s)h^r\left((F^{\overline{\eta}_s^{(N)}})^{-1}(\overline{Q}(s) + C_1^{(N)}(s))\right)ds \\
&+ \int_0^t \widehat{Z}_\eta^{(N)}(s,0)h^r\left((F^{\overline{\eta}_s^{(N)}})^{-1})(\overline{\eta}_s^{(N)}[0,\chi_*] + C_2^{(N)}(s))\right)ds.
\end{aligned} \tag{3.8.20}
$$

Note that $h^r \in B^+[0,\infty)$ by Assumption 3.6.5. Also note that, using Theorem 3.3.7, and equations (3.2.11) and (3.3.10), we have $\overline{Q}^{(N)}(s) \to \overline{Q}(s)$. Since $F^{\overline{\eta}_s} = F^{\lambda\eta^*} = \lambda\eta_*[0,\cdot] = \lambda\int_0^\cdot(1-G^r(y))dy$ is strictly increasing on its support, $(F^{\overline{\eta}_s})^{-1}$ is a continuous function. Also, since $(F^{\overline{\eta}_s^{(N)}})^{-1} \to (F^{\overline{\eta}_s})^{-1}$ on $[0, F^{\overline{\eta}_s}(H^r-))$ in the M_1 topology (see the first paragraph of the proof of Lemma 7.5 [22]), using the continuity of $(F^{\overline{\eta}_s})^{-1}$, Theorem 12.4.1 [45] and the continuity of h^r, we have

$$h^r\left((F^{\overline{\eta}_s^{(N)}})^{-1})(\overline{Q}(s)+C_1^{(N)}(s))\right) \to h^r(\chi_*) \qquad \text{as} \qquad N\to\infty, \tag{3.8.21}$$

and

$$h^r\left((F^{\overline{\eta}_s^{(N)}})^{-1})(\overline{\eta}_s^{(N)}[0,\chi_*]+C_2^{(N)}(s))\right) \to h^r(\chi_*) \qquad \text{as} \qquad N\to\infty, \tag{3.8.22}$$

Lemma 3.8.4. *Suppose Assumptions 3.6.4 and 3.6.6 hold. If the fluid limit is either subcritical or supercritical, there exists $\Omega^* \in \mathcal{F}$ with $\mathbb{P}(\Omega^*) = 1$ such that for every $T < \infty$ and $\omega \in \Omega^*$, there exists $N_0(\omega, T) < \infty$ such that for all $N \geq N_0(\omega, T)$,*

$$\begin{aligned}
&(\widehat{K}^{(N)}, \widehat{X}^{(N)}, \widehat{Z}_\nu^{(N)}(\cdot,0))(\omega) \qquad\qquad (3.8.23)\\
&\quad = \Lambda\Bigg(\widehat{E}^{(N)} - \widehat{\mathcal{M}}_R^{(N)} + \int_0^t \partial_r \widehat{Z}_\eta^{(N)}(s,0)ds\\
&\qquad\qquad - \int_0^t \widehat{Z}_\eta^{(N)}(s,0)h^r\left((F^{\overline{\eta}_s^{(N)}})^{-1})(\overline{\eta}_s^{(N)}[0,\chi_*]+C_2^{(N)}(s))\right)ds,\\
&\qquad\qquad \widehat{X}^{(N)}(0), \widehat{Z}_\nu^{(N)}(0,\cdot) - \widehat{\mathcal{H}}_\nu^{(N)}(\mathbf{1}), h^r\left((F^{\overline{\eta}_\cdot^{(N)}})^{-1}(\overline{Q}(\cdot)+C_1^{(N)}(\cdot))\right)\Bigg)(\omega),
\end{aligned}$$

on the interval $[0,T]$. Moreover, when the fluid limit is critical, (3.8.23) holds almost surely for every $N \in \mathbb{N}$ and $t \in [0,\infty)$.

Proof. Fix $N \in \mathbb{N}$. By the basic definition of the model, $\widehat{E}^{(N)}$ is càdlàg. Assumption 3.6.4 and Remark 3.6.8 show that $\widehat{Z}_\nu^{(N)}(0,\cdot)$ is continuous. Moreover, since $\widehat{\mathcal{K}}^{(N)}$ and $\widehat{Z}_\nu^{(N)}(\cdot,0)$ are càdlàg, from the representation (3.8.10), it follows that $\widehat{\mathcal{H}}_\nu^{(N)}(\mathbf{1})$ is also càdlàg. By Remark 3.4.2 and its definition $\widehat{\mathcal{M}}_R^{(N)}$ is càdlàg. Note that $\widehat{Z}_\eta^{(N)}(s,0)$ is right-continuous by definition. Thus, for almost surely every $\omega \in \Omega$, $(\widehat{E}^{(N)} - \widehat{\mathcal{M}}_R^{(N)} + \int_0^t \partial_r \widehat{Z}_\eta^{(N)}(s,0)ds - \int_0^t \widehat{Z}_\eta^{(N)}(s,0)h^r \left((F^{\overline{\eta}_s^{(N)}})^{-1})(\overline{\eta}_s^{(N)}[0,\chi_*] + C_2^{(N)}(s))\right) ds, \widehat{X}^{(N)}(0), \widehat{Z}_\nu^{(N)}(0,\cdot) - \widehat{\mathcal{H}}_\nu^{(N)}(\mathbf{1}), h^r\left((F^{\overline{\eta}_\cdot^{(N)}})^{-1}(\overline{Q}(\cdot)+C_1^{(N)}(\cdot))\right))(\omega) \in \mathcal{D}^0_{\mathbb{R}}[0,\infty) \times \mathbb{R} \times \mathcal{D}_{\mathbb{R}}[0,\infty) \times B^+[0,\infty)$. Note that

$$\widehat{X}^{(N)}(s) = \sqrt{N}[\overline{X}^{(N)}(s) - x_*] \leq \sqrt{N}\overline{X}^{(N)}(s) \leq \frac{X^{(N)}}{\sqrt{N}}.$$

Since the right-hand side of the above equation is non-negative, $|\widehat{X}^{(N)}(s)| \leq \frac{X^{(N)}}{\sqrt{N}}$. Now, using (3.2.8), $X^{(N)}(s) \leq X^{(N)}(0) + \sup_{s\in[0,T]} E^{(N)}(s) < \infty$ for all $s \in [0,T]$. Hence, $\sup_{s\in[0,T]} |\widehat{X}^{(N)}(s)| < \infty$. Representation (3.8.23) then follows on comparing the many-server equations (3.6.3),(3.6.4), (3.6.5) and (3.6.6) with the second equation in (3.8.3), equation (3.8.5) along with Remark 6.3 [25], equation (3.8.20) and equations (3.8.7) and (3.8.10) with $f = \mathbf{1}$, $r = 0$ and $g = g^s$. □

3.9 Properties of the Diffusion Many-Server Reneging Map

We now look at continuity and measurability properties of the mapping Λ.

Proposition 3.9.1. Fix $\overline{X} \in [0,\infty)$. For $i = 1,2$, suppose $(E^i, X(0)^i, W^i, Y^i) \in \mathcal{D}^0_{\mathbb{R}}[0,\infty) \times \mathbb{R} \times \mathcal{D}_{\mathbb{R}}[0,\infty) \times B^+[0,\infty)$ and $(K^i, X^i, Z^i, Q^i) = \Lambda(E^i, X(0)^i, W^i, Y^i)$. Let $\nabla F := F^2 - F^1$ for $F = K, X, Z, Q, E, X(0), W$ and Y, and recall $\|f\|_T := \sup_{s\in[0,T]} |f(s)|$. Then for any $T \in [0,\infty)$, there exists $f(T) < \infty$

$$\|\nabla K\|_T \vee \|\nabla X\|_T \vee \|\nabla Z\|_T \vee \|\nabla Q\|_T \leq f(T)\epsilon_T, \tag{3.9.1}$$

and

$$\epsilon_T := (\|\nabla E\|_T \vee |\nabla X(0)| \vee \|W\|_T \vee \|\nabla Y\|_T). \tag{3.9.2}$$

Hence, Λ is continuous with respect to the topology of uniform convergence on compact sets and is single valued on its domain.

Proof. Fix $T < \infty$. We first show that $\|\nabla K\|_T \leq \epsilon_T U(T)$ where U is the renewal function associated with the distribution G whose density is denoted by g. For any $t \in [0, T]$, we consider the following cases.

Case 1: $\overline{X} < 1$.

If $\overline{X} < 1$, then using (3.6.5) $Z^i = X^i$ and using (3.6.6) $Q^i = 0$ for $i = 1, 2$. Hence, $\nabla Z(t) - \nabla X(t) = 0$, and $\int_0^t Q^i(s) Y^i(s) = 0$ for $i = 1, 2$. Now, using the fact that each solution satisfies (3.6.4), we obtain

$$\nabla K(t) = \nabla E(t) \leq \epsilon_T. \tag{3.9.3}$$

By symmetry, the inequality (3.9.3) above also holds with ∇K replaced by $-\nabla K$, and hence we have

$$|\nabla K(t)| \leq \epsilon_T \qquad \text{for every } t \in [0, T]. \tag{3.9.4}$$

Taking the supremum over $t \in [0, T]$, we obtain

$$\|\nabla K\|_T \leq \epsilon_T. \tag{3.9.5}$$

Substituting Z and Q from (3.6.5) and (3.6.6) respectively into (3.6.4) and then substituting K from (3.6.4) into (3.6.3) and rearranging, we see that for $i = 1, 2$, and $t \in [0, T]$,

$$X^i(t) = E^i(t) + X(0)^i - Z^i(0) - \int_0^t g(t-s)K^i(t-s)ds + W^i(t).$$

Taking the difference and using the fact that $X(0)^i = Z^i(0)$ from (3.6.5), we obtain

$$|\nabla X(t)| \leq \|\nabla E\|_T + \int_0^t g(t-s)\|\nabla K\|_T ds + \|\nabla W\|_T.$$

Taking supremum over $t \in [0, T]$ and using (3.9.5), we conclude that

$$\|\nabla X\|_T \leq 2\epsilon_T + \epsilon_T G(T) \leq 3\epsilon_T \tag{3.9.6}$$

Case 2: $\overline{X} > 1$.

If $\overline{X} > 1$, then by (3.6.5), $Z^1(t) = Z^2(t) = 0$ for all $t \in [0, \infty)$. Since, each solution satisfies (3.6.3), taking the difference we have for every $t \in [0, \infty)$,

$$\nabla K(t) = \int_0^t g(t-s)\nabla K(s)ds - \nabla W(t). \tag{3.9.7}$$

Now, define

$$\mathcal{B} := \left\{ t : \int_0^t g(t-s)\nabla K(s)ds > 0 \right\}.$$

Then, we conclude that

$$\nabla K(t) \leq 2\epsilon_T + 1_{\mathcal{B}}(t) \int_0^t g(t-s)\nabla K(s)ds.$$

Applying the same inequality recursively to $K(s)$, $s \in [0, t]$, on the right-hand side, we obtain

$$\begin{aligned}\nabla K(t) &\leq \epsilon_T + 1_{\mathcal{B}}(t) \int_0^t g(t-s) \left(\epsilon_T + 1_{\mathcal{B}}(s) \int_0^s g(s-r) \nabla K(r) dr \right) ds \\ &\leq \epsilon_T (1 + G(t)) + 1_{\mathcal{B}}(t) \int_0^t g(t-s) \left(1_{\mathcal{B}}(s) \int_0^s g(s-r) \nabla K(r) dr \right) ds.\end{aligned}$$

Reiterating this procedure, we obtain

$$\nabla K(t) \leq \epsilon_T (1 + G(t) + G^{*,2}(t) + \cdots) \leq \epsilon_T U(T), \tag{3.9.8}$$

where $G^{*,n}$ denotes the n-fold convolution of G.

Now by the same arguments used to obtain (3.9.4) and (3.9.5) and using the fact that $U(T) \geq 1$, we have

$$\|\nabla K\|_T \leq \epsilon_T U(T). \tag{3.9.9}$$

Using the equations (3.6.3)-(3.6.6), we have for $i = 1, 2$ and $t \in [0, T]$

$$X^i(t) = E^i(t) + x_0^i - \int_0^t g(t-s) K^i(s) ds + W^i(t) - \int_0^t X^i(s) Y^i(s) ds.$$

Taking the difference, we see that

$$\begin{aligned}|\nabla X(t)| \leq& \|\nabla E\|_T + |\nabla X(0)| + \int_0^t g(t-s) \|\nabla K\|_T ds + \|\nabla W\|_T \\ &+ \int_0^t |\nabla X(s)| Y^2(s) ds + \int_0^t |X^1(s)| |\nabla Y(s)| ds\end{aligned}$$

$$\begin{aligned}&\leq \|\nabla E\|_T + |\nabla X(0)| + G(t) \|\nabla K\|_T + \|\nabla W\|_T \\ &+ \int_0^t |\nabla X(s)| Y^2(s) ds + T \|\nabla Y(s)\|_T \sup_{s \in [0,T]} |X^1(s)|.\end{aligned}$$

Using Gronwall's Lemma

$$|\nabla X(t)| \leq \exp(\int_0^t |Y^2(s)|ds) \left(\|\nabla E\|_T + |\nabla X(0)| + \int_0^t g(t-s)\|\nabla K\|_T ds + \|\nabla W\|_T \right.$$
$$\left. + T\|\nabla Y(s)\|_T \sup_{s\in[0,T]} |X^1(s)| \right).$$

Since $Y^i \in B^+[0,\infty)$, and $\sup_{s\in[0,T]} |X^i(s)| < \infty$ there exists $f(T) < \infty$ such that

$$\|\nabla X(t)\| \leq f(T)\epsilon_T. \tag{3.9.10}$$

Case 3: $\overline{X} = 1$.

If $\overline{X} = 1$. then substituting Z and Q from (3.6.5) and by (3.6.6) respectively into (3.6.4) and noting that $x - x \wedge 0 = x^+$, we get for $i = 1, 2$ and $t \in 0, T]$

$$K^i(t) = E^i(t) + (X^i)^+(0) - (X^i)^+(t) - \int_0^t (X^i)^+(s)Y(s)ds. \tag{3.9.11}$$

Now substituting K from (3.9.11) above into (3.6.3), using integration by parts and rearranging, we obtain for $i = 1, 2$ and $t \in 0, T]$

$$\begin{aligned} X^i(t) =& W^i(t) + E^i(t) + (X^i)^+(0)(1 - G(t)) - \int_0^t g(t-s)E^i(s)ds \\ &+ \int_0^t g(t-s)(X^i)^+(s)ds - \int_0^t (1 - G(t-s))(X^i)^+(s)Y^i(s)ds \end{aligned}$$

Taking difference, we see that

$$\begin{aligned}
|\nabla X(t)| =& |\nabla W(t)| + |\nabla E(t)| + |\nabla X^+(0)|(1-G(t)) + \int_0^t g(t-s)|\nabla E(s)|ds \\
&+ \int_0^t |\nabla X^+(s)|g(t-s)ds + \int_0^t (1-G(t-s))|(X^2)^+(s)Y^2(s)ds - (X^1)^+(s)Y^1(s)|ds \\
\le& |\nabla W(t)| + |\nabla E(t)| + |\nabla X(0)|(1-G(t)) + \int_0^t g(t-s)|\nabla E(s)|ds \\
&+ \int_0^t |\nabla X(s)|g(t-s)ds + \sup_{s\in[0,T]}\{Y^1(s), Y^2(s)\} \int_0^t (1-\overline{G}(t-s)|\nabla X(s)|ds \\
\le& 4\epsilon_T + \int_0^t |\nabla X(s)|g(t-s)ds + \sup_{s\in[0,T]}\{Y^1(s), Y^2(s)\} \int_0^t (1-\overline{G}(t-s)|\nabla X(s)|ds.
\end{aligned} \tag{3.9.12}$$

Since integral is an absolutely continuous function, given $\epsilon \in (0,1)$ there exists a $\delta > 0$ such that

$$\int_t^{t+\delta} g(s)ds + \sup_{s\in[0,T]}\{Y^1(s), Y^2(s)\} \int_t^{t+\delta} 1 - G(s)ds < \epsilon. \tag{3.9.13}$$

Note that for $0 \le t < \delta$, using (3.9.12) and (3.9.13) we have

$$\|\nabla X\|_\delta \le 4\epsilon_T + \epsilon \|\nabla X\|_\delta,$$

which implies that

$$\|\nabla X\|_\delta \le \frac{4\epsilon_T}{1-\epsilon}.$$

Similarly for $\delta \le t < 2\delta$, using the inequality above, we have

$$\begin{aligned}
\|\nabla X\|_\delta \le& 4\epsilon_T + \epsilon\|\nabla X\|_\delta + \epsilon\|\nabla X\|_{2\delta} \\
\le& 4\epsilon_T + \epsilon\frac{4\epsilon_T}{1-\epsilon} + \epsilon\|\nabla X\|_{2\delta} \\
\le& \frac{4\epsilon_T}{1-\epsilon} + \epsilon\|\nabla X\|_{2\delta},
\end{aligned}$$

which implies that

$$\|\nabla X\|_{2\delta} \leq \frac{4\epsilon_T}{(1-\epsilon)^2}.$$

Repeating the above process for $\lceil \delta^{-1}T \rceil - 2$ more time intervals, where $\lceil x \rceil$ denotes the smallest integer greater than or equal to x, we obtain

$$\|\nabla X\|_T \leq \frac{4\epsilon_T}{(1-\epsilon)^{\lceil \delta^{-1}T \rceil}}. \tag{3.9.14}$$

Taking the difference in (3.9.11), using (3.9.14) above, and noting that $\sup_{s\in[0,T]}\{Y^1(s), Y^2(s)\}$ there exists $f(T) < \infty$ such that

$$\|\nabla K\|_T \leq 2\epsilon_T + \|\nabla X\|_T + \sup_{s\in[0,T]} \{Y^1(s), Y^2(s)\} T \|\nabla X\|_T \leq \epsilon_T f(T). \tag{3.9.15}$$

Together with (3.9.5) (3.9.6), (3.9.9), (3.9.10), (3.9.14), (3.9.15) and the fact that (3.6.5) and (3.6.6) imply $\|\nabla Z\|_T \leq \|\nabla X\|_T$, and $\|\nabla Q\|_T \leq \|\nabla X\|_T$, this establishes (3.9.1). Since the Skorokhod topology coincides with the uniform topology on the space of continuous functions, (3.9.1) implies that the map Λ is continuous at points $(E, X(0), W, Y) \in \mathcal{C}[0,\infty) \times \mathbb{R} \times \mathcal{C}[0,\infty)$. $\square$

Lemma 3.9.2. *Suppose the fluid limit is subcritical, critical or supercritical. Then the diffusion many-server reneging map* $\Lambda : dom(\Lambda) \subseteq \mathcal{D}_{\mathbb{R}}[0,\infty) \times \mathbb{R} \times \mathcal{D}_{\mathbb{R}}[0,\infty) \times B^+[0,\infty) \mapsto \mathcal{D}_{\mathbb{R}}[0,\infty)^4$*, where* $\mathcal{D}_{\mathbb{R}}$ *is equipped with the Skorokhod topology, is measurable.*

Proof. We first establish the measurability of the mapping that takes $(E, X(0), W, Y) \in \mathrm{dom}(\Lambda)$ to X, where $(K, X, Z, Q) \in \Lambda(E, X(0), W, Y)$. Let G denote the distribution associated with the density g. Substituting Z and Q from (3.6.5) and (3.6.6) respectively into (3.6.4) and then K from (3.6.4) into (3.6.3), we see that X satisfies the integral equation

$$X(t) = W(t) + E(t) - \int_0^t g(t-s)E(s)ds, \qquad t \geq 0, \tag{3.9.16}$$

if $\overline{X} < 1$ is subcritical,

$$\begin{aligned} X(t) =& W(t) + E(t) + (1-G(t))X^+(0) - \int_0^t g(t-s)E(s)ds \\ &+ \int_0^t g(t-s)X^+(s)ds - \int_0^t (1-G(t-s))X^+(s)Y(s)ds, \end{aligned} \tag{3.9.17}$$

if $\overline{X}$ is critical, and

$$\begin{aligned} X(t) + \int_0^t X(s)Y(s)ds =& W(t) + E(t) + (1-G(t))X(0) - \int_0^t g(t-s)E(s)ds \\ &+ \int_0^t g(t-s)\left(X(s) + \int_0^s X(u)Y(u)ds\right) ds, \end{aligned} \tag{3.9.18}$$

if $\overline{X}$ is supercritical. Using (3.9.16), X is clearly a measurable function of $(E, X(0), W, Y)$ when $\overline{X} < 1$. For the critical case, note that for a fixed t the map $(E, W, X(0), Y) \mapsto L(t) = W(t) + E(t) + (1-G(t))X^+(0) - \int_0^t g(t-s)E(s)ds$ from $\mathcal{D}_{\mathbb{R}}[0,\infty)^2 \times \mathbb{R} \times B^+[0,\infty) \mapsto \mathbb{R}$ is clearly measurable. Since the function $x \mapsto x^+$ is Lipschitz, by standard arguments from the theory of Volterra integral equations (see [14, Theorem 3.2.1]) it follows that $X(t) = \lim_{n\to\infty}(\mathcal{T}^{(n)}\mathbf{0})(t)$, where $\mathcal{T}^{(n)}$ is the $\mathbf{n}$-fold composition of the operator $\mathcal{T} : \mathcal{D}_{\mathbb{R}}[0,\infty) \mapsto \mathcal{D}_{\mathbb{R}}[0,\infty)$ given by $(\mathcal{T}\xi)(t) = L(t) + \int_0^t [g(t-s) - (1-G(t-s))Y(s)]\xi^+(s)ds$, $\xi \in \mathcal{D}_{\mathbb{R}}[0,\infty)$. Due to the fact that convergence in the Skorokhod topology implies convergence in $\mathbb{L}^1_{\text{loc}}$ and the map $\theta \mapsto \int_0^{\cdot}[g(\cdot - s) - (1-G(\cdot - s))Y(s)]\theta^+(s)ds$ is a continuous map from $\mathbb{L}^1_{\text{loc}}[0,\infty)$ to $\mathcal{C}[0,\infty)$, it follows that for every $t > 0$, $(L,\xi) \mapsto \mathcal{T}(\xi)(t)$ is a measurable map. Because the Borel algebra associated with the Skorokhod topology is generated by cylinder sets, and measurability is preserved under compositions and limits, this implies that the map from L to A, and hence from $(E, W, X(0), Y)$ to X is measurable. For the supercritical , define

$$A(t) = X(t) + \int_0^t X(s)Y(s)ds. \tag{3.9.19}$$

Then using (3.9.18) we see that

$$A(t) = W(t)+E(t)+(1-G(t))X(0)-\int_0^t g(t-s)E(s)ds+\int_0^t g(t-s)A(s)ds, \tag{3.9.20}$$

For a fixed t, the map $(E, W, X(0), Y) \mapsto L(t) = W(t) + E(t) + (1 - G(t))X(0) - \int_0^t g(t-s)E(s)ds$ from $\mathcal{D}_{\mathbb{R}}[0,\infty)^2 \times \mathbb{R} \times B^+[0,\infty) \mapsto \mathbb{R}$ is clearly measurable. Since the identity function is Lipschitz, using the same argument as the previous paragraph it follows that $A(t) = \lim_{n\to\infty}(\mathcal{T}^{(n)}\mathbf{0})(t)$, where $\mathcal{T}^{(n)}$ is the **n**-fold composition of the operator $\mathcal{T} : \mathcal{D}_{\mathbb{R}}[0,\infty) \mapsto \mathcal{D}_{\mathbb{R}}[0,\infty)$ given by $(\mathcal{T}\xi)(t) = L(t) + \int_0^t g(t-s)\xi(s)ds$, $\xi \in \mathcal{D}_{\mathbb{R}}[0,\infty)$. Due to the fact that convergence in the Skorokhod topology implies convergence in $\mathbb{L}^1_{\text{loc}}$ and the Laplace convolution $\theta \mapsto \int_0^{\cdot} g(\cdot - s)\theta(s)ds$ is a continuous map from $\mathbb{L}^1_{\text{loc}}[0,\infty)$ to $\mathcal{C}[0,\infty)$, it follows that for every $t > 0$, $(L, \xi) \mapsto \mathcal{T}(\xi)(t)$ is a measurable map. Because the Borel algebra associated with the Skorokhod topology is generated by cylinder sets, and measurability is preserved under compositions and limits, this implies that the map from L to A, and hence from $(E, W, X(0), Y)$ to A is measurable. Rearranging (3.9.19), we have

$$X(t) = A(t) - \int_0^t X(s)Y(s)ds.$$

Using the same argument as in the paragraph above, Assumption 3.6.3 and the measurability of A we have that the map $(E, W, X(0), Y) \mapsto X$ is measurable. Now, to complete the proof, note that by (3.6.5) and (3.6.6), Z equals either $X, X \wedge 0$ or 0, and Q equals either $0, X^+, X$ depending on whether the fluid $\overline{X}$ is respectively, subcritical, critical or supercirtical. The maps $f \mapsto (f, f)$, $f \mapsto (f, f \wedge 0)$, $f \mapsto (f, f^+)$ and $f \mapsto (f, 0)$ from $\mathcal{D}_{\mathbb{R}}[0,\infty)$, equipped with the Skorokhod topology, to $\mathcal{D}^2_{\mathbb{R}}[0,\infty)$ are all measurable (in fact, continuous). In addition, for each t, $K(t)$ is a linear combination of $E(t)$, $X(0)$, $X(t)$, $Z(t)$, $Z(0)$, $\{Q(s), s \in [0,t]\}$ and $\{Y(s), s \in [0,t]\}$ by (3.6.4). Therefore, it immediately follows that the map from $(E, X(0), W, Y)$ to (K, X, Z, Q) is also measurable. $\square$

3.10 Convergence Results

Consider the families of operators $\{\Phi^s_t, t \geq 0\}$ and $\{\Phi^r_t, t \geq 0\}$ which are defined as follows: for every $t \geq 0$, let

$$(\Phi^s_t f)(x) := f(x+t)\frac{1 - G^s(x+t)}{1 - G^s(x)}, \qquad x \in [0,\infty), \tag{3.10.1}$$

and

$$(\Phi^r_t f)(x) := f(x+t)\frac{1 - G^r(x+t)}{1 - G^r(x)}, \qquad x \in [0,\infty), \tag{3.10.2}$$

for any function f on $\mathcal{C}_b[0,\infty)$. The operators Φ^s_t and Φ^r_t clearly map the space $\mathcal{C}_b[0,\infty)$ into itself.

Note that the families of functions $\{\alpha^r; r \geq 0\}$ and $\{\beta^r; r \geq 0\}$ have the representation

$$\alpha^r = \Phi^s_r \mathbf{1} \quad \text{and} \quad \beta^r = \Phi^r_r \mathbf{1}, \qquad r \geq 0. \tag{3.10.3}$$

Also, $\Phi^s_0 f = f = \Phi^r_0 f$ and the families $\{\Phi^s_t, t \geq 0\}$ and $\{\Phi^r_t, t \geq 0\}$ satisfy the following semigroup property:

$$\Phi^s_t \Phi^s_s = \Phi^s_{t+s} \quad \text{and} \quad \Phi^r_t \Phi^r_s = \Phi^r_{t+s} \qquad t, s \geq 0. \tag{3.10.4}$$

Furthermore, recalling the definition of Ψ_t^s and Ψ_t^r from (3.4.42) and (3.4.43) respectively, for every function f defined on $[0,\infty)$ and $s,t \geq 0$,

$$(\Psi_s^s \Phi_t^s f)(x,v) = (\Psi_{s+t}^s f)(x,v), \quad (x,v) \in [0,\infty) \times [0,s], \tag{3.10.5}$$

and

$$(\Psi_s^r \Phi_t^r f)(x,v) = (\Psi_{s+t}^r f)(x,v), \quad (x,v) \in [0,\infty) \times [0,s]. \tag{3.10.6}$$

For $t > 0$ define the operators $\Upsilon_t^s : \mathcal{C}_b([0,H^s) \times [0,t]) \mapsto \mathcal{C}_b([0,H^s) \times [0,t])$

$$(\Upsilon_t^s \varphi)(x,u) \;:=\; \int_u^t (\Psi_s^s \varphi(\cdot,s))(x,u)ds, \tag{3.10.7}$$

and $\Upsilon_t^r : \mathcal{C}_b([0,H^r) \times [0,t]) \mapsto \mathcal{C}_b([0,H^r) \times [0,t])$

$$(\Upsilon_t^r \psi)(x,u) \;:=\; \int_u^t (\Psi_s^r \psi(\cdot,s))(x,u)ds, \tag{3.10.8}$$

for $\varphi \in \mathcal{C}_b[0,H^s)$ and $\psi \in \mathcal{C}_b[0,H^r)$.

The two lemmas stated next are analogous to Lemma 8.6[25] along with the observation that for any bounded, Hölder continuous function f, using Lemma 8.5 [25], $1_{[0,\chi_*]}\Psi_t^r f$ is Hölder continuous as well.

Lemma 3.10.1. *If Assumption 3.6.3 is satisfied, the following properties hold:*

1. *Given a bounded and Hölder continuous function f on $[0,H^s)$, the sequence of processes$\{\widehat{\mathcal{H}}_\nu^{(N)}(f)\}_{N\in\mathbb{N}}$ is tight in $\mathcal{D}_{\mathbb{R}}[0,\infty)$ and $\widehat{\mathcal{H}}_\nu(f)$ is $\mathbb{P}$-almost surely continuous.*

2. *There exists* $r \geq 2$ *and a constant* $C_0^s < \infty$ *such that for any* $f \in \mathcal{S}$,

$$\sup_N \mathbb{E}[\sup_{t\in[0,T]} |\widehat{\mathcal{H}}^{(N)}_{t,\nu}(f)|^r] \leq C_0^s \|f\|^r_{\mathbb{H}^1}. \tag{3.10.9}$$

3. *Suppose* $\varphi : [0, H^s) \times [0,\infty) \mapsto \mathbb{R}$ *is a Borel measurable function such that for every* $x \in [0, H^s)$, *the function* $r \mapsto \varphi(x, r)$ *is locally integrable and, for every* $t \in [0, T]$, *the function* $x \mapsto \int_0^t \varphi(x, r)dr$ *is bounded and Hölder continuous with constant* $C^s_{\varphi,T}$ *and exponent* $\gamma^s_{\varphi,T}$ *(that are independent of* t*), Then* $\mathbb{P}$*-almost surely, the random field* $\left\{\widehat{\mathcal{H}}_{s,\nu}\left(\int_0^t \varphi(\cdot, r)dr\right), s, t \geq 0\right\}$ *is jointly continuous* s *and* t.

4. *Suppose* $\varphi \in \mathcal{C}_b([0, H^s) \times [0,\infty))$. *Then the process* $\{\widehat{\mathcal{M}}_{t,\nu}(\Upsilon^s_t\varphi), t \geq 0\}$ *admits a continuous version.*

For the next Lemma in addition to [25, Lemma 8.6] it suffices to note that for any bounded, Hölder continuous function f, using Lemma 8.5 [25], $1_{[0,\chi_*]}\Psi^r_t f$ is Hölder continuous as well.

Lemma 3.10.2. *If Assumption 3.6.5 is satisfied, the following properties hold:*

1. *Given a bounded and Hölder continuous function* f *on* $[0, H^r)$, *the sequence of processes* $\{\widehat{\mathcal{H}}^{(N)}_\eta(f)\}_{N\in\mathbb{N}}$ *is tight in* $\mathcal{D}_{\mathbb{R}}[0,\infty)$ *and* $\widehat{\mathcal{H}}_\eta(f)$ *is* $\mathbb{P}$*-almost surely continuous.*

2. *There exists* $r \geq 2$ *and a constant* $C_0^r < \infty$ *such that for any* $f \in \mathcal{S}$,

$$\sup_N \mathbb{E}[\sup_{t\in[0,T]} |\widehat{\mathcal{H}}^{(N)}_{t,\eta}(f)|^r] \leq C_0^r \|f\|^r_{\mathbb{H}^1}. \tag{3.10.10}$$

3. *Suppose* $\psi : [0, H^r) \times [0,\infty) \mapsto \mathbb{R}$ *is a Borel measurable function such that for every* $x \in [0, H^r)$, *the function* $r \mapsto \psi(x, r)$ *is locally integrable and, for every*

$t \in [0,T]$, the function $x \mapsto \int_0^t \psi(x,r)dr$ is bounded and Hoÿlder continuous with constant $C^r_{\psi,T}$ and exponent $\gamma^r_{\psi,T}$ (that are independent of t), Then $\mathbb{P}$-almost surely, the random field $\left\{\widehat{\mathcal{H}}_{s,\eta}\left(\int_0^t \psi(\cdot,r)dr\right), s,t \geq 0\right\}$ is jointly continuous s and t.

4. *Suppose $\psi \in \mathcal{C}_b([0,H^r) \times [0,\infty))$. Then the process $\{\widehat{\mathcal{M}}_{t,\eta}(1_{[0,\chi_*]}\Upsilon^r_t\psi), t \geq 0\}$ admits a continuous version.*

Lemma 3.10.3. *Suppose Assumption 3.6.3 holds. Then, $\{\widehat{\mathcal{M}}_{t,\nu}(\Phi^s_r\mathbf{1}) : t,r \geq 0\}$, $\{\widehat{\mathcal{M}}_{t,\nu}(\Psi^s_{t+r}\mathbf{1}); t,r \geq 0\}$ and $\{\widehat{\mathcal{M}}_{t,\nu}(\Psi^s_{t+r}h^s); t,r \geq 0\}$ have modifications that are jointly continuous on $[0,\infty) \times [0,\infty)$.*

Proof. This proof is analogous to the proof of Lemma 4.1 [3] replacing g, M_t, Φ and Ψ by $(\lambda \wedge 1)g^s$, $M_{t,\nu}$, Φ^s and Ψ^s respectively. □

Lemma 3.10.4. *Suppose Assumption 3.6.5 holds. Then, $\{\widehat{\mathcal{M}}_{t,\eta}(1_{[0,\chi_*]}\Phi^r_r\mathbf{1}) : t,r \geq 0\}$, $\{\widehat{\mathcal{M}}_{t,\eta}(1_{[0,\chi_*]}\Psi^r_{t+r}\mathbf{1}); t,r \geq 0\}$ and $\{\widehat{\mathcal{M}}_{t,\eta}(1_{[0,\chi_*]}\Psi^r_{t+r}h^r); t,r \geq 0\}$ have modifications that are jointly continuous on $[0,\infty) \times [0,\infty)$.*

Proof. The proof is analogous to the proof of Lemma 4.1 [3] replacing g, M_t, Φ and Ψ by λg^r, $M_{t,\eta}$, Φ^r and Ψ^r respectively, and using the fact that $1_{[0,\chi_*]} \leq \mathbf{1}$. □

Proposition 3.10.5. Let Assumptions 3.6.3 and 3.6.4 hold. Then $\{\widehat{\mathcal{M}}_{t,\nu}(\Psi^s_{t+\cdot}\mathbf{1}); t \geq 0\}$ has a continuous $\mathbb{H}^1(0,\infty)$-valued modification. Also, almost surely, for every $t \geq 0$, the function $r \mapsto \widehat{\mathcal{M}}_{t,\nu}(\Psi^s_{t+r}\mathbf{1})$ has weak derivative $-\widehat{\mathcal{M}}_{t,\nu}(\Psi^s_{t+r}h^s)$ on $(0,\infty)$.

Proof. The proof is analogous to the proof of Proposition 4.3 [3], with $\mathcal{M}_t$, Ψ_t and h replaced by $\widehat{\mathcal{M}}_{t,\nu}$, Ψ^s_t and h^s respectively. □

Proposition 3.10.6. Let Assumptions 3.6.5 and 3.6.6 hold. Then $\{\widehat{\mathcal{M}}_{t,\eta}(1_{[0,\chi_*]}\Psi^r_{t+\cdot}\mathbf{1}); t \geq 0\}$ has a continuous $\mathbb{H}^1(0,\infty)$-valued modification. Also, almost surely, for every $t \geq 0$, the function $r \mapsto \widehat{\mathcal{M}}_{t,\eta}(1_{[0,\chi_*]}\Psi^r_{t+r}\mathbf{1})$ has weak derivative $-\widehat{\mathcal{M}}_{t,\eta}(1_{[0,\chi_*]}\Psi^r_{t+r}h^r)$ on $(0,\infty)$.

Proof. The proof is analogous to the proof of Proposition 4.3 [3] with $\mathcal{M}_t$, Ψ_t and h replaced by $\widehat{\mathcal{M}}_{t,\eta}$, Ψ^r_t and h^r respectively. □

We now state the following lemma which is similar to [3, Lemma 4.4(a)] and we show the proof here for completeness.

Lemma 3.10.7. *Suppose Assumption 3.6.5 holds. Then, almost surely,*

$$\widehat{\mathcal{M}}_{t,\eta}(1_{[0,\chi_*]}\Psi^r_{t+r}\mathbf{1}) = \widehat{\mathcal{M}}_{t,\eta}(1_{[0,\chi_*]}\Phi^r_r\mathbf{1}) - \int_0^t \widehat{\mathcal{M}}_{s,\eta}(1_{[0,\chi_*]}\Psi^r_{s+r}h^r)ds, \qquad \forall t, r \geq 0. \tag{3.10.11}$$

Proof. Let $\psi \in \mathcal{C}_b([0,H^r) \times [0,\infty))$. Fix $s, t \geq 0$. Then by the boundedness assumption on ψ, $\Upsilon^r_t\psi$ is uniformly bounded on $[0,H^r) \times [0,\infty)$. Using the uniform boundedness of the indicator function, we can thus apply stochastic integrals with respect to martingale measures (see [44, Theorem 2.6]) to see that almost surely, for every $t \geq 0$,

$$\widehat{\mathcal{M}}_{t,\eta}(1_{[0,\chi_*]}\Upsilon^r_t\psi) = \int_0^t \widehat{\mathcal{M}}_{s,\eta}(1_{[0,\chi_*]}(\cdot)\Psi^r_s\psi(\cdot,s))ds = \int_0^t \widehat{\mathcal{H}}_{s,\eta}(\psi(\cdot,s))ds. \tag{3.10.12}$$

Now, fix $r \geq 0$, and define

$$\psi_r(x,u) = \Phi_r h^r(x), \qquad (x,u) \in [0,\infty) \times [0,t].$$

Note that for every $t \geq 0$, $\psi_r \in \mathcal{C}_b([0, H^r) \times [0,t])$, due to Assumption 3.6.3,

$$(\Upsilon_t^r \psi_r)(x,u) = \int_u^t \frac{g^r(x+s+r-u)}{1-G^r(x)} ds = \Phi_r^r \mathbf{1}(x) - \Psi_{t+r}^r \mathbf{1}(x,u).$$

Using (3.10.12) with $\psi = \psi_r$, and substituting the previous identity, we see that almost surely,

$$\widehat{\mathcal{M}}_{t,\eta}(1_{[0,\chi_*]}\Phi_r^r \mathbf{1}) - \widehat{\mathcal{M}}_{t,\eta}(1_{[0,\chi_*]}\Psi_{t+r}^r \mathbf{1}) = \int_0^t \widehat{\mathcal{H}}_{s,\eta}(\Phi_r^r h^r) ds, \qquad \forall t \geq 0. \qquad (3.10.13)$$

Using (3.4.46) and (3.10.6), we get

$$\widehat{\mathcal{H}}_{s,\eta}(\Phi_r^r h^r) = \widehat{\mathcal{M}}_{s,\eta}(1_{[0,\chi_*]}\Psi_s^r \Phi_r^r h^r) = \widehat{\mathcal{M}}_{s,\eta}(1_{[0,\chi_*]}\Psi_{s+r}^r h^r), \quad \forall s, r \geq 0.$$

Using the existence of a continuous version of the terms in (3.10.13) from Lemma 3.10.4, we have that both sides of (3.10.13) are jointly continuous in (t,r) and hence the result is strengthened to obtain (3.10.11). □

Lemma 3.10.8. *Suppose Assumption 3.6.3 holds. Then, almost surely,*

$$\widehat{\mathcal{M}}_{t,\nu}(\Psi_{t+r}^s \mathbf{1},) = \widehat{\mathcal{M}}_{t,\nu}(\Phi_r^s \mathbf{1}) - \int_0^t \widehat{\mathcal{M}}_{s,\nu}(\Psi_{s+r}^s h^s) ds, \qquad \forall t, r \geq 0. \qquad (3.10.14)$$

Proof. The prof is analogous to the proof of Lemma 3.10.7 using Lemma 3.10.3. □

Setting $r = 0$ in (3.10.14) and (3.10.11), we see that, almost surely,

$$\widehat{\mathcal{H}}_{t,\nu}(\mathbf{1}) = \widehat{\mathcal{M}}_{t,\nu}(\mathbf{1}) - \int_0^t \widehat{\mathcal{H}}_{s,\nu}(h^s) ds, \qquad t \geq 0, \qquad (3.10.15)$$

and

$$\widehat{\mathcal{H}}_{t,\eta}(\mathbf{1}) = \widehat{\mathcal{M}}_{t,\eta}(1_{[0,\chi_*]}) - \int_0^t \widehat{\mathcal{H}}_{s,\eta}(h^r) ds, \qquad t \geq 0. \qquad (3.10.16)$$

For $f \in \mathcal{C}_b([0,H^s)\times\mathbb{R})$ and $g \in \mathcal{C}_b([0,H^r)\times\mathbb{R})$, define

$$\widehat{Y}_1^{(N)} := \ (\widehat{E}^{(N)}, \widehat{X}^{(N)}(0), \widehat{Z}_\nu^{(N)}(0,\cdot), \widehat{Z}_\eta^{(N)}(0,\cdot), \widehat{\mathcal{M}}_{t,\nu}^{(N)}(f), \widehat{\mathcal{M}}_{t,\eta}^{(N)}(g), \widehat{\mathcal{H}}_\nu^{(N)}(\mathbf{1}), \widehat{\mathcal{H}}_\eta(\mathbf{1})),$$

and let $\widehat{Y}_1$ be the corresponding limit without the superscript N. Also, let $\mathcal{Y}_1$ be the space given by

$$\mathcal{Y}_1 := \mathcal{D}_\mathbb{R}[0,\infty)\times\mathbb{R}\times\mathbb{H}^1(0,\infty)^2\times\mathbb{R}^2\times\mathcal{D}_\mathbb{R}[0,\infty)^2.$$

We now prove the following convergence result,

Proposition 3.10.9. As $N\to\infty$, $\widehat{Y}_1^{(N)} \Rightarrow \widehat{Y}_1$ in $\mathcal{Y}_1$. Also, if for any $r \geq 0$,

$$(\widehat{E}^{(N)}, \widehat{X}^{(N)}(0), \widehat{Z}_\nu^{(N)}(0,r), \widehat{Z}_\eta^{(N)}(0,r)) \Rightarrow (\widehat{E}, \widehat{X}(0), \widehat{Z}_\nu(0,r), \widehat{Z}_\eta(0,r))$$

holds, then, for any bounded and Hölder continuous functions f and g on $[0,H^s)$ and $[0,H^r)$ respectively, $(\widehat{Y}_1^{(N)}, \widehat{\mathcal{H}}_\nu^{(N)}(f), \widehat{\mathcal{H}}_\eta^{(N)}(g)) \Rightarrow (\widehat{Y}_1, \widehat{\mathcal{H}}_\nu(f), \widehat{\mathcal{H}}_\eta(g))$ as $N\to\infty$.

Proof. Fix $N\in\mathrm{N}$. For $\tilde{k}, k \in \mathrm{N}$, $i = 1,\cdots,\tilde{k}$, $j = 1\cdots,k$, let $\tilde{f}_i$ and $\tilde{g}_i$ be bounded continuous functions, and f_j and g_j be bounded Hölder continuous functions. Proposition 3.7.5, Lemma 3.10.1 and Lemma 3.10.2 imply that the sequence

$$\Big\{\Big(\widehat{\mathcal{M}}_\nu^{(N)}(\tilde{f}_1),\cdots,\widehat{\mathcal{M}}_\nu^{(N)}(\tilde{f}_{\tilde{k}}), \widehat{\mathcal{M}}_\eta^{(N)}(\tilde{g}_1),\cdots,\widehat{\mathcal{M}}_\eta^{(N)}(\tilde{g}_{\tilde{k}}),\\ \widehat{\mathcal{H}}_\nu^{(N)}(f_1),\cdots,\widehat{\mathcal{H}}_\nu^{(N)}(f_k), \widehat{\mathcal{H}}_\eta^{(N)}(g_1),\cdots,\widehat{\mathcal{H}}_\eta^{(N)}(g_k)\Big)\Big\}_{N\in\mathrm{N}}$$

is tight in $\mathcal{D}_\mathbb{R}[0,\infty)^{2\tilde{k}+2k}$. Since $\widehat{\mathcal{H}}_{t,\nu}^{(N)}(f) = \widehat{\mathcal{M}}_{t,\nu}^{(N)}(\Psi_t^s f)$ and, likewise, $\widehat{\mathcal{H}}_{t,\nu}(f) = \widehat{\mathcal{M}}_{t,\nu}(\Psi_t^s f)$, and similarly $\widehat{\mathcal{H}}_{t,\eta}^{(N)}(g) = \widehat{\mathcal{M}}_{t,\eta}^{(N)}(1_{[0,\chi_*]}\Psi_t^r g)$, and $\widehat{\mathcal{H}}_{t,\eta}(f) = \widehat{\mathcal{M}}_{t,\eta}(1_{[0,\chi_*]}\Psi_t^r g)$, Proposition 3.7.5 also shows that for $\tilde{t}_i, \tilde{s}_i, t_j, s_i \in [0,\infty)$, $i = 1,\cdots,\tilde{k}$, $j = 1,\cdots,k$, as $N \to \infty$, the following corresponding terms converge:

$$\begin{aligned}
&\Big(\widehat{\mathcal{M}}^{(N)}_{\tilde{t}_1,\nu}(\tilde{f}_1), \cdots, \widehat{\mathcal{M}}^{(N)}_{\tilde{t}_k,\nu}(\tilde{f}_{\tilde{k}}), \widehat{\mathcal{M}}^{(N)}_{\tilde{s}_1,\eta}(\tilde{g}_1), \cdots, \widehat{\mathcal{M}}^{(N)}_{\tilde{s}_k,\eta}(\tilde{g}_{\tilde{k}}),\\
&\qquad \widehat{\mathcal{H}}^{(N)}_{t_1,\nu}(f_1), \cdots, \widehat{\mathcal{H}}^{(N)}_{t_k,\nu}(f_k), \widehat{\mathcal{H}}^{(N)}_{s_1,\eta}(g_1), \cdots, \widehat{\mathcal{H}}^{(N)}_{s_k,\eta}(g_k)\Big)\\
&\Rightarrow \Big(\widehat{\mathcal{M}}_{\tilde{t}_1,\nu}(\tilde{f}_1), \cdots, \widehat{\mathcal{M}}_{\tilde{t}_k,\nu}(\tilde{f}_{\tilde{k}}), \widehat{\mathcal{M}}_{\tilde{s}_1,\eta}(\tilde{g}_1), \cdots, \widehat{\mathcal{M}}_{\tilde{s}_k,\eta}(\tilde{g}_{\tilde{k}}),\\
&\qquad \widehat{\mathcal{H}}_{t_1,\nu}(f_1), \cdots, \widehat{\mathcal{H}}_{t_k,\nu}(f_k), \widehat{\mathcal{H}}_{s_1,\eta}(g_1), \cdots, \widehat{\mathcal{H}}_{s_k,\eta}(g_k)\Big).
\end{aligned}$$

Now, $(\widehat{\mathcal{H}}_\nu(f_1), \widehat{\mathcal{H}}_\eta(g_1))$ is adapted to the filtration generated by $(\widehat{\mathcal{M}}_\nu, \widehat{\mathcal{M}}_\eta)$, and $\widehat{\mathcal{M}}_\nu$ and $\widehat{\mathcal{M}}_\eta$ are independent of $(\widehat{E}, \widehat{X}(0), \widehat{Z}_\nu(0,\cdot), \widehat{Z}_\eta(0,\cdot))$ by (3.7.18) and the construction of $\widehat{M}$ and B. The same argument used to establish asymptotic independence in Proposition 3.7.5 also shows that the convergence above can be strengthened to $\widehat{Y}_1^{(N)} \Rightarrow \widehat{Y}_1$ and $(\widehat{Y}_1^{(N)}, \widehat{\mathcal{H}}_\nu^{(N)}(f), \widehat{\mathcal{H}}_\eta^{(N)}(g)) \Rightarrow (\widehat{Y}_1, \widehat{\mathcal{H}}_\nu(f), \widehat{\mathcal{H}}_\eta(g))$. $\square$

Lemma 3.10.10. *Suppose Assumptions 3.6.5 and 3.6.6 hold, and let $\widehat{Z}_\eta$ satisfy (3.6.7). Then, almost surely, the sample paths of $\{\widehat{Z}_\eta(t,\cdot); t \geq 0\}$ are $\mathbb{H}^1(0,\infty)$-valued and continuous, and for $t \geq 0$, the weak derivative $\partial_r \widehat{Z}_\eta(t,\cdot)$ of $\widehat{Z}_\eta(t,\cdot)$ satisfies for a.e. $r \in (0,\infty)$,*

$$\begin{aligned}
\partial_r \widehat{Z}_\eta(t,r) &= \widehat{Z}'_\eta(0,t+r) + \widehat{\mathcal{M}}_{t,\eta}(1_{[0,\chi_*]}\Psi^r_{t+r}h^r) - g^r(r)\widehat{E}(t) \\
&\quad - \int_0^t \widehat{E}(s)(g^r)'(t+r-s)ds.
\end{aligned} \tag{3.10.17}$$

Also almost surely, $(s,r) \mapsto \partial_r \widehat{Z}_\eta(s,r)$ is locally integrable on $(0,\infty)\times(0,\infty)$ and for every $t, r \geq 0$

$$\begin{aligned}
\int_0^t \partial_r \widehat{Z}_\eta(s,r)ds &= \widehat{Z}_\eta(0,t+r) - \widehat{Z}_\eta(0,r) + \widehat{\mathcal{M}}_{t,\eta}(1_{[0,\chi_*]}\Phi^r_r\mathbf{1}) - \widehat{\mathcal{M}}_{t,\eta}(1_{[0,\chi_*]}\Psi^r_{t+r}\mathbf{1}) \\
&\quad - \int_0^t \widehat{E}(s)g^r(t+r-s)ds.
\end{aligned} \tag{3.10.18}$$

Proof. Let us look at each term in the definition of $\widehat{Z}_\eta$ in equation (3.6.7). Since,

$\widehat{Z}_\eta(0,\cdot) \in \mathbb{H}^1(0,\infty)$, the translation mapping $t \mapsto \widehat{Z}_\eta(0,t+\cdot)$ is continuous in $\mathbb{H}^1(0,\infty)$ by [3, Lemma 3.3] and for each $t>0$, $\widehat{Z}_\eta(0,t+\cdot)$ has weak derivative $\widehat{Z}'_\eta(0,t+\cdot)$. For the second term, by Proposition 3.10.6 $\{\widehat{\mathcal{M}}_{t,\eta}(1_{[0,\chi_*]}\Psi^r_{t+\cdot}\mathbf{1}); t \geq 0\}$ is a continuous $\mathbb{H}^1(0,\infty)$-valued process, and almost surely, for every $t \geq 0$, the function $r \mapsto \widehat{\mathcal{M}}_{t,\eta}(1_{[0,\chi_*]}\Psi^r_{t+r}\mathbf{1})$ has weak derivative $r \mapsto -\widehat{\mathcal{M}}_{t,\eta}(1_{[0,\chi_*]}\Psi^r_{t+r}h^r)$. Finally for the third term, since $\widehat{E}$ is a continuous process, $\{\widehat{\mathcal{E}}_t(\beta^r); t \geq 0\}$ is a continuous $\mathbb{H}^1(0,\infty)$-valued process. Also, by definition (3.6.8), it is easy to see that, for every $t \geq 0$, $r \mapsto \widehat{\mathcal{E}}_t(\Phi^r_r\mathbf{1})$ has weak derivative $-g^r(r)\widehat{E}(t) - \int_0^t \widehat{E}(s)(g^r)'(t+r-s)ds$. Thus we have (3.10.17).

We prove (3.10.18) by treating each term on the right-hand side of equation (3.10.17) separately. First, since $\widehat{Z}_\eta(0,\cdot) \in \mathbb{H}^1(0,\infty)$, the mapping $(s,r) \mapsto \widehat{Z}'_\eta(0,s+r)$ is clearly locally integrable on $(0,\infty)\times(0,\infty)$ and for every $t \geq 0$, and almost every $r \in (0,\infty)$,

$$\int_0^t \widehat{Z}'_\eta(0,s+r)ds = \widehat{Z}_\eta(0,t+r) - \widehat{Z}_\eta(0,r). \tag{3.10.19}$$

Moreover, for every $t \geq 0$, since $\widehat{Z}_\eta(0,\cdot) \in \mathbb{H}^1(0,\infty)$, $\widehat{Z}_\eta(0,t+\cdot)$ also lies in $\mathbb{H}^1(0,\infty)$, and therefore by [3, Lemma 3.3], there exists a continuous function on $[0,\infty)$, that is equal to $\widehat{Z}_\eta(0,t+r) - \widehat{Z}_\eta(0,r)$ almost everywhere on $(0,\infty)$. Next, Lemma 3.10.4 implies that almost surely, $(s,r) \mapsto \widehat{\mathcal{M}}_{s,\eta}(1_{[0,\chi_*]}\Psi^r_{s+r}h^r)$ is jointly continuous and hence, locally integrable on $(0,\infty)\times(0,\infty)$. Also, for every $t \geq 0$, by equation (3.10.11),

$$\int_0^t \widehat{\mathcal{M}}_{s,\eta}(1_{[0,\chi_*]}\Psi^r_{s+r}h^r)ds = \widehat{\mathcal{M}}_{t,\eta}(1_{[0,\chi_*]}\Phi^r_r\mathbf{1}) - \widehat{\mathcal{M}}_{t,\eta}(1_{[0,\chi_*]}\Psi^r_{t+r}\mathbf{1}), \qquad r \geq 0. \tag{3.10.20}$$

By Lemma 3.10.4, $r \mapsto \widehat{\mathcal{M}}_{t,\eta}(1_{[0,\chi_*]}\Phi^r_r\mathbf{1}) - \widehat{\mathcal{M}}_{t,\eta}(1_{[0,\chi_*]}\Psi^r_{t+r}\mathbf{1})$ is continuous on $[0,\infty)$. Finally, it follows from the continuity of $\widehat{E}$, g^r and $(g^r)'$ (see Assumption 3.6.5) that the mapping $(s,r) \mapsto g^r(r)\widehat{E}(s) + \int_0^s \widehat{E}(v)(g^r)'(s+r-v)dv$ is jointly continuous and hence, locally integrable on $(0,\infty)\times(0,\infty)$. Also, for every $t \geq 0$, using Fubini's theorem,

$$\begin{aligned}
&\int_0^t \left(g^r(r)\widehat{E}(s) + \int_0^s \widehat{E}(v)(g^r)'(s+r-v)dv \right) ds \\
=&g^r(r)\int_0^t \widehat{E}(s)ds + \int_0^t \int_v^t \widehat{E}(v)(g^r)'(s+r-v)dsdv \\
=&\int_0^t \widehat{E}(s)g^r(t+r-s)ds.
\end{aligned} \tag{3.10.21}$$

Again, by the continuity of $\widehat{E}$ and g^r, the mapping $r \mapsto \int_0^t \widehat{E}(s)g^r(t+r-s)ds$ is continuous on $[0,\infty)$. Equation (3.10.18) then follows from the three equations above along with (3.10.17). $\square$

We now prove Theorem 3.6.10.

Proof of Theorem 3.6.10. For this proof, we use the convention that all continuity results on $\mathcal{D}[0,\infty)$ are with respect to the topology of uniform convergence on compact sets. Fix $t \geq 0$ and $k \in \mathbb{N}$.

Next we compute the projections of $\widehat{Z}^{(N)}_\eta(t,\cdot)$ and $\widehat{Z}_\eta(t,\cdot)$ onto the directions $e_1,\cdots,e_k$. Recall the representations (3.8.11) of $\widehat{Z}^{(N)}_\eta(t,\cdot)$, (3.8.13) of $\partial_r\widehat{Z}^{(N)}_\eta(t,\cdot)$, (3.6.7) of $\widehat{Z}_\eta(t,\cdot)$ and (3.10.17) of $\partial_r\widehat{Z}_\eta(t,\cdot)$. Note that, since the translation mapping $f \mapsto f(t+\cdot)$ is continuous from $\mathbb{L}^2(0,\infty)$ to itself, and for any $u \in \mathbb{L}^2(0,\infty)$, the projection mapping $f \mapsto \langle f,u\rangle_{\mathbb{L}^2}$ is continuous from $\mathbb{L}^2(0,\infty)$ to $\mathbb{R}$, there exist continuous mappings $F^1_j, F^2_j : \mathbb{L}^2(0,\infty) \mapsto \mathbb{R}$, $j=1,\cdots,k$, such that

$$\langle \widehat{Z}^{(N)}_\eta(0,t+\cdot), e_j\rangle_{\mathbb{L}^2} = F^1_j(\widehat{Z}^{(N)}_\eta(0,\cdot)), \qquad \langle \widehat{Z}'_\eta(0,t+\cdot), e_j\rangle_{\mathbb{L}^2} = F^1_j(\widehat{Z}_\eta(0,\cdot)), \tag{3.10.22}$$

and

$$\langle \partial_r \widehat{Z}^{(N)}_\eta(0,t+\cdot), e'_j\rangle_{\mathbb{L}^2} = F^2_j(\partial_r \widehat{Z}^{(N)}_\eta(0,\cdot)), \qquad \langle \partial_r \widehat{Z}_\eta(0,t+\cdot), e'_j\rangle_{\mathbb{L}^2} = F^2_j(\partial_r \widehat{Z}_\eta(0,\cdot)). \tag{3.10.23}$$

Next, for every $j = 1,\cdots,k$ define the function $\psi_j : [0,\infty)\times[0,\infty) \mapsto \mathbb{R}$ as

$$\begin{aligned} \psi_j(x,s) := \langle \Psi^r_{t+\cdot}\mathbf{1}(x,s), e_j\rangle_{\mathbb{L}^2} &= \int_0^\infty \Psi^r_{t+r}\mathbf{1}(x,s)e_j(r)dr \\ &= \int_0^\infty \frac{1-G^r(x+t-s+r)}{1-G^r(x)} e_j(r)dr, \end{aligned} \tag{3.10.24}$$

which is bounded by $\|e_j\|_{\mathbb{L}^1} < \infty$ (recall that e_j has compact support by choice). Also, by the continuity of $1-G^r$ and boundedness of $1/(1-G^r)$ on finite intervals, the dominated convergence theorem shows that ψ_j is continuous. Since $\psi_j \in \mathcal{C}_b([0,\infty)\times[0,\infty))$, by the stochastic Fubini theorem for orthogonal martingale measures $\widehat{\mathcal{M}}^{(N)}_\eta$ (see e.g. [44, Theorem 2.6]), we have

$$\begin{aligned} \langle \widehat{\mathcal{M}}^{(N)}_{t,\eta}(1_{[0,\chi_*]}\Psi^r_{t+\cdot}\mathbf{1}), e_j\rangle_{\mathbb{L}^2} &= \int_0^\infty \widehat{\mathcal{M}}^{(N)}_{t,\eta}(1_{[0,\chi_*]}\Psi^r_{t+r}\mathbf{1})e_j(r)dr \\ &= \widehat{\mathcal{M}}^{(N)}_{t,\eta}\left(1_{[0,\chi_*]}\int_0^\infty \Psi^r_{t+r}\mathbf{1}e_j(r)dr\right) \\ &= \widehat{\mathcal{M}}^{(N)}_{t,\eta}(1_{[0,\chi_*]}\psi_j). \end{aligned} \tag{3.10.25}$$

Similarly, the function $\tilde{\psi}_j : [0,\infty)\times[0,\infty) \mapsto \mathbb{R}$ defined as

$$\begin{aligned} \tilde{\psi}_j(x,s) := \langle \Psi^r_{t+\cdot}h^r(x,s), e'_j\rangle_{\mathbb{L}^2} &= \int_0^\infty \Psi^r_{t+r}h^r(x,s)e'_j(r)dr \\ &= \int_0^\infty \frac{g^r(x+t-s+r)}{1-G^r(x)} e'_j(r)dr, \end{aligned} \tag{3.10.26}$$

is also continuous and bounded (by $H^r_1\|e'_j\|_{\mathbb{L}^1} < \infty$) due to Assumption 3.6.5. Another application of the stochastic Fubini theorem for orthogonal martingales shows that

$$\langle \widehat{\mathcal{M}}^{(N)}_{t,\eta}(1_{[0,\chi_*]}\Psi^r_{t+\cdot}h^r), e'_j\rangle_{\mathbb{L}^2} = \int_0^\infty \widehat{\mathcal{M}}^{(N)}_{t,\eta}(1_{[0,\chi_*]}\Psi^r_{t+r}h^r)e'_j(r)dr$$
$$= \widehat{\mathcal{M}}^{(N)}_{t,\eta}\left(1_{[0,\chi_*]}\int_0^\infty \Psi^r_{t+r}h^r e'_j(r)dr\right)$$
$$= \widehat{\mathcal{M}}^{(N)}_{t,\eta}(1_{[0,\chi_*]}\tilde{\psi}_j). \qquad (3.10.27)$$

Similarly, replacing $\widehat{\mathcal{M}}^{(N)}_{t,\eta}$ with $\widehat{\mathcal{M}}_{t,\eta}$, and noting that $\widehat{\mathcal{H}}_{t,\eta}(\beta^r) = \widehat{\mathcal{M}}_{t,\eta}(1_{[0,\chi_*]}\Psi^r_{t+r}\mathbf{1})$ from (3.4.46) we have

$$\langle \widehat{\mathcal{M}}_{t,\eta}(1_{[0,\chi_*]}\Psi^r_{t+\cdot}\mathbf{1}), e_j\rangle_{\mathbb{L}^2} = \widehat{\mathcal{M}}_{t,\eta}(1_{[0,\chi_*]}\psi_j), \quad \langle \widehat{\mathcal{M}}_{t,\eta}(1_{[0,\chi_*]}\Psi^r_{t+\cdot}h^r), e'_j\rangle_{\mathbb{L}^2} = \widehat{\mathcal{M}}_{t,\eta}(1_{[0,\chi_*]}\tilde{\psi}_j). \qquad (3.10.28)$$

Finally, since $\widehat{E}^{(N)}$ is almost surely bounded on finite intervals, by Fubini's theorem,

$$\langle \widehat{E}^{(N)}(t)(1-G^r(\cdot)) - \int_0^t \widehat{E}^{(N)}(s)g^r(t-s+\cdot)ds, e_j\rangle_{\mathbb{L}^2}$$
$$= \widehat{E}^{(N)}(t)\langle 1-G^r, e_j\rangle_{\mathbb{L}^2} - \int_0^t \widehat{E}^{(N)}(s)\langle \widehat{E}^{(N)}(s)\langle g^r(t-s+\cdot), e_j\rangle_{\mathbb{L}^2}ds,$$

for $j-1,\cdots,k$. By Assumption 3.6.5, both $1-G^r$ and g^r lie in $\mathbb{L}^2(0,\infty)$, and hence $\langle 1-G^r, e_j\rangle_{\mathbb{L}^2}$ is finite, and the function $s\mapsto \langle g^r(t-s+\cdot), e_j\rangle$ is bounded and continuous on $[0,t]$. Moreover, for every $t\geq 0$ and $u\in\mathcal{C}_b[0,\infty)$, the mapping $f\mapsto \int_0^t f(s)u(s)ds$ from $\mathcal{D}[0,\infty)$ to $\mathbb{R}$ is continuous. Therefore, there exist continuous functions $F^3_j : \mathcal{D}[0,\infty)\mapsto\mathbb{R}$ such that

$$\langle \widehat{E}^{(N)}(t)(1-G^r(\cdot)) - \int_0^t \widehat{E}^{(N)}(s)g^r(t-s+\cdot)ds, e_j\rangle_{\mathbb{L}^2} = F^3_j(\widehat{E}^{(N)}). \qquad (3.10.29)$$

By the same argument,

$$\langle \widehat{E}(t)(1-G^r(\cdot)) - \int_0^t \widehat{E}(s)g^r(t-s+\cdot)ds, e_j\rangle_{\mathbb{L}^2} = F^3_j(\widehat{E}). \qquad (3.10.30)$$

Similarly, by Assumption 3.6.5, $(g^r)'$ also lies in $\mathbb{L}^2$, and hence $\langle g^r, e_j'\rangle$ is finite and the mapping $s \mapsto \langle (g^r)'(t-s+\cdot), e_j'\rangle$ is bounded and continuous on $[0,t]$. Therefore, there exists a continuous function $F_j^4 : \mathcal{D}[0,\infty) \mapsto \mathbb{R}$ such that

$$\langle \widehat{E}^{(N)}(t) g^s(\cdot) + \int_0^t \widehat{E}^{(N)}(s)(g^s)'(t-s+\cdot)ds, e_j'\rangle_{\mathbb{L}^2} = F_j^4(\widehat{E}^{(N)}), \tag{3.10.31}$$

and

$$\langle \widehat{E}(t) g^s(\cdot) + \int_0^t \widehat{E}(s)(g^s)'(t-s+\cdot)ds, e_j'\rangle_{\mathbb{L}^2} = F_j^4(\widehat{E}). \tag{3.10.32}$$

Now combining the representations (3.8.11) of $\widehat{Z}_\eta^{(N)}(t,\cdot)$, (3.8.13) of $\partial_r \widehat{Z}_\eta^{(N)}(t,\cdot)$, (3.6.7) of $\widehat{Z}_\eta(t,\cdot)$ and (3.10.17) of $\partial_r \widehat{Z}_\eta(t,\cdot)$, equations (3.10.22)-(3.10.32), and the fact that addition is continuous on $\mathcal{D}[0,\infty)$, we conclude that there exists a continuous functions

$$F : \mathcal{I} := \mathbb{H}^1(0,\infty) \times \mathcal{D}[0,\infty)^2 \times \mathbb{R}^{2k} \mapsto \mathbb{R}^k,$$

such that

$$\left(\langle \widehat{Z}_\eta^{(N)}(t,\cdot), e_1\rangle_{\mathbb{H}^1}, \cdots, \langle \widehat{Z}_\eta^{(N)}(t,\cdots), e_k\rangle\right) = F(\widehat{I}^{(N)}), \tag{3.10.33}$$

and

$$\left(\langle \widehat{Z}_\eta(t,\cdot), e_1\rangle_{\mathbb{H}^1}, \cdots, \langle \widehat{Z}_\eta(t,\cdots), e_k\rangle\right) = F(\widehat{I}), \tag{3.10.34}$$

where

$$\widehat{I}^{(N)} = \left(\widehat{Z}_\eta^{(N)}(0,\cdot), \widehat{E}^{(N)}, \widehat{\mathcal{H}}_\eta^{(N)}(\mathbf{1}), \widehat{\mathcal{M}}_{t,\eta}^{(N)}(1_{0,\chi*]}\psi_j), \widehat{\mathcal{M}}_{t,\eta}^{(N)}(1_{[0,\chi*]}\tilde{\psi}_j); j = 1,\cdots,k\right),$$

and

$$\widehat{I} = \left(\widehat{Z}_\eta(0,\cdot), \widehat{E}, \widehat{\mathcal{H}}_\eta(\mathbf{1}), \widehat{\mathcal{M}}_{t,\eta}(1_{[0,\chi*]}\psi_j), \widehat{\mathcal{M}}_{t,\eta}(1_{[0,\chi*]}\tilde{\psi}_j); j = 1,\cdots,k\right).$$

Thus, by Proposition 3.10.9 with ψ_j and $\tilde{\psi}_j; j = 1, \cdots, k$, defined in (3.10.24) and (3.10.26), we have

$$\widehat{I}^{(N)} \Rightarrow \widehat{I}, \tag{3.10.35}$$

in $\mathcal{I}$. The result in (3.6.11) follows from (3.10.33), (3.10.34), (3.10.35) and the continuous mapping theorem. □

Lemma 3.10.11. *Suppose Assumptions 3.6.3, 3.6.4, and 3.6.7 hold, and let $\widehat{Z}_\nu$ satisfy (3.6.9). Then almost, surely, the sample paths of $\{\widehat{Z}_\nu(t,\cdot); t \geq 0\}$ are $\mathbb{H}^1(0,\infty)$-valued and continuous, and for $t \geq 0$, the weak derivative $\partial_r \widehat{Z}_\nu(t,\cdot)$ of $\widehat{Z}_\nu(t,\cdot)$ satisfies for a.e. $r \in (0,\infty)$,*

$$\partial_r \widehat{Z}_\nu(t,r) = \widehat{Z}'_\nu(0,t+r) + \widehat{\mathcal{M}}_{t,\nu}(\Psi^s_{t+r}h^s) - g^s(r)\widehat{K}(t) - \int_0^t \widehat{K}(s)(g^s)'(t+r-s)ds. \tag{3.10.36}$$

Also almost surely, $(s,r) \mapsto \partial_r \widehat{Z}_\nu(s,r)$ is locally integrable on $(0,\infty)\times(0,\infty)$ and for every $t, r \geq 0$

$$\begin{aligned}\int_0^t \partial_r \widehat{Z}_\nu(s,r)ds &= \widehat{Z}_\nu(0,t+r) - \widehat{Z}_\nu(0,r) + \widehat{\mathcal{M}}_{t,\nu}(\Phi^s_r\mathbf{1}) - \widehat{\mathcal{M}}_{t,\nu}(\Psi^s_{t+r}\mathbf{1}) \\ &\quad - \int_0^t \widehat{K}(s)g^s(t+r-s)ds. \end{aligned}\tag{3.10.37}$$

Proof. The proof is analogous to the proof of Lemma 3.10.10. □

We now prove Theorem 3.6.11.

Proof of Theorem 3.6.11. For this proof, we again use the convention that all continuity results on $\mathcal{D}[0,\infty)$ are with respect to the topology of uniform convergence on compact sets. Fix $t \geq 0$ and $k \in \mathbb{N}$. First, recall that the diffusion many-server reneging mapping Λ is continuous by by Proposition 3.9.1 and measurable by Lemma 3.9.2.

Next we compute the projections of $\widehat{Z}_\nu^{(N)}(t,\cdot)$ and $\widehat{Z}_\nu(t,\cdot)$ onto the directions $e_1,\cdots,e_k$. Recall the representations (3.8.10) of $\widehat{Z}_\nu^{(N)}(t,\cdot)$ and (3.8.12) of $\partial_r\widehat{Z}_\nu^{(N)}(t,\cdot)$, (3.6.9) $\widehat{Z}_\nu(t,\cdot)$ and (3.10.36) of $\partial_r\widehat{Z}_\nu(t,\cdot)$. First, since the translation mapping $f\mapsto f(t+\cdot)$ is continuous from $\mathbb{L}^2(0,\infty)$ to itself, and for any $u\in\mathbb{L}^2(0,\infty)$, the projection mapping $f\mapsto\langle f,u\rangle_{\mathbb{L}^2}$ is continuous from $\mathbb{L}^2(0,\infty)$ to $\mathbb{R}$, there exist continuous mappings $F_j^1,F_j^2:\mathbb{L}^2(0,\infty)\mapsto\mathbb{R}$, $j=1,\cdots,k$, such that

$$\langle\widehat{Z}_\nu^{(N)}(0,t+\cdot),e_j\rangle_{\mathbb{L}^2}=F_j^1(\widehat{Z}_\nu^{(N)}(0,\cdot)),\qquad\langle\widehat{Z}_\nu(0,t+\cdot),e_j\rangle_{\mathbb{L}^2}=F_j^1(\widehat{Z}_\nu(0,\cdot)),\tag{3.10.38}$$

and

$$\langle\partial_r\widehat{Z}_\nu^{(N)}(0,t+\cdot),e_j'\rangle_{\mathbb{L}^2}=F_j^2(\partial_r\widehat{Z}_\nu^{(N)}(0,\cdot)),\qquad\langle\partial_r\widehat{Z}_\nu(0,t+\cdot),e_j'\rangle_{\mathbb{L}^2}=F_j^2(\partial_r\widehat{Z}_\nu(0,\cdot)).\tag{3.10.39}$$

Next, for every $j=1,\cdots,k$ define the function $\varphi_j:[0,\infty)\times[0,\infty)\mapsto\mathbb{R}$ as

$$\begin{aligned}\varphi_j(x,s):=\langle\Psi_{t+\cdot}^s\mathbf{1}(x,s),e_j\rangle_{\mathbb{L}^2}&=\int_0^\infty\Psi_{t+r}^s\mathbf{1}(x,s)e_j(r)dr\\&=\int_0^\infty\frac{1-G^s(x+t-s+r)}{1-G^s(x)}e_j(r)dr,\end{aligned}\tag{3.10.40}$$

which is bounded by $\|e_j\|_{\mathbb{L}^1}<\infty$ (recall that e_j has compact support by choice). Also, by the continuity of $1-G^s$ and boundedness of $1/(1-G^s)$ on finite intervals, the dominated convergence theorem shows that φ_j is continuous. Since $\varphi_j\in\mathcal{C}_b([0,\infty)\times[0,\infty))$, by the stochastic Fubini theorem for orthogonal martingale measures $\widehat{M}_\nu^{(N)}$, we have

$$\langle \widehat{\mathcal{M}}_{t,\nu}^{(N)}(\Psi_{t+\cdot}^{s}\mathbf{1}), e_j\rangle_{\mathbb{L}^2} = \int_0^\infty \widehat{\mathcal{M}}_{t,\nu}^{(N)}(\Psi_{t+r}^{s}\mathbf{1})e_j(r)dr$$
$$= \widehat{\mathcal{M}}_{t,\nu}^{(N)}\left(\int_0^\infty \Psi_{t+r}^{s}\mathbf{1}e_j(r)dr\right) = \widehat{\mathcal{M}}_{t,\nu}^{(N)}(\varphi_j). \tag{3.10.41}$$

Similarly, the function $\tilde{\varphi}_j : [0,\infty)\times[0,\infty) \mapsto \mathbb{R}$ defined as

$$\tilde{\varphi}_j(x,s) := \langle \Psi_{t+\cdot}^{s}h^s(x,s), e_j'\rangle_{\mathbb{L}^2} = \int_0^\infty \Psi_{t+r}^{s}h^s(x,s)e_j'(r)dr$$
$$= \int_0^\infty \frac{g^s(x+t-s+r)}{1-G^s(x)}e_j'(r)dr, \tag{3.10.42}$$

is also continuous and bounded (by $H_1^s\|e_j'\|_{\mathbb{L}^1} < \infty$) due to Assumption 3.6.3. Another application of the stochastic Fubini theorem for orthogonal martingales shows that

$$\langle \widehat{\mathcal{M}}_{t,\nu}^{(N)}(\Psi_{t+\cdot}^{s}h^s), e_j'\rangle_{\mathbb{L}^2} = \int_0^\infty \widehat{\mathcal{M}}_{t,\nu}^{(N)}(\Psi_{t+r}^{s}h^s)e_j'(r)dr$$
$$= \widehat{\mathcal{M}}_{t,\nu}^{(N)}\left(\int_0^\infty \Psi_{t+r}^{s}h^s e_j'(r)dr\right) = \widehat{\mathcal{M}}_{t,\nu}^{(N)}(\tilde{\varphi}_j). \tag{3.10.43}$$

Similarly, replacing $\widehat{\mathcal{M}}_{t,\nu}^{(N)}$ with $\widehat{\mathcal{M}}_{t,\nu}$, we have

$$\langle \widehat{\mathcal{M}}_{t,\nu}(\Psi_{t+\cdot}^{s}\mathbf{1}), e_j\rangle_{\mathbb{L}^2} = \widehat{\mathcal{M}}_{t,\nu}(\varphi_j), \qquad \langle \widehat{\mathcal{M}}_{t,\nu}(\Psi_{t+\cdot}^{s}h^s), e_j'\rangle_{\mathbb{L}^2} = \widehat{\mathcal{M}}_{t,\nu}(\tilde{\varphi}_j). \tag{3.10.44}$$

Finally, since $\widehat{K}^{(N)}$ is almost surely bounded on finite intervals, by Fubini's theorem,

$$\langle \widehat{K}^{(N)}(t)(1-G^s(\cdot)) - \int_0^t \widehat{K}^{(N)}(s)g^s(t-s+\cdot)ds, e_j\rangle_{\mathbb{L}^2}$$
$$= \widehat{K}^{(N)}(t)\langle 1-G^s, e_j\rangle_{\mathbb{L}^2} - \int_0^t \widehat{K}^{(N)}(s)\langle \widehat{K}^{(N)}(s)\langle g^s(t-s+\cdot), e_j\rangle_{\mathbb{L}^2}ds,$$

for $j-1,\cdots,k$. By Assumption (3.6.3) both $1-G^s$ and g^s lie in $\mathbb{L}^2(0,\infty)$, and hence $\langle 1-G^s, e_j\rangle_{\mathbb{L}^2}$ is finite, and the function $s \mapsto \langle g^s(t-s+\cdot), e_j\rangle$ is bounded and continuous on $[0,t]$. Moreover, for every $t \geq 0$ and $u \in \mathcal{C}_b[0,\infty)$, the mapping $f \mapsto \int_0^t f(s)u(s)ds$ from $\mathcal{D}[0,\infty)$ to $\mathbb{R}$ is continuous. Therefore, by equation (3.8.23), there exist continuous functions $\tilde{F}_j^3 : \mathcal{D}[0,\infty) \mapsto \mathbb{R}$ and $F_j^3 : \mathcal{D}[0,\infty)\times\mathbb{R}\times\mathcal{D}[0,\infty)\times B^+[0,\infty) \mapsto \mathbb{R}$, such that

$$\begin{aligned}
&\langle \widehat{K}^{(N)}(t)(1-G^s(\cdot)) - \int_0^t \widehat{K}^{(N)}(s)g^s(t-s+\cdot)ds, e_j\rangle_{\mathbb{L}^2}\\
&= \tilde{F}_j^3(\widehat{K}^{(N)})\\
&= F_j^3(\widehat{E}^{(N)}, \widehat{Z}_\eta^{(N)}(\cdot,0), \int_0^\cdot \partial_r \widehat{Z}_\eta^{(N)}(s,0)ds, \widehat{X}^{(N)}(0), \widehat{Z}_\nu^{(N)}(0,\cdot), \widehat{\mathcal{H}}_\nu^{(N)}(\mathbf{1}), \widehat{M}_R^{(N)}, \widehat{Y}^{(N)}),
\end{aligned} \tag{3.10.45}$$

$$\widehat{Y}^{(N)} = h^r\left((F^{\overline{\eta}_\cdot^{(N)}})^{-1}(\overline{Q}(\cdot) + C_1^{(N)}(\cdot))\right) \tag{3.10.46}$$

By the same argument,

$$\begin{aligned}
&\langle \widehat{K}(t)(1-G^s(\cdot)) - \int_0^t \widehat{K}(s)g^s(t-s+\cdot)ds, e_j\rangle_{\mathbb{L}^2}\\
=&F_j^3(\widehat{E}, \widehat{Z}_\eta(\cdot,0), \int_0^\cdot \partial_r \widehat{Z}_\eta(s,0)ds, \widehat{X}(0), \widehat{Z}_\nu(0,\cdot), \widehat{\mathcal{H}}_\nu(\mathbf{1}), \widehat{M}_R, \widehat{Y}),
\end{aligned} \tag{3.10.47}$$

where

$$\widehat{Y} = h^r(\chi_*) \tag{3.10.48}$$

Similarly, by Assumption 3.6.3, $(g^s)'$ also lies in $\mathbb{L}^2$, and hence $\langle g^s, e_j'\rangle$ is finite and the mapping $s \mapsto \langle (g^s)'(t-s+\cdot), e_j'\rangle$ is bounded and continuous on $[0,t]$. Therefore, there exists a continuous function $F_j^4 : \mathcal{D}[0,\infty)\times\mathbb{R}\times\mathcal{D}[0,\infty)\times B^+[0,\infty) \mapsto \mathbb{R}$ such that

$$\begin{aligned}
&\langle \widehat{K}^{(N)}(t)g^s(\cdot) + \int_0^t \widehat{K}^{(N)}(s)(g^s)'(t-s+\cdot)ds, e_j'\rangle_{\mathbb{L}^2}\\
=&F_j^4(\widehat{E}^{(N)}, \widehat{Z}_\eta^{(N)}, \widehat{X}^{(N)}(0), \widehat{Z}_\nu^{(N)}(0,\cdot) - \widehat{\mathcal{H}}_\nu^{(N)}(\mathbf{1}), \widehat{Y}^{(N)}),
\end{aligned} \tag{3.10.49}$$

and

$$\langle \widehat{K}(t)g^s(\cdot) + \int_0^t \widehat{K}(s)(g^s)'(t-s+\cdot)ds, e_j'\rangle_{\mathbb{L}^2} = F_j^4(\widehat{E}, \widehat{Z}_\eta, \widehat{X}(0), \widehat{Z}_\nu(0,\cdot) - \widehat{\mathcal{H}}_\nu(\mathbf{1}), \widehat{Y}). \tag{3.10.50}$$

Now combining representations (3.8.10) of $\widehat{Z}_\nu^{(N)}(t,\cdot)$ and (3.8.12) of $\partial_r \widehat{Z}_\nu^{(N)}(t,\cdot)$, (3.6.9) $\widehat{Z}_\nu(t,\cdot)$ and (3.10.36) of $\partial_r \widehat{Z}_\nu(t,\cdot)$, equations (3.10.38)-(3.10.50), and the fact that addition is continuous on $\mathcal{D}[0,\infty)$, we conclude that there exists a continuous functions

$$F : \mathcal{J} := \mathbb{R}\times\mathbb{H}^1(0,\infty)\times\mathcal{D}_{\mathbb{R}}[0,\infty)\times\mathbb{R}\times\mathcal{D}[0,\infty)^2\times B^+[0,\infty)\times\mathcal{D}[0,\infty)\times\mathbb{R}^{2k} \mapsto \mathbb{R}^{k+2},$$

such that

$$\left(\widehat{X}^{(N)}(t), \widehat{Q}^{(N)}(t), \langle \widehat{Z}_\nu^{(N)}(t,\cdot), e_1\rangle_{\mathbb{H}^1}, \cdots, \langle \widehat{Z}_\nu^{(N)}(t,\cdots), e_k\rangle\right) = F(\widehat{J}^{(N)}), \tag{3.10.51}$$

and

$$\left(\widehat{X}(t), \widehat{Q}(t), \langle \widehat{Z}_\nu(t,\cdot), e_1\rangle_{\mathbb{H}^1}, \cdots, \langle \widehat{Z}_\nu(t,\cdots), e_k\rangle\right) = F(\widehat{J}), \tag{3.10.52}$$

where

$$\widehat{J}^{(N)} = \Big(\widehat{X}^{(N)}(0), \widehat{Z}_\nu^{(N)}(0,\cdot), \widehat{Z}_\eta^{(N)}(\cdot,0), \int_0^\cdot \partial_r \widehat{Z}_\eta^{(N)}(s,0)ds, \widehat{E}^{(N)}, \widehat{M}_R^{(N)}, \widehat{Y}^{(N)},$$
$$\widehat{\mathcal{H}}_\nu^{(N)}(\mathbf{1}), \widehat{\mathcal{M}}_{t,\nu}^{(N)}(\varphi_j), \widehat{\mathcal{M}}_{t,\nu}^{(N)}(\tilde{\varphi}_j); j=1,\cdots,k\Big),$$

and

$$\widehat{J} = \Big(\widehat{X}(0), \widehat{Z}_\nu(0,\cdot), \widehat{Z}_\eta(\cdot,0), \int_0^\cdot \partial_r \widehat{Z}_\eta(s,0)ds, \widehat{E}, \widehat{M}_R, \widehat{Y},$$
$$\widehat{\mathcal{H}}_\nu(\mathbf{1}), \widehat{\mathcal{M}}_{t,\nu}(\varphi_j), \widehat{\mathcal{M}}_{t,\nu}(\tilde{\varphi}_j); j=1,\cdots,k\Big).$$

Note that subtracting (3.8.10) from (3.8.2), rearranging and using (3.10.18), (3.10.3) and the definition (3.4.46) of $\widehat{\mathcal{H}}_\eta$, we see that $\int_0^t \partial_r \widehat{Z}_\eta^{(N)}(s,0)ds \to \int_0^t \partial_r \widehat{Z}_\eta(s,0)ds$ as $N \to \infty$. Thus, by Proposition 3.10.9 with φ_j and $\tilde{\varphi}_j; j = 1, \cdots, k$, defined in (3.10.40) and (3.10.42), Theorem 3.6.10, Corollary 3.7.6, (3.8.21) and (3.8.22) we have

$$\widehat{J}^{(N)} \Rightarrow \widehat{J}, \tag{3.10.53}$$

in $\mathcal{J}$. The result in (3.6.12) follows from (3.10.51), (3.10.52), (3.10.53) and the continuous mapping theorem. □

We now prove Theorem 3.6.13 which gives an alternate characterization of the limit process.

Proof of Theorem 3.6.13. Since $(\widehat{K}, \widehat{X}, \widehat{Z}_\nu(\cdot, 0), \widehat{Q})$ ia a solution to the DMSRE equation, using (3.6.4) and (3.6.5) it is easy to check that (3.6.18) and (3.6.15) hold. Now, note that subtracting (3.8.10) from (3.8.1), rearranging and using (3.10.37), (3.10.3) and the definition (3.4.45) of $\widehat{\mathcal{H}}_\nu$, we see that $\int_0^t \partial_r \widehat{Z}_\nu^{(N)}(s,0)ds \to \int_0^t \partial_r \widehat{Z}_\nu(s,0)ds$ as $N \to \infty$. Similarly $\int_0^t \partial_r \widehat{Z}_\eta^{(N)}(s,0)ds \to \int_0^t \partial_r \widehat{Z}_\eta(s,0)ds$ as $N \to \infty$. Now using (3.7.14) and (3.8.20) along with the convergence (3.6.11), (3.6.12) and Corollary 3.7.6, we have (3.6.14).

Substituting $\int_0^t \partial_r \widehat{Z}_\nu(s,r)ds$ from equation (3.10.37), and using the definition of $\widehat{\mathcal{K}}$, the right-hand side of equation (3.6.16) can be written as

$$\begin{aligned}
&\widehat{Z}_\nu(0,r) + \widehat{Z}_\nu(0,t+r) - \widehat{Z}_\nu(0,r) + \widehat{\mathcal{M}}_{t,\nu}(\Phi_r^s \mathbf{1}) - \widehat{\mathcal{M}}_{t,\nu}(\Psi_{t+r}^s \mathbf{1})) \\
&\quad - \int_0^t \widehat{K}(s) g^s(t+r-s)ds - \widehat{\mathcal{M}}_{t,\nu}(\Phi_r^s \mathbf{1}) + (1 - G^s(r))\widehat{K}(t) \\
&\quad = \widehat{Z}_\nu(0,t+r) - \widehat{\mathcal{M}}_{t,\nu}(\Psi_{t+r}^s \mathbf{1}) + \widehat{\mathcal{K}}_t(\alpha^r).
\end{aligned}$$

By the definition of $\widehat{\mathcal{H}}_\nu$ the above equation is equal to $\widehat{Z}_\nu(t,r)$ by (3.6.9) proving (3.6.16).

Substituting $\int_0^t \partial_r \widehat{Z}_\eta(s,r)ds$ from equation (3.10.18), and using the definition of $\widehat{\mathcal{E}}$, the right-hand side of equation (3.6.17) can be written as

$$\begin{aligned}
&\widehat{Z}_\eta(0,r) + \widehat{Z}_\eta(0,t+r) - \widehat{Z}_\eta(0,r) + \widehat{\mathcal{M}}_{t,\eta}(1_{[0,\chi_*]}\Phi_r^r\mathbf{1}) - \widehat{\mathcal{M}}_{t,\eta}(1_{[0,\chi_*]}\Psi_{t+r}^r\mathbf{1}) \\
&\quad - \int_0^t \widehat{E}(s)g^r(t+r-s)ds - \widehat{\mathcal{M}}_{t,\eta}(1_{[0,\chi_*]}\Phi_r^r\mathbf{1}) + (1-G^r(r))\widehat{E}(t) \\
&\quad = \widehat{Z}_\eta(0,t+r) - \widehat{\mathcal{M}}_{t,\eta}(1_{[0,\chi_*]}\Psi_{t+r}^r\mathbf{1}) + \widehat{\mathcal{E}}_t(\beta^r).
\end{aligned}$$

By the definition of $\widehat{\mathcal{H}}_\eta$ the above equation is equal to $\widehat{Z}_\eta(t,r)$ by (3.6.7) proving (3.6.17). □

We now look at the proof of uniqueness of solution to the diffusion SPDE. We begin by stating this Lemma from [3, Lemma 4.19] with g replaced by g^s.

Lemma 3.10.12. *Suppose Assumption 3.6.3 holds, and a function $F \in \mathcal{C}[0,\infty)$ with $F(0) = 0$ is given. Let $\xi : [0,\infty) \times [0,\infty) \mapsto \mathbb{R}$ be a function that satisfies the following two properties:*

1. *The mapping $t \mapsto \xi(t,\cdot)$ lies in $\mathcal{C}([0,\infty);\mathbb{H}^1(0,\infty))$. For each $t > 0$, let $\partial_r\xi(t,\cdot)$ denote the weak derivative of $\xi(t,\cdot)$.*
2. *The function $\partial_r\xi : (0,\infty) \times (0,\infty) \mapsto \mathbb{R}$ is locally integrable.*

Then ξ satisfies the equation

$$\xi(t,r) = \int_0^t \partial_r\xi(s,r)ds + (1-G^s(r))F(t), \qquad a.e. \quad r \in (0,\infty), \tag{3.10.54}$$

for every $t \geq 0$ if and only if

$$\xi(t,r) = \mathcal{F}_t(\alpha^r), \qquad t, r \geq 0, \tag{3.10.55}$$

with $\mathcal{F}_t(\alpha^r)$ is defined as

$$\mathcal{F}_t(\alpha^r) = (1 - G^s(r))F(t) - \int_{[0,t]} F(s) g^s(r + t - s) ds, \qquad t \in [0, \infty),$$

.

Proof of Theorem 3.6.14. For $i = 1, 2$, let $Y_i = (X_i, \widehat{Z}_\nu^i, \widehat{Z}_\eta^i)$ be solutions of the equations (3.6.14)-(3.6.17) with the same initial condition $\widehat{Y}(0)$. Define

$$\widehat{K}_i(t) = \widehat{E}(t) - \widehat{R}_i(t) - \int_0^t \widehat{Q}_i(s) h^r(\chi_*) ds + \widehat{Q}_i(0) - \widehat{Q}_i(t), \tag{3.10.56}$$

where

$$\widehat{R}_i(t) = \widehat{\mathcal{M}}_R(t) - \int_0^t \partial_r \widehat{Z}_\eta^i(s, 0) ds + \int_0^t \widehat{Z}_\eta^i(s, 0) h^r(\chi_*) ds.$$

and

$$\widehat{Q}_i(t) = \begin{cases} 0 & \text{if } x_* < 1, \\ \widehat{X}_i^+(t) & \text{if } x_* = 1, \\ \widehat{X}_i(t) & \text{if } x_* > 1, \end{cases}$$

Let $\Delta H = H_1 - H_2$ for $H = \widehat{X}, \widehat{Q}, \widehat{Z}_\nu, \widehat{Z}_\eta, \widehat{K}, \widehat{R}$. Using (3.6.17),

$$\Delta \widehat{Z}_\eta(t, r) = \int_0^t \partial_r \Delta \widehat{Z}_\eta(s, r) ds$$

Since $\Delta\widehat{Z}_\eta(0,\cdot) \equiv 0$, the above equation forces

$$\Delta\widehat{Z}_\eta(t,r) = 0 \qquad \forall t, r. \tag{3.10.57}$$

By (3.6.16)

$$\Delta\widehat{Z}_\nu(t,r) = \int_0^t \partial_r \Delta\widehat{Z}_\nu(s,r)ds + (1 - G^s(r))\Delta\widehat{K}(t), \tag{3.10.58}$$

where $\Delta\widehat{Z}_\nu(0,\cdot) \equiv 0$ and

$$\Delta\widehat{K}(t) = \int_0^t \Delta\widehat{Q}(s)h^r(\chi_*)ds - \Delta\widehat{Q}(t). \tag{3.10.59}$$

By the above definition $\Delta\widehat{Z}_\nu$ satisfies properties 1 and 2 of Lemma 3.10.12, and hence

$$\Delta\widehat{Z}_\nu(t,r) = \Delta\widehat{\mathcal{K}}_t(\alpha^r), \qquad t, r \geq 0. \tag{3.10.60}$$

Since Y_i for $i = 1, 2$ is a continuous $\mathbb{Y}$-valued random element, $\Delta\widehat{K}$ is continuous almost surely. So, for every $t \geq 0$, $\Delta\widehat{Z}_v(t,\cdot)$ is continuously differentiable on $[0,\infty)$, with

$$\partial_r \Delta\widehat{Z}_\nu(t,r) = -g^s(r)\Delta\widehat{K}(t) - \int_0^t \Delta\widehat{K}(u)(g^s)'(r+t-u)du, \qquad r, t \in [0,\infty). \tag{3.10.61}$$

Using equation (3.6.14) and the fact that $|\widehat{Q}(t)| \leq |\widehat{X}(t)|$, we get

$$\begin{aligned} |\Delta\widehat{X}(t)| &= |\int_0^t \partial_r \Delta\widehat{Z}_\nu(s,0)ds - \int_0^t \Delta\widehat{Q}(s)h^r(\chi_*)ds| \\ &\leq \int_0^t |\partial_r \Delta\widehat{Z}_\nu(s,0)|ds + \int_0^t |\Delta\widehat{X}(s)|h^r(\chi_*)ds. \end{aligned} \tag{3.10.62}$$

By Gronwall's lemma

$$|\Delta\widehat{X}(t)| \leq \exp(th^r(\chi_*)) \int_0^t |\partial_r \Delta\widehat{Z}_\nu(s,0)| ds \tag{3.10.63}$$

Using equation (3.10.59), this implies that

$$\begin{aligned}
|\Delta\widehat{K}(t)| &\leq |\Delta\widehat{X}(t)| + \int_0^t |\Delta\widehat{X}(s)| h^r(\chi_*) ds \\
&\leq \exp(th^r(\chi_*)) \int_0^t |\partial_r \Delta\widehat{Z}_\nu(s,0)| ds + \int_0^t h^r(\chi_*) \exp(sh^r(\chi_*)) \int_0^s |\partial_r \Delta\widehat{Z}_\nu(s,0)| du ds \\
&\leq 2\exp(th^r(\chi_*)) \int_0^t |\partial_r \Delta\widehat{Z}_\nu(s,0)| ds
\end{aligned} \tag{3.10.64}$$

Also,

$$|\Delta\widehat{Q}(t)| \leq |\Delta\widehat{X}(t)| \leq \exp(th^r(\chi_*)) \int_0^t |\partial_r \Delta\widehat{Z}_\nu(s,0)| ds, \qquad t \in [0,\infty). \tag{3.10.65}$$

The above equation when combined with equation (3.10.61) with $r = 0$ shows that for all $t \in [0,T]$,

$$\begin{aligned}
|\partial_r \Delta\widehat{Z}_\nu(t,0)| \leq& g^s(0)|\Delta\widehat{K}(t)| + \int_0^t |(g^s)'(t-s)||\Delta\widehat{K}(s)| ds \\
\leq& 2H_1^s \exp(Th^r(\chi_*)) \int_0^t |\partial_r \Delta\widehat{Z}_\nu(s,0)| ds \\
&+ 2H_2^s \int_0^t \exp(sh^r(\chi_*)) \left(\int_0^v |\partial_r \Delta\widehat{Z}_\nu(s,0)| ds \right) dv \\
\leq& 2\left(H_1^s + H_2^s \frac{\exp(Th^r(\chi_*))}{h^r(\chi_*)} \right) \int_0^t |\partial_r \Delta\widehat{Z}_\nu(s,0)| ds.
\end{aligned}$$

Since $\partial_r \Delta\widehat{Z}_\nu(0,0) = 0$, by Gronwall's lemma this implies $\partial_r \Delta\widehat{Z}_\nu(s,0) = 0$ for all $t \geq 0$. Using equations (3.10.63), (3.10.64) and (3.10.65), this implies $\Delta\widehat{X}(t) = \Delta\widehat{Q}(t) = \Delta\widehat{K}(t) = 0$ for all $t \geq 0$. Using equation (3.10.60), this implies $\Delta\widehat{Z}_\nu(t,\cdot) = 0$. Thus $Y_1(t) = Y_2(t)$ for all $t \geq 0$. □

Bibliography

[1] Aghajani, R., Li, X., and Ramanan, K. Mean-field dynamics of load-balancing networks with general service distributions. Arxiv preprint, https://arxiv.org/abs/1512.05056, 2015.

[2] Aghajani, R., Li, X., and Ramanan, K. The PDE method for the analysis of randomized load balancing networks. *Proc. ACM Meas. Anal. Comput. Syst. 1*, 2 (Dec. 2017), 38:1–38:28.

[3] Aghajani, R., and Ramanan, K. Ergodicity of an SPDE associated with a many-server queue. *Ann. Appl. Probab. 29*, 2 (2019), 994–1045.

[4] Aghajani, R., and Ramanan, K. The hydrodynamic limit of a randomized load balancing network. *Ann. Appl. Probab. 29*, 4 (2019), 2114–2174.

[5] Baccelli, F., and Hebuterne, G. On queues with impatient customers. In *Performance '81 (Amsterdam, 1981).* North-Holland, Amsterdam-New York, 1981, pp. 159–179.

[6] Boxma, O. J., and de Waal, P. R. Multiserver queues with impatient customers. In *Proceedings of ITC* (1994), vol. 14, Elsevier, Amsterdam, p. 8.

[7] Bramson, M. Stability of join the shortest queue networks. *Ann. Appl. Probab. 21*, 4 (2011), 1568–1625.

[8] Bramson, M., Lu, Y., and Prabhakar, B. Randomized load balancing with general service time distributions. *ACM SIGMETRICS performance evaluation review 38*, 1 (2010), 275–286.

[9] Bramson, M., Lu, Y., and Prabhakar, B. Asymptotic independence of queues under randomized load balancing. *Queueing Syst. 71*, 3 (2012), 247–292.

[10] Bramson, M., Lu, Y., and Prabhakar, B. Decay of tails at equilibrium for FIFO join the shortest queue networks. *Ann. Appl. Probab. 23*, 5 (2013), 1841–1878.

[11] Brezis, H. *Functional analysis, Sobolev spaces and partial differential equations.* Universitext. Springer, New York, 2011.

[12] Brightwell, G., Fairthorne, M., and Luczak, M. J. The supermarket model with bounded queue lengths in equilibrium. *J. Stat. Phys. 173*, 3-4 (2018), 1149–1194.

[13] Brown, L., Gans, N., Mandelbaum, A., Sakov, A., Shen, H., Zeltyn, S., and Zhao, L. Statistical analysis of a telephone call center: a queueing-science perspective. *J. Amer. Statist. Assoc. 100*, 469 (2005), 36–50.

[14] Burton, T. A. *Volterra integral and differential equations*, second ed., vol. 202 of *Mathematics in Science and Engineering*. Elsevier B. V., Amsterdam, 2005.

[15] Chen, S., Sun, Y., Huang, L., Sinha, P., Liang, G., Liu, X., Shroff, N. B., et al. When queueing meets coding: Optimal-latency data retrieving scheme in storage clouds. In *IEEE INFOCOM 2014-IEEE Conference on Computer Communications* (2014), IEEE, pp. 1042–1050.

[16] Dai, J. G., He, S., and Tezcan, T. Many-server diffusion limits for $G/Ph/n+GI$ queues. *Ann. Appl. Probab. 20*, 5 (2010), 1854–1890.

[17] Ethier, S. N., and Kurtz, T. G. *Markov processes.* Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, 1986. Characterization and convergence.

[18] Folland, G. B. *Real analysis*, second ed. Pure and Applied Mathematics (New York). John Wiley & Sons, Inc., New York, 1999. Modern techniques and their applications, A Wiley-Interscience Publication.

[19] Gans, N., Koole, G., and Mandelbaum, A. Telephone call centers: Tutorial, review, and research prospects. *Manufacturing & Service Operations Management 5*, 2 (2003), 79–141.

[20] Garnett, O., Mandelbaum, A., and Reiman, M. Designing a call center with impatient customers. *Manufacturing & Service Operations Management 4*, 3 (2002), 208–227.

[21] Jacod, J., and Shiryaev, A. N. *Limit theorems for stochastic processes*, vol. 288 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences].* Springer-Verlag, Berlin, 1987.

[22] Kang, W., and Ramanan, K. Fluid limits of many-server queues with reneging. *Ann. Appl. Probab. 20*, 6 (2010), 2204–2260.

[23] Kang, W., and Ramanan, K. Asymptotic approximations for stationary distributions of many-server queues with abandonment. *Ann. Appl. Probab. 22*, 2 (2012), 477–521.

[24] Kaspi, H., and Ramanan, K. Law of large numbers limits for many-server queues. *Ann. Appl. Probab. 21*, 1 (2011), 33–114.

[25] Kaspi, H., and Ramanan, K. SPDE limits of many-server queues. *Ann. Appl. Probab. 23*, 1 (2013), 145–229.

[26] Kolesar, P. Stalking the endangered cat: A queueing analysis of congestion at automatic teller machines. *Interfaces 14*, 6 (1984), 16–26.

[27] Kruk, L. u., Lehoczky, J., Shreve, S., and Yeung, S.-N. Earliest-deadline-first service in heavy-traffic acyclic networks. *Ann. Appl. Probab. 14*, 3 (2004), 1306–1352.

[28] Li, Q.-L., Lui, J. C., and Wang, Y. A matrix-analytic solution for randomized load balancing models with ph service times. In *International Workshop on Performance Evaluation of Computer and Communication Systems* (2010), Springer, pp. 240–253.

[29] Liang, G., and Kozat, U. C. Tofec: Achieving optimal throughput-delay trade-off of cloud storage using erasure codes. In *IEEE INFOCOM 2014-IEEE Conference on Computer Communications* (2014), IEEE, pp. 826–834.

[30] Mandelbaum, A., Massey, W. A., and Reiman, M. I. Strong approximations for Markovian service networks. *Queueing Systems Theory Appl. 30*, 1-2 (1998), 149–201.

[31] Mandelbaum, A., and Momčilović, P. Queues with many servers and impatient customers. *Math. Oper. Res. 37*, 1 (2012), 41–65.

[32] Mitzenmacher, M. The power of two choices in randomized load balancing. *IEEE Transactions on Parallel and Distributed Systems 12*, 10 (2001), 1094–1104.

[33] Mukherjee, D., Borst, S. C., van Leeuwaarden, J. S., and Whiting, P. A. Asymptotic optimality of power-of-d load balancing in large-scale systems. *Mathematics of Operations Research* (2020).

[34] Mukherjee, D., Borst, S. C., van Leeuwaarden, J. S. H., and Whiting, P. A. Universality of power-of-d load balancing in many-server systems. *Stoch. Syst. 8*, 4 (2018), 265–292.

[35] Mukhopadhyay, A., Karthik, A., and Mazumdar, R. R. Randomized assignment of jobs to servers in heterogeneous clusters of shared servers for low delay. *Stoch. Syst. 6*, 1 (2016), 90–131.

[36] Narasimhan, R., and Nievergelt, Y. *Complex analysis in one variable*, second ed. Birkhäuser Boston, Inc., Boston, MA, 2001.

[37] Parthasarathy, K. R. *Probability measures on metric spaces.* AMS Chelsea Publishing, Providence, RI, 2005. Reprint of the 1967 original.

[38] Puha, A. L., and Williams, R. J. Asymptotic behavior of a critical fluid model for a processor sharing queue via relative entropy. *Stoch. Syst. 6*, 2 (2016), 251–300.

[39] Puhalskii, A. A., and Reiman, M. I. The multiclass $GI/PH/N$ queue in the Halfin-Whitt regime. *Adv. in Appl. Probab. 32*, 2 (2000), 564–595.

[40] Reed, J., and Tezcan, T. Hazard rate scaling of the abandonment distribution for the $GI/M/n+GI$ queue in heavy traffic. *Oper. Res. 60*, 4 (2012), 981–995.

[41] Stanley, R. P. *Enumerative combinatorics. Volume 1*, second ed., vol. 49 of *Cambridge Studies in Advanced Mathematics.* Cambridge University Press, Cambridge, 2012.

[42] Vasantam, T., Mukhopadhyay, A., and Mazumdar, R. R. The mean-field behavior of processor sharing systems with general job lengths under the sq (d) policy. *Performance Evaluation 127* (2018), 120–153.

[43] Vvedenskaya, N. D., Dobrushin, R. L., and Karpelevich, F. I. A queueing system with a choice of the shorter of two queues—an asymptotic approach. *Problemy Peredachi Informatsii 32*, 1 (1996), 20–34.

[44] Walsh, J. B. An introduction to stochastic partial differential equations. In *École d'été de probabilités de Saint-Flour, XIV—1984*, vol. 1180 of *Lecture Notes in Math.* Springer, Berlin, 1986, pp. 265–439.

[45] Whitt, W. *Stochastic-process limits.* Springer Series in Operations Research. Springer-Verlag, New York, 2002. An introduction to stochastic-process limits and their application to queues.

[46] Whitt, W. Heavy-traffic limits for the $G/H_2^*/n/m$ queue. *Math. Oper. Res. 30*, 1 (2005), 1–27.

[47] Zeltyn, S., and Mandelbaum, A. Call centers with impatient customers: many-server asymptotics of the $M/M/n+G$ queue. *Queueing Syst. 51*, 3-4 (2005), 361–402.

[48] Zwart, B. Heavy-traffic asymptotics for the single-server queue with random order of service. *Oper. Res. Lett. 33*, 5 (2005), 511–518.